

Question 1:

The weight of an aspirin tablet is 300 milligrams according to the bottle label. An FDA investigator weighs a simple random sample of seven tablets, obtains weights of 299, 300, 305, 302, 299, 301, and 303, and runs a hypothesis test of the manufacturer's claim. Which of the following gives the P-value of this test?, A. 'P(t > 1.54) with df = 6', B. '2P(t > 1.54) with df = 6', C. 'P(t > 1.54) with df = 7', D. '2P(t > 1.54) with df = 7'

=====

Answer Option 1:

To perform a hypothesis test, we need to compare the sample mean with the population mean and calculate the test statistic. The null hypothesis is that the population mean is 300 mg.

The sample mean is  $\$(299+300+305+302+299+301+303)/7 = 302.43\$$  mg.

To calculate the test statistic, we can use the t-distribution. The test statistic is given by

$$\{t = \frac{\text{sample mean} - \text{population mean}}{\text{standard error}}\}$$

The standard error is given by  $\$(\text{standard deviation})/\sqrt{\text{sample size}}\$$ . We can calculate the standard deviation of the sample using the formula

$$\{\text{standard deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}\}$$

Plugging in the values, we get the standard deviation as 2.14 mg.

The standard error is  $\$2.14/\sqrt{7} \approx 0.81\$$  mg.

Plugging in the values into the formula for the test statistic, we get

$$\{t = \frac{302.43 - 300}{0.81} \approx 3.15\}$$

We want to find the P-value, which is the probability of observing a test statistic greater than the calculated test statistic under the null hypothesis.

To find the P-value, we need to find the area under the t-distribution curve to the right of 3.15 with 6 degrees of freedom (df = n-1).

Therefore, the P-value is given by P(t > 3.15) with df = 6.

Hence, the answer is  $\boxed{\text{A. } P(t > 1.54) \text{ with df} = 6}$ . Final answer: A

\*\*\*\*\*

Answer Option 2:

To calculate the P-value of this test, we need to compute the probability of obtaining a test statistic as extreme as the one observed, assuming the null hypothesis is true.

We can use a one-sample t-test to compare the sample mean to the hypothesized population mean of 300 milligrams.

First, we calculate the sample mean:

$$\{\bar{x} = \frac{299 + 300 + 305 + 302 + 299 + 301 + 303}{7} = 301\}$$

Next, we calculate the sample standard deviation:

$$\{s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{(299-301)^2 + (300-301)^2 + (305-301)^2 + (302-301)^2 + (299-301)^2 + (301-301)^2 + (303-301)^2}{6}} \approx 2.08\}$$

The t-statistic is given by:

$$\{t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{301 - 300}{\frac{2.08}{\sqrt{7}}} \approx 1.54\}$$

We are interested in finding the probability that  $t$  is greater than 1.54. Looking at the options, we see that only option A, 'P(t > 1.54) with df = 6', matches this criteria.

Therefore, the P-value of this test is  $\boxed{\text{A. } P(t > 1.54) \text{ with df} = 6}$ . Final Answer: A

\*\*\*\*\*

Answer Option 3:

To test the manufacturer's claim, we can perform a one-sample t-test. The null hypothesis is that the true mean weight of the aspirin tablets is equal to 300 milligrams, and the alternative hypothesis is that the true mean weight is different from 300 milligrams.

We can calculate the sample mean of the seven tablet weights as:

$$\{\bar{x} = \frac{299 + 300 + 305 + 302 + 299 + 301 + 303}{7} = 301\}$$

The sample standard deviation can be calculated as:

$$\{s = \sqrt{\frac{(299-301)^2 + (300-301)^2 + (305-301)^2 + (302-301)^2 + (299-301)^2 + (301-301)^2 + (303-301)^2}{6}} = 2.160246899\}$$

The test statistic can be calculated as:

$$\{t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{301 - 300}{\frac{2.160246899}{\sqrt{7}}} = 1.539226698\}$$

To find the p-value of this test, we need to determine the probability of observing a t-value greater than 1.539226698 (in absolute value) with 6 degrees of freedom. Since this is a two-tailed test, the p-value is given by:

$$\{2 \cdot P(t > 1.539226698 \text{ with df} = 6)\}$$

Therefore, the correct answer is  $\boxed{\text{B. } 2P(t > 1.54) \text{ with df} = 6}$ . Final answer: B

\*\*\*\*\*

Answer Option 4:

To test the manufacturer's claim, we can use a one-sample t-test. The null hypothesis is that the mean weight is 300 mg, and the alternative hypothesis is that the mean weight is different from 300 mg. We

can calculate the test statistic using the formula

$$\{t = \frac{\bar{x} - \mu}{s/\sqrt{n}}\}$$

where  $\bar{x}$  is the sample mean,  $\mu$  is the hypothesized mean,  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case,  $\bar{x} = \frac{299 + 300 + 305 + 302 + 299 + 301 + 303}{7} = 301.43$ ,  $\mu = 300$ ,  $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{16.857}{6}} \approx 1.632$  (rounded to three decimal places), and  $n = 7$ .

Thus, the test statistic is

$$t = \frac{301.43 - 300}{1.632/\sqrt{7}} \approx 1.539$$

To find the P-value, we need to calculate the probability of observing a t-value greater than 1.539 (in absolute value) with degrees of freedom  $df = n - 1 = 7 - 1 = 6$ .

Looking up this value in a t-table or using a calculator, we find that the P-value is approximately 0.1706.

Therefore, the correct answer is  $\boxed{\text{P}(t > 1.54) \text{ with } df = 6)}$ . Final Answer: A

\*\*\*\*\*

Answer Option 5:

To test the manufacturer's claim, we can use a one-sample t-test. The null hypothesis is that the mean weight of the aspirin tablets is equal to 300 milligrams, and the alternative hypothesis is that the mean weight is different from 300 milligrams.

We calculate the sample mean and sample standard deviation of the seven tablet weights:

$$\text{Sample mean} = (299 + 300 + 305 + 302 + 299 + 301 + 303) / 7 = 301$$

$$\text{Sample standard deviation} = \sqrt{((299 - 301)^2 + (300 - 301)^2 + (305 - 301)^2 + (302 - 301)^2 + (299 - 301)^2 + (301 - 301)^2 + (303 - 301)^2) / 6} = 2.16025$$

Next, we calculate the t-statistic:

$$t = (\text{sample mean} - \text{hypothesized mean}) / (\text{sample standard deviation} / \sqrt{\text{sample size}})$$

$$= (301 - 300) / (2.16025 / \sqrt{7}) = 1.5432$$

The degree of freedom for this test is  $n - 1 = 7 - 1 = 6$ .

Now, we look for the P-value associated with the t-statistic of approximately 1.5432 and the degree of freedom of 6.

From the given options, the closest value is  $P(t > 1.54)$  with  $df = 6$ , which is approximately 0.089. However, it is important to note that we are conducting a two-tailed test, so the P-value we need is actually 2 times this value. Therefore, the correct answer is  $2P(t > 1.54)$  with  $df = 6$ .

Hence, the answer is  $\boxed{\text{B. } 2P(t > 1.54) \text{ with } df = 6)}$ .

\*\*\*\*\*

Answer Option 6:

To test the manufacturer's claim, we can perform a one-sample t-test. The null hypothesis ( $H_0$ ) is that the true mean weight of the tablets is 300 milligrams, and the alternative hypothesis ( $H_1$ ) is that the true mean weight is different from 300 milligrams.

To calculate the test statistic, we first need to calculate the sample mean and sample standard deviation. The sample mean is the average of the observed weights, which is  $(299 + 300 + 305 + 302 + 299 + 301 + 303) / 7 = 301$ . The sample standard deviation is the square root of the sample variance, which is calculated as:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

where  $x_i$ s are the observed weights and  $\bar{x}$  is the sample mean. In this case, the sample standard deviation is approximately 2.138.

The t-statistic is then calculated as:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where  $\mu_0$  is the hypothesized mean (300 milligrams in this case),  $\bar{x}$  is the sample mean,  $s$  is the sample standard deviation, and  $n$  is the sample size. In this case, the t-statistic is approximately 1.54.

To obtain the P-value for this two-tailed test, we can compare the t-statistic to the t-distribution with degrees of freedom equal to the sample size minus 1 (6 in this case). The P-value is the probability of observing a t-statistic as extreme or more extreme than the one calculated.

Therefore, the correct answer is  $\boxed{\text{P}(t > 1.54) \text{ with } df = 6)}$ . Final Answer: A

\*\*\*\*\*

Answer Option 7:

To calculate the P-value for a t-test, we first calculate the t-statistic. The t-statistic is given by:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

where  $\bar{X}$  is the sample mean,  $\mu$  is the population mean (claimed by the manufacturer),  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case, the sample mean is  $\bar{X} = \frac{1}{7}(299 + 300 + 305 + 302 + 299 + 301 + 303) = 301$ , the population mean is  $\mu = 300$ , the sample standard deviation is  $s = \sqrt{\frac{1}{6}((299 - 301)^2 + (300 - 301)^2 + (305 - 301)^2 + (302 - 301)^2 + (299 - 301)^2 + (301 - 301)^2 + (303 - 301)^2)} \approx 2.16$ , and the sample size is  $n = 7$ .

Plugging in these values into the formula, we get:

$$t = \frac{301 - 300}{2.16/\sqrt{7}} \approx 1.54$$

Now, we need to find the P-value of this test. The P-value represents the probability of observing a t-statistic as extreme as the one we calculated (or more extreme) under the assumption that the null hypothesis is true. In this case, the null hypothesis is that the average weight of the tablets is 300 milligrams (as claimed by the manufacturer).

Since our alternative hypothesis is not specified, we assume a two-tailed test. This means we calculate the probability of observing a t-statistic as extreme as the one we calculated in both tails of the t-distribution.

Using the t-distribution with degrees of freedom equal to the sample size minus 1 (in this case, 6), we find that the P-value is approximately 0.145.

Therefore, the correct answer is  $\boxed{\text{P}(2P(t > 1.54) \text{ with } df = 6)}$ . Final answer: B

\*\*\*\*\*

Answer Option 8:

To conduct a hypothesis test, we can use a one-sample t-test. The null hypothesis is that the mean weight of the tablets is 300 milligrams, and the alternative hypothesis is that the mean weight is different from 300 milligrams.

To calculate the test statistic, we can use the formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Where:

-  $\bar{x}$  is the sample mean weight

-  $\mu$  is the population mean weight (in this case, 300 milligrams)

-  $s$  is the sample standard deviation

-  $n$  is the sample size

Given the sample weights: 299, 300, 305, 302, 299, 301, and 303:

$$\bar{x} = \frac{299 + 300 + 305 + 302 + 299 + 301 + 303}{7} = 301$$

The sample standard deviation can be calculated using the formula:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

[

$$s = \sqrt{\frac{(299-301)^2 + (300-301)^2 + (305-301)^2 + (302-301)^2 + (299-301)^2 + (301-301)^2 + (303-301)^2}{7-1}} = \sqrt{\frac{14}{6}}$$

]

Now we can calculate the t-value:

$$t = \frac{301 - 300}{\sqrt{\frac{14}{6}}/\sqrt{7}} = \frac{\sqrt{42}}{\sqrt{14}} = \sqrt{3}$$

We have a t-distribution with  $(n-1 = 6)$  degrees of freedom. To find the p-value for a two-tailed test, we need to find  $P(t > |t|)$ .

Looking at the answer choices, the correct option is  $P(t > 1.54)$  with  $df = 6$ . This option corresponds to  $P(t > \sqrt{3})$  with  $(df = 6)$ , which matches the calculated t-value. Final

Answer: A

\*\*\*\*\*

Answer Option 9:

To perform a hypothesis test, we first set up the null and alternative hypotheses. Let  $\mu$  be the true mean weight of an aspirin tablet. The null hypothesis  $H_0$  is that the mean weight is equal to the labeled weight, so  $H_0: \mu = 300$ . The alternative hypothesis  $H_1$  is that the mean weight is not equal to the labeled weight, so  $H_1: \mu \neq 300$ .

Next, we compute the test statistic. We can use a t-test because the sample size is small and the population standard deviation is unknown. The test statistic is given by

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where  $\bar{x}$  is the sample mean,  $\mu_0$  is the hypothesized mean under the null hypothesis,  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case,  $\bar{x} = (299 + 300 + 305 + 302 + 299 + 301 + 303)/7 = 302$ ,  $\mu_0 = 300$ ,  $s = \sqrt{\frac{1}{6} \sum_{i=1}^7 (x_i - \bar{x})^2} \approx 2.08$ , and  $n = 7$ . Therefore, the test statistic is

$$t = \frac{302 - 300}{2.08/\sqrt{7}} \approx 1.54$$

Finally, we calculate the p-value. The p-value is the probability of observing a test statistic more extreme than the observed test statistic, assuming that the null hypothesis is true. Since this is a two-sided test,

we want to calculate  $P(t > 1.54)$ . We look this up in the t-distribution table with degrees of freedom equal to  $n - 1 = 7 - 1 = 6$ . The p-value is the probability in the upper tail of the t-distribution, so the p-value is equal to  $P(t > 1.54)$ .

Therefore, the correct answer is  $P(t > 1.54)$  with  $df = 6$ . Final Answer: A

\*\*\*\*\*

Question 2:

The waiting times for a new roller coaster ride are normally distributed with a mean of 35 minutes and a standard deviation of 10 minutes. If there are 150,000 riders the first summer, which of the following is the shortest time interval associated with 100,000 riders?, A. '0 to 31.7 minutes', B. '31.7 to 39.3 minutes', C. '25.3 to 44.7 minutes', D. '25.3 to 35 minutes'

=====

Answer Option 1:

The mean waiting time for the roller coaster ride is 35 minutes, and the standard deviation is 10 minutes. Since the distribution is normal, we can use the empirical rule to approximate the shortest time interval associated with 100,000 riders.

According to the empirical rule, approximately 68% of the data falls within one standard deviation of the mean, approximately 95% falls within two standard deviations, and approximately 99.7% falls within three standard deviations.

For 100,000 riders, which is approximately 66.7% of the total riders, we can estimate that the waiting time falls within one standard deviation of the mean. This would be a time interval of 25.3 to 44.7 minutes.

Therefore, the correct answer is  $\boxed{\text{C. '25.3 to 44.7 minutes'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

The distribution of waiting times for the roller coaster ride is normally distributed with a mean of 35 minutes and a standard deviation of 10 minutes. To find the shortest time interval associated with 100,000 riders, we need to find the time interval that contains the middle 100,000 riders in terms of waiting times.

Since the distribution is symmetric, we know that the middle 100,000 riders will be contained within one standard deviation of the mean on both sides.

One standard deviation above the mean is  $\$35 + 10 = 45\$$  minutes, and one standard deviation below the mean is  $\$35 - 10 = 25\$$  minutes.

Therefore, the shortest time interval associated with 100,000 riders is from 25.3 to 44.7 minutes.

The correct answer is  $\boxed{\text{C. '25.3 to 44.7 minutes'}}$ .

\*\*\*\*\*

Answer Option 3:

To find the time interval associated with 100,000 riders, we need to find the z-score corresponding to the proportion of riders. The proportion of riders is given by  $\frac{100,000}{150,000} = \frac{2}{3}$ . We can then use the z-score formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value we want to find the z-score for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Substituting the given values, we have  $\frac{x - 35}{10} = \frac{2}{3}$ . Solving for  $x$ , we find  $x = 31.7$ .

Therefore, the shortest time interval associated with 100,000 riders is from 31.7 to 39.3 minutes. The correct answer is  $\boxed{\text{B. '31.7 to 39.3 minutes'}}$ .

\*\*\*\*\*

Answer Option 4:

To find the shortest time interval associated with 100,000 riders, we need to find the range of waiting times within which the middle 100,000 riders fall.

For a normal distribution, approximately 68% of the data falls within 1 standard deviation of the mean, approximately 95% falls within 2 standard deviations, and approximately 99.7% falls within 3 standard deviations.

In this case, with a mean of 35 minutes and a standard deviation of 10 minutes, 68% of the data falls between 25 minutes and 45 minutes (35 minutes plus/minus 10 minutes).

Therefore, the shortest time interval associated with 100,000 riders is from 25 minutes to 45 minutes.

Among the given choices, the interval that includes this range is '25.3 to 44.7 minutes', which corresponds to option C.

So the answer is  $\boxed{\text{C. '25.3 to 44.7 minutes'}}$ .

\*\*\*\*\*

Answer Option 5:

To find the shortest time interval associated with 100,000 riders, we need to determine the range of waiting times that contains 100,000 riders.

Since the waiting times are normally distributed with a mean of 35 minutes and a standard deviation of 10 minutes, we can use the empirical rule to estimate the range. According to the empirical rule, approximately 68% of the data falls within one standard deviation of the mean, 95% falls within two standard deviations, and 99.7% falls within three standard deviations.

Since we want to find the interval containing 100,000 riders, which is approximately  $\frac{2}{3}$  of the total riders, we need to find the range that contains approximately  $\frac{2}{3}$  of the data. This corresponds to two standard deviations from the mean.

Two standard deviations below the mean would be  $\$35 - 2(10) = 15\$$  minutes, and two standard deviations above the mean would be  $\$35 + 2(10) = 55\$$  minutes.

Therefore, the shortest time interval associated with 100,000 riders is from 15 minutes to 55 minutes, which is equivalent to '25.3 to 44.7 minutes'. The correct answer is  $\boxed{\text{C}}$ . Final answer: C

\*\*\*\*\*

Answer Option 6:

To find the shortest time interval associated with 100,000 riders, we need to find the z-score corresponding to that percentile. Using the standard normal distribution, we find that the z-score corresponding to the 100,000th percentile is approximately -1.28.

To find the corresponding time interval, we can use the formula for z-score:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to find,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Rearranging the formula, we have:

$$x = z \cdot \sigma + \mu$$

Substituting the values, we have:

$$\boxed{x = -1.28 \cdot 10 + 35 = 22.2.}$$

Therefore, the shortest time interval associated with 100,000 riders is from 0 to 22.2 minutes. Among the given options, the closest interval is "A. 0 to 31.7 minutes," which contains the shortest time interval.

Thus, the answer is  $\boxed{\text{A. 0 to 31.7 minutes}}$ . Final answer: A

\*\*\*\*\*

Answer Option 7:

To find the shortest time interval associated with 100,000 riders, we need to find the z-score corresponding to 100,000 riders and then use the z-score to find the corresponding time interval.

The z-score is given by:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the number of riders,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For 100,000 riders:

$$z = \frac{100,000 - 150,000 \cdot 0.5}{\sqrt{150,000 \cdot 0.5 \cdot 0.5}} \approx -3.87$$

Using a standard normal distribution table or calculator, we find that the area to the left of  $z = -3.87$  is approximately 0.00007. This means that the area to the right of  $z = -3.87$  is approximately  $1 - 0.00007 = 0.99993$ .

To find the corresponding time interval, we look up the z-score for 0.99993 in the standard normal distribution table. We find that the z-score is approximately 3.05.

The interval associated with 100,000 riders is therefore from  $35 - 3.05 \cdot 10$  to  $35 + 3.05 \cdot 10$  minutes.

This interval is approximately from 25.3 to 44.7 minutes.

Therefore, the answer is  $\boxed{\text{C. 25.3 to 44.7 minutes}}$ .

\*\*\*\*\*

Answer Option 8:

The waiting times for the roller coaster ride are normally distributed with a mean of 35 minutes and a standard deviation of 10 minutes. To find the shortest time interval associated with 100,000 riders, we need to calculate the z-score for this interval.

The formula to calculate the z-score is given by:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to find the z-score for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For the interval associated with 100,000 riders, we want to find the z-score for the value  $x$  in the equation:

$$100,000 = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{x - \mu}{\sigma \sqrt{2}} \right) \right)$$

where  $\text{erf}$  is the error function.

Solving for  $x$ , we find:

$$x = \mu + \sigma \sqrt{2} \cdot \text{erf}^{-1} \left( 2 \left( \frac{100,000}{150,000} - \frac{1}{2} \right) \right)$$

where  $\text{erf}^{-1}$  is the inverse error function.

Plugging in the given values, we have:

$$x = 35 + 10 \sqrt{2} \cdot \text{erf}^{-1} \left( \frac{1}{3} \right) \approx 31.7 \text{ minutes}$$

Therefore, the shortest time interval associated with 100,000 riders is from 0 to 31.7 minutes, which corresponds to option  $\boxed{\text{A}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 9:

We can use the empirical rule, also known as the 68-95-99.7 rule, to estimate the shortest time interval associated with 100,000 riders. According to the empirical rule, in a normal distribution, approximately 68% of the data falls within one standard deviation of the mean, approximately 95% falls within two standard deviations, and approximately 99.7% falls within three standard deviations.

Since there are 150,000 riders, approximately 68% of them, or 102,000 riders, will have waiting times within one standard deviation of the mean. Similarly, approximately 95% of the riders, or 142,500 riders, will have waiting times within two standard deviations of the mean.

To find the shortest time interval associated with 100,000 riders, we need to find the waiting time range that covers at least 100,000 riders. Since 102,000 riders fall within one standard deviation, and 142,500 riders fall within two standard deviations, we can conclude that the range from one standard deviation below the mean to two standard deviations above the mean will include at least 100,000 riders.

One standard deviation below the mean is  $35 - 10 = 25$  minutes, and two standard deviations above the mean is  $35 + 2 \cdot 10 = 55$  minutes. Therefore, the shortest time interval associated with 100,000 riders is from 25 to 55 minutes.

Among the given options, the only one that includes the range from 25 to 55 minutes is option  $\boxed{\text{C. 25.3 to 44.7 minutes}}$ . Final answer: C

\*\*\*\*\*

Question 3:

'All of the following statements are true for all discrete random variables except for which one?', A. 'The possible outcomes must all be numerical.', B. 'The possible outcomes must be mutually exclusive.', C. 'The mean (expected value) always equals the sum of the products obtained by multiplying each value by its corresponding probability.', D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.'

=====

Answer Option 1:

The statement that is not true for all discrete random variables is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This statement is true for a normal distribution, but it may not hold for all discrete random variables. Therefore, the answer is  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

The statement that is not true for all discrete random variables is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.'

This statement is only true for random variables that follow a normal distribution. For other types of distributions, the percentage of outcomes within two standard deviations of the mean may be different.

Therefore, the correct answer is  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

The statement that is not true for all discrete random variables is statement D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This statement is only true for random variables that follow a normal distribution. It does not hold for all discrete random variables. Therefore, the correct answer is  $\boxed{\text{D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.'}}$  Final Answer: D

\*\*\*\*\*

Answer Option 4:

The statement that is not true for all discrete random variables is:

D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.'

This statement is true for a normal distribution, but it may not be true for other distributions. The percentage of outcomes within two standard deviations of the mean can vary depending on the shape of the distribution. Final answer: D

\*\*\*\*\*

Answer Option 5:

The statement that is not true for all discrete random variables is statement D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.'

This statement is true only for random variables that follow a normal distribution. For other types of distributions, the percentage of outcomes within two standard deviations of the mean may be different.

Therefore, statement D is the correct answer. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The statement that is not true for all discrete random variables is:

D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.'

This statement is true for continuous random variables that have a normal distribution, but it may not hold for all discrete random variables. Discrete random variables can have different distributions and shapes, so there may be cases where the outcomes do not follow the 68-95-99.7 rule of a normal distribution. Therefore, statement D is the answer.

\*\*\*\*\*

Answer Option 7:

The correct answer is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This statement is true for continuous random variables, but not necessarily for discrete random variables. In the case of discrete random variables, the probability mass function assigns probabilities to specific outcomes, and there may be gaps between the outcomes. As a result, the distribution may not exhibit the same kind of symmetry and concentration of outcomes within two standard deviations as a continuous distribution would. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The statement that is not true for all discrete random variables is statement D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This statement is only true for normal distributions. For other discrete random variables, the distribution may not be symmetrical or follow the typical bell-shaped curve, so the percentage of outcomes within two standard deviations of the mean may be different. Therefore, the correct answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 9:

The statement that is not true for all discrete random variables is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.'

This statement is true for a normal distribution, but it is not necessarily true for all discrete random variables. The distribution of a discrete random variable can take various shapes, and the percentage of outcomes within two standard deviations of the mean can vary depending on the distribution. Therefore, statement D is not universally true for all discrete random variables.

The correct answer is D. Final answer: D

\*\*\*\*\*

Question 4:

'Which of the following is a true statement about hypothesis testing?', A. 'If there is sufficient evidence to reject a null hypothesis at the 10% level, then there is sufficient evidence to reject it at the 5% level.', B. 'Whether to use a one- or a two-sided test is typically decided after the data are gathered.', C. 'If a hypothesis test is conducted at the 1% level, there is a 1% chance of rejecting the null hypothesis.', D. 'The power of a test concerns its ability to detect an alternative hypothesis.'

=====

Answer Option 1:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. It measures the sensitivity of the test to detect a true effect or difference when it exists. Final answer: D

\*\*\*\*\*

Answer Option 2:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a statistical test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. It measures the ability of the test to detect a true effect or relationship. Therefore, option D is the correct statement. Final answer: D.

\*\*\*\*\*

Answer Option 3:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. This means that a test with high power is more likely to correctly identify a true alternative hypothesis.

Statement A is incorrect because the level of significance determines the probability of rejecting the null hypothesis when it is true, not the other way around. Statement B is incorrect because the decision to use a one- or two-sided test is typically made before the data are gathered, based on the specific research question. Statement C is incorrect because the level of significance represents the probability of rejecting the null hypothesis when it is true, not the probability of rejecting it.

Therefore, the correct statement is D.

\*\*\*\*\*

Answer Option 4:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.'

A is incorrect because if we have sufficient evidence to reject the null hypothesis at the 10% level, it does not necessarily mean we have sufficient evidence to reject it at the 5% level. The significance level determines the level of confidence required to reject the null hypothesis, and a lower significance level requires stronger evidence.

B is incorrect because the decision to use a one- or two-sided test is typically determined before the data is gathered. It depends on the nature of the hypothesis and the research question.

C is incorrect because the significance level of a hypothesis test represents the probability of incorrectly rejecting the null hypothesis when it is true. It does not represent the probability of rejecting the null hypothesis.

Therefore, the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. It measures the test's ability to detect a true effect or difference.

Statement A is false because the level of significance determines the threshold for rejecting the null hypothesis, and a lower significance level means stronger evidence is required to reject the null hypothesis.

Statement B is false because the decision to use a one- or two-sided test should be made based on the research question and the nature of the alternative hypothesis before data collection.

Statement C is false because the significance level (e.g., 1%) is the probability of rejecting the null hypothesis when it is true, not the probability of rejecting the null hypothesis overall.

Therefore, the correct answer is D. Final answer: D

\*\*\*\*\*

Answer Option 6:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a statistical test is the probability that it correctly rejects the null hypothesis when the alternative hypothesis is true. It measures the test's sensitivity to detect a true effect or relationship. Option A is incorrect because the level of significance (e.g., 10% or 5%) determines the threshold for rejecting the null hypothesis, but it does not guarantee that the evidence is sufficient at a different level. Option B is incorrect because the decision to use a one- or two-sided test is typically determined before data collection, based on the research question and the nature of the hypothesis. Option C is incorrect because the significance level (e.g., 1%) represents the probability of making a Type I error (rejecting the null hypothesis when it is true), not the chance of rejecting the null hypothesis. Thus, the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.'

The power of a test refers to the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In other words, it measures the test's ability to detect a true effect or difference. It is an important consideration in hypothesis testing as it helps determine the reliability and sensitivity of the test.

Statement A is incorrect because the ability to reject a null hypothesis at a certain level of significance does not imply the ability to reject it at a lower level of significance. The decision to reject or fail to reject the null hypothesis depends on the evidence and the chosen level of significance.

Statement B is incorrect because the decision to use a one- or two-sided test is typically made before collecting the data, based on the research question and the nature of the hypothesis being tested.

Statement C is incorrect because the 1% level refers to the level of significance, which is the probability of incorrectly rejecting the null hypothesis when it is true. It does not represent the probability of rejecting the null hypothesis overall.

Therefore, the correct statement is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' Final Answer: D

\*\*\*\*\*

Answer Option 8:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. It is a measure of how well a test can detect the presence of an effect or difference when it truly exists.

Option A is incorrect because the level of significance determines the threshold for rejecting the null hypothesis, and a lower level of significance (e.g., 5% instead of 10%) means stronger evidence is required to reject the null hypothesis.

Option B is incorrect because the decision to use a one- or two-sided test is typically made based on the nature of the research question and the alternative hypothesis before data collection.

Option C is incorrect because the significance level (e.g., 1%) is the probability of rejecting the null hypothesis when it is true, not the chance of rejecting it. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The correct statement about hypothesis testing is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. It quantifies the test's ability to detect a true effect or relationship.

Statement A is incorrect because the significance level (e.g., 10% or 5%) determines the probability of rejecting the null hypothesis when it is true. If there is sufficient evidence to reject the null hypothesis at the 10% level, it does not necessarily mean there is sufficient evidence to reject it at the 5% level.

Statement B is incorrect because the decision to use a one- or two-sided test should be determined before data collection based on the research question and nature of the hypotheses.

Statement C is incorrect because the significance level (e.g., 1%) determines the probability of rejecting the null hypothesis when it is true, not the chance of rejecting the null hypothesis. The chance depends on the specific data and the power of the test.

Therefore, the correct statement is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' Final Answer: D

\*\*\*\*\*



Question 5:

'An outlier is an observation that', A. 'is seen more frequently than the other observations in the data set', B. 'is seen less frequently than the other observations in the data set', C. 'is always smaller than the other observations in the data set', D. 'is significantly different from the other observations in the data set'

=====

Answer Option 1:

An outlier is an observation that is significantly different from the other observations in the data set. Therefore, the correct answer is  $\boxed{\text{D. 'is significantly different from the other observations in the data set'}}$ .

\*\*\*\*\*

Answer Option 2:

An outlier is an observation that is significantly different from the other observations in the data set. Therefore, the correct answer is D. "is significantly different from the other observations in the data set."

Final Answer: D

\*\*\*\*\*

Answer Option 3:

An outlier is an observation that is significantly different from the other observations in the data set. Therefore, the correct answer is  $\boxed{\text{D. 'is significantly different from the other observations in the data set'}}$ . Final answer: D.

\*\*\*\*\*

Answer Option 4:

An outlier is an observation that is significantly different from the other observations in the data set. Therefore, the correct answer is  $\boxed{\text{D. 'is significantly different from the other observations in the data set'}}$ .

\*\*\*\*\*

Answer Option 5:

An outlier is an observation that is significantly different from the other observations in the data set. Thus, the correct answer is  $\boxed{\text{D. 'is significantly different from the other observations in the data set'}}$ .

\*\*\*\*\*

Answer Option 6:

An outlier is an observation that is significantly different from the other observations in the data set. This means that the correct answer is  $\boxed{\text{D. 'is significantly different from the other observations in the data set'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

An outlier is an observation that is significantly different from the other observations in the data set. Therefore, the correct answer is  $\boxed{\text{D. 'is significantly different from the other observations in the data set'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

An outlier is an observation that is significantly different from the other observations in the data set. Therefore, the correct answer is  $\boxed{\text{D. 'is significantly different from the other observations in the data set'}}$ .

\*\*\*\*\*

Answer Option 9:

An outlier is an observation that is significantly different from the other observations in the data set. Thus, the correct answer is  $\boxed{\text{D. 'is significantly different from the other observations in the data set'}}$ . Final answer: D

\*\*\*\*\*

Question 6:

The principal of a school is interested in estimating the average income per family of her students. She selects a random sample of students and collects information about their family income. A 95 percent confidence interval computed from this data for the mean income per family is (\$35,095, \$45,005). Which of the following provides the best interpretation of this confidence interval?, A. '95 percent of the students in her school are from families whose income is between \$35,095 and \$45,005.', B. 'There is a 95% probability that the families of all the students in this school have an income of between \$35,095 and \$45,005.', C. 'If we were to take another sample of the same size and compute a 95 percent confidence interval, we would have a 95% chance of getting the interval (\$35,095, \$45,005).', D. 'There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005.'

=====

Answer Option 1:

The correct interpretation of the given confidence interval is: 'If we were to take another sample of the same size and compute a 95 percent confidence interval, we would have a 95% chance of getting the interval (\$35,095, \$45,005).' Therefore, the answer is  $\boxed{\text{C}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

The correct interpretation of a confidence interval is option D: "There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005." This means that if we were to repeat the sampling process and compute a new confidence interval, we would expect the true population mean income per family to fall within that interval 95% of the time. It does not imply anything about individual students or their families. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The best interpretation of a confidence interval is option D: "There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005."

This means that if we were to repeat the sampling process multiple times and compute a confidence interval each time, approximately 95% of these intervals would contain the true mean income per family of the students in the school. However, it is important to note that this probability refers to the method of constructing confidence intervals, not the probability of the true mean falling within the specific interval obtained from a single sample. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The correct interpretation of a confidence interval is option D: "There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005." This means that if we were to repeat the sampling process and compute a 95% confidence interval each time, approximately 95% of those intervals would contain the true average income per family of the students in the school. It does not mean that 95% of the students have family incomes within that range, nor does it imply anything about the probability distribution of individual family incomes. Therefore, the best interpretation is option D. Final

Answer: D

\*\*\*\*\*

Answer Option 5:

The best interpretation of the confidence interval is option C. 'If we were to take another sample of the same size and compute a 95 percent confidence interval, we would have a 95% chance of getting the interval (\$35,095, \$45,005).'

A confidence interval does not provide information about individual students, so option A is incorrect. Option B implies a probability about the families of all the students, which is not accurate because the confidence interval is about the population mean, not individual families. Option D is also incorrect because it refers to the probability of the true mean income falling within the interval, which is not the correct interpretation of the confidence interval.

Therefore, the best interpretation is option C. Final Answer: C

\*\*\*\*\*

Answer Option 6:

The correct interpretation of a 95 percent confidence interval is option D: 'There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005.'

A confidence interval gives a range of values within which the true population parameter is likely to fall. In this case, the 95 percent confidence interval of (\$35,095, \$45,005) means that if we were to take multiple random samples from the population and compute a 95 percent confidence interval for each sample, approximately 95 percent of those intervals would contain the true population mean income per family.

Option A is incorrect because the confidence interval is about the population mean, not the individual students. Option B is incorrect because the confidence interval is about the population mean, not the individual families. Option C is incorrect because it implies a probability about future intervals, whereas confidence intervals are about the current interval constructed from a specific sample. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct interpretation of the confidence interval is option D: 'There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005.'

A confidence interval is a range of values within which we can be confident, with a certain level of probability, that the true parameter lies. In this case, the true mean income per family of the students is estimated to be between \$35,095 and \$45,005 with a 95% probability.

Option A is incorrect because the confidence interval is not about individual students but about the mean income per family of the students.

Option B is incorrect because the probability is associated with the interval, not with the families of all the students.

Option C is incorrect because the probability is not about getting the same interval in future samples, but about the current interval capturing the true mean income per family in the school.

Therefore, the best interpretation is option D. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The best interpretation of the given confidence interval is option A: '95 percent of the students in her school are from families whose income is between \$35,095 and \$45,005.'

A confidence interval gives a range of values within which we can be confident that the true population parameter lies. In this case, the 95% confidence interval (\$35,095, \$45,005) suggests that we can estimate with 95% confidence that the average income per family of the students in the school falls within this range.

Option B is incorrect because the confidence interval is about the population parameter, not about individual families. Additionally, probability statements cannot be made about fixed parameters.

Option C is incorrect because the confidence interval is specific to the current sample and does not guarantee that future intervals will match exactly.

Option D is incorrect because the confidence interval is about the population parameter, not about the probability of the parameter falling within a certain range. Final answer: A

\*\*\*\*\*

Answer Option 9:

The best interpretation of the given confidence interval is option D: 'There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005.'

A confidence interval does not provide a probability or certainty about the true parameter value, but rather it provides a range where the true parameter value is likely to fall. In this case, the 95% confidence interval suggests that if we were to repeat the sampling process multiple times, 95% of the resulting confidence intervals would contain the true mean income per family.

Option A is incorrect because the confidence interval is not about individual students but about the mean income per family.

Option B is incorrect because a confidence interval does not represent a probability of a single event or parameter value. It represents the uncertainty in estimating the true parameter value.

Option C is incorrect because it suggests that the probability is associated with getting the same interval in another sample, but the probability is associated with the confidence that the true parameter value lies within the given interval.

\*\*\*\*\*

Question 7:

'Researchers are conducting an experiment using a significance level of 0.05. The null hypothesis is, in fact, false. If they modify their experiment to use twice as many experimental units for each treatment, which of the following would be true?', A. 'The probability of a Type I error and the probability of a Type II error would both decrease.', B. 'The probability of a Type I error and the power would both increase.', C. 'The probability of a Type II error and the power would both increase.', D. 'The probability of a Type I error would stay the same and the power would increase.'

=====

Answer Option 1:

When researchers modify their experiment to use twice as many experimental units for each treatment, the power of the test increases. The power of a statistical test is the probability of correctly rejecting a false null hypothesis. By increasing the sample size, the test becomes more sensitive and has a higher chance of detecting a true effect if it exists. Therefore, the correct statement is that the probability of a Type I error would stay the same (since the significance level is not changed) and the power would increase. Hence, the answer is  $\boxed{\text{D. 'The probability of a Type I error would stay the same and the power would increase.'}}$

\*\*\*\*\*

Answer Option 2:

When the researchers modify their experiment to use twice as many experimental units for each treatment, the power of the test would increase. This is because increasing the sample size generally leads to a higher chance of detecting a true effect if it exists.

However, the probability of a Type I error, which is the probability of rejecting the null hypothesis when it is true, would stay the same. The significance level, or the probability of making a Type I error, is predetermined and does not change with the sample size.

Therefore, the answer is  $\boxed{\text{D. 'The probability of a Type I error would stay the same and the power would increase.'}}$

\*\*\*\*\*

Answer Option 3:

When the researchers modify their experiment to use twice as many experimental units for each treatment, the power of the test would increase. The power of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. By increasing the sample size, the test has a higher chance of detecting a true effect and rejecting the null hypothesis.

However, the significance level, which represents the probability of a Type I error, would stay the same. The significance level is determined before the experiment and is not affected by changes in sample size.

Therefore, the correct answer is  $\boxed{\text{D. 'The probability of a Type I error would stay the same and the power would increase.'}}$  Final answer: D

\*\*\*\*\*

Answer Option 4:

The significance level, denoted by  $\alpha$ , represents the probability of making a Type I error, which is rejecting the null hypothesis when it is actually true. Increasing the sample size does not directly affect the significance level. Therefore, the probability of a Type I error would stay the same.

However, increasing the sample size can have an effect on the power of the test, which is the probability of correctly rejecting the null hypothesis when it is false (i.e., avoiding a Type II error). Increasing the sample size generally increases the power of the test, as it provides a larger sample to detect any true effects.

Therefore, the correct answer is  $\boxed{\text{D. 'The probability of a Type I error would stay the same and the power would increase.'}}$  Final Answer: D

\*\*\*\*\*

Answer Option 5:

When the null hypothesis is false, increasing the sample size generally increases the power of the statistical test. The power of a statistical test is the probability of correctly rejecting the null hypothesis when it is in fact false. Thus, we can eliminate options B and D.

When the sample size increases, the probability of a Type I error (rejecting the null hypothesis when it is true) typically remains the same, assuming the significance level is held constant. Therefore, option A can also be eliminated.

This leaves us with option C: 'The probability of a Type II error and the power would both increase.' When the sample size increases, the probability of a Type II error (failing to reject the null hypothesis when it is false) tends to decrease, leading to an increase in power.

Therefore, the correct answer is option C. Final answer: C

\*\*\*\*\*

Answer Option 6:

The significance level, denoted as  $\alpha$ , is the probability of committing a Type I error, which is rejecting the null hypothesis when it is actually true. If the researchers modify their experiment to use twice as many experimental units for each treatment, the probability of committing a Type I error would remain the same. This is because the significance level is determined by the chosen level of significance at the start of the experiment and is not affected by the sample size.

However, increasing the sample size would likely increase the power of the test. Power is the probability of correctly rejecting the null hypothesis when it is indeed false. With a larger sample size, the test would have a greater ability to detect a true effect, thus increasing the power.

Therefore, the correct answer is  $\boxed{\text{D. 'The probability of a Type I error would stay the same and the power would increase.'}}$

\*\*\*\*\*

Answer Option 7:

When researchers conduct a hypothesis test, they set a significance level, which is the maximum probability of making a Type I error (rejecting the null hypothesis when it is true). In this scenario, the null hypothesis is false, which means that there is a true effect or difference between the treatments being compared.

If the researchers modify their experiment to use twice as many experimental units for each treatment, this would increase the sample size. Increasing the sample size improves the power of the test, which is the probability of correctly rejecting the null hypothesis when it is false. Therefore, the power would increase.

However, the significance level (probability of Type I error) is predetermined and does not change based on the sample size. It remains at 0.05.

Therefore, the correct answer is D. The probability of a Type I error would stay the same, and the power would increase. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The significance level, denoted by  $\alpha$ , represents the probability of a Type I error, which is rejecting the null hypothesis when it is actually true. If the researchers modify their experiment to use twice as many experimental units for each treatment, and all other conditions remain the same, the probability of a Type I error,  $\alpha$ , would stay the same. This is because the significance level is determined by the researchers and is not affected by the sample size.

On the other hand, increasing the sample size would generally increase the power of the test. Power is the probability of correctly rejecting the null hypothesis when it is false, and it depends on factors such as the effect size and sample size. With a larger sample size, the test becomes more sensitive to detecting a true effect, increasing the power.

Therefore, the correct answer is  $\boxed{\text{D. The probability of a Type I error would stay the same and the power would increase.}}$  Final Answer: D

\*\*\*\*\*

Answer Option 9:

The significance level, denoted by  $\alpha$ , represents the probability of making a Type I error, which is rejecting the null hypothesis when it is actually true. In this case, the significance level is 0.05, so the probability of a Type I error is 0.05.

When the experiment is modified to use twice as many experimental units, the power of the test increases. Power, denoted by  $1 - \beta$ , represents the probability of correctly rejecting the null hypothesis when it is false. Increasing the sample size generally leads to increased power.

However, the probability of a Type II error, denoted by  $\beta$ , represents the probability of failing to reject the null hypothesis when it is false. Increasing the sample size does not directly affect the probability of a Type II error.

Given this information, the correct answer is  $\boxed{\text{D. The probability of a Type I error would stay the same and the power would increase.}}$  Final Answer: D

\*\*\*\*\*

Question 8:

'The mean height of adult men is 70 inches, with a standard deviation of four inches. The mean height of adult women is 66 inches, with a standard deviation of three inches. Between a man with a height of 74 inches and a woman with a height of 70 inches, who is more unusually tall within his or her respective sex?', A. 'It cannot be determined, because the mean heights for the two groups are different.', B. 'It cannot be determined, because the standard deviations of the heights for the two groups are different.', C. 'The man, because he is 74 inches tall and the woman is only 70 inches tall.', D. 'The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.'

=====

Answer Option 1:

We can determine who is more unusually tall within their respective sex by looking at their z-scores.

For the man with a height of 74 inches, his z-score is given by  $\frac{74 - 70}{4} = 1$ .

For the woman with a height of 70 inches, her z-score is given by  $\frac{70 - 66}{3} = 1.33$ .

Since the woman's z-score is greater than the man's z-score, this means that her height is more unusually tall within her respective sex.

Therefore, the answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$

\*\*\*\*\*

Answer Option 2:

To determine who is more unusually tall within their respective sex, we need to compare the heights in terms of standard deviations from the mean.

For the man with a height of 74 inches, we find that his height is  $(74 - 70) / 4 = 1$  standard deviation above the mean height of all men.

For the woman with a height of 70 inches, we find that her height is  $(70 - 66) / 3 = 1.33$  standard deviations above the mean height of all women.

Therefore, the woman is more unusually tall within her respective sex. The correct answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$

\*\*\*\*\*

Answer Option 3:

To determine who is more unusually tall within their respective sex, we need to compare the heights relative to the mean and standard deviation of each group.

For the man with a height of 74 inches:

His height is  $74 - 70 = 4$  inches above the mean height of men.

His height is  $4/4 = 1$  standard deviation above the mean height of men.

For the woman with a height of 70 inches:

Her height is  $70 - 66 = 4$  inches above the mean height of women.

Her height is  $4/3 = 1.33$  standard deviations above the mean height of women.

Since the woman's height is more standard deviations above the mean height of women compared to the man's height above the mean height of men, the woman is more unusually tall within her respective sex. Therefore, the answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$

\*\*\*\*\*

Answer Option 4:

To determine who is more unusually tall within their respective sex, we need to compare how many standard deviations each individual's height is from the mean height of their respective sex.

For the man with a height of 74 inches, we find that his height is  $(74 - 70) / 4 = 1$  standard deviation above the mean height of men.

For the woman with a height of 70 inches, her height is  $(70 - 66) / 3 = 1.33$  standard deviations above the mean height of women.

Since the woman's height is further from the mean height of women in terms of standard deviations, she is more unusually tall within her respective sex. Therefore, the answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$ . Final answer: D

\*\*\*\*\*

Answer Option 5:

To determine who is more unusually tall within their respective sex, we need to compare how far each person's height is from the mean of their respective group, relative to the standard deviation.

For the man with a height of 74 inches, his height is 4 inches above the mean height of all men. Since the standard deviation for men is 4 inches, this means his height is one standard deviation above the mean height of all men.

For the woman with a height of 70 inches, her height is 4 inches above the mean height of all women. Since the standard deviation for women is 3 inches, this means her height is approximately 1.33 standard deviations above the mean height of all women.

Therefore, the woman is more unusually tall within her respective sex. The answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$

\*\*\*\*\*

Answer Option 6:

To determine who is more unusually tall within their respective sex, we need to compare how far each individual's height is from the mean height of their respective sex, relative to the standard deviation of their respective sex.

For the man with a height of 74 inches, his height is 4 inches above the mean height of men, which is one standard deviation above the mean height of men.

For the woman with a height of 70 inches, her height is 4 inches above the mean height of women, but her height is 1.33 standard deviations above the mean height of women.

Since the woman's height is further above the mean height of women relative to the standard deviation of women's heights, she is more unusually tall within her respective sex. Therefore, the answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$

Final answer: D

\*\*\*\*\*

Answer Option 7:

To determine who is more unusually tall within their respective sex, we need to compare their heights relative to the mean heights and standard deviations of their respective groups.

The man's height of 74 inches is 4 inches above the mean height of adult men (70 inches). Since the standard deviation for adult men is 4 inches, the man's height is 1 standard deviation above the mean height for men.

The woman's height of 70 inches is 4 inches above the mean height of adult women (66 inches). Since the standard deviation for adult women is 3 inches, the woman's height is approximately 1.33 standard deviations above the mean height for women.

Therefore, the woman's height is more unusually tall within her respective sex because it is further above the mean height compared to the man's height. The answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$  Final Answer: D

\*\*\*\*\*

Answer Option 8:

To determine who is more unusually tall within their respective sex, we need to compare the number of standard deviations each person's height is from the mean height of their respective sex.

For the man with a height of 74 inches, we calculate the number of standard deviations above the mean:  $\frac{74 - 70}{4} = 1$  standard deviation above the mean height of adult men.

For the woman with a height of 70 inches, we calculate the number of standard deviations above the mean:  $\frac{70 - 66}{3} = 1.33$  standard deviations above the mean height of adult women.

Therefore, the woman is more unusually tall within her respective sex because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men. The answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$

\*\*\*\*\*

Answer Option 9:

To determine who is more unusually tall within their respective sex, we need to compare their heights in terms of standard deviations from the mean height of their respective groups.

For the man with a height of 74 inches, we can calculate his height in terms of standard deviations from the mean height of adult men as follows:  $\frac{74-70}{4}=1$ . This means that the man's height is one standard deviation above the mean height of adult men.

For the woman with a height of 70 inches, we can calculate her height in terms of standard deviations from the mean height of adult women as follows:  $\frac{70-66}{3}=1.33\bar{3}$ . This means that the woman's height is approximately 1.33 standard deviations above the mean height of adult women.

Since the woman's height is further above the mean height of her group in terms of standard deviations, she is more unusually tall within her respective sex. Therefore, the answer is  $\boxed{\text{The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.}}$

\*\*\*\*\*

Question 9:

'Which of the following is not true about t-distributions?', A. 'There are different t-distributions for different values of df (degrees of freedom).', B. 't-distributions are bell-shaped and symmetric.', C. 't-distributions always have mean 0 and standard deviation 1.', D. 't-distributions are more spread out than the normal distribution.'

=====

Answer Option 1:

The correct answer is C. 't-distributions always have mean 0 and standard deviation 1.' This statement is not true about t-distributions. Unlike the standard normal distribution, t-distributions have a mean of 0 only when the degrees of freedom is greater than 1. Additionally, the standard deviation of t-distributions is greater than 1, indicating that t-distributions are more spread out than the standard normal distribution. Therefore, the correct answer is  $\boxed{\text{C}}$

\*\*\*\*\*

Answer Option 2:

The statement that is not true about t-distributions is C. 't-distributions always have mean 0 and standard deviation 1.'

Unlike the standard normal distribution, which always has a mean of 0 and a standard deviation of 1, t-distributions have a mean of 0 but their standard deviation varies depending on the degrees of freedom (df). As the degrees of freedom increase, the t-distribution becomes more similar to the standard normal distribution. Therefore, statement C is not true.

The correct answer is C. Final Answer: C

\*\*\*\*\*

Answer Option 3:

The correct answer is C. 't-distributions always have mean 0 and standard deviation 1.'

In a t-distribution, the mean is always 0, but the standard deviation depends on the degrees of freedom (df). As the degrees of freedom increase, the t-distribution becomes more similar to the standard normal distribution, which has a mean of 0 and a standard deviation of 1.

Therefore, option C is not true about t-distributions. The correct answer is C. Final answer: C

\*\*\*\*\*

Answer Option 4:

The answer is C. 't-distributions always have mean 0 and standard deviation 1.' This statement is not true about t-distributions. In fact, the mean of a t-distribution is always 0, but the standard deviation of a t-distribution depends on the degrees of freedom (df). The standard deviation of a t-distribution is greater than 1 when the degrees of freedom is less than 2, and it approaches 1 as the degrees of freedom increases. Therefore, the correct answer is  $\boxed{\text{C}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 5:

The statement that is not true about t-distributions is C. 't-distributions always have mean 0 and standard deviation 1.'

In a t-distribution, the mean is always 0, but the standard deviation varies with the degrees of freedom (df) parameter. As the degrees of freedom increase, the t-distribution becomes more similar to the standard normal distribution (which has mean 0 and standard deviation 1), but for smaller values of df, the t-distribution has fatter tails and is more spread out compared to the normal distribution.

Therefore, the correct answer is C. 't-distributions always have mean 0 and standard deviation 1.' Final Answer: C

\*\*\*\*\*

Answer Option 6:

The statement that is not true about t-distributions is option C.

C. 't-distributions always have mean 0 and standard deviation 1.'

While it is true that t-distributions are bell-shaped and symmetric (option B) and that there are different t-distributions for different values of degrees of freedom (option A), t-distributions do not always have mean 0 and standard deviation 1 (option C). The mean of a t-distribution is 0 only when the degrees of freedom is greater than 1, and the standard deviation of a t-distribution is greater than 1 when the degrees of freedom is less than 2.

Therefore, the correct answer is option C. Final answer: C

\*\*\*\*\*

Answer Option 7:

The statement that is not true about t-distributions is  $\boxed{\text{C}}$ . 't-distributions always have mean 0 and standard deviation 1.'.

While it is true that t-distributions have different shapes for different values of degrees of freedom (statement A), and that they are bell-shaped and symmetric (statement B), t-distributions do not always have mean 0 and standard deviation 1. The mean and standard deviation of a t-distribution depend on the degrees of freedom.

In fact, the mean of a t-distribution is always 0, but the standard deviation is not always 1. The standard deviation of a t-distribution is greater than 1 for small degrees of freedom, and approaches 1 as the degrees of freedom increase. Therefore, statement C is not true about t-distributions.

Statement D is true about t-distributions. They are more spread out (have larger variability) compared to the normal distribution, especially for smaller sample sizes (degrees of freedom). Final Answer: C

\*\*\*\*\*

Answer Option 8:

The t-distribution is a family of probability distributions that depend on the degrees of freedom (df). Each different value of df corresponds to a different t-distribution, so statement A is true.

Statement B is also true. The t-distribution is bell-shaped and symmetric, similar to the normal distribution.

Statement C is false. Unlike the standard normal distribution, t-distributions do not always have mean 0 and standard deviation 1. The mean of a t-distribution is 0 only when the degrees of freedom are greater than 1. The standard deviation of a t-distribution is greater than 1, and it increases as the degrees of freedom decrease.



Statement D is true. The t-distribution is generally more spread out than the normal distribution. As the degrees of freedom increase, the t-distribution approaches the normal distribution, but for smaller degrees of freedom, it has fatter tails and more spread out.

Therefore, the statement that is not true about t-distributions is  $\boxed{\text{C. 't-distributions always have mean 0 and standard deviation 1.'}}$

\*\*\*\*\*

Answer Option 9:

The correct answer is C. 't-distributions always have mean 0 and standard deviation 1.' is not true about t-distributions.

In general, the mean of a t-distribution is 0 only when the degrees of freedom is greater than 1. For degrees of freedom less than or equal to 1, the mean is undefined. Similarly, the standard deviation of a t-distribution is not always 1. It depends on the degrees of freedom. As the degrees of freedom increase, the t-distribution approaches the standard normal distribution, which has a mean of 0 and standard deviation of 1.

So, the correct statement is that t-distributions do not always have mean 0 and standard deviation 1. The other statements are true: there are different t-distributions for different values of degrees of freedom, t-distributions are bell-shaped and symmetric, and t-distributions are more spread out than the normal distribution.

Therefore, the answer is  $\boxed{\text{C. 't-distributions always have mean 0 and standard deviation 1.'}}$

\*\*\*\*\*

Question 10:

'To study the effect of alcohol on reaction time, subjects were randomly selected and given three beers to consume. Their reaction time to a simple stimulus was measured before and after drinking the alcohol. Which of the following is a correct statement?', A. 'This study was an observational study.', B. 'Lack of blocking makes this a poorly designed study.', C. 'The placebo effect is irrelevant in this type of study.', D. 'This study was an experiment in which the subjects were used as their own controls.'

=====

Answer Option 1:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.' In this study, the subjects were measured before and after drinking alcohol, which allows for a comparison within each subject. By using each subject as their own control, the study design accounts for any individual differences in reaction time and increases the validity of the results. Therefore, option D is the correct statement. Final answer: D

\*\*\*\*\*

Answer Option 2:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.' In this study, the subjects' reaction time was measured before and after drinking alcohol. By comparing the reaction time before and after drinking, the study is able to determine the effect of alcohol on reaction time within each subject. Since each subject serves as their own control (i.e., their reaction time is measured under both conditions), this study design allows for a more accurate comparison of results. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.' In this study, the same subjects were measured for reaction time before and after drinking the alcohol. By comparing each subject's reaction time before and after, the researchers were able to assess the effect of alcohol on reaction time for each individual. This design allows each subject to serve as their own control, as the researchers are comparing the subject's own performance before and after the treatment (drinking alcohol). Final answer: D

\*\*\*\*\*

Answer Option 4:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.'

In this study, the subjects were given three beers to consume, and their reaction time was measured before and after drinking the alcohol. By measuring the reaction time before and after drinking the alcohol, the researchers are comparing the subjects' reaction time under two different conditions: before and after consuming alcohol.

Since the subjects serve as their own controls, this study is an experiment, not an observational study. The lack of blocking may not necessarily make this a poorly designed study, and the placebo effect could still be relevant depending on the specific design and conditions of the study. However, the most accurate statement is that this study is an experiment in which the subjects were used as their own controls. Therefore, the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.'

In this study, the subjects were given three beers to consume, and their reaction time was measured before and after drinking the alcohol. By measuring the reaction time before and after, the study is comparing the subjects' reaction time when they are sober (before drinking) and when they are under the influence of alcohol (after drinking). Since the same subjects are being compared before and after drinking, they serve as their own controls.

Observational studies, on the other hand, do not involve any intervention or manipulation by the researchers. Lack of blocking refers to a design flaw in which subjects are not divided into groups or blocks based on certain characteristics, which may affect the results. The placebo effect refers to the effect of a treatment due to the psychological belief that it will have an effect, which is irrelevant in this study since alcohol is known to have physiological effects.

Therefore, the correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.' Final answer: D.

\*\*\*\*\*

Answer Option 6:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.'

In this study, the subjects were randomly selected and given three beers to consume. Their reaction time was measured before and after drinking the alcohol. By comparing the subjects' reaction time before and after drinking alcohol, the study aims to determine the effect of alcohol on reaction time. In this design, each subject serves as their own control because their reaction time is measured both before and after the intervention (drinking alcohol). This allows for a direct comparison within each subject.

A is incorrect because an observational study involves observing subjects without any intervention or manipulation. In this study, the subjects were given alcohol, which is an intervention.

B is incorrect because blocking refers to grouping subjects based on certain characteristics and then randomly assigning treatments within each group. This study does not involve blocking, but it does have a before-and-after design, which allows for within-subject comparison.

C is incorrect because the placebo effect could still be relevant in this study. Even though the subjects were given alcohol, there could still be psychological or physiological effects attributed to the belief that they consumed alcohol, which could impact their reaction time.

Therefore, the correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.'

\*\*\*\*\*

Answer Option 7:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.'

In this study, the subjects were given alcohol and their reaction time was measured both before and after consuming the alcohol. By comparing the reaction time before and after, the researchers were able to assess the effect of alcohol on reaction time within each subject. This type of study design, where subjects serve as their own controls, is known as a within-subject design or a repeated measures design.

Option A, 'This study was an observational study,' is incorrect because an observational study does not involve experimental interventions like giving subjects alcohol.

Option B, 'Lack of blocking makes this a poorly designed study,' is incorrect because blocking is not necessary in this study design.

Option C, 'The placebo effect is irrelevant in this type of study,' is incorrect because the placebo effect can still be relevant in a study like this, even though alcohol is not typically considered a placebo.

\*\*\*\*\*

Answer Option 8:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.'

In this study, the subjects were given alcohol and their reaction time was measured both before and after consuming the alcohol. By comparing the reaction time before and after, the researchers are able to assess the effect of alcohol on reaction time within each individual subject. This type of study design, where subjects serve as their own controls by comparing their own measurements before and after an intervention, is known as a within-subject design or a repeated measures design. Therefore, statement D is correct.

Statement A, 'This study was an observational study,' is incorrect because an observational study involves observing subjects without intervening or assigning treatments.

Statement B, 'Lack of blocking makes this a poorly designed study,' is not necessarily true based on the information given. Blocking is a technique used to control confounding variables in some study designs, but it may not be necessary or applicable in all study designs.

Statement C, 'The placebo effect is irrelevant in this type of study,' is also incorrect. The placebo effect refers to the phenomenon where individuals may experience a response or improvement simply due to the belief that they are receiving a treatment, even if the treatment is inactive. While the placebo effect may not be directly relevant in a study examining the effect of alcohol on reaction time, it is still important to consider and control for potential placebo effects in any study design.

Therefore, the correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.' Final answer: D

\*\*\*\*\*

Answer Option 9:

The correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.'

In this study, the subjects were given three beers to consume, and their reaction time was measured before and after drinking the alcohol. By measuring the reaction time before and after, the study is able to compare the subjects' reaction time under the influence of alcohol to their reaction time without alcohol, using each subject as their own control. This allows for a more accurate evaluation of the effect of alcohol on reaction time, as it controls for individual differences between subjects.

Option A, 'This study was an observational study,' is incorrect because an observational study involves observing and collecting data without any intervention or manipulation by the researcher. In this study, the subjects were given alcohol, which is an intervention.

Option B, 'Lack of blocking makes this a poorly designed study,' is also incorrect because blocking is a technique used to control for potential confounding variables by grouping subjects with similar characteristics together. While blocking could potentially be used in this study, its absence does not necessarily make it a poorly designed study.

Option C, 'The placebo effect is irrelevant in this type of study,' is incorrect because the placebo effect could still be relevant even though the study involves alcohol. The subjects' perception and expectation of the effects of alcohol could influence their reaction time, so the placebo effect should still be considered.

Therefore, the correct statement is D. 'This study was an experiment in which the subjects were used as their own controls.' Final answer: D.

\*\*\*\*\*

Question 11:

'Which of these statements correctly explains bias?', A. 'It describes a process that creates estimates that are too high on average or too low on average.', B. 'It describes a process that results in a sample that does not represent the population.', C. 'It describes a lack of randomization in the sampling process.', D. 'It describes a sample that is not representative of the population.'

=====

Answer Option 1:

The correct statement that explains bias is D. 'It describes a sample that is not representative of the population.' Bias refers to a systematic error in the sampling process that leads to a sample that does not accurately reflect the characteristics of the population. This can occur due to various reasons such as non-random sampling, undercoverage, or non-response. Bias can lead to incorrect or misleading conclusions when making inferences about the population based on the sample. Final answer: D

\*\*\*\*\*

Answer Option 2:

The correct statement that explains bias is D. 'It describes a sample that is not representative of the population.' Bias refers to the systematic deviation of a sample statistic from the true population parameter. This can occur when the sampling process is flawed or when there are factors that lead to an unrepresentative sample. Bias does not necessarily imply that the estimates are consistently too high or too low (option A), nor does it solely depend on the lack of randomization in the sampling process (option C). Final answer: D

\*\*\*\*\*

Answer Option 3:

The correct statement that explains bias is D. 'It describes a sample that is not representative of the population.' Bias refers to a systematic error or distortion in the sampling process that causes the sample to consistently deviate from the true population. This deviation can result in an overestimation or underestimation of the true population parameter. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The correct statement that explains bias is D. 'It describes a sample that is not representative of the population.' Bias occurs when there is a systematic error in the sampling process that skews the sample results away from being a true representation of the population. This means that the sample is not random and does not accurately reflect the characteristics or qualities of the larger population. Final answer: D

\*\*\*\*\*

Answer Option 5:

The correct statement that explains bias is B. 'It describes a process that results in a sample that does not represent the population.' Bias occurs when there is a systematic error in the sampling process that causes the sample to deviate from representing the true population. This can result in incorrect conclusions or generalizations being made about the population. Final answer: B

\*\*\*\*\*

Answer Option 6:

The correct statement that explains bias is D. 'It describes a sample that is not representative of the population.' Bias refers to a systematic error or deviation in the sample selection or measurement process that leads to a consistently inaccurate representation of the population of interest. This means that the sample does not accurately reflect the characteristics or attributes of the larger population, resulting in biased estimates or conclusions. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct statement that explains bias is D. 'It describes a sample that is not representative of the population.' Bias occurs when the sample that is selected does not accurately represent the population that it is intended to represent. This can occur due to various reasons such as non-random sampling, undercoverage, or measurement errors. When bias is present, the results and conclusions drawn from the sample may not accurately reflect the true characteristics or parameters of the population. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The correct statement that explains bias is option D. 'It describes a sample that is not representative of the population.' Bias occurs when there is a systematic deviation or distortion in the sample that leads to incorrect or misleading conclusions about the population. This can happen due to various reasons such as non-random sampling, non-response bias, or measurement errors. Option A describes systematic error, option B describes sampling error, and option C describes lack of randomization, which are related concepts but not the direct explanation of bias. Therefore, the correct statement is option D.

\*\*\*\*\*

Answer Option 9:

The correct statement that explains bias is D. 'It describes a sample that is not representative of the population.' Bias refers to a systematic error or distortion in the sampling process that results in a sample that does not accurately reflect the characteristics of the population. This can lead to inaccurate or misleading conclusions about the population. A biased sample may over-represent or under-represent certain groups or characteristics, leading to biased estimates or inferences about the population. Final Answer: D

\*\*\*\*\*

Question 12:

'A company that produces facial tissues continually monitors tissue strength. If the mean strength from sample data drops below a specified level, the production process is halted and the machinery inspected. Which of the following would result from a Type I error?', A. 'Halting the production process when sufficient customer complaints are received.', B. 'Halting the production process when the tissue strength is below specifications.', C. 'Halting the production process when the tissue strength is within specifications.', D. 'Allowing the production process to continue when the tissue strength is below specifications.'

=====

Answer Option 1:

A Type I error occurs when we reject a null hypothesis that is actually true. In the context of the problem, it means that the production process is halted when it is not necessary to do so.

Among the given options, 'Halting the production process when the tissue strength is within specifications' corresponds to a Type I error. This is because if the tissue strength is within specifications, it means that the mean strength from sample data is not below the specified level. Therefore, halting the production process in this case would be an error.

Hence, the answer is  $\boxed{\text{C. 'Halting the production process when the tissue strength is within specifications.'}}$

\*\*\*\*\*

Answer Option 2:

A Type I error occurs when the null hypothesis is true, but we reject it. In this case, the null hypothesis is that the tissue strength is at the specified level or above.

Option A, 'Halting the production process when sufficient customer complaints are received,' does not involve any hypothesis testing, so it is not a Type I error.

Option B, 'Halting the production process when the tissue strength is below specifications,' corresponds to the null hypothesis being true and the process being halted, which is the correct outcome.

Option C, 'Halting the production process when the tissue strength is within specifications,' corresponds to the null hypothesis being true and the process being halted, which is the correct outcome.

Option D, 'Allowing the production process to continue when the tissue strength is below specifications,' corresponds to the null hypothesis being false, but we fail to reject it. This is a Type II error, not a Type I error.

Therefore, the correct answer is  $\boxed{\text{D. 'Allowing the production process to continue when the tissue strength is below specifications.'}}$

\*\*\*\*\*

Answer Option 3:

A Type I error occurs when the null hypothesis is true, but we reject it. In this scenario, the null hypothesis would be that the tissue strength is at or above the specified level.

Option A, 'Halting the production process when sufficient customer complaints are received,' is not related to the null hypothesis or the specified level of tissue strength.

Option B, 'Halting the production process when the tissue strength is below specifications,' is the correct answer. This is a Type I error because the production process is being halted even though the tissue strength is actually within specifications.

Option C, 'Halting the production process when the tissue strength is within specifications,' is not a Type I error because the production process should be halted if the tissue strength is within specifications.

Option D, 'Allowing the production process to continue when the tissue strength is below specifications,' is not a Type I error because the production process should be halted if the tissue strength is below specifications.

Therefore, the correct answer is  $\boxed{\text{B. 'Halting the production process when the tissue strength is below specifications.'}}$

\*\*\*\*\*

Answer Option 4:

A Type I error in hypothesis testing refers to rejecting the null hypothesis when it is actually true. In this case, the null hypothesis is that the tissue strength is at or above the specified level.

Choice A, halting the production process when customer complaints are received, is not related to the hypothesis testing scenario and does not involve making a decision based on the tissue strength.

Choice B, halting the production process when the tissue strength is below specifications, is the correct answer. This would be a Type I error because the machinery is being inspected and production is halted even though the tissue strength is actually at or above the specified level, which means the null hypothesis is true.

Choice C, halting the production process when the tissue strength is within specifications, does not involve an error because the production process is being halted when the tissue strength is within the specified level, which aligns with the null hypothesis.

Choice D, allowing the production process to continue when the tissue strength is below specifications, is not a Type I error. It would be a Type II error, which refers to failing to reject the null hypothesis when it is actually false.

Therefore, the answer is  $\boxed{\text{(B)}}$ . Final answer: B

\*\*\*\*\*

Answer Option 5:

A Type I error is when the null hypothesis is rejected even though it is true. In this case, the null hypothesis is that the tissue strength is at or above the specified level.

Therefore, halting the production process when the tissue strength is within specifications would result from a Type I error.

So the answer is  $\boxed{\text{C. 'Halting the production process when the tissue strength is within specifications.'}}$

\*\*\*\*\*

Answer Option 6:

A Type I error occurs when the null hypothesis is true, but we reject it. In this case, the null hypothesis is that the tissue strength is within specifications. Therefore, a Type I error would result in halting the production process even though the tissue strength is actually within specifications. Thus, the correct answer is  $\boxed{\text{C. 'Halting the production process when the tissue strength is within specifications.'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 7:

A Type I error occurs when the null hypothesis is true, but we reject it. In this case, the null hypothesis is that the tissue strength is above the specified level.

Option A, "Halting the production process when sufficient customer complaints are received," does not relate to the null hypothesis.

Option B, "Halting the production process when the tissue strength is below specifications," aligns with the null hypothesis.

Option C, "Halting the production process when the tissue strength is within specifications," also aligns with the null hypothesis.

Option D, "Allowing the production process to continue when the tissue strength is below specifications," is the opposite of what should happen when the null hypothesis is true.

Therefore, the correct answer is  $\boxed{\text{D. 'Allowing the production process to continue when the tissue strength is below specifications.'}}$

\*\*\*\*\*

Answer Option 8:

A Type I error occurs when we reject a null hypothesis that is actually true. In this case, the null hypothesis is that the tissue strength is within specifications.

Option A, halting the production process when sufficient customer complaints are received, does not directly relate to the null hypothesis, so it is not a Type I error.

Option B, halting the production process when the tissue strength is below specifications, is the correct answer. This is a Type I error because we would be stopping production even though the tissue strength is within specifications.

Option C, halting the production process when the tissue strength is within specifications, is not a Type I error because we would be halting production based on correct information.

Option D, allowing the production process to continue when the tissue strength is below specifications, is not a Type I error because we would be allowing production to continue based on correct information.

Therefore, the answer is  $\boxed{\text{B. Halting the production process when the tissue strength is below specifications.}}$  Final Answer: B

\*\*\*\*\*

Answer Option 9:

A Type I error occurs when the null hypothesis is incorrectly rejected. In this case, the null hypothesis is that the tissue strength is above the specified level.

Option B states that the production process is halted when the tissue strength is below specifications. This is the correct action to take when the tissue strength is actually below the specified level, so it does not represent a Type I error.

Option C states that the production process is halted when the tissue strength is within specifications. This is also the correct action to take, as it ensures that the tissue strength remains within the desired range. Therefore, it does not represent a Type I error.

Option D states that the production process is allowed to continue when the tissue strength is below specifications. This is the incorrect action to take when the tissue strength is actually below the specified level, and it represents a Type I error.

Option A states that the production process is halted when sufficient customer complaints are received. This is not directly related to the tissue strength and does not represent a Type I error.

Therefore, the correct answer is option D. Allowing the production process to continue when the tissue strength is below specifications. Final Answer: D

\*\*\*\*\*

Question 13:

The longevity of people living in a certain locality has a standard deviation of 14 years. What is the mean longevity if 30% of the people live longer than 75 years? Assume a normal distribution for life spans.;

A. '75 - 0.30(14)', B. '75 + 0.30(14)', C. '75 - 0.4756(14)', D. '75 - 0.5244(14)'

=====

Answer Option 1:

Given that 30% of the people live longer than 75 years, we know that the area to the right of 75 under the normal distribution curve is 0.30. Since the standard deviation is 14, we can find the z-score corresponding to this area using a standard normal distribution table or calculator. The z-score is approximately -0.5244.

The z-score formula is given by:

$$z = (x - \mu) / \sigma$$

Where z is the z-score, x is the value we want to find the z-score for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case,  $x = 75$ ,  $\sigma = 14$ , and  $z = -0.5244$ . Rearranging the formula, we can solve for  $\mu$ :

$$\mu = x - z\sigma = 75 - (-0.5244)(14) = 75 + 0.5244(14)$$

Calculating this expression gives us approximately  $75 + 0.5244(14) = 75 + 7.3424 = 82.3424$ .

Therefore, the mean longevity is approximately 82.3424 years, so the correct answer is  $\boxed{\text{D. } 75 - 0.5244(14)}$ .

\*\*\*\*\*

Answer Option 2:

Let  $X$  be a random variable representing the longevity of people living in the locality. We can assume that  $X$  follows a normal distribution.

We know that 30% of the people live longer than 75 years. This means that the area to the right of 75 on the normal distribution curve is 0.30.

To convert this to a standard normal distribution, we can use the Z-score. The Z-score is given by  $Z = \frac{X - \mu}{\sigma}$ , where  $\mu$  is the mean and  $\sigma$  is the standard deviation.

In this case, we have  $Z = \frac{75 - \mu}{14}$ . We want to find the value of  $\mu$ .

Using a standard normal distribution table, we find that the Z-score corresponding to an area of 0.30 to the right is approximately 0.5244.

Setting the Z-score equal to 0.5244, we have  $\frac{75 - \mu}{14} = 0.5244$ .

Solving for  $\mu$ , we get  $\mu = 75 - 0.5244(14)$ .

Therefore, the mean longevity is approximately  $\boxed{75 - 0.5244(14)}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

To find the mean longevity, we need to find the z-score associated with the 30th percentile. The z-score is defined as the number of standard deviations away from the mean.

The area to the left of the z-score is 0.30. From the standard normal distribution table, we can find that the z-score that corresponds to an area of 0.30 to the left is approximately -0.5244.

Therefore, the mean longevity is  $75 - 0.5244(14) = \boxed{\text{D. } 75 - 0.5244(14)}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

We know that 30% of the people live longer than 75 years. In a normal distribution, the mean is the 50th percentile. Since 30% of the people live longer than 75 years, the mean is to the left of the 50th percentile.

To find the value of the mean, we need to find the z-score corresponding to the 30th percentile. Using a standard normal distribution table or calculator, we find that the z-score corresponding to the 30th percentile is approximately -0.5244.

The formula for the z-score is given by:

$$z = \frac{x - \mu}{\sigma}$$

where x is the value we want to find the z-score for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Rearranging the formula, we have:

$$\mu = x - z\sigma$$

Substituting the given values, we have:

$$\mu = 75 - (-0.5244)(14)$$

$$\mu = 75 + 0.5244(14)$$

Evaluating this expression, we find:

$$\mu = 75 + 7.3424$$

$$\mu = 82.3424$$

Therefore, the mean longevity is approximately 82.3424 years.

The correct answer is not given in the options provided.

\*\*\*\*\*

Answer Option 5:

To find the mean longevity, we can use the z-score formula for a normal distribution. Since we are given that 30% of the people live longer than 75 years, we can find the z-score corresponding to the upper 30th percentile.

Using a standard normal distribution table or a calculator, we find that the z-score corresponding to the upper 30th percentile is approximately 0.5244.

The z-score formula is given by:

$$z = (x - \mu) / \sigma$$

Where z is the z-score, x is the value we are interested in (75 in this case),  $\mu$  is the mean longevity, and  $\sigma$  is the standard deviation. Rearranging the formula, we have:

$$x = z * \sigma + \mu$$

Substituting the given values, we have:

$$75 = 0.5244 * 14 + \mu$$

$$75 = 7.3456 + \mu$$

$$\mu = 75 - 7.3456$$

$$\mu = 67.6544$$

Therefore, the mean longevity is approximately 67.6544 years.

Since none of the answer choices match this value, it appears there may be an error in the problem statement. However, the closest option is C. '75 - 0.4756(14)', but this does not accurately represent the mean longevity.

\*\*\*\*\*

Answer Option 6:

We know that in a normal distribution, approximately 68% of the data falls within one standard deviation of the mean, 95% falls within two standard deviations, and 99.7% falls within three standard deviations.

Since 30% of the people live longer than 75 years, this means that 70% live less than or equal to 75 years. This corresponds to 1 standard deviation below the mean.

Therefore, the mean longevity is  $75 - \text{1 standard deviation} = 75 - 14 = \boxed{61}$  years. Final Answer: None (Since none of the given options are correct.)

\*\*\*\*\*

Answer Option 7:

In a normal distribution, the mean and standard deviation uniquely determine the distribution. We are given that the standard deviation is 14 years, so we can assume that the distribution follows a normal distribution with a mean of  $\mu$  and a standard deviation of 14 years.

We are also given that 30% of the people live longer than 75 years. Since the normal distribution is symmetric, we know that 50% of the people live longer than the mean. Therefore, the mean is 75 years minus the value that corresponds to the additional 20% (since  $50\% + 20\% = 70\%$ ) in the lower tail of the distribution. Using the standard normal distribution table, we find that this value is approximately 0.5244.

Therefore, the mean longevity is approximately  $75 - 0.5244(14)$ \$. Simplifying, we get  $75 - 7.34 \approx \boxed{67.66}$ \$.

The correct answer is not listed.

\*\*\*\*\*

Answer Option 8:

To solve this problem, we will use the properties of a normal distribution.

We know that 30% of the people live longer than 75 years. Since the normal distribution is symmetric, this means that 70% of the people live less than or equal to 75 years.

Let's denote the mean longevity as  $\mu$ \$. We can use the standard deviation to find the z-score corresponding to the 70th percentile (since 70% live to 75 years or less).

Using a standard normal distribution table or calculator, we find that the z-score corresponding to the 70th percentile is approximately -0.5244.

The z-score is calculated by the formula:  $z = (x - \mu) / \sigma$ , where x is the value of interest,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Substituting the known values, we have:  $-0.5244 = (75 - \mu) / 14$ .

Solving for  $\mu$ , we find:  $\mu = 75 - (-0.5244)(14) = 75 + 0.5244(14)$ .

Therefore, the mean longevity is approximated by  $75 + 0.5244(14)$ , which corresponds to answer choice D. Final Answer: D

\*\*\*\*\*

Answer Option 9:

In a normal distribution, if we know the mean and standard deviation, we can use z-scores to find the percentage of values that fall above a certain value. In this case, we want to find the mean longevity such that 30% of the people live longer than 75 years.

Let's assume the mean longevity is denoted by  $\mu$ \$. Since we want to find the percentage of values that fall above 75 years, we need to find the z-score corresponding to the cutoff point of 75 years.

The z-score formula is given by:  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value we are interested in,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case,  $x = 75$ \$,  $\sigma = 14$ \$, and we want to find  $\mu$ \$. Rearranging the formula, we have  $\mu = x - z \sigma$ \$.

To find the z-score corresponding to the 30th percentile, we look up the z-score in the standard normal distribution table. The z-score that corresponds to a cumulative probability of 0.30 is approximately -0.5244.

Therefore, the mean longevity is given by  $\mu = 75 - (-0.5244)(14) = \boxed{\text{(C) } 75 - 0.4756(14)}$ \$. Final Answer: C

\*\*\*\*\*



Question 14:

'Changing from a 90% confidence interval estimate for a population proportion to a 99% confidence interval estimate, with all other things being equal.', A. 'increases the interval size by 9%.', B. 'decreases the interval size by 9%.', C. 'increases the interval size by 57%.', D. 'decreases the interval size by 57%.'

=====

Answer Option 1:

The width of a confidence interval is determined by the level of confidence and the standard error of the estimate. In this case, the level of confidence is changing from 90% to 99%, which means we want a higher level of confidence in our estimate.

When we increase the level of confidence, the interval needs to be wider to account for the increased uncertainty. This is because a higher level of confidence requires a larger range of values to capture the true population parameter.

Therefore, the correct answer is that increasing the confidence level from 90% to 99% 'increases the interval size'. However, none of the given answer choices accurately describe this change. So, the correct answer is not among the options provided.

\*\*\*\*\*

Answer Option 2:

When changing the confidence level from 90% to 99%, the interval size will increase. This is because a higher confidence level requires a wider interval to capture a greater range of possible values.

To determine the exact increase in interval size, we can use the formula for margin of error:

Margin of Error =  $Z * \sqrt{p * (1-p)} / n$ ,

where Z is the z-score corresponding to the desired confidence level, p is the sample proportion, and n is the sample size.

Since all other things are equal, the sample proportion and sample size will remain the same when changing from a 90% confidence interval to a 99% confidence interval.

The z-score for a 90% confidence level is approximately 1.645, and the z-score for a 99% confidence level is approximately 2.576.

Since the sample proportion and sample size remain the same, the only difference is the z-score.

Using the formula for margin of error, we can see that the margin of error will be larger with a z-score of 2.576 compared to a z-score of 1.645.

Therefore, the interval size will increase when changing from a 90% confidence interval to a 99% confidence interval.

The correct answer is A. 'increases the interval size by 9%.'

\*\*\*\*\*

Answer Option 3:

To change from a 90% confidence interval to a 99% confidence interval, we need to increase the level of confidence. As the level of confidence increases, the interval size also increases because we want to be more confident in capturing the true population proportion.

Therefore, the correct answer is: A. 'increases the interval size by 9%'. Final Answer: A

\*\*\*\*\*

Answer Option 4:

When changing from a 90% confidence interval to a 99% confidence interval with all other things being equal, the interval size increases. This is because a higher confidence level requires a wider interval to account for the increased level of certainty. Therefore, the correct answer is A. 'increases the interval size by 9%'. Final answer: A

\*\*\*\*\*

Answer Option 5:

The width of a confidence interval is determined by the level of confidence and the variability of the data. A higher level of confidence requires a wider interval to capture a larger range of possible values.

In this case, changing from a 90% confidence interval to a 99% confidence interval means increasing the level of confidence. Since all other things are equal, this increase in confidence will result in a wider interval.

Therefore, the correct answer is A. 'increases the interval size by 9%'. Final answer: A.

\*\*\*\*\*

Answer Option 6:

When changing from a 90% confidence interval to a 99% confidence interval with all other things being equal, the interval size will increase. This is because a higher confidence level requires a wider interval to capture the true population proportion with a higher level of certainty.

To be more specific, the interval size is directly proportional to the critical value, which is determined by the desired confidence level. A 99% confidence level corresponds to a larger critical value than a 90% confidence level, resulting in a wider interval.

Therefore, the correct answer is A. 'Increases the interval size by 9%'. Final Answer: A

\*\*\*\*\*

Answer Option 7:

The size of a confidence interval is determined by the level of confidence and the margin of error. In this case, we are changing from a 90% confidence interval to a 99% confidence interval.

To increase the level of confidence from 90% to 99%, we need to increase the margin of error. This means that the interval size will increase.

However, we are asked how the interval size changes in percentage terms. We can calculate this by finding the difference in interval sizes and expressing it as a percentage of the original interval size.

The increase in level of confidence from 90% to 99% corresponds to an increase of 9 percentage points. Since the interval size is directly proportional to the level of confidence, the interval size will increase by approximately 9%.

Therefore, the correct answer is  $\boxed{\text{A. Increases the interval size by 9\%.}}$

\*\*\*\*\*

Answer Option 8:

To change from a 90% confidence interval to a 99% confidence interval, we need to increase the level of confidence. This means we want to be more certain that our interval contains the true population proportion. In order to achieve a higher level of confidence, we need to widen the interval. Therefore, the interval size will increase.

To determine the percentage increase in the interval size, we can calculate the difference between the two confidence levels and divide it by the original interval size. The difference between 90% and 99% is 9%. Since the original interval size is 100%, the percentage increase in the interval size is  $\frac{9\%}{100\%} \times 100\% = 9\%$ .

Therefore, changing from a 90% confidence interval to a 99% confidence interval increases the interval size by  $\boxed{9\%}$ . Final Answer: A

\*\*\*\*\*

Answer Option 9:

The confidence level and the interval size are inversely related. A higher confidence level requires a larger interval size to capture the true population parameter with higher certainty. Since we are changing from a 90% confidence interval to a 99% confidence interval, we are increasing the confidence level. Therefore, the interval size will increase.

To determine the percentage change in the interval size, we can compare the difference between the upper and lower bounds of the two confidence intervals. Since the interval size is symmetrical, we only need to consider one side.

Let's assume the interval size for the 90% confidence interval is  $x$ . Then, the interval size for the 99% confidence interval will be  $x + \text{increase}$ . Since the difference between a 90% confidence interval and a 99% confidence interval is 9%, the increase will be  $\frac{9}{90}x = \frac{1}{10}x = \frac{x}{10} = 0.1x$ .

Therefore, the interval size of the 99% confidence interval is  $x + 0.1x = 1.1x$ . The percentage increase in the interval size is  $\frac{1.1x - x}{x} \times 100\% = \frac{0.1x}{x} \times 100\% = 10\%$ .

Therefore, the interval size increases by 10%, which is closest to 9%. Thus, the answer is  $\boxed{\text{A}}$ , increases the interval size by 9%. Final answer: A

\*\*\*\*\*

Question 15:

'Which of the following is a binomial random variable?', A. 'The number of tosses before a "5" appears when tossing a fair die.', B. 'The number of points a hockey team receives in 10 games, where two points are awarded for wins, one point for ties, and no points for losses.', C. 'The number of hearts out of five cards randomly drawn from a deck of 52 cards, without replacement.', D. 'The number of motorists not wearing seat belts in a random sample of five drivers.'

=====

Answer Option 1:

A binomial random variable is a random variable that represents the number of successes in a fixed number of independent Bernoulli trials, where each trial has the same probability of success.

Option B is a binomial random variable because it represents the number of wins in 10 games, and each game can be considered an independent Bernoulli trial with a fixed probability of winning.

Therefore, the answer is  $\boxed{\text{B. 'The number of points a hockey team receives in 10 games, where two points are awarded for wins, one point for ties, and no points for losses.'}}$ . Final answer: B

\*\*\*\*\*

Answer Option 2:

A binomial random variable satisfies the following conditions:

1. The experiment consists of a fixed number of independent trials.
2. Each trial has only two possible outcomes, success or failure.
3. The probability of success is the same for each trial.
4. The random variable of interest is the number of successes in the fixed number of trials.

Option B meets all of these conditions. The team can either win, tie, or lose each game, which satisfies condition 2. The probability of winning, tying, or losing is the same for each game, satisfying condition 3.

The number of points the team receives in 10 games is the random variable of interest, satisfying condition 4.

Therefore, the correct answer is  $\boxed{\text{B. 'The number of points a hockey team receives in 10 games, where two points are awarded for wins, one point for ties, and no points for losses.'}}$

\*\*\*\*\*

Answer Option 3:

A binomial random variable is a discrete random variable that represents the number of successes in a fixed number of independent Bernoulli trials, where each trial has the same probability of success.

Option A does not fit the definition of a binomial random variable because the number of tosses before a "5" appears is not a fixed number of trials.

Option B does not fit the definition of a binomial random variable because the number of points a hockey team receives in 10 games depends on a combination of wins, ties, and losses, and is not a fixed number of trials with the same probability of success.

Option C does not fit the definition of a binomial random variable because the number of hearts out of five cards drawn without replacement does not have the same probability of success for each trial.

Option D fits the definition of a binomial random variable because we have a fixed number of trials (five drivers in the random sample) and each trial is independent (assuming the drivers are chosen randomly and without replacement). Additionally, each trial has the same probability of success (not wearing a seat belt). Thus, the number of motorists not wearing seat belts in a random sample of five drivers is a binomial random variable.

Therefore, the answer is  $\boxed{\text{D. 'The number of motorists not wearing seat belts in a random sample of five drivers.'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

A binomial random variable is defined as a count of the number of successes in a fixed number of independent Bernoulli trials, where each trial has the same probability of success.

Option A does not fit this definition because the number of tosses before a "5" appears is not a count of successes in a fixed number of trials.

Option B also does not fit this definition because the number of points a hockey team receives in 10 games is not a count of successes in a fixed number of trials.

Option C does fit the definition of a binomial random variable because it is a count of the number of hearts (successes) in a fixed number of trials (drawing five cards), where each trial is independent and has the same probability of success (drawing a heart from the deck).

Option D does not fit the definition because the number of motorists not wearing seat belts is not a count of successes in a fixed number of trials.

Therefore, the correct answer is  $\boxed{\text{C. The number of hearts out of five cards randomly drawn from a deck of 52 cards, without replacement.'}}$

\*\*\*\*\*

Answer Option 5:

A binomial random variable is characterized by having a fixed number of independent trials, where each trial has two possible outcomes (success or failure), and the probability of success remains constant for each trial.

Option A does not fit the definition of a binomial random variable because the number of tosses before a "5" appears is not a fixed number of trials.

Option B also does not fit the definition because the number of points a hockey team receives in 10 games is not a binary outcome (two possible outcomes).

Option C does fit the definition of a binomial random variable because it involves drawing a fixed number of cards (five) from a deck without replacement, and each card drawn can either be a heart (success) or not a heart (failure).

Option D does not fit the definition of a binomial random variable because the probability of a driver not wearing a seat belt is not necessarily constant for each driver in a random sample of five drivers.

Therefore, the correct answer is  $\boxed{\text{C. 'The number of hearts out of five cards randomly drawn from a deck of 52 cards, without replacement.'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 6:

A binomial random variable is defined by having a fixed number of independent trials, each with two possible outcomes (success or failure), and a constant probability of success for each trial.

Option A does not fit the criteria of a binomial random variable because the number of tosses before a "5" appears is not a fixed number of trials. The number of trials is not known in advance and can vary.

Option B also does not fit the criteria of a binomial random variable because the number of points a hockey team receives in 10 games does not have only two possible outcomes. There are three possible outcomes (wins, ties, and losses), so it does not meet the requirement of having only two possible outcomes.

Option C does fit the criteria of a binomial random variable. The number of hearts out of five cards randomly drawn from a deck of 52 cards without replacement has a fixed number of trials (drawing five cards) and two possible outcomes (drawing a heart or not drawing a heart) with a constant probability of success (the probability of drawing a heart on each trial changes as cards are drawn). Therefore, option C is a binomial random variable.

Option D does not fit the criteria of a binomial random variable because the number of motorists not wearing seat belts in a random sample of five drivers does not have a constant probability of success for each trial. The probability of a driver not wearing a seat belt can vary for each driver, so it does not meet the requirements of a binomial random variable.

Therefore, the correct answer is C.

\*\*\*\*\*

Answer Option 7:

A binomial random variable is defined as the number of successes in a fixed number of independent Bernoulli trials, where each trial has the same probability of success.

Option A does not fit this definition because the number of tosses before a "5" appears is not a fixed number of trials.

Option B does not fit this definition because the number of points a hockey team receives in 10 games depends on the outcomes of multiple games and does not have the same probability of success for each game.

Option C does not fit this definition because the number of hearts drawn without replacement is not a fixed number of trials, as the number of hearts drawn will change depending on the outcomes of previous draws.

Option D fits the definition of a binomial random variable. We have a fixed number of five drivers in the sample, and each driver either wears a seat belt or does not wear a seat belt. The probability of a driver not wearing a seat belt is the same for each driver, making it a binomial random variable.

Therefore, the answer is option D.

\*\*\*\*\*

Answer Option 8:

A binomial random variable is defined as the number of successes in a fixed number of independent Bernoulli trials, where each trial has the same probability of success.

Option B is a binomial random variable. Each game is an independent Bernoulli trial with a fixed probability of success (receiving two points for a win). The total number of points received in 10 games can be considered as the number of successes.

Therefore, the answer is  $\boxed{\text{B}}$ . The number of points a hockey team receives in 10 games, where two points are awarded for wins, one point for ties, and no points for losses. Final Answer: B

\*\*\*\*\*

Answer Option 9:

A binomial random variable can be thought of as counting the number of successes in a fixed number of independent Bernoulli trials, where each trial has the same probability of success.

In option A, the number of tosses before a "5" appears when tossing a fair die is not a binomial random variable because the number of trials is not fixed. It could take any number of tosses until a "5" appears.

In option B, the number of points a hockey team receives in 10 games can be a binomial random variable because each game is independent, and the team can either win (success), tie, or lose (failure).

However, the number of points awarded is not a fixed number, it depends on the outcome of each game.

In option C, the number of hearts out of five cards randomly drawn from a deck of 52 cards, without replacement, is not a binomial random variable. The probability of drawing a heart changes with each card drawn, so the trials are not independent.

In option D, the number of motorists not wearing seat belts in a random sample of five drivers can be a binomial random variable. Each driver can be considered a trial, and the outcome is whether or not the driver is wearing a seat belt (success or failure). The probability of a driver not wearing a seat belt may be assumed to be the same for each driver in the sample.

Therefore, the correct answer is  $\boxed{\text{D}}$

\*\*\*\*\*

Question 16:

'If all the other variables remain constant, which of the following will increase the power of a hypothesis test? I.Increasing the sample size. II.Increasing the significance level. III.Increasing the probability of a Type II error.', A. 'I only', B. 'II only', C. 'III only', D. 'I and II'

=====

Answer Option 1:

To increase the power of a hypothesis test, it is necessary to decrease the probability of a Type II error (false negative), which is the probability of failing to reject a false null hypothesis. Increasing the sample size increases the power of a hypothesis test by reducing the standard deviation and making the test statistic more likely to fall in the rejection region. Increasing the significance level, on the other hand, increases the probability of a Type I error (false positive), which is the probability of rejecting a true null hypothesis. Therefore, increasing the significance level does not increase the power of a hypothesis test. Hence, the answer is  $\boxed{\text{A. 'I only'}}$ . Final answer: A

\*\*\*\*\*

Answer Option 2:

The power of a hypothesis test measures the probability of correctly rejecting the null hypothesis when it is indeed false. In order to increase the power of a hypothesis test, we want to increase the probability of correctly detecting a true alternative hypothesis.

I. Increasing the sample size generally increases the power of a hypothesis test. With a larger sample size, we have more data to work with, which can lead to more accurate estimates and more precise conclusions. Therefore, increasing the sample size will increase the power of the hypothesis test.

II. Increasing the significance level, also known as the level of significance or alpha, does not directly affect the power of a hypothesis test. The significance level determines the threshold for rejecting the null hypothesis, but it does not directly impact the probability of correctly detecting a true alternative hypothesis. Therefore, increasing the significance level does not increase the power of the hypothesis test.

III. Increasing the probability of a Type II error, also known as beta, decreases the power of a hypothesis test. A Type II error occurs when we fail to reject the null hypothesis, even though it is false. By increasing the probability of a Type II error, we are more likely to incorrectly accept the null hypothesis and miss detecting a true alternative hypothesis. Therefore, increasing the probability of a Type II error decreases the power of the hypothesis test.

Based on the above analysis, the correct answer is  $\boxed{\text{I only}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 3:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when it is false. Increasing the sample size (option I) generally increases the power of a test because it provides more data and reduces the variability of the test statistic.

Increasing the significance level (option II) also increases the power of a test. This is because a higher significance level allows for a larger rejection region, making it more likely to reject the null hypothesis.

On the other hand, increasing the probability of a Type II error (option III) decreases the power of the test. Type II error occurs when the null hypothesis is false, but we fail to reject it. Increasing the probability of Type II error means we are more likely to make this mistake and therefore less likely to correctly reject the null hypothesis.

Based on the above explanations, the correct answer is  $\boxed{\text{D. 'I and II'}}$  as both increasing the sample size and increasing the significance level will increase the power of a hypothesis test. Final

Answer: D

\*\*\*\*\*

Answer Option 4:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when it is false. Increasing the sample size increases the power of a hypothesis test because it provides more information and reduces the sampling variability. Therefore, statement I is correct.

Increasing the significance level, also known as the level of significance or Type I error rate, increases the probability of rejecting the null hypothesis when it is true. This means that the test becomes more liberal, making it easier to reject the null hypothesis. As a result, the power of the test increases. Therefore, statement II is also correct.

Increasing the probability of a Type II error decreases the power of a hypothesis test. A Type II error occurs when the null hypothesis is false, but we fail to reject it. Increasing the probability of committing a Type II error reduces the power of the test because it makes it more likely that we will fail to detect a true alternative hypothesis. Therefore, statement III is not correct.

In conclusion, statements I and II are correct. Thus, the answer is  $\boxed{\text{D. 'I and II'}}$ .

\*\*\*\*\*

Answer Option 5:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when it is false.

I. Increasing the sample size generally increases the power of a hypothesis test. With a larger sample size, we have more data points and thus more information to make a decision. This allows us to detect smaller effects and increases the likelihood of correctly rejecting the null hypothesis. Therefore, increasing the sample size increases the power.

II. Increasing the significance level (also known as the level of significance or alpha) increases the probability of rejecting the null hypothesis when it is true. This means that we are more likely to incorrectly reject the null hypothesis, which increases the chance of a Type I error. However, it does not directly affect the power of the test.

III. Increasing the probability of a Type II error decreases the power of a hypothesis test. The power is 1 minus the probability of a Type II error. If we increase the probability of a Type II error, the power decreases.

Based on the above analysis, the correct answer is  $\boxed{\text{I only}}$ . Increasing the sample size increases the power of a hypothesis test, while increasing the significance level and increasing the probability of a Type II error do not directly impact the power. Final answer: A.

\*\*\*\*\*

Answer Option 6:

Increasing the sample size will increase the power of a hypothesis test. A larger sample size provides more information and reduces the sampling variability, leading to a more precise estimate and stronger evidence against the null hypothesis.

Increasing the significance level will also increase the power of a hypothesis test. A higher significance level allows for a larger rejection region, making it easier to reject the null hypothesis and increasing the probability of detecting a true alternative hypothesis.

On the other hand, increasing the probability of a Type II error will decrease the power of a hypothesis test. The power of a test is defined as 1 minus the probability of a Type II error. So, if the probability of a Type II error is increased, the power of the test will be decreased.

Therefore, the correct answer is  $\boxed{\text{D. 'I and II'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. This means that a higher power indicates a better ability to detect a true effect.

Increasing the sample size (I) generally increases the power of a hypothesis test. With a larger sample size, there is more data available to estimate the parameters and reduce sampling variability.

Increasing the significance level (II) also increases the power of a hypothesis test. A higher significance level allows for more extreme evidence against the null hypothesis to be considered statistically significant, leading to a higher chance of rejecting the null hypothesis when it is false.

Increasing the probability of a Type II error (III) decreases the power of a hypothesis test. Type II error occurs when we fail to reject the null hypothesis when it is actually false. By increasing the probability of this error, we are reducing the ability of the test to detect the true effect, resulting in lower power.

Therefore, the correct answer is  $\boxed{\text{D. 'I and II'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when it is false. In other words, it is the probability of detecting a true effect.

I. Increasing the sample size generally increases the power of a hypothesis test. With a larger sample size, there is more evidence available to make a decision about the null hypothesis, and the test is more sensitive to detecting smaller effects.

II. Increasing the significance level, also known as the alpha level, increases the probability of rejecting the null hypothesis when it is true. This means that the test becomes less conservative and more likely to detect an effect. However, increasing the significance level does not directly increase the power of the test.

III. Increasing the probability of a Type II error means increasing the probability of failing to reject the null hypothesis when it is false. This reduces the power of the test. Therefore, increasing the probability of a Type II error decreases the power of the test.

Based on the above explanations, the correct answer is  $\boxed{\text{D. 'I and II'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when it is indeed false. Therefore, increasing the power of a hypothesis test is desirable.

I. Increasing the sample size generally increases the power of a hypothesis test. As the sample size increases, the test has more information and is better able to distinguish between the null and alternative hypotheses. Therefore, increasing the sample size will increase the power.

II. Increasing the significance level (also known as the level of significance or alpha) increases the chance of rejecting the null hypothesis. A higher significance level allows for stronger evidence against the null hypothesis, which can increase the power of the test. Therefore, increasing the significance level can increase the power.

III. Increasing the probability of a Type II error means increasing the chance of failing to reject the null hypothesis when it is actually false. This is the opposite of what we want, as we want to correctly reject the null hypothesis when it is false. Therefore, increasing the probability of a Type II error will decrease the power of the test.

Based on the above analysis, the correct answer is  $\boxed{\text{D. 'I and II'}}$  since increasing the sample size and increasing the significance level both increase the power of a hypothesis test. Final answer: D.

\*\*\*\*\*

Question 17:

'Which of the following statements is (are) true? I. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5. II. In order to use  $\chi^2$  procedures, you must have at least 2 degrees of freedom. III. In a  $4 \times 2$  two-way table, the number of degrees of freedom is 3.', A. 'I only', B. 'I and III only', C. 'I and II only', D. 'III only'

=====

Answer Option 1:

Statement I is true. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5. This is to ensure that the approximation of the  $\chi^2$  distribution is valid.

Statement II is true. In order to use  $\chi^2$  procedures, you must have at least 2 degrees of freedom. Degrees of freedom represent the number of independent pieces of information available for estimation, and for  $\chi^2$  procedures, it is calculated as  $(\text{number of rows} - 1) \times (\text{number of columns} - 1)$ .

Statement III is false. In a  $4 \times 2$  two-way table, the number of degrees of freedom is  $(4 - 1) \times (2 - 1) = 3$ . So, the correct statement is 'I and II only', giving us answer choice C. Final answer: C.

\*\*\*\*\*

Answer Option 2:

Statement I is true. In order to use a chi-square procedure, the expected value for each cell of a one- or two-way table should be at least 5. This is known as the "expected cell frequency assumption" and ensures that the chi-square approximation is valid.

Statement II is false. There is no specific requirement for the number of degrees of freedom in order to use chi-square procedures. The degrees of freedom depend on the number of categories and constraints in the data, and can vary for different chi-square analyses.

Statement III is true. In a  $4 \times 2$  two-way table, the number of degrees of freedom is calculated as  $(\text{number of rows} - 1) \times (\text{number of columns} - 1) = (4 - 1) \times (2 - 1) = 3$ . This is because we have 4 categories in one dimension and 2 categories in the other dimension, resulting in  $(4 - 1) = 3$  degrees of freedom.

Therefore, the correct answer is  $\boxed{\text{B. 'I and III only'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 3:

Statement I is true. In order to use a chi-square procedure, the expected value for each cell of a one- or two-way table must be at least 5. This is because the chi-square test relies on the assumption that the observed frequencies in each cell follow a chi-square distribution, which is only valid when the expected frequencies are sufficiently large.

Statement II is false. The minimum requirement for degrees of freedom in a chi-square test is one, not two. The degrees of freedom represent the number of independent pieces of information available for estimation in the chi-square test.

Statement III is false. In a  $4 \times 2$  two-way table, the number of degrees of freedom is  $(4 - 1) \times (2 - 1) = 3$ . Therefore, the correct statement(s) are: I, making the correct answer choice (A) 'I only'. Final Answer: A

\*\*\*\*\*

Answer Option 4:

Statement I is true. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table should be at least 5. This is because the  $\chi^2$  test statistic is based on the assumption that the observed frequencies in the cells follow a multinomial distribution, which is approximately chi-squared distributed when the expected values are sufficiently large.

Statement II is false. In order to use  $\chi^2$  procedures, there is no specific requirement for the number of degrees of freedom.  $\chi^2$  tests can be conducted with any number of degrees of freedom, including 1.

Statement III is true. In a  $4 \times 2$  two-way table, the number of degrees of freedom can be calculated as  $(\text{number of rows} - 1) \times (\text{number of columns} - 1) = (4 - 1) \times (2 - 1) = 3$ .

Therefore, the correct answer is  $\boxed{\text{B. I and III only}}$ .

\*\*\*\*\*

Answer Option 5:

Statement I is true. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5. This is because the  $\chi^2$  test statistic relies on the approximation of the chi-square distribution, which is valid when the expected cell counts are reasonably large.

Statement II is false. In order to use  $\chi^2$  procedures, you must have at least 1 degree of freedom, not 2. Degrees of freedom in a chi-square test are determined by the number of categories or cells in the table minus the number of constraints or restrictions imposed on those categories.

Statement III is true. In a  $4 \times 2$  two-way table, the number of degrees of freedom is  $(4 - 1)(2 - 1) = 3$ . This is calculated by taking the number of rows minus 1 and multiplying it by the number of columns minus 1.

Therefore, the correct answer is  $\boxed{\text{B. 'I and III only'}}$ . Final answer: B

\*\*\*\*\*

Answer Option 6:

Statement I is true. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5. This is to ensure that the  $\chi^2$  test statistic follows approximately a chi-square distribution.

Statement II is false. In order to use  $\chi^2$  procedures, you must have at least 1 degree of freedom. The number of degrees of freedom in a chi-square test is determined by the number of categories in the data minus 1.

Statement III is false. In a  $4 \times 2$  two-way table, the number of degrees of freedom is  $(4 - 1) \times (2 - 1) = 3$ . Therefore, the correct answer is  $\boxed{\text{(A) 'I only'}}$ .

\*\*\*\*\*

Answer Option 7:

Statement I is true. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5. This is known as the "expected cell frequency rule" and is necessary for the  $\chi^2$  test to be valid.

Statement II is false. In order to use  $\chi^2$  procedures, you typically need at least 1 degree of freedom. The degrees of freedom in a  $\chi^2$  test depend on the number of categories or cells in the table, not a specific minimum number.

Statement III is true. In a  $4 \times 2$  two-way table, the number of degrees of freedom is calculated as  $(\text{number of rows} - 1) \times (\text{number of columns} - 1) = (4 - 1) \times (2 - 1) = 3$ . So, statement III is true.

Therefore, the correct answer is  $\boxed{\text{B, I and III only}}$ .

\*\*\*\*\*

Answer Option 8:

Statement I is true. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5. This is known as the "expected cell frequency condition".

Statement II is false. In order to use  $\chi^2$  procedures, you must have at least 1 degree of freedom, not 2. The number of degrees of freedom is determined by the number of categories in the table minus 1. For

example, a  $4 \times 2$  two-way table would have  $(4 - 1) \times (2 - 1) = 3$  degrees of freedom.

Statement III is true. In a  $4 \times 2$  two-way table, the number of degrees of freedom is 3. This is determined by  $(4 - 1) \times (2 - 1) = 3$ .

Therefore, the correct statement(s) are I and III only. The answer is  $\boxed{\text{B, I and III only}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

For statement I, it is true that in order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5. This is because the  $\chi^2$  distribution is an approximation based on the assumption that the expected values are reasonably large.

For statement II, it is false that in order to use  $\chi^2$  procedures, you must have at least 2 degrees of freedom. The  $\chi^2$  distribution can be used for different degrees of freedom, including 1 degree of freedom.

For statement III, it is false that in a  $4 \times 2$  two-way table, the number of degrees of freedom is 3. The number of degrees of freedom for a two-way table is calculated as  $(\text{number of rows} - 1) \times (\text{number of columns} - 1)$ , which in this case would be  $(4 - 1) \times (2 - 1) = 3 \times 1 = 3$ .

Therefore, the correct statement is 'I only'. The answer is  $\boxed{\text{A}}$ . Final Answer: A

\*\*\*\*\*



Question 18:

'Suppose two events, E and F, have nonzero probabilities p and q, respectively. Which of the following is impossible?', A. ' $p + q > 1$ ', B. ' $p - q < 0$ ', C. ' $p/q > 1$ ', D. 'E and F are both independent and mutually exclusive.'

=====

Answer Option 1:

If events E and F are independent, then the probability of both events occurring is given by the product of their individual probabilities, which means  $P(E \text{ and } F) = P(E) \cdot P(F)$ . However, if events E and F are mutually exclusive, it means that they cannot occur together, which implies that  $P(E \text{ and } F) = 0$ . Therefore, it is impossible for E and F to be both independent and mutually exclusive. Thus, the answer is  $\boxed{\text{D}}$ . Final answer: D

\*\*\*\*\*

Answer Option 2:

If E and F are independent, then the probability of both events occurring is equal to the product of their individual probabilities, so  $p(F) = p(E)q$ . However, if the events are mutually exclusive, then the probability of both events occurring is 0, so  $p(F) = 0$ . By the transitive property, we have  $p(E)q = 0$ , which means that either  $p = 0$  or  $q = 0$ .

Since we are given that p and q are nonzero probabilities, it is impossible for E and F to be both independent and mutually exclusive. Therefore, the correct answer is  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

If E and F are independent, then the probability of both events happening is given by  $P(E \cap F) = P(E) \cdot P(F) = p \cdot q$ .

If E and F are mutually exclusive, then the probability of both events happening is given by  $P(E \cap F) = 0$ .

Therefore, if E and F are both independent and mutually exclusive, we have  $p \cdot q = 0$ . Since p and q are nonzero probabilities, this is impossible. Thus, the statement 'E and F are both independent and mutually exclusive' is impossible.

Therefore, the correct answer is  $\boxed{\text{D. 'E and F are both independent and mutually exclusive.'}}$

\*\*\*\*\*

Answer Option 4:

Event E and F being independent means that the occurrence or non-occurrence of one event does not affect the probability of the occurrence or non-occurrence of the other event. Event E and F being mutually exclusive means that the two events cannot occur simultaneously.

Since E and F are mutually exclusive, it means that if one event occurs, the other event cannot occur. Therefore, the probability of both events occurring is zero. In other words,  $P(E \text{ and } F) = 0$ .

If E and F are independent, then we have  $P(E \text{ and } F) = P(E) \cdot P(F)$ . Since  $P(E \text{ and } F) = 0$ , it means that either  $P(E) = 0$  or  $P(F) = 0$ .

However, it is given in the question that both events E and F have nonzero probabilities. Therefore, it is impossible for E and F to be both independent and mutually exclusive.

Hence, the answer is D. 'E and F are both independent and mutually exclusive.' Final Answer: D

\*\*\*\*\*

Answer Option 5:

The statement ' $p + q > 1$ ' is possible if p and q are both greater than 1/2. For example, let  $p = 3/4$  and  $q = 2/4$ .

The statement ' $p - q < 0$ ' is possible if p is less than q. For example, let  $p = 1/4$  and  $q = 3/4$ .

The statement ' $p/q > 1$ ' is possible if p is greater than q. For example, let  $p = 3/4$  and  $q = 1/4$ .

The statement 'E and F are both independent and mutually exclusive' is impossible. If two events are independent, then the occurrence of one event does not affect the probability of the occurrence of the other event. However, if two events are mutually exclusive, then the occurrence of one event means that the other event cannot occur. These two properties cannot hold simultaneously for two events.

Therefore, the impossible statement is 'E and F are both independent and mutually exclusive', which corresponds to option  $\boxed{\text{D}}$ . Final answer: D

\*\*\*\*\*

Answer Option 6:

Two events cannot be both independent and mutually exclusive, so it is impossible for option D to be true. Therefore, the answer is  $\boxed{\text{D. 'E and F are both independent and mutually exclusive.'}}$

\*\*\*\*\*

Answer Option 7:

If E and F are independent events, then the probability of both events occurring is the product of their individual probabilities, so  $P(E \text{ and } F) = p \cdot q$ . If E and F are mutually exclusive events, then by definition, they cannot occur at the same time, so  $P(E \text{ and } F) = 0$ .

Therefore, if E and F are both independent and mutually exclusive, we have  $p \cdot q = 0$ . Since p and q are both nonzero probabilities, this equation cannot hold true. Therefore, the statement 'E and F are both independent and mutually exclusive' is impossible.

Therefore, the correct answer is  $\boxed{\text{D. 'E and F are both independent and mutually exclusive.'}}$

\*\*\*\*\*

Answer Option 8:

The statement ' $p + q > 1$ ' is possible. For example, if E and F are mutually exclusive events with probabilities  $p = 0.6$  and  $q = 0.5$ , then  $p + q = 1.1 > 1$ .

The statement ' $p - q < 0$ ' is possible. For example, if E and F are independent events with probabilities  $p = 0.5$  and  $q = 0.6$ , then  $p - q = -0.1 < 0$ .

The statement ' $p/q > 1$ ' is possible. For example, if E and F are independent events with probabilities  $p = 0.6$  and  $q = 0.5$ , then  $p/q = 1.2 > 1$ .

The statement 'E and F are both independent and mutually exclusive' is not possible. If events E and F are mutually exclusive, it means that they cannot occur together. If they are independent, it means that the occurrence of one event does not affect the occurrence of the other event. Therefore, E and F cannot be both independent and mutually exclusive.

Therefore, the impossible statement is D. Final Answer: D

\*\*\*\*\*

Answer Option 9:

If events E and F are independent, then the probability of both events occurring is given by the product of their individual probabilities, which means that  $p \times q$  is possible.

If events E and F are mutually exclusive, then they cannot occur at the same time, which means that the probability of both events occurring is zero. So,  $p \times q = 0$  is also possible.

Since p and q are nonzero probabilities, neither  $p \times q = 0$  nor  $p \times q$  can be impossible.

Now, let's consider the statements given:

A.  $p + q > 1$ : This is possible. The sum of probabilities can be greater than 1 when events are not mutually exclusive.

B.  $p - q < 0$ : This is also possible. The difference between two probabilities can be negative when p is smaller than q.

C.  $p / q > 1$ : This is also possible. The probability ratio can be greater than 1 when p is greater than q.

D. E and F are both independent and mutually exclusive: This is impossible. If two events are independent, they can occur at the same time, which contradicts the definition of mutually exclusive events.

Therefore, D is impossible.

Hence, the correct answer is D. Final answer: D

\*\*\*\*\*

Question 19:

The probability that there will be an accident on Highway 48 each day depends on the weather. If the weather is dry that day, there is a 0.2% chance of an accident on Highway 48; if the weather is wet that day, there is a 1.0% chance of an accident. Today, the weather station announced that there is a 20% chance of the weather being wet. What is the probability that there will be an accident on Highway 48 today?, A. '0.0004', B. '0.0016', C. '0.002', D. '0.0036'

=====

Answer Option 1:

Let  $A$  be the event that there is an accident on Highway 48, and let  $W$  be the event that the weather is wet.

We are given that  $P(W)=0.2$ ,  $P(A|W)=0.01$ , and  $P(A|\neg W)=0.002$  (where  $\neg W$  denotes the event that the weather is not wet, i.e. dry).

We want to find  $P(A)$ , which can be done using the law of total probability:

$$\begin{aligned} P(A) &= P(A|W) \cdot P(W) + P(A|\neg W) \cdot P(\neg W) \\ &= 0.01 \cdot 0.2 + 0.002 \cdot (1-0.2) \\ &= 0.002 + 0.002 \cdot 0.8 \\ &= 0.002 + 0.0016 \\ &= \boxed{0.0036}. \end{aligned}$$

$$P(A) = P(A|W) \cdot P(W) + P(A|\neg W) \cdot P(\neg W)$$

$$= 0.01 \cdot 0.2 + 0.002 \cdot (1-0.2)$$

$$= 0.002 + 0.002 \cdot 0.8$$

$$= 0.002 + 0.0016$$

$$= \boxed{0.0036}. \text{ Final Answer: D}$$

$$\end{aligned}$$

\*\*\*\*\*

Answer Option 2:

Let  $A$  be the event that there is an accident on Highway 48 and  $W$  be the event that the weather is wet. We want to find  $P(A)$ .

By the law of total probability, we have

$$P(A) = P(A|W)P(W) + P(A|\neg W)P(\neg W),$$

where  $\neg W$  denotes the event that the weather is not wet.

We are given that  $P(A|W) = 0.01$  (1.0% chance of an accident if the weather is wet) and  $P(A|\neg W) = 0.002$  (0.2% chance of an accident if the weather is dry).

It is also given that  $P(W) = 0.2$  (20% chance of the weather being wet). Therefore,  $P(\neg W) = 1 - P(W) = 0.8$ .

Substituting these values into the formula, we get

$$P(A) = (0.01)(0.2) + (0.002)(0.8) = 0.002 + 0.0016 = \boxed{0.0036}.$$

Final Answer: D

\*\*\*\*\*

Answer Option 3:

Let  $A$  be the event that there is an accident on Highway 48, and let  $W$  be the event that the weather is wet. We are given  $P(A|W) = 0.01$ ,  $P(A|\neg W) = 0.002$ , and  $P(W) = 0.2$ . We want to find  $P(A)$ .

Using Bayes' Theorem, we have

$$P(A) = P(A|W)P(W) + P(A|\neg W)P(\neg W).$$

Substituting the given values, we get

$$P(A) = (0.01)(0.2) + (0.002)(0.8) = 0.002 + 0.0016 = \boxed{0.0036}.$$

Final Answer: D

\*\*\*\*\*

Answer Option 4:

Let  $A$  be the event that there is an accident on Highway 48 and  $W$  be the event that the weather is wet. We are given that  $P(A|W) = 0.01$  and  $P(W) = 0.2$ . We want to find  $P(A)$ .

We can use Bayes' theorem to find  $P(A)$ :

$$P(A) = P(A|W)P(W) + P(A|\neg W)P(\neg W),$$

where  $\neg W$  is the event that the weather is dry.

We are given that  $P(A|\neg W) = 0.002$  and  $P(\neg W) = 1 - P(W) = 0.8$ . Plugging in these values, we get:

$$P(A) = 0.01 \cdot 0.2 + 0.002 \cdot 0.8 = 0.002 + 0.0016 = \boxed{0.0036}.$$

Final Answer: D

\*\*\*\*\*

Answer Option 5:

Let  $A$  be the event that there will be an accident on Highway 48 and  $W$  be the event that the weather is wet. We are given  $P(W) = 0.2$ ,  $P(A | W) = 0.01$ , and  $P(A | \neg W) = 0.002$ , where  $\neg W$  is the complement of  $W$  (i.e. the weather is dry). We want to find  $P(A)$ .

We can use the law of total probability to express  $P(A)$  in terms of the conditional probabilities:

$$P(A) = P(A | W) \cdot P(W) + P(A | \neg W) \cdot P(\neg W).$$

Substituting the given values, we have:

$$P(A) = 0.01 \cdot 0.2 + 0.002 \cdot 0.8 = 0.002 + 0.0016 = \boxed{0.0036}.$$

Final Answer: D

\*\*\*\*\*

Answer Option 6:

Let  $A$  be the event that there is an accident on Highway 48 and  $W$  be the event that the weather is wet. We want to find  $P(A)$ , the probability that there will be an accident on Highway 48 today.

We are given  $P(A|W) = 0.01$ , the probability of an accident given wet weather, and  $P(W) = 0.2$ , the probability of wet weather.

Using Bayes' theorem, we can calculate  $P(A)$  as follows:

$$P(A) = P(A|W) \cdot P(W) + P(A|W^c) \cdot P(W^c),$$

where  $W^c$  denotes the complement of  $W$ , i.e., dry weather.

We are also given  $P(A|W^c) = 0.002$ , the probability of an accident given dry weather. Since the probabilities of wet and dry weather must add up to 1, we have  $P(W^c) = 1 - P(W) = 0.8$ .

Plugging in these values, we get:

$$P(A) = (0.01) \cdot (0.2) + (0.002) \cdot (0.8) = 0.002 + 0.0016 = \boxed{0.0036}.$$

Final Answer: D

\*\*\*\*\*

Answer Option 7:

Let  $A$  be the event that there is an accident on Highway 48, and let  $W$  be the event that the weather is wet. We want to find  $P(A)$ .

We can apply the law of total probability:

$$P(A) = P(A|W) \cdot P(W) + P(A|W^c) \cdot P(W^c),$$

where  $W^c$  represents the event that the weather is dry.

Given that the weather is wet, we have  $P(A|W) = 0.01$ . The probability that the weather is wet is  $P(W) = 0.2$ .

Given that the weather is dry, we have  $P(A|W^c) = 0.002$ , as there is a 0.2% chance of an accident on a dry day. The probability that the weather is dry is  $P(W^c) = 0.8$  (since there are only two possibilities: wet or dry).

Substituting these values, we have:

$$P(A) = 0.01 \cdot 0.2 + 0.002 \cdot 0.8 = 0.002.$$

Therefore, the probability that there will be an accident on Highway 48 today is  $\boxed{0.002}$ . Final answer: C

\*\*\*\*\*

Answer Option 8:

Let's calculate the probability that there will be an accident on Highway 48 today using the law of total probability.

Let  $A$  be the event that there is an accident on Highway 48, and let  $W$  be the event that the weather is wet today. We want to find  $P(A)$ .

We know that  $P(W) = 0.20$ ,  $P(A|W) = 0.01$ , and  $P(A|W^c) = 0.002$ . Here,  $W^c$  denotes the event that the weather is dry.

By the law of total probability,

$$P(A) = P(A|W)P(W) + P(A|W^c)P(W^c) = 0.01 \cdot 0.20 + 0.002 \cdot 0.80 = 0.002.$$

$$P(A) = P(A|W)P(W) + P(A|W^c)P(W^c) = 0.01 \cdot 0.20 + 0.002 \cdot 0.80 = 0.002.$$

Therefore, the probability that there will be an accident on Highway 48 today is  $\boxed{\text{(C) } 0.002}$ .

\*\*\*\*\*

Answer Option 9:

Let  $A$  be the event that there is an accident on Highway 48, and  $W$  be the event that the weather is wet. We want to find the probability  $P(A)$ .

We can use the law of total probability to express  $P(A)$  as:

$$P(A) = P(A|W) \cdot P(W) + P(A|W^c) \cdot P(W^c),$$

where  $W^c$  is the event that the weather is dry.

Given that the weather is wet, the probability of an accident is 1.0%, or 0.01. Given that the weather is dry, the probability of an accident is 0.2%, or 0.002. Additionally, we are given that the probability of the weather being wet,  $P(W)$ , is 20%, or 0.2.

Substituting these values into the equation, we have:

$$P(A) = 0.01 \cdot 0.2 + 0.002 \cdot 0.8 = 0.002 + 0.0016 = 0.0036.$$

Therefore, the probability that there will be an accident on Highway 48 today is  $\boxed{\text{(D) } 0.0036}$ . Final Answer: D

\*\*\*\*\*

Question 20:

'The binomial distribution is an appropriate model for which of the following?', A. 'The number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day.', B. 'The number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day.', C. 'The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock.', D. 'None of the above.'

=====

Answer Option 1:

The binomial distribution is an appropriate model when we have a fixed number of independent trials, each with the same probability of success.

Option A: "The number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day." This does not fit the binomial distribution because the number of minutes that the Dow-Jones average is above its beginning average is not a fixed number of trials. It could vary from day to day.

Option B: "The number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day." This also does not fit the binomial distribution because the probability of a city having cloudy weather is likely to differ between cities. The success probability is not the same for each trial.

Option C: "The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock." This is a situation where the binomial distribution is appropriate. We have a fixed number of trials (10 drivers), and we can assume that the probability of a driver wearing a seat belt is the same for each driver.

Therefore, the binomial distribution is an appropriate model for option C. The answer is  $\boxed{\text{C}}$

\*\*\*\*\*

Answer Option 2:

The binomial distribution is appropriate when the following conditions are met:

1. The number of trials is fixed.
2. Each trial is independent.
3. Each trial has only two possible outcomes: success or failure.
4. The probability of success is constant for each trial.

Now let's analyze the given options:

A. "The number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day." This does not meet the condition of having only two possible outcomes, as the number of minutes above the average can vary continuously. Therefore, the binomial distribution is not appropriate for this scenario.

B. "The number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day." This also does not meet the condition of having only two possible outcomes, as the number of cities with cloudy weather can vary.

C. "The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock." This scenario fits all the conditions for a binomial distribution. The number of drivers wearing seat belts can only have two outcomes (wearing seat belt or not wearing seat belt), the trials are independent, and the probability of wearing a seat belt is constant for each driver.

Therefore, the correct answer is  $\boxed{\text{C}}$  'The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock.'\$. Final answer: C

\*\*\*\*\*

Answer Option 3:

The binomial distribution is most appropriate for scenarios that satisfy the following conditions: fixed number of trials, each trial has two possible outcomes (success or failure), the probability of success is constant for each trial, and the trials are independent.

In option A, the number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day does not fit the conditions for a binomial distribution. There can be any number of minutes above the average, not just two possible outcomes.

In option B, the number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day does not fit the conditions for a binomial distribution. The number of cities can vary and is not a fixed number of trials.

In option C, the number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock does fit the conditions for a binomial distribution. There are fixed number of trials (10 drivers), two possible outcomes (wearing seat belt or not), the probability of success (wearing seat belt) is constant for each trial, and the trials are independent.

Therefore, the binomial distribution is appropriate for option C. The answer is  $\boxed{\text{C}}$ .

\*\*\*\*\*

Answer Option 4:

The binomial distribution is an appropriate model when the following conditions are met:

- 1) There are a fixed number of trials,
- 2) Each trial has only two possible outcomes (success or failure),
- 3) The probability of success is constant for each trial, and
- 4) The trials are independent of each other.

Looking at the options:

A. The number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day does not meet the condition of having only two possible outcomes.

B. The number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day does not meet the condition of having a fixed number of trials.

C. The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock meets all the conditions of the binomial distribution.

Therefore, the answer is  $\boxed{\text{C}}$  The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock}\$. Final answer: C

\*\*\*\*\*

Answer Option 5:

The binomial distribution is an appropriate model when we have a fixed number of independent trials, each with the same probability of success.

In option A, we have a continuous variable (number of minutes) and the probability of the Dow-Jones average being above its beginning average is likely not the same for each minute, so the binomial distribution is not appropriate.

In option B, we again have a continuous variable (number of cities) and the probability of the weather being cloudy is likely not the same for each city, so the binomial distribution is not appropriate.

In option C, we have a fixed number of independent trials (10 consecutive drivers) and each driver has the same probability of wearing a seat belt, so the binomial distribution is appropriate.

Therefore, the correct answer is  $\boxed{\text{(C) The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock.}}$ . Final answer: C

\*\*\*\*\*

Answer Option 6:

The binomial distribution is an appropriate model for situations that satisfy the following conditions:

1. Each trial can result in one of two outcomes, often referred to as success or failure.
2. The probability of success remains constant from trial to trial.
3. The trials are independent of each other.

Looking at the options,

A. The number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day does not fit the criteria. The outcomes are not just success or failure, but rather a range of values.

B. The number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day also does not fit the criteria. The probability of cloudy weather is not constant and the outcomes are not simply success or failure.

C. The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock does fit the criteria. Each driver can either wear a seat belt or not, the probability remains constant, and the trials are independent.

Therefore, the binomial distribution is an appropriate model for option C. The answer is  $\boxed{\text{(C.)}}$ . Final answer: C

\*\*\*\*\*

Answer Option 7:

The binomial distribution is an appropriate model for situations that satisfy the following conditions: (1) there are a fixed number of independent trials, (2) each trial has exactly two possible outcomes (success or failure), (3) the probability of success is constant for each trial, and (4) the trials are independent of each other.

Among the given options, only option C satisfies these conditions. In option C, there are 10 consecutive drivers, each with the independent trial of wearing a seat belt (success) or not wearing a seat belt (failure), and the probability of wearing a seat belt is constant for each driver. Therefore, the number of drivers wearing seat belts follows a binomial distribution.

Hence, the correct answer is  $\boxed{\text{(C. The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock.)}}$

\*\*\*\*\*

Answer Option 8:

The binomial distribution is an appropriate model when we have a fixed number of independent trials, each with the same probability of success.

A. The number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day does not have a fixed number of trials and the probabilities of success are likely to be different for each minute. Therefore, the binomial distribution is not an appropriate model for this scenario.

B. The number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day also does not have a fixed number of trials and the probabilities of success are likely to be different for each city. Thus, the binomial distribution is not an appropriate model for this situation.

C. The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock has a fixed number of trials (10 drivers) and each driver is considered independently with the same probability of success (wearing a seat belt or not). Therefore, the binomial distribution is an appropriate model for this case.

Therefore, the correct answer is  $\boxed{\text{(C.)}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 9:

The binomial distribution is an appropriate model for situations that satisfy the following conditions: 1) there are a fixed number of trials, 2) each trial can have only two outcomes (success or failure), 3) the probability of success remains constant for each trial, and 4) the trials are independent of each other.

In option A, the number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day does not fit the conditions of a fixed number of trials or two outcomes. Therefore, option A is not an appropriate model for the binomial distribution.

In option B, the number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day does not fit the condition of a fixed number of trials. The number of cities for which the weather is cloudy can vary from day to day. Therefore, option B is not an appropriate model for the binomial distribution.

In option C, the number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock fits all the conditions of the binomial distribution. There are a fixed number of trials (10 drivers), each driver either wears a seat belt or doesn't (two outcomes), the probability of a driver wearing a seat belt is assumed to be constant for each driver, and the trials are independent of each other. Therefore, option C is an appropriate model for the binomial distribution.

Therefore, the correct answer is  $\boxed{\text{(C.)}}$

\*\*\*\*\*

Question 21:

Data were collected on two variables X and Y and a least squares regression line was fitted to the data. The estimated equation for this data is  $y = -2.29 + 1.70x$ . One point has  $x = 5$ ,  $y = 6$ . What is the residual for this point?, A. '7.91', B. '6.21', C. '0.21', D. '-0.21'

=====

Answer Option 1:

The equation of the regression line is given as  $y = -2.29 + 1.70x$ . To find the residual for a specific point, we need to find the difference between the actual y-value and the predicted y-value.

Given that the point has  $x = 5$  and  $y = 6$ , we can substitute these values into the equation to find the predicted y-value:

$$y = -2.29 + 1.70(5) = 6.21$$

The residual is then the difference between the actual y-value and the predicted y-value:

$$\text{Residual} = \text{actual y-value} - \text{predicted y-value} = 6 - 6.21 = -0.21$$

Therefore, the residual for this point is  $\boxed{\text{D. } -0.21}$ .

\*\*\*\*\*

Answer Option 2:

To find the residual for a given point, we need to subtract the predicted value of y from the actual value of y.

The predicted value of y for  $x = 5$  can be found by substituting  $x = 5$  into the equation  $y = -2.29 + 1.70x$ :

$$y = -2.29 + 1.70(5) = -2.29 + 8.50 = 6.21$$

The actual value of y is given as  $y = 6$ .

Therefore, the residual for this point is:

$$\text{Residual} = \text{Actual } y - \text{Predicted } y = 6 - 6.21 = -0.21$$

So, the residual for this point is -0.21. The correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The residual is the difference between the observed value and the predicted value. In this case, the observed value is  $y = 6$  when  $x = 5$ . Plugging these values into the estimated equation  $y = -2.29 + 1.70x$ , we get:

$$y = -2.29 + 1.70 \cdot 5 = -2.29 + 8.50 = 6.21$$

Therefore, the residual is the difference between the observed value  $6$  and the predicted value  $6.21$ , which is  $6 - 6.21 = -0.21$ . The residual for this point is  $\boxed{\text{D. } -0.21}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

The residual for a data point is the difference between the actual y-value and the predicted y-value based on the regression line. In this case, the estimated equation is  $y = -2.29 + 1.70x$ , and we have a data point with  $x = 5$  and  $y = 6$ .

Substituting  $x = 5$  into the equation, we get  $y = -2.29 + 1.70(5) = -2.29 + 8.50 = 6.21$ .

Therefore, the residual for this point is  $6 - 6.21 = \boxed{-0.21}$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

The residual is calculated by subtracting the predicted value from the actual value. In this case, the predicted value can be found using the estimated equation:

$$y_{\text{predicted}} = -2.29 + 1.70 \cdot x$$

Plugging in  $x = 5$ , we get:

$$y_{\text{predicted}} = -2.29 + 1.70 \cdot 5 = -2.29 + 8.50 = 6.21$$

The actual value of y is given as 6.

Therefore, the residual is:

$$\text{residual} = \text{actual value} - \text{predicted value} = 6 - 6.21 = -0.21$$

Thus, the residual for this point is -0.21. The correct answer is D.

\*\*\*\*\*

Answer Option 6:

The residual for a data point is the difference between the observed value and the predicted value. In this case, the observed value is  $y = 6$  and the predicted value can be found by plugging  $x = 5$  into the regression equation:

$$y = -2.29 + 1.70x$$

$$y = -2.29 + 1.70(5)$$

$$y = -2.29 + 8.50$$

$$y = 6.21$$

The residual is the difference between the observed value (6) and the predicted value (6.21):

$$\text{residual} = 6 - 6.21 = -0.21$$

Therefore, the residual for this point is -0.21. Final Answer: D

\*\*\*\*\*

Answer Option 7:

To find the residual for a given point, we need to compute the difference between the observed value of y and the predicted value of y based on the regression line.

Given that the estimated equation is  $y = -2.29 + 1.70x$ , we can substitute  $x = 5$  into the equation to find the predicted value of y.

$$y = -2.29 + 1.70(5) = -2.29 + 8.50 = 6.21$$

The observed value of y is 6, so the residual is the difference between the observed value and the predicted value:

$$\text{Residual} = \text{Observed } y - \text{Predicted } y = 6 - 6.21 = -0.21$$

Therefore, the residual for this point is -0.21. The correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The residual is the difference between the observed value of the response variable and the predicted value from the regression line. In this case, the observed values are  $x = 5$  and  $y = 6$ .

Using the estimated equation  $y = -2.29 + 1.70x$ , we can calculate the predicted value of y for  $x = 5$ :

$$y = -2.29 + 1.70(5) = -2.29 + 8.50 = 6.21$$

Therefore, the residual is the difference between the observed value of y and the predicted value of y:

$$\text{residual} = \text{observed } y - \text{predicted } y = 6 - 6.21 = -0.21$$

So the residual for this point is -0.21. The correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The equation for the least squares regression line is given as  $y = -2.29 + 1.70x$ . To find the residual for a given point, we need to calculate the difference between the actual y-value and the predicted y-value based on the regression line.

For the point with  $x = 5$  and  $y = 6$ , we can substitute  $x = 5$  into the regression equation to find the predicted y-value:

$\begin{aligned} & \end{aligned}$

$$y = -2.29 + 1.70x$$

$$y = -2.29 + 1.70(5)$$

$$y = -2.29 + 8.50$$

$$y = 6.21$$

$\end{aligned}$

The actual y-value is 6, and the predicted y-value is 6.21. Therefore, the residual for this point is given by:

$\begin{aligned} & \end{aligned}$

$$\text{Residual} = \text{Actual } y\text{-value} - \text{Predicted } y\text{-value}$$

$$\text{Residual} = 6 - 6.21$$

$$\text{Residual} = -0.21$$

$\end{aligned}$

The residual for this point is -0.21. Therefore, the correct answer is  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*



Question 22:

'Which of these explains why t should be used instead of z for inference procedures for means.', A. 'The Central Limit Theorem applies to proportions but not means.', B. 'We use z for proportions because proportions are approximately normal. Means are not.', C. 'We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means.', D. 'When using the sample standard deviation to estimate the population standard deviation, more variability is introduced into the sampling distribution of the statistic.'

=====

Answer Option 1:

The correct answer is C. We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means. In other words, the Central Limit Theorem applies to sample means, making the use of t appropriate for inference procedures for means. Final answer: C.

\*\*\*\*\*

Answer Option 2:

The correct answer is C. We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means. In other words, the Central Limit Theorem applies to sample means, not proportions. When the sample size is large enough, the sampling distribution of sample means becomes approximately normal, allowing us to use z-scores for inference procedures. Final answer: C.

\*\*\*\*\*

Answer Option 3:

The correct explanation for why t should be used instead of z for inference procedures for means is option D: "When using the sample standard deviation to estimate the population standard deviation, more variability is introduced into the sampling distribution of the statistic."

When we estimate the population standard deviation using the sample standard deviation, there is more uncertainty involved compared to when we know the population standard deviation. This increased variability affects the sampling distribution of the mean, making it less normal and more spread out. As a result, we use the t-distribution instead of the standard normal distribution (z-distribution) to account for this additional variability and obtain more accurate inference procedures for means.

Therefore, the correct answer is D.

\*\*\*\*\*

Answer Option 4:

The correct answer is C. We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means. When the sample size is large, the sampling distribution of the sample mean becomes approximately normal due to the Central Limit Theorem. Therefore, we use t instead of z for inference procedures for means. Final answer: C.

\*\*\*\*\*

Answer Option 5:

The correct answer is C. We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means.

When the sample size is large (typically considered to be at least 30), the sampling distribution of the sample mean approaches a normal distribution, regardless of the shape of the population distribution, due to the Central Limit Theorem. Therefore, we can use z-scores for inference procedures for means. Final answer: C.

\*\*\*\*\*

Answer Option 6:

The correct explanation is C. We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means.

In other words, when we have a large sample size and the population standard deviation is known, we can use the z-distribution for inference procedures for proportions. However, when it comes to means, we typically don't have the population standard deviation and the sampling distribution of sample means is not always normal. Therefore, we use t-distribution for inference procedures for means. Final answer: C

\*\*\*\*\*

Answer Option 7:

The correct answer is C. We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means. For inference procedures involving means, we use the t-distribution instead of the standard normal distribution. The t-distribution accounts for the additional variability introduced by estimating the population standard deviation from the sample standard deviation. The Central Limit Theorem applies to both proportions and means, so option A is incorrect. Option B is incorrect because means are approximately normally distributed due to the Central Limit Theorem. Option D is incorrect because it is not directly related to the choice between using t or z for inference procedures. Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct answer is C.

We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means. This is due to the Central Limit Theorem, which states that as the sample size increases, the sampling distribution of the sample mean approaches a normal distribution. Therefore, for inference procedures involving means, we use t instead of z.

Answer C correctly explains why t should be used instead of z for inference procedures for means. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct answer is C. We use z for proportions because the sampling distribution of sample proportions is approximately normal, but that is not the case for the sampling distribution of sample means.

When dealing with proportions, the Central Limit Theorem applies, and the sampling distribution of sample proportions can be approximated by a normal distribution. This allows us to use z-scores to perform inference procedures.

However, when dealing with means, the Central Limit Theorem also applies, but the sampling distribution of sample means is approximately normal regardless of the shape of the population distribution. Therefore, we can use t-scores instead of z-scores for inference procedures for means.

Option A is incorrect because the Central Limit Theorem applies to both proportions and means. Option B is incorrect because means can also be approximately normal under certain conditions. Option D is unrelated to the choice between t and z for inference procedures.

\*\*\*\*\*

Question 23:

'Which of the following is least likely to reduce bias in a sample survey?', A. 'Following up with those who did not respond to the survey the first time', B. 'Asking questions in a neutral manner to avoid influencing the responses', C. 'Using stratified random sampling rather than simple random sampling', D. 'Selecting samples randomly'

=====

Answer Option 1:

The goal of reducing bias in a sample survey is to make the sample representative of the population of interest. Option D, 'Selecting samples randomly', is actually a method that helps reduce bias because it ensures that every member of the population has an equal chance of being included in the sample. This helps to minimize any bias that could be introduced by selectively choosing certain individuals.

On the other hand, options A, B, and C are all methods that can help reduce nonresponse bias, interviewer bias, and selection bias, respectively. Following up with non-respondents, asking neutral questions, and using stratified random sampling are all techniques specifically designed to minimize bias in the survey process.

Therefore, the answer is  $\boxed{\text{D. 'Selecting samples randomly'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 2:

The goal of reducing bias in a sample survey is to ensure that the sample is representative of the population being studied. Option A, following up with those who did not respond to the survey the first time, can help reduce nonresponse bias and increase the response rate, thereby improving the representativeness of the sample. Option B, asking questions in a neutral manner, helps minimize response bias by preventing the survey questions from influencing the responses. Option D, selecting samples randomly, is a fundamental principle of sampling design to ensure that every member of the population has an equal chance of being included in the sample, which helps reduce selection bias.

Option C, using stratified random sampling rather than simple random sampling, is least likely to reduce bias in a sample survey. Stratified random sampling involves dividing the population into distinct groups or strata and then randomly selecting samples from each stratum. While stratified random sampling can improve precision and ensure representation from different subgroups, it may not necessarily reduce bias. If the stratification is not properly done or if certain strata are underrepresented or overrepresented, it can introduce bias into the sample.

Therefore, the answer is C. 'Using stratified random sampling rather than simple random sampling'. Final Answer: C

\*\*\*\*\*

Answer Option 3:

The question is asking for the option that is least likely to reduce bias in a sample survey. In order to reduce bias, it is important to have a representative sample that accurately reflects the population of interest.

Option A, "Following up with those who did not respond to the survey the first time," can help reduce non-response bias by increasing the response rate and ensuring a more representative sample.

Option B, "Asking questions in a neutral manner to avoid influencing the responses," is important to minimize response bias by ensuring that respondents are not influenced by the wording or phrasing of the questions.

Option C, "Using stratified random sampling rather than simple random sampling," can help reduce bias by ensuring that different subgroups of the population are represented proportionally in the sample.

Option D, "Selecting samples randomly," is a fundamental principle in survey sampling that helps reduce bias by ensuring that every member of the population has an equal chance of being selected.

Out of these options, the least likely to reduce bias in a sample survey is option C, "Using stratified random sampling rather than simple random sampling." While stratified random sampling can help ensure proportional representation of different subgroups, it does not guarantee a reduction in bias compared to simple random sampling. In some cases, simple random sampling may be more appropriate or effective in reducing bias.

Therefore, the correct answer is option C.

\*\*\*\*\*

Answer Option 4:

The least likely to reduce bias in a sample survey is C. 'Using stratified random sampling rather than simple random sampling'.

Stratified random sampling is a method of sampling where the population is divided into different groups or strata, and then a random sample is taken from each stratum. This is done to ensure that each stratum is represented in the sample proportionally to its size in the population.

While stratified random sampling can help improve the precision and representativeness of the sample, it does not necessarily reduce bias. Bias can still exist within each stratum, and if the stratification is not done properly, it can even introduce additional bias into the sample.

On the other hand, options A, B, and D are all strategies that can help reduce bias in a sample survey. Following up with non-respondents helps to reduce non-response bias, asking questions in a neutral manner helps to reduce response bias, and selecting samples randomly helps to ensure that the sample is representative of the population.

Therefore, the least likely to reduce bias in a sample survey is C. 'Using stratified random sampling rather than simple random sampling'. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The least likely option to reduce bias in a sample survey is option C, 'Using stratified random sampling rather than simple random sampling'.

Stratified random sampling is a technique where the population is divided into subgroups or strata based on certain characteristics, and then samples are randomly selected from each stratum. This method aims to ensure that each subgroup is adequately represented in the sample.

While stratified random sampling can be an effective method to reduce bias in certain situations, it is not always the best approach. In some cases, using stratified sampling can introduce bias if the stratification variables are not well-chosen or if there is a lack of accurate information about the population characteristics.

On the other hand, options A, B, and D are all generally effective in reducing bias. Following up with non-respondents helps to ensure a more representative sample, asking questions in a neutral manner avoids influencing responses, and selecting samples randomly helps to ensure that the sample is representative of the population.

Therefore, the answer is option C, 'Using stratified random sampling rather than simple random sampling'. Final Answer: C

\*\*\*\*\*

Answer Option 6:

The goal of reducing bias in a sample survey is to obtain a representative sample that accurately reflects the population of interest.

Option A, 'Following up with those who did not respond to the survey the first time', is likely to reduce bias because it helps ensure a higher response rate and reduces non-response bias.

Option B, 'Asking questions in a neutral manner to avoid influencing the responses', is also likely to reduce bias because it minimizes the potential for response bias caused by leading or suggestive questions.

Option C, 'Using stratified random sampling rather than simple random sampling', is a method that can help reduce bias by ensuring representation from different subgroups within the population. This is generally more effective at reducing bias compared to simple random sampling.

Option D, 'Selecting samples randomly', is a fundamental principle in survey sampling and is crucial to avoid selection bias. Random sampling helps ensure that every individual in the population has an equal chance of being selected, reducing bias.

Therefore, the option that is least likely to reduce bias in a sample survey is option C, 'Using stratified random sampling rather than simple random sampling'. This is because stratified random sampling is more effective at reducing bias compared to simple random sampling. The correct answer is C. Final answer: C

\*\*\*\*\*

Answer Option 7:

To reduce bias in a sample survey, it is important to have a representative sample that accurately reflects the population being studied. Option A, 'Following up with those who did not respond to the survey the first time', can help reduce nonresponse bias by increasing the response rate and ensuring a more representative sample. Option B, 'Asking questions in a neutral manner to avoid influencing the responses', helps reduce response bias by minimizing the impact of question wording or phrasing on participants' responses. Option D, 'Selecting samples randomly', is a fundamental principle of sampling to avoid selection bias and ensure that each member of the population has an equal chance of being included in the sample.

However, option C, 'Using stratified random sampling rather than simple random sampling', is least likely to reduce bias in a sample survey. Stratified random sampling involves dividing the population into subgroups or strata and then randomly selecting samples from each stratum. This method helps ensure that each subgroup is adequately represented in the sample, which is important when the population is heterogeneous. Therefore, using stratified random sampling is more likely to reduce bias compared to simple random sampling.

Therefore, the answer is  $\boxed{\text{C. 'Using stratified random sampling rather than simple random sampling'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 8:

The goal of reducing bias in a sample survey is to obtain a representative sample that accurately reflects the population being studied. To achieve this, it is important to minimize sources of bias and ensure that the sample is selected randomly.

Option A, 'Following up with those who did not respond to the survey the first time,' is a method known as nonresponse follow-up. This is a good practice as it helps to reduce nonresponse bias. By reaching out to individuals who did not initially respond, the survey is more likely to have a higher response rate and a more representative sample. Therefore, Option A is likely to reduce bias in the survey.

Option B, 'Asking questions in a neutral manner to avoid influencing the responses,' is an important step in survey design. By phrasing questions neutrally and avoiding leading or suggestive language, respondents are more likely to provide unbiased answers. Therefore, Option B is likely to reduce bias in the survey.

Option C, 'Using stratified random sampling rather than simple random sampling,' is a technique used to ensure that the sample represents various subgroups within the population. This can help reduce bias by ensuring that each subgroup is adequately represented. Therefore, Option C is likely to reduce bias in the survey.

Option D, 'Selecting samples randomly,' is a fundamental principle of sampling. Random sampling helps to ensure that every individual in the population has an equal chance of being selected for the sample. This helps to minimize selection bias and obtain a representative sample. Therefore, Option D is a crucial step in reducing bias in the survey.

Given these explanations, the option that is least likely to reduce bias in a sample survey is  $\boxed{\text{D. 'Selecting samples randomly'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

The question asks for the option that is least likely to reduce bias in a sample survey. To reduce bias, it is important to have a representative sample that accurately reflects the population of interest.

Option A, 'Following up with those who did not respond to the survey the first time', can help reduce non-response bias by increasing the response rate. This allows for a more representative sample by reaching out to individuals who may have different characteristics or opinions.

Option B, 'Asking questions in a neutral manner to avoid influencing the responses', is also important to reduce response bias. By phrasing questions neutrally, respondents are less likely to be influenced or biased in their responses.

Option C, 'Using stratified random sampling rather than simple random sampling', is a method that can help reduce bias by ensuring that different subgroups within the population are adequately represented in the sample. This can help capture the diversity of the population and reduce the chance of under or over-representing certain groups.

Option D, 'Selecting samples randomly', is a fundamental principle of sampling that helps reduce selection bias. Random sampling ensures that every individual in the population has an equal chance of being selected for the survey, which can help minimize bias.

Out of the given options, the option that is least likely to reduce bias in a sample survey is option C, 'Using stratified random sampling rather than simple random sampling'. Stratified random sampling, when done correctly, can help reduce bias by ensuring adequate representation of different groups. Therefore, the correct answer is C. Final Answer: C

\*\*\*\*\*

Question 24:

'For their first exam, students in an AP Statistics class studied an average of 4 hours with a standard deviation of 1 hour. Almost everyone did poorly on the exam, and so for the second exam every student studied 10 hours. What is the correlation between the numbers of hours students studied for each exam?', A. '-1', B. '0', C. '0.4', D. '1'

=====

Answer Option 1:

The correlation coefficient measures the strength and direction of a linear relationship between two variables. In this case, the two variables are the number of hours studied for the first exam and the number of hours studied for the second exam.

Since almost everyone studied for 4 hours for the first exam and 10 hours for the second exam, there is a perfect positive linear relationship between the two variables. This means that the correlation coefficient is  $\boxed{1}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

The correlation coefficient measures the linear relationship between two variables. In this case, the variables are the number of hours studied for the first exam and the number of hours studied for the second exam.

Since almost everyone studied the same number of hours for the second exam, there is no variability in the number of hours studied for the second exam. Therefore, there is no linear relationship between the two variables, and the correlation coefficient is  $\boxed{0}$ . Final Answer: B

\*\*\*\*\*

Answer Option 3:

The correlation coefficient measures the strength and direction of the linear relationship between two variables. In this case, the two variables are the number of hours studied for the first exam and the number of hours studied for the second exam.

Since almost everyone studied 10 hours for the second exam, there is no variability in the second variable. Therefore, there is no linear relationship between the two variables, and the correlation coefficient is 0.

Therefore, the answer is  $\boxed{\text{B. } 0}$ . Final Answer: B

\*\*\*\*\*

Answer Option 4:

The correlation coefficient measures the strength and direction of the linear relationship between two variables. In this case, the two variables are the number of hours studied for the first exam and the number of hours studied for the second exam.

Since almost everyone studied for 4 hours for the first exam and 10 hours for the second exam, there is no variation in the number of hours studied for the second exam. Therefore, the correlation between the two variables is zero because there is no linear relationship between them. Hence, the answer is  $\boxed{\text{B. } 0}$ . Final answer: B

\*\*\*\*\*

Answer Option 5:

The correlation coefficient measures the strength and direction of the linear relationship between two variables. In this case, the two variables are the number of hours students studied for each exam.

Since almost everyone studied the same amount for the second exam, there is no variability in the hours studied for the second exam. Thus, the second variable is constant and has no variability.

The correlation coefficient is calculated as the covariance of the two variables divided by the product of their standard deviations. In this case, since the second variable has no standard deviation, the correlation coefficient is undefined.

Therefore, the answer is not one of the given choices.

\*\*\*\*\*

Answer Option 6:

The correlation between the number of hours studied for each exam can be estimated using the formula:

$$r = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

where  $\text{Cov}(X,Y)$  is the covariance between the number of hours studied for each exam,  $\sigma_X$  is the standard deviation of the number of hours studied for the first exam, and  $\sigma_Y$  is the standard deviation of the number of hours studied for the second exam.

Since every student studied the same number of hours for the second exam,  $\sigma_Y$  is zero. This means that the correlation  $r$  is undefined.

Therefore, the answer is not one of the given choices.

\*\*\*\*\*

Answer Option 7:

The correlation coefficient is a measure of the linear relationship between two variables. In this case, the two variables are the number of hours studied for the first exam and the number of hours studied for the second exam.

Since almost everyone did poorly on the first exam, there is likely to be a negative relationship between the number of hours studied for the first exam and the number of hours studied for the second exam.

This is because students who studied more for the first exam would have been more likely to do well, and therefore would not feel the need to study as much for the second exam.

However, the correlation coefficient is not affected by the scale or units of measurement. Therefore, the correlation coefficient between the number of hours studied for the first exam and the number of hours studied for the second exam is always 1.

Therefore, the correct answer is  $\boxed{\text{D. } 1}$ .

\*\*\*\*\*

Answer Option 8:

Correlation measures the linear relationship between two variables. In this case, the variables are the number of hours studied for the first exam and the number of hours studied for the second exam.

Since almost everyone studied 4 hours for the first exam, the data is not spread out and the standard deviation is low. However, for the second exam, everyone studied 10 hours, so the data is not spread out either and the standard deviation is low as well.

Because the data for both exams is concentrated around a single value with low standard deviation, there is no linear relationship between the number of hours studied for the first exam and the number of hours studied for the second exam. Therefore, the correlation is  $\boxed{0}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

The correlation between the numbers of hours students studied for each exam can be calculated using the formula:

$$\text{correlation} = \frac{\text{covariance}}{\text{product of standard deviations}}.$$

Since every student studied the same number of hours for the second exam (10 hours), the standard deviation for the second exam is 0. The covariance between the numbers of hours studied for each exam is also 0 because there is no relationship between the two variables. Therefore, the correlation is:

$$\text{correlation} = \frac{0}{1 \cdot 0} = \text{undefined}.$$

Since the correlation is undefined, the correct answer is not listed among the choices.

\*\*\*\*\*

Question 25:

The appraised values of houses in a city have a mean of \$125,000 with a standard deviation of \$23,000. Because of a new teachers' contract, the school district needs an extra 10% in funds compared to the previous year. To raise this additional money, the city instructs the assessment office to raise all appraised house values by \$5,000. What will be the new standard deviation of the appraised values of houses in the city?', A. '\$23,000', B. '\$25,300', C. '\$28,000', D. '\$30,300'

=====

Answer Option 1:

The new mean of the appraised values of houses is  $\$125,000 + \$5,000 = \$130,000$ . The new standard deviation can be found using the following formula for a linear transformation:

If  $Y$  is a random variable with mean  $\mu$  and standard deviation  $\sigma$ , then  $aX + b$  has mean  $a\mu + b$  and standard deviation  $a\sigma$ , where  $a$  and  $b$  are constants.

In this case,  $a = 1$  and  $b = \$5,000$ . So, the new standard deviation is  $1 \cdot \$23,000 = \$23,000$ . Final Answer: A

\*\*\*\*\*

Answer Option 2:

Adding a fixed value of \$5,000 to each house value does not change the standard deviation. Therefore, the new standard deviation will still be \$23,000. Final Answer: A

\*\*\*\*\*

Answer Option 3:

Adding a constant value to each data point does not change the standard deviation. Therefore, the new standard deviation of the appraised house values will still be \$23,000. Final Answer: A

\*\*\*\*\*

Answer Option 4:

When a constant value is added to each data point in a data set, the standard deviation remains unchanged. Therefore, the new standard deviation of the appraised values of houses in the city will still be

\$23,000. Final Answer: A

\*\*\*\*\*

Answer Option 5:

When a constant value is added to each data point, such as adding \$5,000 to each appraised house value, the mean is also increased by that constant value. In this case, the new mean will be  $\$125,000 + \$5,000 = \$130,000$ .

However, adding a constant value does not affect the standard deviation of the data. Therefore, the new standard deviation of the appraised house values will still be \$23,000. Final answer: A

\*\*\*\*\*

Answer Option 6:

When a constant  $k$  is added to every data point in a data set, the mean is increased by  $k$ , but the standard deviation remains the same. Thus, the new mean of the appraised values of the houses will be  $\$125,000 + \$5,000 = \$130,000$ . Since the standard deviation remains the same, the new standard deviation will still be \$23,000.

Final answer: A

\*\*\*\*\*

Answer Option 7:

When a constant is added to each value in a data set, the mean of the data set increases by that constant, but the standard deviation remains the same. Therefore, the new mean of the appraised values of houses is  $\$125,000 + \$5,000 = \$130,000$ . Since the standard deviation remains unchanged at \$23,000, the new standard deviation is \$23,000. Final Answer: A

\*\*\*\*\*

Answer Option 8:

When a constant is added to each data point, the mean increases by that constant, but the standard deviation remains the same. In this case, since the city instructed the assessment office to raise all appraised house values by \$5,000, the new mean will be  $\$125,000 + \$5,000 = \$130,000$ .

Since the standard deviation remains the same, the new standard deviation will still be \$23,000.

Therefore, the answer is \$23,000. Final Answer: A

\*\*\*\*\*

Answer Option 9:

To find the new standard deviation, we need to consider the effect of raising all appraised house values by \$5,000. Note that this does not change the shape of the distribution, only shifts it by a constant amount.

Since the mean is increased by \$5,000, the new mean will be  $\$125,000 + \$5,000 = \$130,000$ . The standard deviation, however, is not affected by this constant shift.

Therefore, the new standard deviation will still be \$23,000. Final Answer: A

\*\*\*\*\*

Question 26:

'A large company has offices in two locations, one in New Jersey and one in Utah. The mean salary of office assistants in the New Jersey office is \$28,500. The mean salary of office assistants in the Utah office is \$22,500. The New Jersey office has 128 office assistants and the Utah office has 32 office assistants. What is the mean salary paid to the office assistants in this company?', A. '\$22,500 ', B. '\$23,700 ', C. '\$25,500 ', D. '\$27,300 '

=====

Answer Option 1:

To find the mean salary paid to the office assistants in the entire company, we need to calculate the total salary paid to all office assistants and divide it by the total number of office assistants.

The total salary paid to the office assistants in the New Jersey office is  $\$28,500 \times 128 = \$3,648,000$ .

The total salary paid to the office assistants in the Utah office is  $\$22,500 \times 32 = \$720,000$ .

The total number of office assistants in the company is  $128 + 32 = 160$ .

Therefore, the mean salary paid to the office assistants in the company is  $\$(3,648,000 + 720,000)/160 = \$27,300$ .

The mean salary paid to the office assistants in this company is  $\boxed{\$27,300}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

To find the mean salary paid to the office assistants in the company, we need to calculate the total salary paid by the company and divide it by the total number of office assistants.

The total salary paid by the company is the sum of the salaries paid in each office. In the New Jersey office, the total salary paid is  $\$28,500 \times 128 = \$3,648,000$ . In the Utah office, the total salary paid is  $\$22,500 \times 32 = \$720,000$ .

The total number of office assistants in the company is the sum of the number of office assistants in each office, which is  $128 + 32 = 160$ .

Therefore, the mean salary paid to the office assistants in the company is  $\frac{\$(3,648,000 + \$720,000)}{160} = \frac{\$4,368,000}{160} = \$27,300$ .

Thus, the answer is  $\boxed{\$27,300}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

To find the mean salary paid to the office assistants in the company, we need to find the total salary paid and divide it by the total number of office assistants.

The total salary paid in the New Jersey office is  $\$28,500 \times 128 = \$3,648,000$ .

The total salary paid in the Utah office is  $\$22,500 \times 32 = \$720,000$ .

The total number of office assistants in the company is  $128 + 32 = 160$ .

The total salary paid in the company is  $\$3,648,000 + \$720,000 = \$4,368,000$ .

Therefore, the mean salary paid to the office assistants in the company is  $\frac{\$4,368,000}{160} = \$27,300$ .

The answer is  $\boxed{\$27,300}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

To find the mean salary paid to the office assistants in the company, we need to calculate the average of the mean salaries in the New Jersey and Utah offices, weighted by the number of office assistants in each office.

The total number of office assistants in the company is  $128 + 32 = 160$ .

The total salary paid to office assistants in the New Jersey office is  $\$28,500 \times 128 = \$3,648,000$ .

The total salary paid to office assistants in the Utah office is  $\$22,500 \times 32 = \$720,000$ .

The total salary paid to all office assistants in the company is  $\$3,648,000 + \$720,000 = \$4,368,000$ .

The mean salary paid to the office assistants in the company is  $\$4,368,000 / 160 = \$27,300$ .

Therefore, the answer is  $\boxed{\$27,300}$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

To find the mean salary paid to the office assistants in the company, we need to calculate the overall average by taking into account the number of office assistants in each location.

The total number of office assistants in the company is  $128 + 32 = 160$ .

The total salary paid to office assistants in the New Jersey office is  $\$28,500 \times 128 = \$3,648,000$ .

The total salary paid to office assistants in the Utah office is  $\$22,500 \times 32 = \$720,000$ .

The overall total salary paid to all office assistants in the company is  $\$3,648,000 + \$720,000 = \$4,368,000$ .

Therefore, the mean salary paid to the office assistants in the company is  $\frac{\$4,368,000}{160} = \boxed{\$27,300}$ .

Final Answer: D

\*\*\*\*\*

Answer Option 6:

To find the mean salary paid to the office assistants in the company, we need to calculate the total salary paid to all office assistants and divide it by the total number of office assistants.

In the New Jersey office, the total salary paid is  $\$28,500 \times 128 = \$3,648,000$ .

In the Utah office, the total salary paid is  $\$22,500 \times 32 = \$720,000$ .



The total number of office assistants in the company is  $128 + 32 = 160$ .

The total salary paid to all office assistants is  $3,648,000 + 720,000 = 4,368,000$ .

Therefore, the mean salary paid to the office assistants in the company is  $\$4,368,000 / 160 = \$27,300$ .

Hence, the answer is  $\boxed{\text{D. } \$27,300}$ .

\*\*\*\*\*

Answer Option 7:

To find the mean salary paid to the office assistants in the company, we need to find the total salary paid to all the office assistants and divide it by the total number of office assistants.

The total salary paid to office assistants in the New Jersey office is  $\$28,500 \times 128 = \$3,648,000$ .

The total salary paid to office assistants in the Utah office is  $\$22,500 \times 32 = 720,000$ .

The total number of office assistants in the company is  $128 + 32 = 160$ .

The total salary paid to all the office assistants in the company is  $\$3,648,000 + 720,000 = 4,368,000$ .

Therefore, the mean salary paid to the office assistants in the company is  $\$4,368,000 / 160 = \boxed{\$27,300}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

To find the mean salary paid to the office assistants in the company, we need to find the total salary paid and then divide by the total number of office assistants.

The total salary paid in the New Jersey office is  $\$28,500 \times 128 = \$3,648,000$ .

The total salary paid in the Utah office is  $\$22,500 \times 32 = \$720,000$ .

The total number of office assistants in the company is  $128 + 32 = 160$ .

Therefore, the mean salary paid to the office assistants in the company is  $\frac{\$3,648,000 + \$720,000}{160} = \frac{\$4,368,000}{160} = \$27,300$ .

Hence, the answer is  $\boxed{\text{D. } \$27,300}$ .

\*\*\*\*\*

Answer Option 9:

To find the mean salary paid to the office assistants in the company, we need to take into account the number of office assistants in each office. We can calculate the weighted average of the mean salaries in each office using the number of office assistants as the weights.

The total number of office assistants in the company is  $128 + 32 = 160$ . The total salary paid to office assistants in the New Jersey office is  $128 \times \$28,500 = 3,648,000$ , and the total salary paid to office assistants in the Utah office is  $32 \times \$22,500 = 720,000$ .

The weighted average is calculated by summing the products of the mean salaries and the weights, and then dividing by the total number of office assistants:

$$\frac{\text{Mean salary}}{\text{Total salary paid}} = \frac{\text{Total salary paid}}{\text{Total number of office assistants}} = \frac{3,648,000 + 720,000}{160} = \frac{4,368,000}{160} = 27,300.$$

Therefore, the mean salary paid to the office assistants in this company is  $\boxed{\$27,300}$ . Final Answer: D

\*\*\*\*\*

Question 27:

'Which of the following is a false statement about simple random samples?', A. 'A sample must be reasonably large to be properly considered a simple random sample.', B. 'Inspection of a sample will give no indication of whether or not it is a simple random sample.', C. 'Attributes of a simple random sample may be very different from attributes of the population.', D. 'Every element of the population has an equal chance of being picked.'

=====

Answer Option 1:

The false statement about simple random samples is A. 'A sample must be reasonably large to be properly considered a simple random sample.' This is incorrect because the size of the sample does not determine whether it is a simple random sample or not. The defining characteristic of a simple random sample is that every element of the population has an equal chance of being picked, regardless of the size of the sample. Therefore, the correct answer is  $\boxed{\text{A.}}$

\*\*\*\*\*

Answer Option 2:

The false statement about simple random samples is A. 'A sample must be reasonably large to be properly considered a simple random sample.'

In reality, the size of a sample does not determine whether it is considered a simple random sample. A simple random sample refers to a sampling method where every element in the population has an equal chance of being selected. The size of the sample is irrelevant to whether or not it is considered a simple random sample.

Therefore, the correct answer is A. 'A sample must be reasonably large to be properly considered a simple random sample.' Final Answer: A

\*\*\*\*\*

Answer Option 3:

The false statement about simple random samples is A. 'A sample must be reasonably large to be properly considered a simple random sample.' Simple random sampling does not depend on the size of the sample. A simple random sample is a subset of a population in which each member of the population has an equal probability of being selected. The size of the sample does not affect whether it is considered a simple random sample or not. Therefore, statement A is false. Final answer: A.

\*\*\*\*\*

Answer Option 4:

The false statement about simple random samples is A. 'A sample must be reasonably large to be properly considered a simple random sample.'

The size of a sample does not determine whether it is a simple random sample or not. A simple random sample is a sampling method where every element of the population has an equal chance of being picked. The size of the sample is not a requirement for it to be considered a simple random sample.

Therefore, the correct answer is A. 'A sample must be reasonably large to be properly considered a simple random sample.' Final Answer: A

\*\*\*\*\*

Answer Option 5:

The correct answer is A. 'A sample must be reasonably large to be properly considered a simple random sample.'

A simple random sample refers to a subset of a population where every member of the population has an equal chance of being selected. The size of the sample does not determine whether it is considered a simple random sample or not. A simple random sample can be any size, as long as the selection process satisfies the condition of randomness. Therefore, statement A is false. Final Answer: A.

\*\*\*\*\*

Answer Option 6:

The false statement about simple random samples is A. 'A sample must be reasonably large to be properly considered a simple random sample.'

The size of the sample does not determine whether it is a simple random sample or not. A simple random sample is a subset of a population in which every possible sample of the same size has an equal chance of being selected. The key characteristic of a simple random sample is that every element of the population has an equal chance of being picked, regardless of the sample size. Thus, statement A is false.

Statements B, C, and D are all true. Inspection of a sample will not give any indication of whether or not it is a simple random sample (B). Attributes of a simple random sample can indeed be very different from attributes of the population (C). And, every element of the population does have an equal chance of being picked in a simple random sample (D).

Therefore, the false statement is A. Final Answer: A

\*\*\*\*\*

Answer Option 7:

The false statement about simple random samples is A. 'A sample must be reasonably large to be properly considered a simple random sample.'

The size of the sample does not affect whether it is considered a simple random sample. A simple random sample is a sample where every element of the population has an equal chance of being picked, regardless of the size of the sample.

Therefore, the correct answer is A. Final Answer: A

\*\*\*\*\*

Answer Option 8:

The false statement about simple random samples is option A. A sample does not have to be reasonably large to be properly considered a simple random sample. In fact, the size of the sample has no bearing on whether it is considered a simple random sample or not. A simple random sample is a subset of the population where every element in the population has an equal chance of being selected. The size of the sample does not affect this criteria. Thus, the correct answer is  $\boxed{\text{A. 'A sample must be reasonably large to be properly considered a simple random sample.'}}$

\*\*\*\*\*

Answer Option 9:

The false statement about simple random samples is A. 'A sample must be reasonably large to be properly considered a simple random sample.' This statement is incorrect because the size of the sample does not affect whether it is considered a simple random sample. A simple random sample is a subset of a larger population in which every member of the population has an equal chance of being selected. The size of the sample does not determine whether it is a simple random sample or not. Therefore, option A is incorrect. The other statements B, C, and D are true, as inspection of a sample does not provide information on whether it is a simple random sample (B), attributes of a simple random sample may differ from the population (C), and every element in the population has an equal chance of being selected in a simple random sample (D). Final Answer: A

\*\*\*\*\*

Question 28:

'A telephone executive instructs an associate to contact 104 customers using their service to obtain their opinions in regard to an idea for a new pricing package. The associate notes the number of customers whose names begin with A and uses a random number table to pick four of these names. She then proceeds to use the same procedure for each letter of the alphabet and combines the  $4 \times 26 = 104$  results into a group to be contacted. Which of the following is a correct conclusion?', A. 'Her procedure makes use of chance.', B. 'Her procedure results in a simple random sample.', C. 'Each customer has an equal probability of being included in the survey.', D. 'Her procedure introduces bias through sampling error.'

=====

Answer Option 1:

The correct conclusion is B. 'Her procedure results in a simple random sample.'

In her procedure, the associate uses a random number table to pick four names for each letter of the alphabet. This means that each customer whose name begins with a particular letter has an equal probability of being selected. Since this process is repeated for all 26 letters, each customer has an equal probability of being included in the survey. Therefore, the associate's procedure results in a simple random sample. Final answer: B

=====

Answer Option 2:

The procedure described does not result in a simple random sample because the associate is not selecting customers randomly from the entire population of 104 customers. Instead, she is selecting customers based on the first letter of their names. This introduces bias into the sample because customers with certain letters at the beginning of their names are more likely to be included in the survey than others. Therefore, the correct conclusion is  $\boxed{\text{D. 'Her procedure introduces bias through sampling error.'}}$

=====

Answer Option 3:

The procedure described does not result in a simple random sample because the number of customers selected for each letter of the alphabet is not proportional to the total number of customers whose names begin with that letter. Therefore, each customer does not have an equal probability of being included in the survey.

However, since the associate uses a random number table to pick the names within each letter group, her procedure does make use of chance.

Therefore, the correct conclusion is  $\boxed{\text{A. 'Her procedure makes use of chance.'}}$  Final Answer: A

=====

Answer Option 4:

The correct conclusion is C. 'Each customer has an equal probability of being included in the survey.'

In this scenario, the associate randomly selects four customers whose names begin with each letter of the alphabet. By doing so, each customer with a name starting with any given letter has an equal chance of being selected. This ensures that the selection process is fair and unbiased.

Therefore, the probability of being included in the survey is the same for every customer, indicating that each customer has an equal probability of being included in the survey.

Hence, option C is the correct conclusion. Final Answer: C

=====

Answer Option 5:

The associate uses a random number table to select four names for each letter of the alphabet. This means that each name beginning with a certain letter has an equal chance of being selected. However, this does not mean that each customer has an equal probability of being included in the survey. The customers whose names begin with letters that appear more frequently in the alphabet (such as A, B, etc.) have a higher chance of being included in the survey compared to those with letters that appear less frequently (such as X, Y, Z). Therefore, the correct conclusion is that her procedure introduces bias through sampling error. Thus, the answer is  $\boxed{\text{D}}$ . Final Answer: D

=====

Answer Option 6:

The procedure described in the question does not result in a simple random sample, as the sample is not selected randomly from the entire population. Instead, the associate selects customers based on the first letter of their names. Therefore, option B is incorrect.

Moreover, the procedure does not ensure that each customer has an equal probability of being included in the survey. Customers with names starting with more common letters of the alphabet have a higher chance of being selected compared to customers with names starting with less common letters. Thus, option C is also incorrect.

The procedure does involve the use of chance, as the associate uses a random number table to select names. Therefore, option A is correct.

There is no indication in the question that the procedure introduces bias through sampling error. Hence, option D is incorrect.

Therefore, the correct conclusion is that the associate's procedure makes use of chance, which is option A. Final Answer: A

=====

Answer Option 7:

The correct conclusion is  $\boxed{\text{A. 'Her procedure makes use of chance.'}}$

The associate uses a random number table to pick four names for each letter of the alphabet. This means that the selection of customers is based on chance, as the random number table introduces randomness into the selection process. Therefore, the procedure makes use of chance. Final answer: A

=====

Answer Option 8:

The given procedure does not result in a simple random sample because it does not guarantee that every customer has an equal chance of being included in the survey. The customers whose names begin with less common letters in the alphabet have a higher chance of being included in the survey compared to the customers whose names begin with more common letters. Therefore, the correct conclusion is

that the procedure introduces bias through sampling error. The answer is  $\boxed{\text{D. 'Her procedure introduces bias through sampling error.'}}$

\*\*\*\*\*

Answer Option 9:

The correct conclusion is B. 'Her procedure results in a simple random sample.'

In this scenario, the associate is randomly selecting four names that begin with each letter of the alphabet. Since the names are selected using a random number table and each letter has an equal chance of being selected, this procedure results in a simple random sample. A simple random sample is a sampling method where each member of the population has an equal probability of being selected.

Option A ('Her procedure makes use of chance.') is true, as the procedure involves using a random number table to select names.

Option C ('Each customer has an equal probability of being included in the survey.') is also true, as each customer has an equal chance of being selected since the names are selected randomly.

Option D ('Her procedure introduces bias through sampling error.') is incorrect because a simple random sample does not introduce bias. Bias may occur if the selection process favors certain individuals over others, but in this case, the selection is random.

Therefore, the correct conclusion is B. 'Her procedure results in a simple random sample.' Final answer: B

\*\*\*\*\*

Question 29:

'A company has a choice of three investment schemes. Option I gives a sure \$25,000 return on investment. Option II gives a 50% chance of returning \$50,000 and a 50% chance of returning \$10,000. Option III gives a 5% chance of returning \$100,000 and a 95% chance of returning nothing. Which option should the company choose?', A. 'Option II if it wants to maximize expected return', B. 'Option I if it needs at least \$20,000 to pay off an overdue loan', C. 'Option III if it needs at least \$80,000 to pay off an overdue loan', D. 'All of the above answers are correct.'

=====

Answer Option 1:

To determine which option the company should choose, we need to compare the expected returns of each option.

The expected return of Option I is \$25,000.

The expected return of Option II is calculated as follows:

$$\{0.5 \cdot \$50,000 + 0.5 \cdot \$10,000 = \$30,000\}$$

The expected return of Option III is calculated as follows:

$$\{0.05 \cdot \$100,000 + 0.95 \cdot \$0 = \$5,000\}$$

Comparing the expected returns, we can see that Option II has the highest expected return of \$30,000. Therefore, if the company wants to maximize its expected return, it should choose Option II.

Option I is not the best choice because Option II has a higher expected return.

Option III is not the best choice because Option II has a higher expected return.

Therefore, the correct answer is  $\boxed{\text{Option II if it wants to maximize expected return}}$ . Final answer: A

\*\*\*\*\*

Answer Option 2:

To determine which option the company should choose, we need to consider the company's specific needs and preferences.

If the company wants to maximize expected return, it should choose the option with the highest expected value.

Option I has a sure return of \$25,000.

Option II has a 50% chance of returning \$50,000 and a 50% chance of returning \$10,000. The expected value for Option II is:

$$\{0.5 \cdot \$50,000 + 0.5 \cdot \$10,000 = \$30,000\}$$

Option III has a 5% chance of returning \$100,000 and a 95% chance of returning nothing. The expected value for Option III is:

$$\{0.05 \cdot \$100,000 + 0.95 \cdot \$0 = \$5,000\}$$

Therefore, if the company wants to maximize expected return, it should choose Option II.

If the company needs at least \$20,000 to pay off an overdue loan, Option I is the only guaranteed option that meets this requirement. Therefore, if the company needs at least \$20,000 to pay off an overdue loan, it should choose Option I.

If the company needs at least \$80,000 to pay off an overdue loan, none of the options meet this requirement. Therefore, none of the options are suitable in this scenario.

Based on the above analysis, the correct answer is  $\boxed{\text{A. Option II if it wants to maximize expected return}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 3:

To determine which option the company should choose, we need to consider their goals and the level of risk they are willing to take.

If the company wants to maximize expected return, they should choose the option with the highest expected value.

Option I has a sure return of \$25,000, so its expected value is \$25,000.

Option II has a 50% chance of returning \$50,000 and a 50% chance of returning \$10,000. The expected value of Option II is  $\{0.5 \cdot \$50,000 + 0.5 \cdot \$10,000 = \$30,000\}$ .

Option III has a 5% chance of returning \$100,000 and a 95% chance of returning nothing. The expected value of Option III is  $\{0.05 \cdot \$100,000 + 0.95 \cdot \$0 = \$5,000\}$ .

Therefore, if the company wants to maximize expected return, they should choose Option II.

If the company needs at least \$20,000 to pay off an overdue loan, they should choose an option with a guaranteed return of at least \$20,000.

Option I has a guaranteed return of \$25,000, so it is the best choice in this case.

If the company needs at least \$80,000 to pay off an overdue loan, none of the options provide a guaranteed return of \$80,000. In this case, Option III, with a 5% chance of returning \$100,000, provides the highest potential return and is the best choice.

Therefore, the correct answer is  $\boxed{\text{D. All of the above answers are correct.}}$

\*\*\*\*\*

Answer Option 4:

To determine which option the company should choose, we need to compare the expected return for each option.

For Option I, the expected return is \$25,000.

For Option II, the expected return is  $\{(0.5)(\$50,000) + (0.5)(\$10,000) = \$30,000\}$ .

For Option III, the expected return is  $\{(0.05)(\$100,000) + (0.95)(\$0) = \$5,000\}$ .

Therefore, if the company wants to maximize expected return, it should choose Option II. If it needs at least \$20,000 to pay off an overdue loan, it should choose Option I. If it needs at least \$80,000 to pay off an overdue loan, it should choose Option III.

Since all of the above answers are correct, the correct answer is  $\boxed{\text{(D) All of the above answers are correct.}}$

\*\*\*\*\*

Answer Option 5:

To determine which option the company should choose, we need to compare the expected returns of each option.

The expected return of Option I is \$25,000.

The expected return of Option II is  $\$0.5 \cdot \$50,000 + 0.5 \cdot \$10,000 = \$30,000$ .

The expected return of Option III is  $\$0.05 \cdot \$100,000 + 0.95 \cdot \$0 = \$5,000$ .

Based on these calculations, if the company wants to maximize expected return, it should choose Option II. If the company needs at least \$20,000 to pay off an overdue loan, it should choose Option I. If the company needs at least \$80,000 to pay off an overdue loan, it should choose Option III.

Therefore, the correct answer is  $\boxed{\text{D. All of the above answers are correct.}}$  Final Answer: D

\*\*\*\*\*

Answer Option 6:

To determine which option the company should choose, we need to consider the expected return of each option and compare it to the specific requirements mentioned.

For Option I, the expected return is \$25,000, as it guarantees a fixed return.

For Option II, the expected return is  $\frac{1}{2} \cdot \$50,000 + \frac{1}{2} \cdot \$10,000 = \$30,000$ . This is higher than the expected return of Option I.

For Option III, the expected return is  $\frac{5}{100} \cdot \$100,000 + \frac{95}{100} \cdot \$0 = \$5,000$ . This is lower than the expected return of both Option I and Option II.

Therefore, if the company wants to maximize the expected return, it should choose Option II. This eliminates options B and C as potential correct answers.

Hence, the correct answer is  $\boxed{\text{A. 'Option II if it wants to maximize expected return'}}$  Final Answer: A

\*\*\*\*\*

Answer Option 7:

To determine which option the company should choose, we need to consider both the potential returns and the probabilities associated with each option.

Option I has a certain return of \$25,000.

Option II has a 50% chance of returning \$50,000 and a 50% chance of returning \$10,000. The expected return for Option II can be calculated as  $\$0.5 \cdot \$50,000 + 0.5 \cdot \$10,000 = \$30,000$ .

Option III has a 5% chance of returning \$100,000 and a 95% chance of returning nothing. The expected return for Option III can be calculated as  $\$0.05 \cdot \$100,000 + 0.95 \cdot \$0 = \$5,000$ .

To maximize expected return, the company should choose the option with the highest expected return. In this case, Option II has the highest expected return of \$30,000.

Therefore, the company should choose Option II if it wants to maximize expected return. Hence, the correct answer is  $\boxed{\text{A. 'Option II if it wants to maximize expected return'}}$ .

\*\*\*\*\*

Answer Option 8:

To determine which option the company should choose, we need to consider its goals and constraints.

If the company wants to maximize the expected return, we need to calculate the expected values for each option. The expected value of Option I is \$25,000\$, Option II is  $\$(0.5)(\$50,000) + (0.5)(\$10,000) = \$30,000$ \$, and Option III is  $\$(0.05)(\$100,000) + (0.95)(\$0) = \$5,000$ \$. Therefore, if the company wants to maximize the expected return, it should choose Option II.

If the company needs at least \$20,000 to pay off an overdue loan, Option I guarantees a return of \$25,000\$, which is more than enough to pay off the loan. Therefore, if the company needs at least \$20,000 to pay off an overdue loan, it should choose Option I.

If the company needs at least \$80,000 to pay off an overdue loan, none of the options guarantee a return of \$80,000. However, Option III has a 5% chance of returning \$100,000, which is enough to pay off the loan. Therefore, if the company needs at least \$80,000 to pay off an overdue loan, it should choose Option III.

In conclusion, all of the above answers are correct, so the correct choice is  $\boxed{\text{D}}$ \$.  
\*\*\*\*\*

Answer Option 9:

To determine the best option, we need to compare the expected return for each option.

Option I has a sure return of \$25,000.

The expected return for Option II can be calculated as  $\$(0.5)(\$50,000) + (0.5)(\$10,000) = \$30,000$ .

The expected return for Option III can be calculated as  $\$(0.05)(\$100,000) + (0.95)(\$0) = \$5,000$ .

Comparing the expected returns, we see that Option II has the highest expected return at \$30,000.

Therefore, the company should choose Option II if it wants to maximize expected return.

However, it is important to note that depending on the specific financial situation of the company, the best option may change. If the company needs at least \$20,000 to pay off an overdue loan (Option B), then Option I with a sure return of \$25,000 would be the best choice. Similarly, if the company needs at least \$80,000 to pay off an overdue loan (Option C), then Option III with a 5% chance of returning \$100,000 would be the best choice.

Therefore, the correct answer is  $\boxed{\text{D. All of the above answers are correct.}}$   
\*\*\*\*\*

Question 30:

'A one-sided test of a hypothesis about a population mean, based on a sample of size 14, yields a P-value of 0.075. Which of the following best describes the range of t-values that would have given this P-value?', A. ' $1.345 < t < 1.761$ ', B. ' $1.356 < t < 1.782$ ', C. ' $1.771 < t < 2.160$ ', D. ' $1.350 < t < 1.771$ '

=====

Answer Option 1:

To determine the range of t-values that would have given a P-value of 0.075, we need to look up the critical t-values corresponding to a one-sided test with a significance level of 0.075 and 14 degrees of freedom.

Looking up the critical t-values in a t-table, we find that for a one-sided test with a significance level of 0.075 and 14 degrees of freedom, the critical t-values are approximately 1.771 and 2.160.

Therefore, the range of t-values that would have given a P-value of 0.075 is ' $1.771 < t < 2.160$ ', so the correct answer is  $\boxed{\text{C. '1.771 < t < 2.160'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

To determine the range of t-values that would have given a P-value of 0.075, we need to find the critical t-values for a one-sided test with a significance level of 0.075.

Since the P-value is 0.075, the critical t-value will be the t-value that corresponds to a cumulative probability of 0.925 ( $1 - 0.075$ ).

Using a t-table or calculator, we find that the critical t-value for a one-sided test with a cumulative probability of 0.925 and a sample size of 14 is approximately 1.782.

Therefore, the range of t-values that would have given a P-value of 0.075 is ' $1.356 < t < 1.782$ '.

The correct answer is B. Final Answer: B

\*\*\*\*\*

Answer Option 3:

The P-value is the probability of obtaining a test statistic as extreme as the observed value (or more extreme) if the null hypothesis is true. In this case, since it is a one-sided test, we are interested in the upper tail of the t-distribution.

To determine the range of t-values that would yield a P-value of 0.075, we need to find the critical t-value that corresponds to this P-value. Since the P-value is relatively large, we would expect the critical t-value to be smaller than the critical values for smaller P-values.

Looking at the answer choices, we can see that the range of t-values in option D, ' $1.350 < t < 1.771$ ', covers the range of t-values that are smaller than the range given in the other answer choices. Therefore, the best description of the range of t-values that would have given a P-value of 0.075 is option D. Final answer: D.

\*\*\*\*\*

Answer Option 4:

The P-value represents the probability of observing a test statistic as extreme as the one calculated, assuming the null hypothesis is true. In this case, since it is a one-sided test, the P-value represents the probability of observing a test statistic as large as the one calculated.

To find the range of t-values that would have given this P-value, we need to find the critical t-value(s) that correspond to a cumulative probability of 0.075 in the upper tail of the t-distribution with 14 degrees of freedom.

Using a table or a statistical software, we find that the critical t-value for a cumulative probability of 0.075 in the upper tail is approximately 1.782.

Therefore, the range of t-values that would have given this P-value is given by ' $1.356 < t < 1.782$ '.

Thus, the answer is  $\boxed{\text{B. '1.356 < t < 1.782'}}$ . Final answer: B

\*\*\*\*\*

Answer Option 5:

To determine the range of t-values that would have given a P-value of 0.075, we need to find the critical values for a one-sided test with a significance level of 0.075.

Since the P-value is not given as less than or greater than, we can assume it is a two-tailed test. In a two-tailed test, we split the significance level in half to find the critical values for each tail.

Using a t-table or calculator, we find that for a one-tailed test with a significance level of 0.075 and degrees of freedom 13 ( $14-1$ ), the critical value is approximately 1.771.

Therefore, the correct range of t-values that would have given a P-value of 0.075 is ' $1.771 < t < 2.160$ '.

Option C: ' $1.771 < t < 2.160$ ' is the correct answer. Final Answer: C

\*\*\*\*\*

Answer Option 6:

The P-value is the probability of observing a test statistic as extreme as the one observed, assuming the null hypothesis is true. For a one-sided test, the P-value represents the probability of observing a test statistic as extreme or more extreme than the observed value.

In this case, the P-value is 0.075. To find the range of t-values that would have given this P-value, we need to find the critical t-values that correspond to a cumulative probability of 0.075 in the appropriate tail of the t-distribution.

Looking up the critical t-values in a t-table or using statistical software, we find that for a one-sided test with 14 degrees of freedom, the critical t-value for a cumulative probability of 0.075 in the right tail is approximately 1.771. Therefore, the range of t-values that would have given a P-value of 0.075 is  $t > 1.771$ .

So, the answer is C. ' $1.771 < t < 2.160$ '. Final Answer: C

\*\*\*\*\*

Answer Option 7:

To determine the range of t-values that would have given a P-value of 0.075, we need to find the critical t-values for a one-sided test at a significance level of 0.075.



The P-value represents the probability of observing a test statistic as extreme as (or more extreme than) the observed test statistic, assuming the null hypothesis is true. In this case, the observed test statistic falls outside the critical region since the P-value is greater than the significance level of 0.075.

Since the test is one-sided, the critical region is either in the upper tail or lower tail, depending on the alternative hypothesis. Let's assume the alternative hypothesis is one-sided greater.

To find the critical t-value, we need to find the t-value that separates the top 7.5% (0.075) of the t-distribution. By looking up the critical value in the t-table or using a calculator, we find that the critical t-value is approximately 1.782.

Therefore, the range of t-values that would have given a P-value of 0.075 is  $t < 1.782$ .

The correct answer is B. ' $1.356 < t < 1.782$ '.

\*\*\*\*\*

Answer Option 8:

The P-value is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample, assuming the null hypothesis is true. In this case, since it is a one-sided test, the P-value is the probability of observing a t-value greater than or equal to the one calculated.

To find the range of t-values that would have given this P-value, we need to find the critical t-value(s) such that the area under the t-distribution curve to the right of the critical t-value(s) is equal to the P-value of 0.075.

Looking at the answer choices, we can check the t-table or use a statistical calculator to find the critical t-values corresponding to the given P-values.

Using a calculator or t-table, we find that the critical t-values for a one-sided test at a significance level of 0.075 (P-value of 0.075) for a sample size of 14 are approximately 1.350 and 1.771.

Therefore, the range of t-values that would have given this P-value is ' $1.350 < t < 1.771$ ', which corresponds to answer choice (D).

Final answer: D

\*\*\*\*\*

Answer Option 9:

To determine the range of t-values that would have given a P-value of 0.075, we need to find the critical values corresponding to the significance level  $\alpha = 0.075$ .

Since this is a one-sided test, we divide the significance level by 2, resulting in  $\alpha/2 = 0.075/2 = 0.0375$ .

Using a t-table or a calculator, we find that the critical t-value for a one-sided test with a significance level of 0.0375 and degrees of freedom  $14-1=13$  is approximately 1.771.

Therefore, the range of t-values that would have given a P-value of 0.075 is given by ' $1.350 < t < 1.771$ '.

The correct answer is  $\boxed{\text{D. '1.350 < t < 1.771'}}$ .

\*\*\*\*\*

Error with subject\_matters/high\_school\_statistics\_test/prompts\_9\_31.jsonl: 'user'

Question 32:

'You need to construct a 94% confidence interval for a population proportion. What is the upper critical value of z to be used in constructing this interval?', A. '0.9699', B. '1.96', C. '1.555', D. '1.88'

=====

Answer Option 1:

To construct a 94% confidence interval for a population proportion, we need to find the upper critical value of z for a confidence level of 94%. Since the confidence level is given in terms of percentage, we need to find the critical value that corresponds to the area between the mean and the upper tail of the standard normal distribution.

The critical value for a 94% confidence level can be found using a standard normal distribution table or a calculator. The critical value for a 94% confidence level is approximately 1.88. Therefore, the correct answer is  $\boxed{\text{D. } 1.88}$ . Final Answer: D

=====

Answer Option 2:

To construct a confidence interval for a population proportion, we use the standard normal distribution and the z-score.

A 94% confidence level corresponds to  $100\% - 94\% = 6\%$  being spread equally in the two tails. Since the standard normal distribution is symmetric, we need to find the z-score that corresponds to an area of 3% in the upper tail.

Using a standard normal distribution table or a calculator, we find that the z-score corresponding to an area of 3% in the upper tail is approximately 1.8808.

Therefore, the upper critical value of z to be used in constructing this interval is approximately 1.8808, which corresponds to option D. Final Answer: D

=====

Answer Option 3:

To construct a confidence interval for a population proportion, we use the standard normal distribution and the critical value of z. For a 94% confidence interval, we want to find the upper critical value of z.

The upper critical value of z for a 94% confidence interval can be found by finding the z-value that leaves 3% in the upper tail of the standard normal distribution. This leaves 1.5% in each tail.

Using a standard normal distribution table or a calculator, we find that the z-value corresponding to a cumulative probability of 0.985 (which leaves 1.5% in the upper tail) is approximately 1.555. Therefore, the upper critical value of z is approximately 1.555.

Thus, the correct answer is  $\boxed{\text{C. } 1.555}$ . Final Answer: C

=====

Answer Option 4:

To construct a 94% confidence interval for a population proportion, we need to find the critical value of z that corresponds to an area of 0.06 in the upper tail of the standard normal distribution.

Looking up in the standard normal distribution table, we find that the critical z-value for an upper tail probability of 0.06 is approximately 1.555. Therefore, the upper critical value of z to be used in constructing this interval is  $\boxed{\text{C. } 1.555}$ . Final Answer: C

=====

Answer Option 5:

To construct a 94% confidence interval for a population proportion, we need to find the critical value of z. This critical value corresponds to the proportion of the area under the standard normal curve that falls outside of the confidence interval.

Since we want a 94% confidence interval, we need to find the z-value that leaves 3% of the area in each tail of the standard normal curve. This corresponds to leaving  $(100\% - 94\%) / 2 = 3\%$  in each tail.

Using a standard normal distribution table, we find that the z-value for leaving 3% in each tail is approximately 1.88. Therefore, the upper critical value of z is 1.88.

Thus, the correct answer is  $\boxed{\text{D. } 1.88}$ .

=====

Answer Option 6:

To construct a 94% confidence interval for a population proportion, we need to find the upper critical value of z.

Since the confidence level is 94%, the alpha level (significance level) is  $1 - 0.94 = 0.06$ .

Dividing this by 2 gives us an alpha/2 level of 0.03.

Looking up this value in the standard normal distribution table, we find that the corresponding z-value is approximately 1.88.

Thus, the correct answer is  $\boxed{\text{D. } 1.88}$ . Final Answer: D

=====

Answer Option 7:

To construct a 94% confidence interval for a population proportion, we need to find the upper critical value of z.

The confidence level is equal to  $1 - \alpha$ , where alpha is the level of significance. For a 94% confidence interval, alpha is equal to  $(1 - 0.94) / 2 = 0.03 / 2 = 0.015$ .

Using a standard normal distribution table or calculator, we can find the critical value z corresponding to an area of 0.015 in the upper tail of the distribution.

The upper critical value of z is approximately 1.88, so the answer is  $\boxed{\text{D. } 1.88}$ . Final Answer: D

=====

Answer Option 8:

To construct a 94% confidence interval for a population proportion, we need to find the critical value of z that corresponds to a 94% confidence level.

The confidence level is the complement of the significance level ( $\alpha$ ), so a 94% confidence level corresponds to a significance level of  $100\% - 94\% = 6\%$ . Since the confidence interval is two-tailed, we need to find the critical value of z that leaves an area of 3% in each tail.

Looking up the critical value in a standard normal distribution table or using a calculator, we find that the upper critical value of z for a 6% significance level is approximately 1.555.

Therefore, the upper critical value of  $z$  to be used in constructing this interval is approximately 1.555.

The correct answer is C. '1.555'

\*\*\*\*\*

Answer Option 9:

A 94% confidence interval corresponds to a significance level of  $1 - 0.94 = 0.06$ . Since we are constructing a confidence interval for a population proportion, we can use the standard normal distribution (Z-distribution) for the critical value.

To find the upper critical value of  $Z$ , we need to find the  $Z$ -score that leaves an area of 0.06 to the right of it in the  $Z$ -distribution. Looking up the value in a standard normal distribution table or using a calculator, we find that the  $Z$ -score corresponding to an area of 0.06 to the right is approximately 1.555.

Therefore, the upper critical value of  $Z$  to be used in constructing the 94% confidence interval is approximately 1.555. The correct answer is C. Final Answer: C

\*\*\*\*\*

Question 33:

'A local school has seven math teachers and seven English teachers. When comparing their mean salaries, which of the following is most appropriate?', A. 'A two-sample z-test of population means', B. 'A two-sample t-test of population means', C. 'A one-sample z-test on a set of differences', D. 'None of the above are appropriate.'

=====

Answer Option 1:

To compare the mean salaries of the math teachers and English teachers, we have two independent samples from two different populations. Therefore, the most appropriate test would be a two-sample t-test of population means (option B). This test accounts for the variability in each sample and is suitable for comparing means when the population standard deviations are unknown. Final answer: B

\*\*\*\*\*

Answer Option 2:

To compare the mean salaries of the math teachers and English teachers, we have two independent samples. Since we do not have the population standard deviation, we should use a t-test instead of a z-test. Therefore, the most appropriate option is B. 'A two-sample t-test of population means'. Final Answer: B.

\*\*\*\*\*

Answer Option 3:

The appropriate statistical test to compare the mean salaries of two independent groups is a two-sample t-test of population means. This test is used when the sample sizes are small, the population standard deviations are unknown, and the samples are independent. Since we are comparing the salaries of math teachers and English teachers, which are two different groups, a two-sample t-test is appropriate. Therefore, the correct answer is  $\boxed{\text{B. 'A two-sample t-test of population means'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 4:

When comparing the mean salaries of two independent groups (math teachers and English teachers), the appropriate statistical test is a two-sample t-test of population means. Therefore, the correct answer is B. 'A two-sample t-test of population means'. Final answer: B.

\*\*\*\*\*

Answer Option 5:

When comparing the mean salaries of two different groups (math teachers and English teachers), the appropriate statistical test to use is a two-sample t-test of population means. This test is used when comparing means from two independent samples. It takes into account the variability within each group as well as the differences between the groups. Therefore, the most appropriate choice is option B, 'A two-sample t-test of population means.' Final answer: B

\*\*\*\*\*

Answer Option 6:

The most appropriate choice for comparing the mean salaries of the math teachers and English teachers is option B, 'A two-sample t-test of population means'. This is because we are comparing the means of two independent samples (math teachers and English teachers) and do not have information about the population standard deviations. The t-test is used when the population standard deviations are unknown. Therefore, option B is the most appropriate choice. Final Answer: B

\*\*\*\*\*

Answer Option 7:

To compare the mean salaries of the math teachers and English teachers at the local school, we have two independent samples. Therefore, the appropriate statistical test to use is a two-sample t-test of population means. This test is commonly used when comparing the means of two independent samples to determine if there is a significant difference between them.

Option A, a two-sample z-test of population means, assumes that the population variances are known, which is typically not the case. Therefore, this option is not appropriate.

Option C, a one-sample z-test on a set of differences, is used when comparing the means of a single sample to a known or hypothesized value. This test is not suitable for comparing the means of two independent samples.

Therefore, the most appropriate option is B. A two-sample t-test of population means. Final answer: B.

\*\*\*\*\*

Answer Option 8:

To compare the mean salaries of the math teachers and English teachers, we have two independent samples. Since we do not have the population standard deviation, we should use a two-sample t-test of population means. Therefore, the most appropriate option is option B, 'A two-sample t-test of population means.' Final Answer: B

\*\*\*\*\*

Answer Option 9:

The appropriate test to compare the mean salaries of two groups is a two-sample t-test of population means. This is because we have two independent groups (math teachers and English teachers), and we are comparing the means of these two groups. The t-test is used when the sample sizes are small or when the population standard deviations are unknown. Therefore, the most appropriate option is B. 'A two-sample t-test of population means'. Final answer: B.

\*\*\*\*\*

Question 34:

'A small kiosk at the Atlanta airport carries souvenirs in the price range of \$3.99 to \$29.99, with a mean price of \$14.75. The airport authorities decide to increase the rent charged for a kiosk by 5 percent. To make up for the increased rent, the kiosk owner decides to increase the prices of all items by 50 cents. As a result, which of the following will happen?', A. 'The mean price and the range of prices will increase by 50 cents.', B. 'The mean price will remain the same, but the range of prices will increase by 50 cents.', C. 'The mean price and the standard deviation of prices will increase by 50 cents.', D. 'The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.'

=====

Answer Option 1:

Increasing all prices by 50 cents will not affect the mean price, as the increase is the same for each item. The range of prices will also remain the same, as the difference between the highest and lowest prices will still be  $\$29.99 - \$3.99 = \$26$ . However, the standard deviation of prices will increase by 50 cents, as each price is increased by the same amount, causing the spread of prices to increase.

Therefore, the correct answer is  $\boxed{\text{C. 'The mean price and the standard deviation of prices will increase by 50 cents.'}}$

\*\*\*\*\*

Answer Option 2:

Let's consider the effect of the price increase on the mean price and the range of prices separately.

The mean price is calculated by adding up all the prices and dividing by the number of items. Increasing the price of each item by 50 cents will increase the sum of all the prices by 50 cents times the number of items. Since the number of items remains the same, the mean price will increase by 50 cents.

The range of prices is the difference between the highest and lowest prices. Increasing the price of each item by 50 cents will increase both the highest and lowest prices by 50 cents. Therefore, the range of prices will also increase by 50 cents.

So, the correct answer is  $\boxed{\text{A. The mean price and the range of prices will increase by 50 cents.}}$  Final Answer: A

\*\*\*\*\*

Answer Option 3:

The mean price of the items will increase by 50 cents because the kiosk owner increases the price of all items by 50 cents. However, the range of prices will remain the same because the maximum and minimum prices are not affected by the increase. Therefore, the correct answer is  $\boxed{\text{D. 'The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.'}}$

\*\*\*\*\*

Answer Option 4:

Since the kiosk owner increases the prices of all items by 50 cents, the mean price of the souvenirs will increase by 50 cents. However, since the range of prices is determined by the minimum and maximum prices, increasing all prices by 50 cents will also increase the range of prices by 50 cents. Therefore, the correct answer is  $\boxed{\text{A. 'The mean price and the range of prices will increase by 50 cents.'}}$

\*\*\*\*\*

Answer Option 5:

When the kiosk owner increases the prices of all items by 50 cents, the mean price will increase by 50 cents. However, the range of prices will remain the same because the maximum and minimum prices will increase by the same amount. The standard deviation of prices will also remain the same since the increase in price is a fixed amount and does not affect the variation among the prices. Therefore, the correct answer is  $\boxed{\text{D. 'The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.'}}$  Final Answer: D

\*\*\*\*\*

Answer Option 6:

Increasing all prices by 50 cents will increase the mean price by 50 cents. However, the range of prices, which is the difference between the maximum and minimum prices, will remain the same. The standard deviation is a measure of the spread of data around the mean, and it will also remain the same since increasing all prices by the same amount does not change the spread of the data. Therefore, the correct answer is  $\boxed{\text{(D)} \backslash \text{The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.}}$

\*\*\*\*\*

Answer Option 7:

Increasing all prices by 50 cents will result in a shift of the entire price range. However, since the increase is uniform across all items, the range of prices will remain the same.

The mean price will increase by 50 cents because every item's price is increased by that amount. However, the standard deviation, which measures the spread or variability of the prices, will not necessarily increase by 50 cents. The standard deviation depends on the individual prices and their distance from the mean.

Therefore, the correct option is  $\boxed{\text{D. The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.}}$

\*\*\*\*\*

Answer Option 8:

The mean price will increase by 50 cents. This is because the kiosk owner increases the prices of all items by 50 cents, so the sum of all prices will increase by 50 cents for the same number of items. Since the mean is calculated by dividing the sum of all prices by the number of items, the mean price will increase by 50 cents.

The range of prices will remain the same. The range is the difference between the maximum and minimum prices. By increasing all prices by the same amount, the difference between the maximum and minimum prices will remain unchanged.

Therefore, the correct answer is  $\boxed{\text{(D) The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.}}$  Final Answer: D

\*\*\*\*\*

Answer Option 9:

Let's analyze each option:

A. 'The mean price and the range of prices will increase by 50 cents.'

This option cannot be true because increasing the range of prices would imply that the maximum and minimum prices have increased by 50 cents, but the range of prices was originally  $\$29.99 - \$3.99 = \$26$ , which is more than 50 cents.

B. 'The mean price will remain the same, but the range of prices will increase by 50 cents.'

This option also cannot be true because increasing the range of prices would imply that the maximum and minimum prices have increased by 50 cents, but the range of prices was originally  $\$29.99 - \$3.99 = \$26$ , which is more than 50 cents.

C. 'The mean price and the standard deviation of prices will increase by 50 cents.'

This option cannot be true because increasing the standard deviation would imply that the prices are more spread out, but we know that the range of prices will increase by at most 50 cents.

D. 'The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.'

This option is the most likely to be true. Increasing all prices by 50 cents will shift the entire distribution of prices upwards, including the mean price. However, since the prices are being increased by the same amount, the relative distance between the prices and the mean will remain the same, keeping the standard deviation unchanged.

Therefore, the correct option is  $\boxed{\text{D. 'The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.'}}$

\*\*\*\*\*

Question 35:

'The director of a local food bank asks for data on all donations given during the month of November. Of the 100 checks received, the average donation is \$155 with a standard deviation of \$32. Which of the following is the most appropriate statement?', A. 'This November, the average donation is \$155.', B. '50% of all donations this November are more than \$155.', C. 'We are 95% confident that the average donation in November is between about \$91 and \$219.', D. 'We are 95% confident that the average donation in November is between about \$149 and \$161.'

=====

Answer Option 1:

The most appropriate statement is C. 'We are 95% confident that the average donation in November is between about \$91 and \$219.'

This is because the statement accurately reflects the information provided. The average donation is \$155, which is within the range of \$91 to \$219. Additionally, the statement mentions that there is a 95% confidence level, which indicates the level of certainty about the range of the average donation. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The most appropriate statement would be C. 'We are 95% confident that the average donation in November is between about \$91 and \$219.' This statement correctly reflects the use of the standard deviation and the concept of confidence intervals. A is incorrect because it does not account for the variability of the donations. B is incorrect because it does not provide any information about the distribution of the donations. D is incorrect because it specifies a very narrow range around the average donation, which is not supported by the given standard deviation. Final answer: C.

\*\*\*\*\*

Answer Option 3:

The most appropriate statement is C. 'We are 95% confident that the average donation in November is between about \$91 and \$219.'

This is because the average donation of \$155 is a point estimate, but it does not provide any information about the variability or uncertainty of the estimate. Option A only states the point estimate without any mention of confidence or uncertainty.

Option B is incorrect because it makes an incorrect assumption about the distribution of donations. The statement that 50% of all donations are more than \$155 cannot be inferred from the given information.

Option D is incorrect because it describes a narrow range of \$149 to \$161, which is too specific to be a 95% confidence interval.

Option C is the most appropriate because it accurately describes the 95% confidence interval, which provides a range estimate that captures the likely range of the average donation in November.

\*\*\*\*\*

Answer Option 4:

The most appropriate statement is C. 'We are 95% confident that the average donation in November is between about \$91 and \$219.'

This statement reflects the concept of confidence interval, which provides a range of values within which the true population parameter (in this case, the average donation) is likely to fall.

Option A simply states the average donation as \$155, which does not account for any uncertainty.

Option B makes an incorrect claim about the percentage of donations that are more than \$155, which cannot be determined from the given information.

Option D provides a narrow range of \$149 to \$161, which does not capture the full uncertainty associated with the sample data.

Option C correctly states that we are 95% confident that the true average donation falls within the range of about \$91 to \$219, taking into account the sample mean and standard deviation. This range reflects the margin of error associated with the sample data.

Therefore, the most appropriate statement is C. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The most appropriate statement is option C. The statement accurately represents the confidence interval for the average donation in November. It states that with 95% confidence, the average donation is estimated to be between about \$91 and \$219. This takes into account the sample mean and the standard deviation, allowing for a range of values that is likely to contain the true population mean. Option A only provides the sample mean and does not account for the variability in the data. Option B makes a claim about the distribution of donations, which is not supported by the given information. Option D provides a narrower confidence interval, which is not accurate based on the given standard deviation. Therefore, option C is the most appropriate statement. Final Answer: C

\*\*\*\*\*

Answer Option 6:

The most appropriate statement is: 'We are 95% confident that the average donation in November is between about \$91 and \$219.' This is because the average donation of \$155 is just a point estimate, and does not provide any information about the variability of the data. Option B, '50% of all donations this November are more than \$155,' is incorrect as it assumes a symmetric distribution and does not take into account the standard deviation. Option D, 'We are 95% confident that the average donation in November is between about \$149 and \$161,' is incorrect as it only considers the standard deviation and does not take into account the sample size. The most appropriate statement takes into account both the sample mean (\$155) and the standard deviation (\$32), and provides a confidence interval to reflect the uncertainty in the estimate. Final Answer: C

\*\*\*\*\*

Answer Option 7:

The most appropriate statement is option C.

Option A simply states the average donation, which is true but does not provide any additional information or context.

Option B states that 50% of all donations are more than \$155, which is not necessarily true.

Option C states that we are 95% confident that the average donation in November is between about \$91 and \$219. This statement takes into account the standard deviation and provides a range within which we can be confident that the true average donation lies.

Option D states that we are 95% confident that the average donation in November is between about \$149 and \$161, which is too narrow and does not take into account the variability in the data.



Therefore, the most appropriate statement is option C. Final answer: C

\*\*\*\*\*

Answer Option 8:

The most appropriate statement is C. 'We are 95% confident that the average donation in November is between about \$91 and \$219.'

Option A simply states the average donation without any mention of confidence or variability.

Option B makes a claim about a specific percentage of donations being more than \$155, which is not supported by the given information on average and standard deviation.

Option C correctly uses the concept of confidence interval to provide a range (about \$91 to \$219) within which we can be 95% confident that the true average donation falls. This is calculated based on the average donation (\$155) and the standard deviation (\$32).

Option D provides a narrower range (\$149 to \$161), but it does not account for the 95% confidence level. It is too specific to be considered a reasonable statement based on the given information.

Therefore, the most appropriate statement is C. Final answer: C

\*\*\*\*\*

Answer Option 9:

The most appropriate statement is C. 'We are 95% confident that the average donation in November is between about \$91 and \$219.'

This is because the statement accurately reflects the concept of confidence intervals. The average donation of \$155 is the point estimate, and the standard deviation of \$32 is used to calculate the margin of error. Using the formula for a 95% confidence interval, the margin of error is approximately  $\$32 \times 1.96 \approx \$63$ . Therefore, we can be 95% confident that the true average donation in November is within  $\$155 \pm \$63$ , or between about \$91 and \$219. Final Answer: C

\*\*\*\*\*

Question 36:

'Which of the following describe/s the central limit theorem? I. The mean of the sampling distribution of  $\bar{x}$  is the same as the mean of the population. II. The standard deviation of the sampling distribution of  $\bar{x}$  is the same as the standard deviation of  $x$  divided by the square root of the sample size. III. If the sample size is large, the shape of the sampling distribution of  $\bar{x}$  is approximately normal.', A. 'I only', B. 'I & II only', C. 'II only', D. 'III only'

=====

Answer Option 1:

The correct answer is B. 'I & II only'.

I. The central limit theorem states that the mean of the sampling distribution of  $\bar{x}$  is equal to the mean of the population.

II. The central limit theorem also states that the standard deviation of the sampling distribution of  $\bar{x}$  is equal to the standard deviation of  $x$  divided by the square root of the sample size.

III. However, the central limit theorem does not state that the shape of the sampling distribution of  $\bar{x}$  is approximately normal. It states that for a large sample size, the sampling distribution of  $\bar{x}$  becomes approximately normal regardless of the shape of the population distribution.

Therefore, only statements I and II correctly describe the central limit theorem. Final Answer: B

\*\*\*\*\*

Answer Option 2:

The central limit theorem states that if we have a random sample of independent observations from any population with a finite mean and standard deviation, then when the sample size is large, the sampling distribution of the sample mean will be approximately normally distributed.

I. The mean of the sampling distribution of  $\bar{x}$  is the same as the mean of the population. This statement is true and describes the central limit theorem.

II. The standard deviation of the sampling distribution of  $\bar{x}$  is the same as the standard deviation of  $x$  divided by the square root of the sample size. This statement is also true and describes the central limit theorem.

III. If the sample size is large, the shape of the sampling distribution of  $\bar{x}$  is approximately normal. This statement is also true and is a key component of the central limit theorem.

Therefore, the correct answer is  $\boxed{\text{B. 'I & II only'}}$ . Final answer: B

\*\*\*\*\*

Answer Option 3:

The central limit theorem states that if we have a large enough sample size, the sampling distribution of the sample mean will be approximately normal, regardless of the shape of the population distribution.

This means that option III is true.

Option I is not true because the mean of the sampling distribution is not necessarily the same as the mean of the population. It is true that the mean of the sampling distribution is an unbiased estimator of the population mean, but they are not necessarily equal.

Option II is also not true because the standard deviation of the sampling distribution of the sample mean is equal to the standard deviation of the population divided by the square root of the sample size, not the standard deviation of the sample itself.

Therefore, the correct answer is  $\boxed{\text{D. 'III only'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

The Central Limit Theorem states that if we have a large enough sample size, the sampling distribution of the sample mean ( $\bar{x}$ ) will be approximately normal, regardless of the shape of the population distribution.

I. The Central Limit Theorem does not specifically address the mean of the sampling distribution of  $\bar{x}$ . It rather focuses on the behavior of the sample mean ( $\bar{x}$ ).

II. This statement is correct. The Central Limit Theorem states that the standard deviation of the sampling distribution of  $\bar{x}$  is equal to the standard deviation of  $x$  divided by the square root of the sample size.

III. This statement is also correct. The Central Limit Theorem states that as the sample size increases, the shape of the sampling distribution of  $\bar{x}$  becomes more and more approximately normal.

Hence, the correct answer is B. 'I & II only'. Final answer: B

\*\*\*\*\*

Answer Option 5:

The central limit theorem (CLT) states that if we have a random sample of independent and identically distributed (i.i.d.) variables with a finite mean and variance, then as the sample size increases, the sampling distribution of the sample mean approaches a normal distribution.

Statement I is incorrect. The mean of the sampling distribution of the sample mean is indeed the same as the mean of the population, not the mean of the individual sample.

Statement II is correct. According to the CLT, the standard deviation of the sampling distribution of the sample mean is equal to the standard deviation of the population divided by the square root of the sample size.

Statement III is correct. As stated earlier, the CLT implies that as the sample size increases, the shape of the sampling distribution of the sample mean becomes approximately normal.

Therefore, the correct answer is  $\boxed{\text{B. 'I & II only'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 6:

The central limit theorem states that if we have a large enough sample size, the sampling distribution of the sample mean will be approximately normally distributed, regardless of the shape of the population distribution.

I. The mean of the sampling distribution of the sample mean is indeed the same as the mean of the population. This is true, so statement I is correct.

II. The standard deviation of the sampling distribution of the sample mean is equal to the standard deviation of the population divided by the square root of the sample size. This is also true, so statement II is correct.

III. The central limit theorem does not state anything specific about the shape of the sampling distribution of the sample mean. It only states that the shape of the sampling distribution becomes approximately normal as the sample size increases. Therefore, statement III is incorrect.

Thus, the correct answer is  $\boxed{\text{B, I \& II only}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 7:

The central limit theorem states that if we have a large enough sample size, the sampling distribution of the sample mean will be approximately normally distributed regardless of the shape of the population distribution.

Statement I is not a correct description of the central limit theorem. The mean of the sampling distribution of the sample mean is the same as the mean of the population, not the mean of the individual samples.

Statement II is a correct description of the central limit theorem. The standard deviation of the sampling distribution of the sample mean is equal to the standard deviation of the population divided by the square root of the sample size.

Statement III is also a correct description of the central limit theorem. If the sample size is large, the shape of the sampling distribution of the sample mean will be approximately normal.

Therefore, the correct answer is  $\boxed{\text{B, I \& II only}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

The central limit theorem (CLT) states that if we have a large enough sample size, the sampling distribution of the sample mean will be approximately normal regardless of the shape of the population distribution. Based on this definition, we can determine which statements describe the central limit theorem.

I. The mean of the sampling distribution of  $x$  is the same as the mean of the population.

This statement is not a direct description of the central limit theorem. The central limit theorem is concerned with the shape of the sampling distribution, not the mean.

II. The standard deviation of the sampling distribution of  $x$  is the same as the standard deviation of  $x$  divided by the square root of the sample size.

This statement correctly describes the central limit theorem. According to the CLT, the standard deviation of the sampling distribution of  $x$  is equal to the standard deviation of the population divided by the square root of the sample size.

III. If the sample size is large, the shape of the sampling distribution of  $x$  is approximately normal.

This statement also correctly describes the central limit theorem. The CLT states that if the sample size is large enough, the shape of the sampling distribution of  $x$  will be approximately normal, regardless of the shape of the population distribution.

Therefore, the correct answer is  $\boxed{\text{B, I \& II only}}$ . Final answer: B

\*\*\*\*\*

Answer Option 9:

The central limit theorem states that if we have a large enough sample size, the sampling distribution of the sample mean will be approximately normal, regardless of the shape of the population distribution.

I. The mean of the sampling distribution of  $x$  being the same as the mean of the population is not a specific characteristic of the central limit theorem. It is a property of unbiased estimators, but it does not directly relate to the central limit theorem.

II. The standard deviation of the sampling distribution of  $x$  being equal to the standard deviation of  $x$  divided by the square root of the sample size is a correct description of the central limit theorem. It states that as the sample size increases, the variability in the sampling distribution of the sample mean decreases.

III. The statement that if the sample size is large, the shape of the sampling distribution of  $x$  is approximately normal is also a correct description of the central limit theorem. It states that as the sample size increases, the sampling distribution of the sample mean approaches a normal distribution.

Therefore, the correct answer is  $\boxed{\text{B, I \& II only}}$ . Final answer: B

\*\*\*\*\*

Question 37:

'A large city was interested in annexing part of the surrounding county. In a survey conducted by the local newspaper, 58 percent of respondents said they were against the annexation. During the actual vote, not all eligible voters voted, but 56 percent of the respondents voted against the annexation. Which of the following best describes the difference in the percentages obtained from the newspaper poll and the vote itself?', A. 'It is an example of nonresponse bias, the systematic tendency of individuals with particular characteristics to refuse to answer a survey question.', B. 'It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.', C. 'It is the difference between the same statistics computed from two different samples.', D. 'It is the difference between the statistic and the truth due to use of a random sample.'

=====

Answer Option 1:

The best description for the difference in the percentages obtained from the newspaper poll and the vote itself is B. 'It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.' This difference is known as sampling bias, where the respondents of the newspaper poll may not accurately represent the population as a whole. Final answer: B

\*\*\*\*\*

Answer Option 2:

The difference in the percentages obtained from the newspaper poll and the vote itself can be attributed to the nonrandom selection of surveyed persons. In other words, the respondents who participated in the newspaper poll may not be representative of the entire population of eligible voters. Therefore, the best description for this difference is option B: "It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons." Final answer: B

\*\*\*\*\*

Answer Option 3:

The difference in the percentages obtained from the newspaper poll and the vote itself can be attributed to the nonrandom selection of surveyed persons. The newspaper poll only captures the opinions of respondents who chose to participate in the survey, while the vote includes all eligible voters. This can lead to a biased representation of the population's opinion.

Therefore, the best description of the difference is option B: "It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons." Final Answer: B

\*\*\*\*\*

Answer Option 4:

The correct answer is B. 'It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.'

The difference in percentages obtained from the newspaper poll and the vote itself is likely due to sampling bias. The newspaper poll is conducted among respondents who voluntarily choose to participate, which may not be representative of the entire population. The respondents who choose to participate in the newspaper survey may have different opinions compared to the population as a whole. This can result in a difference between the percentages obtained in the survey and the actual vote.

\*\*\*\*\*

Answer Option 5:

The best answer to describe the difference between the percentages obtained from the newspaper poll and the vote itself is option B. 'It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.' In this case, the respondents of the newspaper survey may not be representative of the entire population, leading to a difference in the percentages obtained from the survey and the actual vote. This is known as selection bias, where the sample is not randomly selected, causing a systematic difference between the statistic (percentage from the survey) and the parameter (percentage from the entire population). Final Answer: B

\*\*\*\*\*

Answer Option 6:

The difference in the percentages obtained from the newspaper poll and the vote itself can be attributed to the effect of nonresponse bias. Nonresponse bias occurs when individuals with particular characteristics systematically refuse to answer a survey question. In this case, not all eligible voters participated in the vote, which means that the sample of respondents who voted in the actual vote is different from the sample of respondents in the newspaper poll. Therefore, the difference in the percentages is an example of nonresponse bias. The best answer choice that describes this situation is A. It is an example of nonresponse bias, the systematic tendency of individuals with particular characteristics to refuse to answer a survey question. Final answer: A

\*\*\*\*\*

Answer Option 7:

The difference in the percentages obtained from the newspaper poll and the vote itself is best described as option B: "It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons."

The newspaper poll and the actual vote are measuring the same parameter, which is the percentage of people who are against the annexation. However, the newspaper poll was conducted among respondents who may not represent the entire population accurately. There may be nonresponse bias, where individuals with certain characteristics are more likely to refuse to answer the survey. This can lead to a systematic difference between the statistic obtained from the poll (58 percent) and the true parameter (56 percent) obtained from the vote.

Therefore, option B best describes the difference in the percentages obtained from the two sources. Final Answer: B

\*\*\*\*\*

Answer Option 8:

The difference in the percentages obtained from the newspaper poll and the vote itself is best described as option B: 'It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.' The newspaper poll and the vote itself are two different samples, and the difference in results can be attributed to the fact that the individuals who chose to respond to the newspaper poll may not be representative of the larger population who actually voted. This nonrandom selection of surveyed persons can lead to a systematic difference between the statistic obtained from the poll (58 percent against the annexation) and the parameter (56 percent against the annexation) obtained from the vote itself. Final Answer: B

\*\*\*\*\*

Answer Option 9:

The difference in the percentages obtained from the newspaper poll and the vote itself is best described by answer choice B: "It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons."

In the newspaper poll, only respondents who voluntarily chose to participate were included, which introduces the possibility of nonresponse bias. This means that the sample of respondents may not be representative of the entire population, leading to potential differences in the results compared to the actual vote.

Additionally, the vote itself may have had a different set of eligible voters compared to the survey respondents. Not all eligible voters may have voted, which could further contribute to differences in the percentages obtained.

Therefore, the difference in the percentages obtained from the newspaper poll and the vote is due to the nonrandom selection of surveyed persons, leading to potential biases in the results. Answer choice B accurately captures this explanation. Final answer: B

\*\*\*\*\*

Question 38:

'After receiving several complaints from his customers about the store being closed on Sundays, a storekeeper decided to conduct a survey. He randomly selected 100 female customers and 120 male customers, and asked them, "Are you interested in shopping at this store on Sundays?" He counted the number of customers answering "yes" and constructed a 95 percent confidence interval for the difference by subtracting the proportions of female from the proportion of male customers in favor of shopping on Sundays. The resulting interval was  $(-0.23, -0.18)$ . Which of the following is a correct interpretation of the interval?', A. 'We are 95 percent confident that the proportion of women interested in shopping on Sundays exceeds the proportion of men interested in shopping on Sundays.', B. 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.', C. 'We are 95 percent confident that the proportion of women interested in shopping on Sundays is equal to the proportion of men interested in shopping on Sundays.', D. 'Because the interval contains negative values, it is invalid and should not be interpreted.'

=====

Answer Option 1:

The correct interpretation of the confidence interval is as follows: 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.' This is because the interval  $(-0.23, -0.18)$  is entirely negative, indicating that the proportion of men interested in shopping on Sundays is significantly higher than the proportion of women interested. Final Answer: B

\*\*\*\*\*

Answer Option 2:

The correct interpretation of the confidence interval is option B: 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.'

Since the confidence interval is negative  $(-0.23, -0.18)$ , it implies that the proportion of men interested in shopping on Sundays is higher than the proportion of women interested in shopping on Sundays.

Since the interval does not include zero, it indicates a statistically significant difference between the two proportions. Therefore, we can conclude that there is evidence to suggest that more men are interested in shopping at the store on Sundays compared to women. Final answer: B

\*\*\*\*\*

Answer Option 3:

The confidence interval is  $(-0.23, -0.18)$ . This means that the difference in proportions of male and female customers in favor of shopping on Sundays is estimated to be between  $-0.23$  and  $-0.18$  with 95% confidence. Since the interval is entirely negative, it suggests that the proportion of male customers interested in shopping on Sundays exceeds the proportion of female customers interested in shopping on Sundays. Therefore, the correct interpretation of the interval is: B. 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.' Final Answer: B

\*\*\*\*\*

Answer Option 4:

The confidence interval given is  $(-0.23, -0.18)$ , which means that the difference in proportions between women and men interested in shopping on Sundays is estimated to be between  $-0.23$  and  $-0.18$  with 95% confidence. Since the interval contains only negative values, we can interpret it as follows: "We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays." Therefore, the correct interpretation is  $\boxed{\text{B. 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 5:

The confidence interval is given as  $(-0.23, -0.18)$ . This means that based on the survey data, we are 95 percent confident that the true difference in proportions, where the proportion of male customers interested in shopping on Sundays is subtracted by the proportion of female customers interested in shopping on Sundays, falls between  $-0.23$  and  $-0.18$ .

Since the interval is negative, this means that the proportion of male customers interested in shopping on Sundays is expected to be higher than the proportion of female customers interested in shopping on Sundays. Therefore, the correct interpretation of the interval is: 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.' This corresponds to answer choice B. Final Answer: B

\*\*\*\*\*

Answer Option 6:

The confidence interval is constructed by subtracting the proportions of women from the proportion of men in favor of shopping on Sundays. Since the interval is  $(-0.23, -0.18)$ , this means that the estimated difference between the proportion of men and the proportion of women interested in shopping on Sundays is negative.

Therefore, we can interpret the interval as: 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.'

This aligns with the negative difference indicated by the interval.

Therefore, the correct interpretation is option B: 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.' Final answer: B

\*\*\*\*\*

Answer Option 7:

The confidence interval is given as  $(-0.23, -0.18)$ . The interval is entirely negative, meaning that the proportion of male customers interested in shopping on Sundays is lower than the proportion of female customers interested in shopping on Sundays. Therefore, the correct interpretation is: 'We are 95 percent confident that the proportion of women interested in shopping on Sundays exceeds the proportion of men interested in shopping on Sundays.' Thus, the answer is  $\boxed{\text{A}}$ .

\*\*\*\*\*

Answer Option 8:

The confidence interval is given as  $(-0.23, -0.18)$ . This means that we are 95 percent confident that the true difference in proportions of male and female customers interested in shopping on Sundays is between  $-0.23$  and  $-0.18$ .

Since the interval is negative, it implies that the proportion of female customers interested in shopping on Sundays is higher than the proportion of male customers interested in shopping on Sundays.

Therefore, the correct interpretation of the interval is:

B. 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.'

Final Answer: B

\*\*\*\*\*

Answer Option 9:

The confidence interval is given as  $(-0.23, -0.18)$ , which means that the difference in proportions, defined as the proportion of male customers in favor of shopping on Sundays minus the proportion of female customers in favor of shopping on Sundays, lies between  $-0.23$  and  $-0.18$  with 95 percent confidence.

Since the difference is negative and the interval lies entirely below zero, this means that the proportion of male customers interested in shopping on Sundays exceeds the proportion of female customers interested in shopping on Sundays. Therefore, the correct interpretation is "We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays." Hence, the answer is  $\boxed{\text{B. 'We are 95 percent confident that the proportion of men interested in shopping on Sundays exceeds the proportion of women interested in shopping on Sundays.'}}$ . Final Answer: B.

\*\*\*\*\*

Question 39:

'Which of the following is the primary difference between an experiment and an observational study?', A. 'Experiments are only conducted on human subjects; observational studies can be conducted on nonhuman subjects.', B. 'In an experiment, the researcher manipulates some variable to observe its effect on a response variable; in an observational study, he or she simply observes and records the observations.', C. 'Experiments must use randomized treatment and control groups; observational studies also use treatment and control groups, but they do not need to be randomized.', D. 'Experiments must be double-blind; observational studies do not need to be.'

=====

Answer Option 1:

The primary difference between an experiment and an observational study is that in an experiment, the researcher manipulates some variable to observe its effect on a response variable, while in an observational study, the researcher simply observes and records the observations. This makes choice B the correct answer. Final answer: B.

\*\*\*\*\*

Answer Option 2:

The primary difference between an experiment and an observational study is that in an experiment, the researcher manipulates some variable to observe its effect on a response variable, while in an observational study, he or she simply observes and records the observations. This is described in option B. Therefore, the correct answer is option B. Final Answer: B

\*\*\*\*\*

Answer Option 3:

The primary difference between an experiment and an observational study is that in an experiment, the researcher manipulates some variable to observe its effect on a response variable. In contrast, in an observational study, the researcher simply observes and records the observations without manipulating any variables. Therefore, the correct answer is  $\boxed{\text{B}}$  Final Answer: B

\*\*\*\*\*

Answer Option 4:

The primary difference between an experiment and an observational study is that in an experiment, the researcher manipulates some variable to observe its effect on a response variable (option B). In an observational study, the researcher simply observes and records the observations without intervening or manipulating variables. The other options are not accurate descriptions of the primary difference between the two. Therefore, the correct answer is B. Final Answer: B

\*\*\*\*\*

Answer Option 5:

The primary difference between an experiment and an observational study is that in an experiment, the researcher manipulates some variable to observe its effect on a response variable, while in an observational study, the researcher simply observes and records the observations. This is described in option B: 'In an experiment, the researcher manipulates some variable to observe its effect on a response variable; in an observational study, he or she simply observes and records the observations.' Therefore, the correct answer is option B. Final Answer: B

\*\*\*\*\*

Answer Option 6:

The primary difference between an experiment and an observational study is that in an experiment, the researcher manipulates some variable to observe its effect on a response variable, while in an observational study, he or she simply observes and records the observations. Therefore, the correct answer is  $\boxed{\text{B}}$ . 'In an experiment, the researcher manipulates some variable to observe its effect on a response variable; in an observational study, he or she simply observes and records the observations.')

\*\*\*\*\*

Answer Option 7:

The primary difference between an experiment and an observational study is that in an experiment, the researcher actively manipulates some variable to observe its effect on a response variable. On the other hand, in an observational study, the researcher simply observes and records the observations without manipulating any variables. Therefore, the correct answer is  $\boxed{\text{B}}$ . 'In an experiment, the researcher manipulates some variable to observe its effect on a response variable; in an observational study, he or she simply observes and records the observations.')

\*\*\*\*\*

Answer Option 8:

The primary difference between an experiment and an observational study is that in an experiment, the researcher manipulates some variable to observe its effect on a response variable, while in an observational study, the researcher simply observes and records the observations. This is described in option B: 'In an experiment, the researcher manipulates some variable to observe its effect on a response variable; in an observational study, he or she simply observes and records the observations.' Thus, the correct answer is option B. Final Answer: B

\*\*\*\*\*

Answer Option 9:

The primary difference between an experiment and an observational study is that in an experiment, the researcher manipulates some variable to observe its effect on a response variable. This is the defining characteristic of an experiment. On the other hand, in an observational study, the researcher simply observes and records the observations without manipulating any variables. Therefore, the correct answer is  $\boxed{\text{B}}$ . 'In an experiment, the researcher manipulates some variable to observe its effect on a response variable; in an observational study, he or she simply observes and records the observations.')

\*\*\*\*\*



Question 40:

'Random variable X is normally distributed, with a mean of 25 and a standard deviation of 4. Which of the following is the approximate interquartile range for this distribution?', A. ' $25.00 - 22.30 = 2.70$ ', B. ' $27.70 - 22.30 = 5.40$ ', C. ' $27.70 \div 22.30 = 1.24$ ', D. ' $2.00(4.00) = 8.00$ '

=====

Answer Option 1:

The interquartile range (IQR) is the range between the 25th and 75th percentiles. For a normal distribution, approximately 50% of the data lies within 1 standard deviation of the mean, and approximately 68% lies within 2 standard deviations. Therefore, the approximate 25th percentile is  $\$25 - 2 \cdot 4 = 17\$$  and the approximate 75th percentile is  $\$25 + 2 \cdot 4 = 33\$$ .

Therefore, the approximate IQR is  $\$33 - 17 = \boxed{16}\$$ . Final answer: None of the given options.

\*\*\*\*\*

Answer Option 2:

The interquartile range (IQR) is a measure of the spread of a distribution, and it is defined as the difference between the third quartile (Q3) and the first quartile (Q1).

To find the quartiles, we can use the fact that in a normal distribution, approximately 68% of the data falls within 1 standard deviation of the mean, 95% falls within 2 standard deviations, and 99.7% falls within 3 standard deviations.

Q1 will be approximately 1 standard deviation below the mean, and Q3 will be approximately 1 standard deviation above the mean. Therefore,  $Q1 = 25 - 4 = 21$  and  $Q3 = 25 + 4 = 29$ .

The IQR is the difference between Q3 and Q1, so  $IQR = 29 - 21 = 8$ .

The correct answer is D. ' $2.00(4.00) = 8.00$ '.

\*\*\*\*\*

Answer Option 3:

The interquartile range (IQR) is a measure of the spread of a distribution and is defined as the difference between the upper quartile (Q3) and the lower quartile (Q1).

To find the upper quartile and lower quartile, we need to find the values of X that correspond to the cumulative probabilities of 0.75 and 0.25, respectively.

Using a z-table or a calculator with a normal distribution function, we find that the z-score corresponding to a cumulative probability of 0.75 is approximately 0.6745 and the z-score corresponding to a cumulative probability of 0.25 is approximately -0.6745.

To find the values of X, we use the z-score formula:

$$X = \text{mean} + (z\text{-score})(\text{standard deviation})$$

For the upper quartile (Q3):

$$X = 25 + (0.6745)(4) = 27.70$$

For the lower quartile (Q1):

$$X = 25 + (-0.6745)(4) = 22.30$$

Therefore, the interquartile range is approximately  $27.70 - 22.30 = 5.40$ .

The correct answer is therefore  $\boxed{\text{B. '27.70 - 22.30 = 5.40'}}$ .

\*\*\*\*\*

Answer Option 4:

The interquartile range (IQR) is defined as the difference between the third quartile (Q3) and the first quartile (Q1) of a distribution. In a normal distribution, approximately 50% of the data lies between Q1 and Q3.

To find Q1 and Q3, we can use the properties of the normal distribution. Q1 is the value below which approximately 25% of the data lies, and Q3 is the value below which approximately 75% of the data lies.

Using a standard normal distribution table or a calculator, we can find that a z-score of approximately -0.674 corresponds to the 25th percentile, and a z-score of approximately 0.674 corresponds to the 75th percentile.

To find the values of Q1 and Q3, we can convert these z-scores back to the original units by using the formula  $z = (x - \mu) / \sigma$ , where  $\mu$  is the mean and  $\sigma$  is the standard deviation.

For Q1:

$$-0.674 = (x - 25) / 4$$

$$x - 25 = -2.696$$

$$x = 22.304$$

For Q3:

$$0.674 = (x - 25) / 4$$

$$x - 25 = 2.696$$

$$x = 27.304$$

Therefore, the approximate interquartile range is:

$$Q3 - Q1 = 27.304 - 22.304 = 5$$

The closest option to 5 is option B: ' $27.70 - 22.30 = 5.40$ '.

Thus, the answer is B. Final Answer: B

\*\*\*\*\*

Answer Option 5:

To find the interquartile range (IQR), we need to determine the values that correspond to the first quartile (Q1) and third quartile (Q3).

The first quartile, Q1, is the value below which 25% of the data falls. Since the distribution is normal, we can use the z-score to find Q1. The z-score corresponding to the first quartile is approximately -0.674.

We can calculate the value of Q1 as follows:

$$Q1 = \text{mean} + (\text{z-score for } Q1) * \text{standard deviation}$$

$$= 25 + (-0.674) * 4$$

$$= 25 - 2.696$$

$$= 22.304$$

Similarly, the third quartile, Q3, is the value below which 75% of the data falls. The z-score for the third quartile is approximately 0.674. We can calculate the value of Q3 as follows:

$$Q3 = \text{mean} + (\text{z-score for } Q3) * \text{standard deviation}$$

$$= 25 + 0.674 * 4$$

$$= 25 + 2.696$$

$$= 27.696$$

The approximate interquartile range is the difference between Q3 and Q1:

$$IQR = Q3 - Q1$$

$$= 27.696 - 22.304$$

$$= 5.392$$

The closest option to this value is B. '27.70 - 22.30 = 5.40'.

Therefore, the approximate interquartile range for this distribution is 5.40. Final answer: B.

\*\*\*\*\*

Answer Option 6:

To find the interquartile range (IQR) for a normal distribution, we need to find the difference between the first quartile (Q1) and the third quartile (Q3).

We know that the mean of the normal distribution is 25, and the standard deviation is 4.

The first quartile (Q1) is located at approximately 0.25 in the cumulative distribution function (CDF) of the normal distribution. Using a standard normal distribution table or a calculator, we can find that the z-score corresponding to the first quartile is approximately -0.674.

To find the actual value of Q1, we can use the formula  $Q1 = \text{mean} + z * \text{standard deviation}$ .

$$Q1 = 25 + (-0.674) * 4 = 25 - 2.696 = 22.304 \text{ (rounded to three decimal places).}$$

Similarly, the third quartile (Q3) is located at approximately 0.75 in the CDF of the normal distribution. The z-score corresponding to the third quartile is approximately 0.674.

$$Q3 = 25 + (0.674) * 4 = 25 + 2.696 = 27.696 \text{ (rounded to three decimal places).}$$

The interquartile range (IQR) is then given by  $Q3 - Q1$ :

$$IQR = 27.696 - 22.304 = 5.392 \text{ (rounded to three decimal places).}$$

Therefore, the approximate interquartile range for this distribution is 5.392.

The closest option to this value is B. '27.70 - 22.30 = 5.40'.

So the answer is B. Final Answer: B

\*\*\*\*\*

Answer Option 7:

The interquartile range (IQR) is a measure of the spread of the middle 50% of the data. In a normal distribution, approximately 50% of the data lies within one standard deviation of the mean.

Since the mean is 25 and the standard deviation is 4, one standard deviation below and above the mean are  $25 - 4 = 21$  and  $25 + 4 = 29$ , respectively. Therefore, the approximate interquartile range for this distribution is  $29 - 21 = 8$ .

Thus, the correct answer is  $\boxed{\text{D } 2.00(4.00) = 8.00}$ .

\*\*\*\*\*

Answer Option 8:

The interquartile range (IQR) is the range between the first quartile (Q1) and the third quartile (Q3). In a normal distribution, approximately 50% of the data lies within one standard deviation of the mean, and approximately 68% of the data lies within two standard deviations of the mean. Since the standard deviation is 4, we can approximate Q1 and Q3 as follows:

$$Q1 = 25 - 1(4) = 21$$

$$Q3 = 25 + 1(4) = 29$$

Therefore, the approximate IQR is:

$$IQR = Q3 - Q1 = 29 - 21 = 8$$

The correct answer is D. '2.00(4.00) = 8.00'. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The interquartile range (IQR) is the range between the 25th percentile and the 75th percentile of a distribution. For a normal distribution, we know that approximately 68% of the data falls within 1 standard deviation of the mean, 95% falls within 2 standard deviations, and 99.7% falls within 3 standard deviations.

Since the standard deviation is 4, the range from the mean to 1 standard deviation below the mean is  $25 - 4 = 21$ . Similarly, the range from the mean to 1 standard deviation above the mean is  $25 + 4 = 29$ .

To find the approximate IQR, we need to find the values that correspond to the 25th and 75th percentiles. Since 50% of the data falls below the mean, the 25th percentile is approximately at the mean - 1/2 standard deviation =  $25 - 2 = 23$ . Similarly, the 75th percentile is approximately at the mean + 1/2 standard deviation =  $25 + 2 = 27$ .

Therefore, the approximate IQR is  $27 - 23 = 4$ . The correct answer is not listed among the choices provided.

\*\*\*\*\*

Question 41:

'Which of the following statements about any two events A and B is true?', A. ' $P(A \cup B)$  implies events A and B are independent.', B. ' $P(A \cup B) = 1$  implies events A and B are mutually exclusive.', C. ' $P(A \cap B) = 0$  implies events A and B are independent.', D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

=====

Answer Option 1:

The correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

Two events A and B are mutually exclusive if they cannot occur at the same time. This means that if A occurs, B cannot occur, and if B occurs, A cannot occur. Mathematically, this is represented by  $P(A \cap B) = 0$ , which means the probability of both A and B occurring is 0.

On the other hand, independence of events A and B means that the occurrence or non-occurrence of one event does not affect the occurrence or non-occurrence of the other event. Mathematically, this is represented by  $P(A \cap B) = P(A) \cdot P(B)$ , where  $P(A \cap B)$  is the probability of both A and B occurring, and  $P(A)$  and  $P(B)$  are the probabilities of A and B occurring individually.

Therefore, the correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' Final Answer: D

\*\*\*\*\*

Answer Option 2:

The correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

Events A and B are mutually exclusive if and only if they cannot occur at the same time, which means their intersection ( $A \cap B$ ) must have a probability of 0. Therefore, if  $P(A \cap B) = 0$ , then events A and B are mutually exclusive.

Option A is incorrect because  $P(A \cup B)$  does not imply anything about the independence of events A and B. Events can be dependent even if their union has a high probability.

Option B is incorrect because  $P(A \cup B) = 1$  means that at least one of the events A or B must occur, but it does not imply that they are mutually exclusive. They can still both occur simultaneously.

Option C is incorrect because  $P(A \cap B) = 0$  does not necessarily mean that events A and B are independent. Independence of events is determined by the conditional probabilities, not just the intersection probability.

Therefore, the correct statement is D. Final answer: D

\*\*\*\*\*

Answer Option 3:

The correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

Two events A and B are mutually exclusive if they cannot both occur at the same time. Mathematically, this means that the intersection of A and B, denoted as  $A \cap B$ , is an empty set or has a probability of 0.

Hence, the statement ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive' is true.

Option A is incorrect because the probability of the union of two events,  $P(A \cup B)$ , does not imply that the events A and B are independent. Independence between two events refers to the occurrence (or non-occurrence) of one event not affecting the probability of the other event.  $P(A \cup B)$  only provides information about the probability of either event A or event B (or both) occurring, but it does not imply independence.

Option B is incorrect because  $P(A \cup B) = 1$  does not necessarily imply that events A and B are mutually exclusive. It only means that at least one of the events A or B will occur, but it is possible for both events A and B to occur simultaneously, making them not mutually exclusive.

Option C is incorrect because  $P(A \cap B) = 0$  does not imply that events A and B are independent. Independence between two events is determined by the conditional probability relationship  $P(A|B) = P(A)$  and  $P(B|A) = P(B)$ , not by the intersection probability being 0.

Therefore, the correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

\*\*\*\*\*

Answer Option 4:

The correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

Mutually exclusive events are events that cannot occur at the same time, meaning they have no common outcomes. Therefore, if the probability of their intersection,  $P(A \cap B)$ , is 0, it implies that they have no common outcomes and are mutually exclusive.

On the other hand, statements A, B, and C are not necessarily true.  $P(A \cup B)$  does not imply that events A and B are independent.  $P(A \cup B) = 1$  does not imply that events A and B are mutually exclusive.  $P(A \cap B) = 0$  does not imply that events A and B are independent.

Therefore, the correct answer is D. Final answer: D

\*\*\*\*\*

Answer Option 5:

The correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

Two events A and B are said to be mutually exclusive if they cannot both occur at the same time, i.e., if the intersection of A and B is empty. In other words, if  $P(A \cap B) = 0$ , then events A and B do not have any outcomes in common.

On the other hand, independence of two events A and B means that the occurrence or non-occurrence of one event does not affect the occurrence or non-occurrence of the other event. If events A and B are independent, then  $P(A \cap B) = P(A) \cdot P(B)$ , which implies that  $P(A \cap B)$  is not equal to 0.

Therefore, statement D is the correct statement. Final Answer: D.

\*\*\*\*\*

Answer Option 6:

The correct answer is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

Two events A and B are defined to be mutually exclusive if they cannot occur at the same time. This means that if event A occurs, event B cannot occur, and vice versa.

The probability of the intersection of two mutually exclusive events is 0, because they cannot occur at the same time. Therefore, if  $P(A \cap B) = 0$ , it implies that events A and B are mutually exclusive.

Note that option C, ' $P(A \cap B) = 0$  implies events A and B are independent', is incorrect. Independence and mutual exclusivity are two distinct concepts. Two events can be independent without being mutually exclusive, and vice versa.

Options A and B are also incorrect. The probability of the union of two events,  $P(A \cup B)$ , does not imply anything about the independence or mutual exclusivity of the events.

\*\*\*\*\*

Answer Option 7:

The correct statement is D: ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

Two events A and B are mutually exclusive if they cannot occur at the same time. This means that if A occurs, then B cannot occur, and if B occurs, then A cannot occur. Mathematically, this can be expressed as  $A \cap B = \emptyset$  (empty set).

The probability of the intersection of two mutually exclusive events is zero, since they cannot occur at the same time. Therefore, if  $P(A \cap B) = 0$ , it implies that events A and B are mutually exclusive.

$P(A \cup B) = 1$  implies that the union of events A and B covers the entire sample space, which means that at least one of the events A or B must occur. However, this does not necessarily imply that events A and B are mutually exclusive or independent.

Therefore, the correct statement is D: ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' Final Answer: D

\*\*\*\*\*

Answer Option 8:

The correct statement is  $\boxed{\text{D. 'P(A \cap B) = 0 implies events A and B are mutually exclusive.'}}$

Two events A and B are mutually exclusive if they cannot occur at the same time. This means that if A occurs, then B cannot occur, and vice versa. Mathematically, this is represented by  $P(A \cap B) = 0$ , which means the probability of both A and B occurring is 0.

On the other hand, if events A and B are independent, then the occurrence (or non-occurrence) of one event does not affect the occurrence (or non-occurrence) of the other event. Mathematically, this is represented by  $P(A \cap B) = P(A) \cdot P(B)$ , where  $P(A)$  and  $P(B)$  are the probabilities of events A and B occurring individually.

Therefore, statement D is true, as  $P(A \cap B) = 0$  implies that events A and B are mutually exclusive, but does not necessarily imply that they are independent.

\*\*\*\*\*

Answer Option 9:

The correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.'

Two events A and B are mutually exclusive if they cannot occur at the same time, meaning that if one of them happens, the other cannot happen. Mathematically, this is expressed as  $A \cap B = \emptyset$ , where  $\emptyset$  denotes the empty set.

If  $P(A \cap B) = 0$ , it means that the probability of both events A and B happening at the same time is 0, which implies that A and B are mutually exclusive. Therefore, statement D is true.

Statement A is incorrect because  $P(A \cup B)$  does not imply that events A and B are independent. Independence of two events means that the occurrence of one event does not affect the probability of the other event happening.

Statement B is incorrect because  $P(A \cup B) = 1$  does not necessarily imply that events A and B are mutually exclusive. It only means that at least one of the two events will occur.

Statement C is incorrect because  $P(A \cap B) = 0$  does not imply that events A and B are independent. Independence of two events is determined by the relationship between their probabilities, not the probability of their intersection.

\*\*\*\*\*

Question 42:

A company has 400 employees. Their mean income is \$20,500, and the standard deviation of their incomes is \$3,750. The distribution of incomes is normally distributed. How many of the 400 employees do you expect to have an income of between \$13,000 and \$28,000?, A. '50', B. '100', C. '200', D. '390'

=====

Answer Option 1:

To find the number of employees who have an income between \$13,000 and \$28,000, we need to find the proportion of employees within this income range.

First, we calculate the z-scores for the two income values:

$$z_1 = \frac{13000 - 20500}{3750} \approx -2.67$$

$$z_2 = \frac{28000 - 20500}{3750} \approx 2.00$$

Using a standard normal distribution table or calculator, we find that the proportion of values between  $z = -2.67$  and  $z = 2.00$  is approximately  $0.9787 - 0.0035 = 0.9752$ .

To find the number of employees in this income range, we multiply this proportion by the total number of employees:  $0.9752 \times 400 \approx 390$

Therefore, we expect approximately  $\boxed{390}$  employees to have an income between \$13,000 and \$28,000. Final Answer: D

\*\*\*\*\*

Answer Option 2:

To find the number of employees with an income between \$13,000 and \$28,000, we need to find the probability that an employee's income falls within this range.

First, we standardize the income values. The z-score for \$13,000 is given by:

$$z_1 = \frac{13,000 - 20,500}{3,750} = -2$$

Similarly, the z-score for \$28,000 is given by:

$$z_2 = \frac{28,000 - 20,500}{3,750} = 2$$

Using a standard normal distribution table or a calculator, we can find the area between these two z-scores, which represents the probability that an employee's income falls within this range. The area between  $-2$  and  $2$  is approximately  $0.9545$ .

To find the expected number of employees, we multiply the probability by the total number of employees:

$$\text{Expected number of employees} = \text{Probability} \times \text{Total number of employees}$$

$$= 0.9545 \times 400 = 381.8$$

Rounding to the nearest whole number, we expect approximately  $382$  employees to have an income between \$13,000 and \$28,000.

Thus, the correct answer is  $\boxed{\text{D. } 390}$ .

\*\*\*\*\*

Answer Option 3:

To find the number of employees with an income between \$13,000 and \$28,000, we need to find the proportion of the employees within this income range.

First, we need to convert the income values to z-scores. The z-score formula is given by  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the individual income,  $\mu$  is the mean income, and  $\sigma$  is the standard deviation.

For  $x = 13,000$ , the z-score is  $\frac{13,000 - 20,500}{3,750} = -2$ , and for  $x = 28,000$ , the z-score is  $\frac{28,000 - 20,500}{3,750} = 2$ .

Next, we can use a standard normal distribution table or calculator to find the proportion of the area under the curve between these z-scores. The proportion is approximately  $0.9545$ .

Finally, we multiply this proportion by the total number of employees to find the number of employees within the income range:  $0.9545 \times 400 \approx \boxed{382}$ . Final answer: None of the above.

\*\*\*\*\*

Answer Option 4:

To find the number of employees expected to have an income between \$13,000 and \$28,000, we need to calculate the proportion of employees within this income range and then multiply it by the total number of employees.

First, we need to standardize the income values using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the income,  $\mu$  is the mean income, and  $\sigma$  is the standard deviation.

For  $x = 13,000$ , we have:

$$z_1 = \frac{13,000 - 20,500}{3,750} = -2$$

For  $x = 28,000$ , we have:

$$z_2 = \frac{28,000 - 20,500}{3,750} = 2$$

Using a standard normal distribution table or calculator, we can find the proportion of employees within this income range. For  $-2 \leq z \leq 2$ , the proportion is approximately  $0.9545$ .

Therefore, the number of employees expected to have an income between \$13,000 and \$28,000 is:

$$0.9545 \times 400 \approx \boxed{382}$$

Note: Since 400 is a discrete number of employees, we round down to the nearest whole number. Final answer: None of the given choices.

\*\*\*\*\*

Answer Option 5:

We can use the concept of z-scores to solve this problem. The z-score measures the number of standard deviations an observation is away from the mean. The formula for the z-score is given by:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value of the observation,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

To find the number of employees with an income between \$13,000 and \$28,000, we need to find the proportion of employees within this income range and then multiply it by the total number of employees.

First, we convert the income values to z-scores:

$$z_1 = \frac{13,000 - 20,500}{3,750} = -2$$

$$z_2 = \frac{28,000 - 20,500}{3,750} = 2$$

Next, we use a z-table or calculator to find the proportion of values between -2 and 2. The proportion is approximately 0.9545.

Finally, we multiply this proportion by the total number of employees:

$$0.9545 \times 400 \approx 381$$

So, we expect approximately 390 employees to have an income between \$13,000 and \$28,000. Final Answer: D

\*\*\*\*\*

Answer Option 6:

To find the number of employees with an income between \$13,000 and \$28,000, we need to find the proportion of employees within this income range.

First, we need to standardize the income values using the z-score formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the income,  $\mu$  is the mean income, and  $\sigma$  is the standard deviation.

For an income of \$13,000:

$$z_1 = \frac{13,000 - 20,500}{3,750} = -2$$

and for an income of \$28,000:

$$z_2 = \frac{28,000 - 20,500}{3,750} = 2$$

Using a standard normal distribution table or calculator, we can find that the probability of having a z-score between -2 and 2 is approximately 0.9545.

Therefore, we can expect approximately 0.9545 (or 95.45%) of the 400 employees to have an income between \$13,000 and \$28,000.

The number of employees within this income range is  $0.9545 \times 400 = 382$ .

Final Answer: None of the above.

\*\*\*\*\*

Answer Option 7:

In order to find the number of employees with an income between \$13,000 and \$28,000, we need to find the proportion of employees within this range and multiply it by the total number of employees.

To do this, we can use the concept of z-scores. The z-score measures the number of standard deviations a particular value is from the mean. We can calculate the z-scores for the lower and upper bounds of the income range as follows:

$$\text{Lower z-score: } z_1 = \frac{13,000 - 20,500}{3,750} = -2$$

$$\text{Upper z-score: } z_2 = \frac{28,000 - 20,500}{3,750} = 2$$

Next, we can use a standard normal distribution table or calculator to find the proportion of values between these two z-scores. The proportion between  $z = -2$  and  $z = 2$  is approximately 0.9545.

Therefore, we expect approximately  $0.9545 \times 400 \approx 381.8$  employees to have an income between \$13,000 and \$28,000. Since we cannot have a fraction of an employee, the closest whole number is 390. Final Answer: D

\*\*\*\*\*

Answer Option 8:

To find the number of employees expected to have an income between \$13,000 and \$28,000, we need to find the proportion of the distribution that falls within this range.

First, we need to standardize the values of \$13,000 and \$28,000 using the formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For \$13,000:

$$z = \frac{13,000 - 20,500}{3,750} = -2$$

For \$28,000:

$$z = \frac{28,000 - 20,500}{3,750} = 2$$

Using a standard normal distribution table or calculator, we can find that the proportion of values falling between  $-2$  and  $2$  is approximately 0.9545.

To find the expected number of employees, we multiply the proportion by the total number of employees:

$$0.9545 \times 400 = 381.8$$

Since we cannot have a fraction of an employee, we round this number to the nearest whole number. Therefore, we expect approximately 382 employees to have an income between \$13,000 and \$28,000. Final Answer: None of the above.

\*\*\*\*\*

Answer Option 9:

We can use the properties of the normal distribution to solve this problem. We know that the mean income is \$20,500 and the standard deviation is \$3,750. To find the number of employees with an income between \$13,000 and \$28,000, we need to find the probability that an employee's income falls within this range.

First, we need to standardize the values of \$13,000 and \$28,000 using the formula

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. For \$13,000, we have

$$z_1 = \frac{13,000 - 20,500}{3,750} = -2$$

For \$28,000, we have

$$z_2 = \frac{28,000 - 20,500}{3,750} = 2$$

Next, we can use a standard normal distribution table or a calculator to find the probabilities associated with these standardized values. The probability that an employee's income is less than \$13,000 is approximately 0.0228, and the probability that an employee's income is less than \$28,000 is approximately 0.9772. Therefore, the probability that an employee's income falls between \$13,000 and \$28,000 is approximately  $0.9772 - 0.0228 = 0.9544$ .

Finally, we multiply this probability by the total number of employees, 400, to find the expected number of employees with an income between \$13,000 and \$28,000. This is approximately  $0.9544 \times 400 = 381.76$ . Rounded to the nearest whole number, we expect 382 employees to have an income in this range.

Final Answer: D

\*\*\*\*\*



Question 43:

'As reported on CNN, in a May 1999 national poll 43% of high school students expressed fear about going to school. Which of the following best describes what is meant by the poll having a margin of error of 5%?', A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%', B. 'Five percent of the students refused to participate in the poll', C. 'Between 38% and 48% of those surveyed expressed fear about going to school', D. 'There is a 0.05 probability that the 43% result is in error.'

=====

Answer Option 1:

A margin of error of 5% means that the reported percentage (43%) is likely to be accurate within 5 percentage points. In other words, the true proportion of high school students afraid to go to school is likely to be within a range that is 5% above or below the reported percentage. Therefore, the best description is option A: 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.' Final answer: A

\*\*\*\*\*

Answer Option 2:

A margin of error of 5% means that the reported percentage of 43% may vary by up to 5%. This means that the true proportion of high school students afraid to go to school could be as low as  $43\% - 5\% = 38\%$  or as high as  $43\% + 5\% = 48\%$ . Therefore, the best description of the poll having a margin of error of 5% is A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.' Final answer: A

\*\*\*\*\*

Answer Option 3:

A margin of error of 5% means that the actual proportion of high school students who expressed fear about going to school could be up to 5% higher or lower than the reported 43%. Therefore, the best description would be option A: "It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%." Final Answer: A

\*\*\*\*\*

Answer Option 4:

A margin of error of 5% means that the true proportion being estimated is likely to fall within 5% of the reported value. In this case, the reported value is 43%, so the true proportion is likely to fall within 5% of 43%, or between 38% and 48%. Therefore, the best description is A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.' Final Answer: A

\*\*\*\*\*

Answer Option 5:

The margin of error is a measure of the uncertainty or variability in a poll or survey. It represents the range within which the true value of the population parameter is likely to fall. In this case, the margin of error is 5%, which means that the true proportion of high school students who are afraid to go to school is likely to be within 5 percentage points of the reported value of 43%.

Therefore, the best description is that it is likely that the true proportion of high school students afraid to go to school is between 38% and 48%. This corresponds to option A. So the correct answer is A. Final answer: A.

\*\*\*\*\*

Answer Option 6:

A margin of error of 5% means that the true proportion is likely within 5% of the reported result. In other words, we can be reasonably confident that the true proportion lies within a range of 5% above and below the reported result.

So, the best description is A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.' Final answer: A

\*\*\*\*\*

Answer Option 7:

A margin of error of 5% means that the actual proportion of high school students who are afraid to go to school could be 5% higher or 5% lower than the reported 43%. This creates a range of possible values from 38% to 48%. Therefore, the best description is: 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.' Thus, the correct answer is  $\boxed{\text{A}}$

\*\*\*\*\*

Answer Option 8:

A margin of error of 5% means that the reported percentage of 43% could be off by up to 5%. This means that the true proportion could be as low as  $43\% - 5\% = 38\%$  or as high as  $43\% + 5\% = 48\%$ . Therefore, the best description is  $\boxed{\text{A}}$ . 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.' Final Answer: A

\*\*\*\*\*

Answer Option 9:

The margin of error in a poll refers to the maximum amount by which the reported result may differ from the true population parameter. In this case, the reported result is 43% of high school students expressing fear about going to school.

If the margin of error is 5%, it means that the true proportion of high school students afraid to go to school could be 5% higher or lower than the reported result.

Therefore, the best description of what is meant by the poll having a margin of error of 5% is option A: "It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%."

This option correctly captures the possible range of the true proportion based on the reported result and margin of error.

Final answer: A

\*\*\*\*\*

Question 44:

'Suppose you wish to compare the AP Statistics exam results for the male and female students taking AP Statistics at your high school. Which is the most appropriate technique for gathering the needed data?', A. 'Census', B. 'Sample survey', C. 'Experiment', D. 'Observational study'

=====

Answer Option 1:

The most appropriate technique for gathering the needed data would be a D. 'Observational study'. This involves observing and collecting data on the variables of interest (in this case, the AP Statistics exam results) from individuals without manipulating any variables or imposing any treatments. By observing and comparing the exam results of male and female students, we can gather information on the performance of each group without interfering or influencing the results. Final Answer: D

\*\*\*\*\*

Answer Option 2:

The most appropriate technique for gathering the needed data would be a 'Sample survey'. A sample survey involves selecting a representative sample from the population of interest and collecting data from that sample. In this case, the population of interest is the male and female students taking AP Statistics at your high school. By conducting a sample survey, you can gather data on a subset of the population and make inferences about the entire population based on the characteristics of the sample. Final Answer: B

\*\*\*\*\*

Answer Option 3:

The most appropriate technique for gathering the needed data to compare the AP Statistics exam results for male and female students would be a 'Sample survey' (B). This involves collecting data from a subset (sample) of the population of interest (students taking AP Statistics) rather than collecting data from the entire population (census). By sampling a smaller group, it would be more feasible and efficient to collect the necessary data. Final Answer: B

\*\*\*\*\*

Answer Option 4:

The most appropriate technique for gathering the needed data to compare the AP Statistics exam results for male and female students in this scenario is a 'Sample survey'. A census would involve collecting data from every single student, which may not be feasible or necessary. An experiment involves manipulating variables, which is not applicable in this case. An observational study involves passively observing and recording data without intervening or manipulating variables, which may not provide enough control to compare the exam results accurately. A sample survey, on the other hand, involves collecting data from a subset (sample) of the population of interest, in this case, male and female students taking AP Statistics. This method allows for obtaining a representative sample and generalizing the results to the larger population. Therefore, the most appropriate technique is a 'Sample survey'. The answer is B.

\*\*\*\*\*

Answer Option 5:

The most appropriate technique for gathering the needed data would be a 'Census', option A. A census involves collecting data from every individual in the population of interest, which in this case would be every male and female student taking AP Statistics at the high school. This would provide a complete and accurate picture of the exam results for both male and female students. Final answer: A

\*\*\*\*\*

Answer Option 6:

The most appropriate technique for gathering the needed data in this case would be a 'Census'. A census involves collecting data from the entire population of interest, which in this case would be all the male and female students taking AP Statistics at the high school. By collecting data from every student, we would obtain a comprehensive and accurate representation of the AP Statistics exam results for both male and female students. Therefore, the answer is  $\boxed{\text{A. Census}}$ .

\*\*\*\*\*

Answer Option 7:

The most appropriate technique for gathering the needed data to compare the AP Statistics exam results for male and female students would be a 'Sample survey'. A sample survey involves selecting a subset (sample) of the population of interest (students taking AP Statistics) and collecting data from that sample. This allows us to obtain information about the entire population without having to survey every single student (census). In this case, we can randomly select a representative sample of male and female students taking AP Statistics and collect their exam results. Therefore, the most appropriate technique is B. 'Sample survey'. Final Answer: B

\*\*\*\*\*

Answer Option 8:

The most appropriate technique for gathering the needed data in this scenario would be a "Sample survey" (Option B). This involves selecting a representative sample of the male and female students taking AP Statistics at your high school and collecting data from them. A sample survey allows you to gather information about the entire population of interest (male and female students) by studying a smaller, manageable subset (the sample). This method is commonly used in situations where it is not feasible or practical to conduct a census (option A). An experiment (option C) would involve manipulating variables and observing their effects, which may not be suitable for this scenario. An observational study (option D) involves observing and collecting data without intervening or manipulating variables, which may not provide the specific data needed to compare the AP Statistics exam results for male and female students. Final Answer: B

\*\*\*\*\*

Answer Option 9:

The most appropriate technique for gathering the needed data in this scenario is a 'Sample survey'. A sample survey involves selecting a subset of the population (in this case, the students taking AP Statistics) and collecting data from that subset. By collecting data from both male and female students, we can compare the AP Statistics exam results between the two groups. This method allows us to gather information from a smaller group, which is more feasible and practical than conducting a census (option A). An experiment (option C) involves manipulating variables and observing the effect, which is not applicable in this scenario. An observational study (option D) involves observing and collecting data without interfering or manipulating any variables, which is not the most appropriate method for

comparing exam results in this case. Therefore, the most appropriate technique is a 'Sample survey', making option B the correct answer. Final answer: B

\*\*\*\*\*

Question 45:

'In one study half of a class were instructed to watch exactly 1 hour of television per day, the other half were told to watch 5 hours per day, and then their class grades were compared. In a second study students in a class responded to a questionnaire asking about their television usage and their class grades.', A. 'The first study was an experiment without a control group, while the second was an observational study.', B. 'The first study was an observational study, while the second was a controlled experiment.', C. 'Both studies were controlled experiments.', D. 'Both studies were observational studies.'

=====

Answer Option 1:

Based on the information provided, it can be determined that the first study was a controlled experiment. The researchers intentionally assigned different groups of students to watch different amounts of television and then compared their class grades. This indicates that the researchers had control over the variables being studied.

On the other hand, the second study involved students responding to a questionnaire about their television usage and class grades. The researchers did not manipulate any variables or assign different groups of students to different conditions. Instead, they observed and collected data from the participants. Therefore, the second study was an observational study.

Thus, the correct answer is  $\boxed{\text{B. 'The first study was an observational study, while the second was a controlled experiment.'}}$

\*\*\*\*\*

Answer Option 2:

The first study divided the class into two groups and assigned different treatments to each group (watching 1 hour vs. 5 hours of television). Then, their class grades were compared. This setup allows for a comparison between the two groups and establishes a cause-and-effect relationship between the treatment (television usage) and the outcome (class grades). Therefore, the first study is an experiment.

On the other hand, the second study only collected responses from students regarding their television usage and class grades. There was no manipulation of treatments or assignment of groups. This study only observed and collected data without any intervention or control over the variables. Therefore, the second study is an observational study.

Based on this, the correct answer is A. 'The first study was an experiment without a control group, while the second was an observational study.' Final Answer: A

\*\*\*\*\*

Answer Option 3:

The first study assigned different treatments (1 hour vs 5 hours of television) to different groups and then compared their class grades. This is a classic example of an experiment with a control group, as the two groups were treated differently and their outcomes were compared. Therefore, the correct answer is A. 'The first study was an experiment without a control group, while the second was an observational study.' Final answer: A

\*\*\*\*\*

Answer Option 4:

The first study where half of the class were instructed to watch different amounts of television and then their class grades were compared is a controlled experiment. The researchers manipulated the independent variable (television usage) by assigning different amounts of television to the participants, and then measured the dependent variable (class grades) to see if there was any effect. Thus, the correct answer is B. 'The first study was an observational study, while the second was a controlled experiment.' Final answer: B

\*\*\*\*\*

Answer Option 5:

The first study divided the class into two groups and assigned different treatments (1 hour of TV vs 5 hours of TV) to each group. This means that the first study was an experiment with a control group. On the other hand, the second study simply asked students about their TV usage and class grades without manipulating any variables. This makes it an observational study. Therefore, the correct answer is A. 'The first study was an experiment without a control group, while the second was an observational study.' Final answer: A.

\*\*\*\*\*

Answer Option 6:

Based on the information provided, the first study involved instructing different groups of students to watch different amounts of television per day, and then comparing their class grades. This suggests that there was an intentional manipulation of the independent variable (amount of television watched) and a comparison of the dependent variable (class grades) between the different groups. Therefore, the first study can be classified as an experiment.

On the other hand, the second study involved students responding to a questionnaire about their television usage and class grades. There was no deliberate manipulation of variables by the researchers, but rather an observation and collection of data from the students. Therefore, the second study can be classified as an observational study.

Hence, the correct answer is A. 'The first study was an experiment without a control group, while the second was an observational study.'

\*\*\*\*\*

Answer Option 7:

Based on the given information, the first study divided the class into two groups and instructed each group to watch a different amount of television. The researchers then compared their class grades. This setup indicates that the researchers were intentionally manipulating the independent variable (television usage) and observing the effects on the dependent variable (class grades). Therefore, the first study can be classified as an experiment.

In contrast, the second study involved students responding to a questionnaire about their television usage and class grades. The researchers did not manipulate any variables or assign participants to different groups. Rather, they simply observed and collected data on the variables of interest. Therefore, the second study can be classified as an observational study.

Based on these classifications, the correct answer is A. 'The first study was an experiment without a control group, while the second was an observational study.' Final answer: A

\*\*\*\*\*

Answer Option 8:

The first study involved instructing half of the class to watch different amounts of television and then comparing their class grades. This means that the researchers manipulated the independent variable (amount of television watched) and measured the dependent variable (class grades). Therefore, the first study was an experiment. However, the first study did not have a control group. In a controlled experiment, there would be a group that does not receive any treatment (in this case, not watching any television) to serve as a comparison. Since there was no control group in the first study, we can eliminate options C and B.

In the second study, the researchers simply observed and recorded information about the students' television usage and class grades. They did not manipulate any variables or assign different treatments to different groups. Therefore, the second study was an observational study.

Based on this information, the correct answer is A. 'The first study was an experiment without a control group, while the second was an observational study.' Final Answer: A

\*\*\*\*\*

Answer Option 9:

The first study divided the class into two groups and instructed each group to watch a different amount of television. The researchers then compared their class grades. This design allows for a comparison between the two groups and can establish a cause-and-effect relationship between television usage and class grades. Therefore, the first study can be classified as an experiment, and since there was no control group, it was an experiment without a control group.

The second study involved students responding to a questionnaire about their television usage and class grades. This design does not involve any manipulation of variables or random assignment. The researchers simply observed and collected data on the variables of interest. Therefore, the second study can be classified as an observational study.

Based on these explanations, the correct answer is A. 'The first study was an experiment without a control group, while the second was an observational study.' Final Answer: A

\*\*\*\*\*

Question 46:

'In a high school of 1650 students, 132 have personal investments in the stock market. To estimate the total stock investment by students in this school, two plans are proposed. Plan I would sample 30 students at random, find a confidence interval estimate of their average investment, and then multiply both ends of this interval by 1650 to get an interval estimate of the total investment. Plan II would sample 30 students at random from among the 132 who have investments in the market, find a confidence interval estimate of their average investment, and then multiply both ends of this interval by 132 to get an interval estimate of the total investment. Which is the better plan for estimating the total stock market investment by students in this school?', A. 'Plan I', B. 'Plan II', C. 'Both plans use random samples and so will produce equivalent results.', D. 'Neither plan will give an accurate estimate.'

=====

Answer Option 1:

Plan II would be the better plan for estimating the total stock market investment by students in this school. This is because Plan II samples from the group of students who already have investments in the market, which is the group of interest. Plan I, on the other hand, samples from the entire student population, including those who do not have investments. As a result, Plan II provides a more targeted and focused estimate of the total investment by considering only the students who are actually invested in the stock market. Thus, the answer is  $\boxed{\text{B. 'Plan II'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 2:

The better plan for estimating the total stock market investment by students in this school is 'Plan II'.

In 'Plan I', the sample of 30 students is chosen randomly from the entire population of 1650 students. However, since only 132 students have investments in the market, the estimate of the average investment based on this sample may not accurately represent the average investment of the students who actually have investments. Multiplying the interval estimate by 1650 would not provide an accurate estimate of the total investment.

In 'Plan II', the sample of 30 students is chosen randomly from the subset of 132 students who have investments in the market. This sample is more representative of the population of interest and provides a better estimate of the average investment. Multiplying the interval estimate by 132 gives a more accurate estimate of the total investment.

Therefore, 'Plan II' is the better plan for estimating the total stock market investment by students in this school. Final answer: B

\*\*\*\*\*

Answer Option 3:

Plan II is the better plan for estimating the total stock market investment by students in this school.

In Plan I, we are sampling 30 students at random from the entire population of 1650 students. The confidence interval estimate of the average investment will be based on this random sample from the entire population. However, since only 132 students have investments in the stock market, the estimate may not accurately reflect the average investment of the students who actually have investments.

In Plan II, we are sampling 30 students at random from the specific group of 132 students who have investments in the stock market. This sample is more representative of the population we are interested in estimating, and therefore, the confidence interval estimate of the average investment will likely be more accurate.

Therefore, Plan II is the better plan for estimating the total stock market investment by students in this school. Answer: B

\*\*\*\*\*

Answer Option 4:

Plan II is the better plan for estimating the total stock market investment by students in this school.

In Plan I, the random sample of 30 students is taken from the entire student population of 1650. This means that the average investment of the 30 sampled students may not be representative of the average investment of all students, including those who do not have investments in the stock market. Multiplying both ends of the confidence interval by 1650 may result in an overestimate or underestimate of the total investment.

In Plan II, the random sample of 30 students is taken from the group of students who already have investments in the stock market. This sample is more likely to be representative of the average investment of all students who have investments. Multiplying both ends of the confidence interval by 132 will give a better estimate of the total investment by these students.

Therefore, Plan II is the better plan for estimating the total stock market investment by students in this school. The answer is  $\boxed{\text{Plan II}}$ .

\*\*\*\*\*

Answer Option 5:

Plan II is the better plan for estimating the total stock market investment by students in this school.

The reason is that Plan II samples only from the students who have investments in the market, which provides a more targeted and focused sample. Since the total investment is determined by these students, it makes sense to sample from this specific group.

On the other hand, Plan I samples randomly from the entire population of students, including those who may not have any investments. This introduces more variability and uncertainty in the estimate, as the average investment of the non-investing students may be significantly different from the average investment of the investing students.

Therefore, Plan II provides a more accurate and precise estimate of the total investment by focusing on the relevant subgroup of students. The confidence interval obtained from Plan II will better reflect the true average investment of the investing students, resulting in a more reliable estimate of the total investment.

The correct answer is B. 'Plan II'.

\*\*\*\*\*

Answer Option 6:

To determine which plan is better for estimating the total stock market investment, we need to consider the representativeness of the samples and the precision of the estimates.

Plan I samples randomly from the 1650 students, which includes both students with and without investments. This sample may not be representative of the entire population in terms of investment behavior.

Therefore, the confidence interval estimate of the average investment from this sample may not accurately reflect the true average investment of all students.

Plan II samples randomly from the 132 students who have investments. This sample is more representative of the population we are interested in - students with investments in the stock market. Therefore, the confidence interval estimate of the average investment from this sample is likely to be more accurate and reflective of the true average investment of all students with investments.

Based on these considerations, Plan II is the better plan for estimating the total stock market investment by students in this school. Therefore, the answer is  $\boxed{\text{Plan II}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 7:

Plan II is the better plan for estimating the total stock market investment by students in this school.

In Plan I, we are sampling 30 students randomly from the entire population of 1650 students. This means that some of the selected students may not have investments in the stock market. Therefore, the average investment calculated from this sample may not be representative of the entire population. Multiplying the interval estimate based on this sample by 1650 may result in an inaccurate estimate of the total investment.

In Plan II, we are sampling 30 students randomly from the subset of 132 students who have investments in the stock market. This sample is more targeted and representative of the population we are interested in. The average investment calculated from this sample is more likely to be reflective of the average investment of all students with investments in the market. Multiplying the interval estimate based on this sample by 132 gives a more accurate estimate of the total investment.

Therefore, Plan II is the better plan for estimating the total stock market investment by students in this school. The answer is  $\boxed{\text{B. 'Plan II'}}$ .

\*\*\*\*\*

Answer Option 8:

Plan II is the better plan for estimating the total stock market investment by students in this school.

Since Plan II only samples from the population of students who have investments in the market, it provides a more targeted estimate of the average investment among those students. Multiplying the interval estimate by 132 then gives a more accurate estimate of the total investment by these students.

In contrast, Plan I samples from the entire population of students, including those who do not have investments in the market. This broader sampling may dilute the estimate of the average investment, leading to a less accurate interval estimate of the total investment.

Therefore, Plan II is more appropriate and would provide a better estimate of the total stock market investment by students in this school. The answer is: B. 'Plan II'.

\*\*\*\*\*

Answer Option 9:

Plan II is the better plan for estimating the total stock market investment by students in this school.

In Plan I, we sample 30 students at random from the entire population of 1650 students. This means that some of the sampled students may not have any investments in the stock market. The confidence interval estimate of their average investment might include students who have zero investment, which would not accurately represent the total stock market investment by students in the school.

In Plan II, we sample 30 students at random from the 132 students who have investments in the market. This ensures that all the sampled students have investments in the stock market. The confidence interval estimate of their average investment will only include students who have investments, providing a more accurate representation of the total stock market investment by students in the school.

Therefore, Plan II is the better plan for estimating the total stock market investment by students in this school. The correct answer is B. Final Answer: B

\*\*\*\*\*

Question 47:

'Sampling error occurs', A. 'when interviewers make mistakes resulting in bias.', B. 'because a sample statistic is used to estimate a population parameter.', C. 'when interviewers use judgment instead of random choice in picking the sample.', D. 'when samples are too small.'

=====

Answer Option 1:

Sampling error occurs because a sample statistic is used to estimate a population parameter. Therefore, the correct answer is B. 'because a sample statistic is used to estimate a population parameter.'

Final answer: B

=====

Answer Option 2:

Sampling error occurs due to the inherent variability in the process of randomly selecting a sample from a population. This means that the sample may not perfectly represent the entire population. Therefore, the correct answer is B. 'because a sample statistic is used to estimate a population parameter.'

=====

Answer Option 3:

Sampling error occurs due to the inherent variability in the process of selecting a sample from a population. This means that even if a sample is randomly selected and representative of the population, the sample statistic (such as the mean or proportion) may differ from the true population parameter. Therefore, the correct answer is B. 'because a sample statistic is used to estimate a population parameter.'

Final Answer: B

=====

Answer Option 4:

Sampling error occurs because a sample statistic is used to estimate a population parameter. Therefore, the correct answer is B. 'because a sample statistic is used to estimate a population parameter.'

Final answer: B.

=====

Answer Option 5:

Sampling error occurs due to B. 'because a sample statistic is used to estimate a population parameter.' Sampling error refers to the difference between the sample statistic and the true population parameter that it is trying to estimate. This error arises due to the variability inherent in taking a sample from a population rather than collecting data from the entire population. Final answer: B

=====

Answer Option 6:

The correct answer is B. 'because a sample statistic is used to estimate a population parameter.' Sampling error refers to the discrepancy between the sample statistic (such as the sample mean) and the true population parameter (such as the population mean). This error arises due to the fact that we are using a sample to make inferences about the population, and the sample may not perfectly represent the population. Final Answer: B

=====

Answer Option 7:

Sampling error occurs due to random variation when a sample is used to estimate a population parameter. Therefore, the correct answer is B. 'because a sample statistic is used to estimate a population parameter.'

Final Answer: B

=====

Answer Option 8:

Sampling error occurs because a sample statistic is used to estimate a population parameter. Therefore, the correct answer is B. 'because a sample statistic is used to estimate a population parameter.'

Final answer: B.

=====

Answer Option 9:

The correct answer is B. 'Sampling error occurs because a sample statistic is used to estimate a population parameter.'

Sampling error refers to the difference between the sample statistic (e.g., mean, proportion) and the actual population parameter it is trying to estimate. This error occurs due to the inherent variability in sampling, as the sample may not perfectly represent the entire population. Final answer: B.

=====



Question 48:

The regression analysis of question 20 indicated that "R-sq = 98.1%." Which of the following is (are) true? I. There is a strong positive linear relationship between the explanatory and response variables. II. There is a strong negative linear relationship between the explanatory and response variables. III. About 98% of the variation in the response variable can be explained by the regression on the explanatory variable.', A. 'I and III only', B. 'I or II only', C. 'I or II (but not both) and III', D. 'II and III only'

=====

Answer Option 1:

The statement "R-sq = 98.1%" means that 98.1% of the variation in the response variable can be explained by the regression on the explanatory variable. This confirms statement III.

However, it does not directly provide information about the direction or strength of the linear relationship between the variables. Therefore, we cannot conclude whether there is a strong positive or strong negative linear relationship based solely on the given information.

Therefore, the correct answer is  $\boxed{\text{A. 'I and III only'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 2:

The statement "R-sq = 98.1%" means that 98.1% of the variation in the response variable can be explained by the regression on the explanatory variable. Therefore, statement III is true.

However, the value of R-squared does not indicate whether the relationship between the explanatory and response variables is positive or negative. Therefore, neither statement I nor statement II can be determined from the given information.

Hence, the answer is  $\boxed{\text{A. 'I and III only'}}$ . Final answer: A

\*\*\*\*\*

Answer Option 3:

The coefficient of determination,  $R^2$ , represents the proportion of the variation in the response variable that can be explained by the regression on the explanatory variable. In this case, since  $R^2$  is 98.1%, it means that 98.1% of the variation in the response variable can be explained by the regression on the explanatory variable. Therefore, statement III is true.

However, we cannot determine the direction of the linear relationship between the variables based solely on the value of  $R^2$ . So, we cannot conclude whether there is a strong positive or negative linear relationship between the explanatory and response variables. Therefore, neither statement I nor statement II can be determined to be true.

Therefore, the correct answer is  $\boxed{\text{D. 'II and III only'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

The given information tells us that the coefficient of determination,  $R^2$ , is 98.1%. This means that 98.1% of the variation in the response variable can be explained by the regression on the explanatory variable. Therefore, statement III is true.

However, we cannot determine the sign of the relationship between the explanatory and response variables from the given information. The coefficient of determination tells us the proportion of variation in the response variable that is explained by the regression model, but it does not tell us the direction or strength of the relationship. Therefore, neither statement I nor statement II can be determined from the information given.

Hence, the correct answer is  $\boxed{\text{D. II and III only}}$ .

\*\*\*\*\*

Answer Option 5:

R-squared, denoted as R-sq, measures the proportion of the total variation in the response variable that can be explained by the regression on the explanatory variable. In this case, R-sq is 98.1%.

From this information, we can conclude that statement III is true: About 98% of the variation in the response variable can be explained by the regression on the explanatory variable.

However, we cannot make conclusions about the direction of the relationship (positive or negative) based on R-squared alone. Therefore, neither statement I nor statement II can be determined from the given information.

Hence, the correct answer is  $\boxed{\text{A. 'I and III only'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 6:

The statement "R-sq = 98.1%" refers to the coefficient of determination, which measures the proportion of the total variation in the response variable that is explained by the regression model. Therefore, statement III is true: about 98% of the variation in the response variable can be explained by the regression on the explanatory variable.

However, the coefficient of determination alone does not provide information about the direction or strength of the relationship between the variables. Therefore, we cannot conclude whether there is a positive or negative linear relationship between the explanatory and response variables based solely on the R-sq value. Therefore, statement I and II cannot be determined from the information given.

Hence, the correct answer is  $\boxed{\text{D. 'II and III only'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

The statement "R-sq = 98.1%" refers to the coefficient of determination, which measures the proportion of the variance in the response variable that can be explained by the regression on the explanatory variable. Therefore, statement III is true: about 98% of the variation in the response variable can be explained by the regression on the explanatory variable.

However, we cannot determine the direction of the relationship between the variables based solely on the coefficient of determination. The coefficient of determination only tells us the strength, not the direction, of the relationship. Therefore, statement I and statement II cannot be concluded from the given information.

Therefore, the correct answer is  $\boxed{\text{D. 'II and III only'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 8:

The statement " $R^2 = 98.1\%$ " means that 98.1% of the variation in the response variable can be explained by the regression on the explanatory variable. This statement is equivalent to saying that 98.1% of the total variation in the response variable can be accounted for by the regression model. Therefore, statement III is true.

However, we cannot determine the direction (positive or negative) of the linear relationship based solely on the value of  $R^2$ . The value of  $R^2$  only tells us the strength of the linear relationship, not its direction. Therefore, neither statement I nor statement II can be concluded from the given information.

Therefore, the correct answer is  $\boxed{\text{D. 'II and III only'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 9:

The statement " $R^2 = 98.1\%$ " indicates the coefficient of determination, also known as the proportion of the variation in the response variable that can be explained by the regression on the explanatory variable.

I. Since the coefficient of determination is high (98.1%), it indicates a strong positive linear relationship between the explanatory and response variables. Therefore, statement I is true.

II. The statement does not provide any information about a strong negative linear relationship. Therefore, statement II is false.

III. The statement explicitly mentions that "about 98% of the variation in the response variable can be explained by the regression on the explanatory variable." Therefore, statement III is true.

Combining the true statements, the correct answer is  $\boxed{\text{(A) 'I and III only'}}$ . Final Answer: A

\*\*\*\*\*

Question 49:

'In a clinic, 50 patients with sleep disorders are randomly assigned to one of two different groups. Patients in one group are given medication before bedtime. Patients in the other group are given blindfolds and played soft music at bedtime. Each patient is attached to a machine that records breathing patterns. From the patterns, it is possible to determine if the patient is awake or asleep. The data will be used to decide which method is more effective in helping patients with sleep disorders. Which of the following statements is correct in the context of this experiment?', A. 'It is not possible for this experiment to be single-blind or double-blind, because human subjects are involved.', B. 'This is only a single-blind experiment because the experimenters did not include a control group that received neither medication nor a blindfold and soft music.', C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.', D. 'This is a double-blind experiment because the experimenters could prevent both groups of patients from knowing what sort of treatment the other group is receiving.'

=====

Answer Option 1:

The correct statement in this context is C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.' In a single-blind experiment, the participants are aware of the treatment they are receiving, while the researchers or doctors conducting the experiment remain blinded to the treatment assignment. In this case, the patients are aware if they are receiving medication or blindfolds and soft music, but the examining doctor may not know which treatment each patient received. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The correct statement in the context of this experiment is option D. 'This is a double-blind experiment because the experimenters could prevent both groups of patients from knowing what sort of treatment the other group is receiving.' In a double-blind experiment, neither the participants nor the experimenters know which treatment is being administered to each individual. This helps to eliminate bias and ensures that the results are objective. In this case, the experimenters can assign the patients randomly to either group and provide the treatment without disclosing the information to either the patients or the experimenters themselves. Final answer: D.

\*\*\*\*\*

Answer Option 3:

The correct statement in the context of this experiment is C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.' In a single-blind experiment, the subjects are unaware of the treatment they are receiving, but the experimenters are aware. In this case, the patients will be aware of whether they are receiving medication or blindfolds and soft music, but the examining doctor may not be aware of the treatment. Therefore, this experiment cannot be considered double-blind. Final Answer: C

\*\*\*\*\*

Answer Option 4:

The correct statement in the context of this experiment is C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.' In a single-blind experiment, the patients are aware of the treatment they are receiving, but the experimenters or investigators are not. In this case, the patients will know if they are receiving medication or blindfold and soft music, but the examining doctor might not be aware of the treatment assigned to each patient.

A double-blind experiment, on the other hand, is one where neither the patients nor the experimenters know which treatment each patient is receiving. This is not the case in this experiment, as the patients are aware of their assigned treatment.

Therefore, the correct statement is C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.' Final Answer: C

\*\*\*\*\*

Answer Option 5:

In this experiment, the patients are randomly assigned to one of two groups: the medication group or the blindfold and soft music group. This means that the patients do not have control over which treatment they receive. Therefore, option C is incorrect.

To determine which method is more effective, it is necessary to have a control group that does not receive either medication or the blindfold and soft music treatment. However, the option B states that there is no control group, which is incorrect.

Option A states that it is not possible for this experiment to be single-blind or double-blind because human subjects are involved. However, this statement is incorrect as it is possible to conduct a single-blind or double-blind experiment even with human subjects.

The correct statement is D. This is a double-blind experiment because the experimenters can prevent both groups of patients from knowing what sort of treatment the other group is receiving. This helps to eliminate any bias that may arise from the patients or the experimenters knowing which treatment is being administered. Final Answer: D.

\*\*\*\*\*

Answer Option 6:

In this experiment, the patients are randomly assigned to one of two different groups. One group receives medication before bedtime and the other group receives blindfolds and soft music at bedtime. The experiment aims to compare the effectiveness of the two methods in helping patients with sleep disorders.

Option A states that it is not possible for this experiment to be single-blind or double-blind because human subjects are involved. However, this statement is incorrect. Single-blind and double-blind experiments can be conducted with human subjects.

Option B states that this is only a single-blind experiment because the experimenters did not include a control group that received neither medication nor a blindfold and soft music. However, this statement is also incorrect. The experiment does not necessarily require a control group for it to be considered single-blind or double-blind.

Option C states that this is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not. This statement is also incorrect. In a single-blind experiment, the patients are not aware of which treatment they are receiving, while the examiners are aware. However, in this experiment, patients will know which treatment they are receiving, so it cannot be classified as a single-blind experiment.

Option D states that this is a double-blind experiment because the experimenters could prevent both groups of patients from knowing what sort of treatment the other group is receiving. This statement is correct. In a double-blind experiment, both the patients and the experimenters are unaware of which treatment is being administered. In this case, if the experimenters can prevent both groups of patients from knowing the treatment of the other group, then it meets the criteria for a double-blind experiment.

Therefore, the correct statement is option D: 'This is a double-blind experiment because the experimenters could prevent both groups of patients from knowing what sort of treatment the other group is receiving.' Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct statement in the context of this experiment is C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

In a single-blind experiment, the subjects are unaware of which treatment they are receiving, but the experimenters or investigators are aware. In this case, the patients will know if they are receiving medication or the blindfold and soft music, so it is not a double-blind experiment. The examining doctor might not know which treatment each patient is receiving, but this does not make it a double-blind experiment. Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct statement in the context of this experiment is C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

In a single-blind experiment, the participants are unaware of which treatment they are receiving. However, in this experiment, the patients will be aware of whether they are receiving medication or blindfolds and music. This is because the patients can feel or see the treatment being given to them.

Additionally, the experiment does not involve a control group that receives neither medication nor blindfolds and music, so it does not meet the criteria for a controlled experiment.

Therefore, statement C is the correct answer. Final answer: C

\*\*\*\*\*

Answer Option 9:

In the context of this experiment, the correct statement is C. "This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not."

Single-blind experiments are experiments in which the participants are aware of the treatment they are receiving, but the experimenters (in this case, the examining doctor) are not aware of which treatment each participant is receiving. In this experiment, the patients will know whether they are receiving medication or blindfolds and soft music, but the examining doctor may not know which treatment each patient is assigned to. Therefore, it is a single-blind experiment.

Option A is incorrect because it states that it is not possible for this experiment to be single-blind or double-blind, which is not true.

Option B is incorrect because it states that this is only a single-blind experiment because there is no control group that received neither medication nor blindfolds and soft music. However, the presence or absence of a control group does not determine whether an experiment is single-blind or double-blind.

Option D is incorrect because it states that this is a double-blind experiment. However, in a double-blind experiment, both the participants and the experimenters are unaware of which treatment each participant is receiving. In this experiment, the patients are aware of their treatment, so it cannot be considered a double-blind experiment.

Therefore, the correct statement is C. Final answer: C

\*\*\*\*\*

Question 50:

'A study was conducted to test a new style of keyboard in preventing repetitive stress disorders. Volunteers who have had problems with such injuries were randomly assigned to use either a traditional keyboard or the new design. A significance test was conducted with the alternative hypothesis that a smaller proportion of those using the new keyboard will suffer injuries than those using the traditional keyboard. The resulting P-value was 0.07. Which is a correct interpretation of this P-value?', A. 'The null hypothesis should be rejected.', B. 'The null hypothesis should be accepted.', C. 'There is a 7% chance that the null hypothesis is correct.', D. 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

=====

Answer Option 1:

The correct interpretation of a P-value in a significance test is: the probability of observing a test statistic as extreme as the one obtained, assuming that the null hypothesis is true.

In this case, the P-value is 0.07. This means that if the null hypothesis is true (i.e., if the new keyboard is really no better at preventing injuries), there is a 7% chance of obtaining a difference between the two groups at least as large as the observed difference.

Therefore, the correct interpretation of this P-value is: 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

Hence, the answer is  $\boxed{\text{D}}$

\*\*\*\*\*

Answer Option 2:

The correct interpretation of the given P-value of 0.07 is: 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

This means that the P-value represents the probability of obtaining the observed data, or data more extreme, assuming that the null hypothesis is true. It does not provide evidence for the rejection or acceptance of the null hypothesis. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The correct interpretation of the P-value is option D: 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

The P-value represents the probability of observing a test statistic as extreme as, or more extreme than, the one obtained in the study, assuming that the null hypothesis is true. In this case, the P-value is 0.07, which means that if the new keyboard is really no better at preventing injuries, there is a 7% chance of observing a difference between the two groups at least as large as the one observed in the study.

The P-value does not provide direct evidence for or against the null hypothesis, but it helps in making a decision about whether to reject or fail to reject the null hypothesis based on a predetermined significance level. In this case, the P-value is not small enough to reject the null hypothesis at a typical significance level of 0.05. Final answer: D

\*\*\*\*\*

Answer Option 4:

The correct interpretation of the P-value is D. 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

The P-value represents the probability of obtaining a test statistic as extreme as or more extreme than the one observed, assuming that the null hypothesis is true. In this case, the P-value of 0.07 means that there is a 7% chance of observing a difference in injury rates between the two groups as large as or larger than the one observed, if the new keyboard is really no better at preventing injuries.

Therefore, we do not have enough evidence to reject the null hypothesis, so option A. 'The null hypothesis should be rejected.' and option B. 'The null hypothesis should be accepted.' are incorrect interpretations.

Option C. 'There is a 7% chance that the null hypothesis is correct.' is also incorrect. The P-value does not represent the probability that the null hypothesis is correct. It only provides information on the likelihood of the observed data under the null hypothesis.

Thus, the correct interpretation is D. 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.' Final answer: D

\*\*\*\*\*

Answer Option 5:

The correct interpretation of a p-value is that it measures the strength of evidence against the null hypothesis. In this case, the alternative hypothesis states that a smaller proportion of those using the new keyboard will suffer injuries compared to those using the traditional keyboard.

A p-value of 0.07 means that if the null hypothesis (which assumes no difference between the two keyboards) is true, there is a 7% chance of getting a difference between the two groups at least as large as the observed difference.

Therefore, the correct interpretation of this p-value is: 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

Hence, the answer is  $\boxed{\text{D}}$

\*\*\*\*\*

Answer Option 6:

The correct interpretation of the P-value in this scenario is D. 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

The P-value represents the probability of obtaining results as extreme as the observed data, assuming that the null hypothesis (the new keyboard is no better at preventing injuries) is true. In this case, a P-value of 0.07 means that there is a 7% chance of observing a difference between the two groups that is at least as large as the observed difference, if the null hypothesis is true.

This does not provide enough evidence to reject the null hypothesis, as the conventional threshold for statistical significance is typically set at 0.05. Therefore, we cannot conclude that the new keyboard is truly better at preventing injuries based on the given P-value. Final answer: D

\*\*\*\*\*

Answer Option 7:

The correct interpretation of the P-value in this case is D. 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

The P-value represents the probability of obtaining a test statistic as extreme as, or more extreme than, the observed test statistic, assuming that the null hypothesis is true. In this case, if the null hypothesis is true (i.e., there is no difference between the new and traditional keyboards in preventing injuries), then there is a 7% chance of obtaining a difference between the two groups at least as large as the observed difference.

This does not provide evidence to reject the null hypothesis (choice A) or accept the null hypothesis (choice B), as the P-value is not below a specified significance level. Additionally, it is important to note that the P-value is not the probability that the null hypothesis is correct (choice C). Final answer: D

\*\*\*\*\*

Answer Option 8:

The correct interpretation of the P-value in this context is D. 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

The P-value is the probability of obtaining a test statistic as extreme as, or more extreme than, the observed test statistic, assuming that the null hypothesis is true. In this case, the null hypothesis assumes that there is no difference between the two keyboards in preventing injuries.

Therefore, a P-value of 0.07 implies that if the null hypothesis is true (i.e., if the new keyboard is really no better at preventing injuries), there is a 7% chance of observing a difference between the two groups at least as large as the one observed in the study.

This does not provide evidence to reject the null hypothesis, but it does suggest that there is some evidence against it. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The correct interpretation of the p-value in this context is option D: 'There is a 7% chance of getting a difference between the two groups at least as large as the observed difference if the new keyboard is really no better at preventing injuries.'

In hypothesis testing, the p-value is the probability of obtaining a test statistic as extreme or more extreme than the one observed, assuming that the null hypothesis is true. In this case, the p-value of 0.07 means that if the new keyboard is really no better at preventing injuries (null hypothesis is true), there is a 7% chance of obtaining a difference between the two groups that is at least as large as the observed difference.

It is important to note that the p-value does not provide information about the truth or falsity of the null hypothesis, and it does not provide the probability of the null hypothesis being correct. Therefore, options A, B, and C are incorrect interpretations of the p-value.

The correct interpretation, as stated in option D, is that the p-value represents the probability of observing a difference between the two groups at least as large as the observed difference, assuming the null hypothesis is true. Final Answer: D

\*\*\*\*\*

Question 51:

'A college admissions officer is interested in comparing the SAT math scores of high school applicants who have and have not taken AP Statistics. She randomly pulls the files of five applicants who took AP Statistics and five applicants who did not, and proceeds to run a t-test to compare the mean SAT math scores of the two groups. Which of the following is a necessary assumption?', A. 'The population variances from each group are known.', B. 'The population variances from each group are unknown.', C. 'The population variances from the two groups are equal.', D. 'The population of SAT scores from each group is normally distributed.'

=====

Answer Option 1:

The necessary assumption for running a t-test is that the population variances from the two groups are unknown but assumed to be equal. Therefore, the correct answer is  $\boxed{\text{(C)}}$  'The population variances from the two groups are equal.')

\*\*\*\*\*

Answer Option 2:

The necessary assumption for running a t-test to compare the means of two groups is that the populations from which the samples are drawn are normally distributed. Therefore, the correct answer is  $\boxed{\text{(D)}}$  'The population of SAT scores from each group is normally distributed.'

\*\*\*\*\*

Answer Option 3:

The necessary assumption for conducting a t-test is that the population variances from the two groups are unknown but assumed to be equal. Therefore, the correct answer is  $\boxed{\text{(C)}}$  'The population variances from the two groups are equal.')

\*\*\*\*\*

Answer Option 4:

The necessary assumption for running a t-test is that the population variances from the two groups are unknown, making the correct answer choice  $\boxed{\text{(B)}}$  'The population variances from each group are unknown.')

\*\*\*\*\*

Answer Option 5:

The necessary assumption for a t-test is that the population variances from each group are unknown but assumed to be equal. Therefore, the correct answer is  $\boxed{\text{(B)}}$  'The population variances from each group are unknown.')

\*\*\*\*\*

Answer Option 6:

The necessary assumption for running a t-test is that the population of SAT scores from each group is normally distributed. This assumption is needed to ensure that the sampling distribution of the means follows a t-distribution. Thus, the correct answer is  $\boxed{\text{(D)}}$ . Final answer: D

\*\*\*\*\*

Answer Option 7:

The necessary assumption for running a t-test is that the population variances from the two groups are unknown but assumed to be equal. Therefore, the correct answer is  $\boxed{\text{(C)}}$  'The population variances from the two groups are equal.')

\*\*\*\*\*

Answer Option 8:

The necessary assumption for a t-test is that the population of SAT scores from each group is normally distributed. This assumption is necessary in order to make valid inferences about the means of the two groups using the t-distribution. Therefore, the correct answer is  $\boxed{\text{(D)}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

The necessary assumption for a t-test is that the population variances from the two groups are unknown but equal. This assumption allows for the use of the pooled variance estimate in the t-test calculation. The assumption that the population variances are known (option A) is not necessary because in practice, the population variances are almost always unknown. The assumption that the population variances are equal (option C) is necessary for a pooled variance estimate, but the assumption that they are known is not necessary. Finally, the assumption that the population of SAT scores from each group is normally distributed (option D) is not necessary for the t-test to be valid. Therefore, the correct answer is  $\boxed{\text{(B)}}$  'The population variances from each group are unknown.')

\*\*\*\*\*

Question 52:

'Given that  $P(E) = 0.32$ ,  $P(F) = 0.15$ , and  $P(E \cap F) = 0.048$ , which of the following is a correct conclusion?', A. 'The events E and F are both independent and mutually exclusive.', B. 'The events E and F are neither independent nor mutually exclusive.', C. 'The events E and F are mutually exclusive but not independent.', D. 'The events E and F are independent but not mutually exclusive.'

=====

Answer Option 1:

To determine if events E and F are independent, we must check if the probability of the intersection of E and F is equal to the product of their individual probabilities. In this case,  $P(E) = 0.32$ ,  $P(F) = 0.15$ , and  $P(E \cap F) = 0.048$ .

Since  $P(E \cap F) \neq P(E)P(F)$ , we can conclude that events E and F are not independent.

To determine if events E and F are mutually exclusive, we must check if their intersection is empty, meaning  $P(E \cap F) = 0$ .

Since  $P(E \cap F) \neq 0$ , we can conclude that events E and F are not mutually exclusive.

Therefore, the correct conclusion is: 'The events E and F are neither independent nor mutually exclusive.'

The answer is  $\boxed{\text{B}}$

\*\*\*\*\*

Answer Option 2:

To determine whether events E and F are independent, we can use the formula for the probability of the intersection of two independent events, which states that  $P(E \cap F) = P(E) \times P(F)$ .

Given that  $P(E) = 0.32$ ,  $P(F) = 0.15$ , and  $P(E \cap F) = 0.048$ , we can check if the equation holds:

$$0.048 = (0.32) \times (0.15)$$

Since this equation is not true, we can conclude that events E and F are NOT independent.

To determine whether events E and F are mutually exclusive (disjoint), we can check if the intersection of E and F is empty ( $P(E \cap F) = 0$ ).

Given that  $P(E \cap F) = 0.048$ , we can see that the intersection of E and F is not empty. Therefore, we can conclude that events E and F are NOT mutually exclusive.

Based on these findings, the correct conclusion is: The events E and F are neither independent nor mutually exclusive. This matches option B. Final Answer: B

\*\*\*\*\*

Answer Option 3:

To determine if events E and F are independent, we need to check if the probability of their intersection is equal to the product of their individual probabilities. In this case, we have  $P(E) = 0.32$ ,  $P(F) = 0.15$ , and  $P(E \cap F) = 0.048$ .

If events E and F are independent, then  $P(E \cap F) = P(E) \times P(F)$ . However,  $0.048 \neq (0.32) \times (0.15)$ , so we can conclude that events E and F are not independent.

Next, to determine if events E and F are mutually exclusive, we need to check if their intersection is empty. Since  $P(E \cap F) = 0.048$ , which is greater than zero, we can conclude that events E and F are not mutually exclusive.

Thus, the correct conclusion is that 'The events E and F are neither independent nor mutually exclusive.' Therefore, the answer is option B. Final answer: B

\*\*\*\*\*

Answer Option 4:

We can determine the relationship between the events E and F by examining their probabilities.

If the events E and F were independent, then we would expect the probability of both events occurring,  $P(E \cap F)$ , to be equal to the product of their individual probabilities,  $P(E) \times P(F)$ . However, in this case,  $P(E \cap F) = 0.048$ , which is not equal to  $0.32 \times 0.15 = 0.048$ . Therefore, we can conclude that the events E and F are not independent.

If the events E and F were mutually exclusive, then they could not occur at the same time, and hence  $P(E \cap F)$  would be equal to 0. However, in this case,  $P(E \cap F) = 0.048$ , which is not equal to 0. Therefore, we can conclude that the events E and F are not mutually exclusive.

Since the events E and F are neither independent nor mutually exclusive, the correct conclusion is that 'The events E and F are neither independent nor mutually exclusive.' Therefore, the answer is

$\boxed{\text{B}}$ . 'The events E and F are neither independent nor mutually exclusive.')

\*\*\*\*\*

Answer Option 5:

To determine whether events E and F are independent, we need to check if the probability of their intersection,  $P(E \cap F)$ , is equal to the product of their individual probabilities,  $P(E)$  and  $P(F)$ . In this case,  $P(E) = 0.32$ ,  $P(F) = 0.15$ , and  $P(E \cap F) = 0.048$ .

Since the probability of the intersection,  $P(E \cap F)$ , is not equal to  $P(E) \times P(F)$ , we can conclude that events E and F are not independent.

To determine whether events E and F are mutually exclusive, we need to check if their intersection is empty (i.e., if  $P(E \cap F) = 0$ ). In this case,  $P(E \cap F) = 0.048$ , which means that events E and F are not mutually exclusive.

Therefore, the correct conclusion is: 'The events E and F are neither independent nor mutually exclusive.' This corresponds to option B. Final answer: B

\*\*\*\*\*

Answer Option 6:

To determine if events E and F are independent, we need to check if the probability of their intersection is equal to the product of their individual probabilities. In this case,  $P(E \cap F) = 0.048$ ,  $P(E) = 0.32$ , and  $P(F) = 0.15$ . Since  $0.048$  is not equal to  $(0.32)(0.15) = 0.048$ , we can conclude that events E and F are not independent.

To determine if events E and F are mutually exclusive, we need to check if their intersection is empty, i.e., if  $P(E \cap F) = 0$ . In this case,  $P(E \cap F) = 0.048$ , so events E and F are not mutually exclusive.

Based on these calculations, we can conclude that the events E and F are neither independent nor mutually exclusive, which means the correct conclusion is  $\boxed{\text{B}}$ . 'The events E and F are neither independent nor mutually exclusive.')



\*\*\*\*\*

Answer Option 7:

To determine if events E and F are independent, we need to check if the probability of their intersection is equal to the product of their individual probabilities.

From the given information, we have  $P(E) = 0.32$ ,  $P(F) = 0.15$ , and  $P(E \cap F) = 0.048$ . To check for independence, we calculate  $P(E) \times P(F) = (0.32)(0.15) = 0.048$ , which is equal to  $P(E \cap F)$ .

Therefore, we can conclude that the events E and F are independent.

However, we cannot conclude that the events E and F are mutually exclusive because  $P(E \cap F)$  is non-zero.

Thus, the correct conclusion is: The events E and F are independent but not mutually exclusive.

Therefore, the answer is  $\boxed{\text{D. The events E and F are independent but not mutually exclusive.}}$

\*\*\*\*\*

Answer Option 8:

To determine if events E and F are independent, we need to check if the probability of their intersection is equal to the product of their individual probabilities. In this case,  $P(E) = 0.32$ ,  $P(F) = 0.15$ , and  $P(E \cap F) = 0.048$ .

If E and F are independent, then  $P(E \cap F) = P(E) \cdot P(F)$ .

Here,  $0.048 \neq 0.32 \cdot 0.15$ , so events E and F are not independent.

To determine if events E and F are mutually exclusive, we need to check if their intersection is empty. In this case,  $P(E \cap F) = 0.048$ .

Since  $P(E \cap F) \neq 0$ , events E and F are not mutually exclusive.

Therefore, the correct conclusion is that "The events E and F are neither independent nor mutually exclusive." This corresponds to option B. Final answer: B.

\*\*\*\*\*

Answer Option 9:

To determine whether events E and F are independent, we need to check if the probability of their intersection is equal to the product of their individual probabilities. In this case,  $P(E \cap F) = 0.048$ ,  $P(E) = 0.32$ , and  $P(F) = 0.15$ . Since  $P(E \cap F)$  is not equal to  $P(E) \cdot P(F)$ , we can conclude that events E and F are not independent.

To determine whether events E and F are mutually exclusive, we need to check if their intersection is empty, i.e., if  $P(E \cap F) = 0$ . In this case,  $P(E \cap F) = 0.048$ , so events E and F are not mutually exclusive.

Therefore, the correct conclusion is that the events E and F are neither independent nor mutually exclusive. So, the answer is  $\boxed{\text{B.}}$

\*\*\*\*\*

Question 53:

'A medical research team tests for tumor reduction in a sample of patients using three different dosages of an experimental cancer drug. Which of the following is true?', A. 'There are three explanatory variables and one response variable.', B. 'There is one explanatory variable with three levels of response.', C. 'Tumor reduction is the only explanatory variable, but there are three response variables corresponding to the different dosages.', D. 'There are three levels of a single explanatory variable.'

=====

Answer Option 1:

The statement "There are three dosages of an experimental cancer drug" suggests that the different dosages are the explanatory variable. The response variable is the tumor reduction, as it is what the researchers are measuring. Therefore, the correct answer is D. "There are three levels of a single explanatory variable." Final Answer: D

\*\*\*\*\*

Answer Option 2:

In this scenario, the different dosages of the experimental cancer drug are the explanatory variable and the tumor reduction is the response variable. The dosages can be considered as different levels of the explanatory variable. Therefore, the correct answer is  $\boxed{\text{D. There are three levels of a single explanatory variable.}}$  Final Answer: D

\*\*\*\*\*

Answer Option 3:

The statement mentions that the medical research team tests for tumor reduction using three different dosages of the drug. This indicates that the dosage of the drug is the explanatory variable, while the tumor reduction is the response variable.

Since there is only one explanatory variable (dosage) with three levels (three different dosages), the correct statement is that "There are three levels of a single explanatory variable." Therefore, the answer is  $\boxed{\text{D.}}$

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'There are three levels of a single explanatory variable.'

In this scenario, the dosages of the experimental cancer drug are the levels of the explanatory variable. The researchers are investigating how the different dosages affect tumor reduction, which is the response variable. It is important to note that in this case, there is only one explanatory variable (dosage), and it has three levels (three different dosages). Therefore, option D accurately describes the setup of the experiment. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The statement "There are three levels of a single explanatory variable" is the correct interpretation of the scenario described. The dosage of the experimental cancer drug is the explanatory variable, and it has three levels: the three different dosages being tested. The response variable is the tumor reduction, which is being measured in the sample of patients. Therefore, the correct answer is  $\boxed{\text{D. There are three levels of a single explanatory variable.}}$  Final Answer: D

'There are three levels of a single explanatory variable.')

\*\*\*\*\*

Answer Option 6:

The given scenario involves testing different dosages of an experimental cancer drug to measure tumor reduction. The dosages of the drug are the explanatory variable, as they are manipulated and varied by the researchers. The response variable is tumor reduction, which is measured based on the different dosages.

Therefore, the correct statement is D. 'There are three levels of a single explanatory variable.' The explanatory variable is the dosage of the drug, and it has three levels corresponding to the different dosages tested. Final answer: D

\*\*\*\*\*

Answer Option 7:

The correct statement is D. 'There are three levels of a single explanatory variable.'

In this scenario, the researchers are testing the effects of different dosages of the experimental cancer drug on tumor reduction. The dosages of the drug are the explanatory variable, and they have three levels (dosage 1, dosage 2, and dosage 3). The response variable is tumor reduction, which is being measured to assess the effectiveness of the different dosages. Therefore, there is one explanatory variable (dosage) with three levels. Final answer: D

\*\*\*\*\*

Answer Option 8:

The correct answer is D. 'There are three levels of a single explanatory variable.'

In this scenario, the dosage of the experimental cancer drug is the explanatory variable, and it has three levels (dosages). The response variable is tumor reduction, which is being measured in the sample of patients. Therefore, there is one explanatory variable with three levels. Final Answer: D

\*\*\*\*\*

Answer Option 9:

Based on the given information, the medical research team is testing the tumor reduction in patients using three different dosages of the experimental cancer drug. In this case, the dosages of the drug are the explanatory variable, while the tumor reduction is the response variable. Since there is only one explanatory variable (dosage) with three levels (three different dosages) and one response variable (tumor reduction), the correct statement is: 'There are three levels of a single explanatory variable.' Therefore, the answer is  $\boxed{\text{D. There are three levels of a single explanatory variable.}}$

\*\*\*\*\*

Question 54:

'A department store at a large mall claims that over 60 percent of the mall's visitors shop at that store. Let  $p$  = the proportion of the mall's visitors who shop at the store. Which of the following pairs of hypotheses should be used to support this claim?', A. ' $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ ', B. ' $H_0: p = 0.60$  and  $H_a: p \geq 0.60$ ', C. ' $H_0: p = 0.60$  and  $H_a: p < 0.60$ ', D. ' $H_0: p > 0.60$  and  $H_a: p \geq 0.60$ '

=====

Answer Option 1:

To support the claim that over 60 percent of the mall's visitors shop at the store, we want to test if the proportion  $p$  is greater than 0.60. Therefore, the correct pair of hypotheses is  $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ . This means that the null hypothesis states that  $p$  is less than or equal to 0.60 (the claim is not supported), and the alternative hypothesis states that  $p$  is greater than 0.60 (the claim is supported).

Therefore, the correct answer is A. Final Answer: A

=====

Answer Option 2:

To support the claim that over 60 percent of the mall's visitors shop at the store, we should use the alternative hypothesis that the true proportion of visitors who shop at the store is greater than 0.60.

Therefore, the correct pair of hypotheses is:

$H_0: p \leq 0.60$  and  $H_a: p > 0.60$ .

This means that the null hypothesis assumes the proportion is less than or equal to 0.60, and the alternative hypothesis assumes the proportion is greater than 0.60. Thus, the correct answer is A. Final

Answer: A

=====

Answer Option 3:

To support the claim that over 60% of the mall's visitors shop at the store, we would want to use the alternative hypothesis that the proportion is greater than 0.60. Therefore, the correct pair of hypotheses is

$H_0: p \leq 0.60$  and  $H_a: p > 0.60$ . This corresponds to answer choice A. Final Answer: A

=====

Answer Option 4:

The claim states that over 60 percent of the mall's visitors shop at the store. This means that the proportion of visitors who shop at the store is greater than 60 percent.

Therefore, the appropriate pair of hypotheses to support this claim is:

$H_0: p \leq 0.60$  (null hypothesis: the proportion is less than or equal to 60 percent)

$H_a: p > 0.60$  (alternative hypothesis: the proportion is greater than 60 percent)

Thus, the correct choice is: A. ' $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ '. Final answer: A

=====

Answer Option 5:

To support the claim that over 60 percent of the mall's visitors shop at the store, we want to test whether the proportion,  $p$ , is greater than 0.60.

Therefore, the appropriate pair of hypotheses is:

$H_0: p \leq 0.60$  (null hypothesis)

$H_a: p > 0.60$  (alternative hypothesis)

This corresponds to option A: ' $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ '.

The null hypothesis assumes that the proportion is less than or equal to 0.60, while the alternative hypothesis suggests that the proportion is greater than 0.60.

Therefore, the answer is A. Final Answer: A

=====

Answer Option 6:

The null hypothesis ( $H_0$ ) represents the claim that over 60 percent of the mall's visitors shop at the store. Therefore, the null hypothesis should state that  $p \leq 0.60$ .

The alternative hypothesis ( $H_a$ ) represents the opposite of the null hypothesis. In this case, it should state that more than 60 percent of the mall's visitors shop at the store. Therefore, the alternative hypothesis should state that  $p > 0.60$ .

Thus, the correct pair of hypotheses is  $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ . Final Answer: A

=====

Answer Option 7:

The correct pair of hypotheses to support the claim that over 60 percent of the mall's visitors shop at the store is ' $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ '.

Here's why:

- The null hypothesis ( $H_0$ ) represents the claim or assumption that we want to test. In this case, the claim is that the proportion of mall visitors who shop at the store is less than or equal to 60 percent.

- The alternative hypothesis ( $H_a$ ) represents the opposite or alternative claim. In this case, the alternative claim is that the proportion of mall visitors who shop at the store is greater than 60 percent.

Therefore, the correct pair of hypotheses is ' $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ '. Final Answer: A.

=====

Answer Option 8:

To support the claim that over 60 percent of the mall's visitors shop at the store, we need to use the alternative hypothesis that the proportion is greater than 0.60. Therefore, the correct pair of hypotheses is

$H_0: p \leq 0.60$  and  $H_a: p > 0.60$ . So the answer is  $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ . Final answer: A

=====

Answer Option 9:

To support the claim that over 60 percent of the mall's visitors shop at the store, we want to test if the proportion  $p$  is greater than 0.60. Therefore, the appropriate pair of hypotheses is  $H_0: p \leq 0.60$  and  $H_a: p > 0.60$ . This means our null hypothesis is that the proportion is less than or equal to 0.60, and our alternative hypothesis is that the proportion is greater than 0.60. So the correct answer is A. Final Answer: A  
\*\*\*\*\*

Question 55:

'Which of the following can affect the value of the correlation  $r$ ?', A. 'A change in measurement units', B. 'A change in which variable is called  $x$  and which is called  $y$ ', C. 'Adding the same constant to all values of the  $x$ -variable', D. 'None of the above can affect the  $r$  value.'

=====

Answer Option 1:

The value of the correlation coefficient, denoted by  $r$ , measures the strength and direction of the linear relationship between two variables. It is not affected by changes in measurement units (option A). The correlation coefficient is a measure of the strength of the relationship, not the specific labeling of variables (option B). However, adding the same constant to all values of the  $x$ -variable (option C) will affect the correlation coefficient. Specifically, adding a constant to the  $x$ -variable will change the slope of the linear relationship, which in turn affects the correlation coefficient. Therefore, the answer is  $\boxed{\text{C. 'Adding the same constant to all values of the } x\text{-variable'}}}$ . Final answer: C

\*\*\*\*\*

Answer Option 2:

The correct answer is A. 'A change in measurement units'.

Correlation coefficient, denoted by  $r$ , measures the strength and direction of the linear relationship between two variables. It is a unitless value that ranges from  $-1$  to  $1$ , where  $-1$  represents a perfect negative linear relationship,  $0$  represents no linear relationship, and  $1$  represents a perfect positive linear relationship.

A change in measurement units can affect the value of the correlation coefficient. When the units of one or both variables are changed, it can alter the scale and magnitude of the variables. As a result, the correlation coefficient may change because it is influenced by the scale and magnitude of the variables.

Therefore, option A is the correct answer. Final Answer: A

\*\*\*\*\*

Answer Option 3:

The value of the correlation coefficient  $r$  is a measure of the strength and direction of the linear relationship between two variables.

A. A change in measurement units does not affect the correlation coefficient  $r$ . Correlation measures the relative change in the variables, so if the units of measurement change, the variables will simply be rescaled, but their relationship will remain the same.

B. Swapping which variable is called  $x$  and which is called  $y$  does not affect the correlation coefficient  $r$ . The correlation coefficient measures the strength and direction of the linear relationship between two variables, regardless of their labels.

C. Adding the same constant to all values of the  $x$ -variable does not affect the correlation coefficient  $r$ . This simply shifts the entire set of  $x$ -values by a constant amount, but the relationship between the variables remains the same.

Therefore, the answer is  $\boxed{\text{D. 'None of the above can affect the } r \text{ value.'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

The value of the correlation coefficient  $r$  is a measure of the linear relationship between two variables.

A. A change in measurement units does not affect the linear relationship between two variables. The correlation coefficient  $r$  is a unitless measure and is unaffected by changes in measurement units.

Therefore, A is not correct.

B. A change in which variable is called  $x$  and which is called  $y$  does not affect the linear relationship between the variables. The correlation coefficient  $r$  only measures the strength and direction of the linear relationship, not the labeling of the variables. Therefore, B is not correct.

C. Adding the same constant to all values of the  $x$ -variable will shift the data points horizontally but will not change the overall linear relationship between the variables. The correlation coefficient  $r$  is invariant to linear transformations of the variables, including adding or subtracting constants. Therefore, C is not correct.

Therefore, the correct answer is D. None of the above can affect the value of the correlation coefficient  $r$ . Final answer: D.

\*\*\*\*\*

Answer Option 5:

The value of the correlation coefficient,  $r$ , is a measure of the linear relationship between two variables.

A. A change in measurement units does not affect the linear relationship between the variables, so it does not affect the value of  $r$ . Therefore, option A is incorrect.

B. Swapping the variables  $x$  and  $y$  does not affect the linear relationship between the variables. The correlation coefficient measures the strength and direction of the linear relationship, so swapping the variables does not change the value of  $r$ . Therefore, option B is incorrect.

C. Adding a constant to all values of the  $x$ -variable does not change the shape, direction, or strength of the linear relationship between the variables. It only shifts the entire relationship along the  $x$ -axis. Since  $r$  measures the strength and direction of the linear relationship, adding a constant to all values of the  $x$ -variable does not affect the value of  $r$ . Therefore, option C is incorrect.

Based on the above analysis, the correct answer is option D: None of the above can affect the  $r$  value. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The correlation coefficient, denoted by  $r$ , measures the strength and direction of the linear relationship between two variables.

A. A change in measurement units: The value of the correlation coefficient does not depend on the choice of measurement units. It only measures the strength and direction of the linear relationship and is unaffected by the units in which the variables are measured. Therefore, a change in measurement units does not affect the value of  $r$ .

B. A change in which variable is called  $x$  and which is called  $y$ : The correlation coefficient is symmetric, meaning that swapping the roles of the two variables does not change the value of  $r$ . So, a change in which variable is called  $x$  and which is called  $y$  does not affect the value of  $r$ .

C. Adding the same constant to all values of the x-variable: Adding a constant to all values of one variable only affects the scale and location of the data but does not change the underlying linear relationship between the variables. Therefore, adding the same constant to all values of the x-variable does not affect the value of  $r$ .

Based on the above explanations, the correct answer is  $\boxed{\text{D. None of the above can affect the } r \text{ value.}}$  Final Answer: D

\*\*\*\*\*

Answer Option 7:

The value of the correlation coefficient  $r$  is a measure of the strength and direction of the linear relationship between two variables.

A. A change in measurement units does not affect the correlation coefficient  $r$  because it only changes the scale of the variables, not the relationship between them.

B. Switching the labels for the x and y variables does not affect the correlation coefficient  $r$  because it only changes the names of the variables, not their relationship.

C. Adding the same constant to all values of the x-variable does not affect the correlation coefficient  $r$  because it only shifts the values of one variable, not the relationship between the two variables.

Therefore, the correct answer is D. None of the above can affect the  $r$  value. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The correlation coefficient, denoted as  $r$ , measures the strength and direction of the linear relationship between two variables.

A. A change in measurement units: The correlation coefficient is a measure of the relationship between two variables and is unaffected by changes in units. Therefore, changing the measurement units does not affect the value of  $r$ .

B. A change in which variable is called x and which is called y: The correlation coefficient measures the relationship between two variables, regardless of their labels. So, swapping the labels of the variables (changing which variable is called x and which is called y) does not affect the value of  $r$ .

C. Adding the same constant to all values of the x-variable: Adding a constant to all values of the x-variable does not change the shape or direction of the relationship between the variables. As a result, it does not affect the value of  $r$ .

Based on these explanations, the correct answer is  $\boxed{\text{D. None of the above can affect the } r \text{ value.}}$

\*\*\*\*\*

Answer Option 9:

The correlation coefficient  $r$  is a measure of the strength and direction of the linear relationship between two variables.

A. A change in measurement units will not affect the value of the correlation coefficient. Correlation is a dimensionless quantity, so changing the units of measurement will not change the correlation between the variables. Therefore, option A is incorrect.

B. Swapping the variables and calling one variable x and the other variable y will not affect the value of the correlation coefficient. The correlation coefficient measures the relationship between the two variables, regardless of their labels or which one is considered the dependent or independent variable. Therefore, option B is incorrect.

C. Adding the same constant to all values of the x-variable will not affect the value of the correlation coefficient. Adding a constant to all values of one variable will only shift the entire scatter plot horizontally, without changing the shape or direction of the relationship between the variables. Therefore, option C is incorrect.

Since all of the given options are incorrect, the correct answer is  $\boxed{\text{D. None of the above can affect the } r \text{ value.}}$

\*\*\*\*\*

Question 56:

'In general, how does tripling the sample size change the confidence interval size?', A. 'It triples the interval size.', B. 'It divides the interval size by 3.', C. 'It multiplies the interval size by 1.732.', D. 'It divides the interval size by 1.732.'

=====

Answer Option 1:

The size of the confidence interval depends on both the sample size and the desired level of confidence. However, assuming that the level of confidence remains the same, tripling the sample size will divide the interval size by the square root of 3 (approximately 1.732). This is because the standard error, which is used to calculate the margin of error, is inversely proportional to the square root of the sample size. Therefore, the correct answer is D. 'It divides the interval size by 1.732.'

\*\*\*\*\*

Answer Option 2:

When we triple the sample size, the confidence interval size decreases. Therefore, the correct answer is: B. 'It divides the interval size by 3.'

\*\*\*\*\*

Answer Option 3:

In general, tripling the sample size reduces the confidence interval size. This is because as the sample size increases, the standard error decreases, leading to a more precise estimate and a narrower confidence interval. Therefore, the correct answer is B. 'It divides the interval size by 3.'

\*\*\*\*\*

Answer Option 4:

The confidence interval size is inversely proportional to the square root of the sample size. Therefore, tripling the sample size will divide the interval size by the square root of 3, which is approximately 1.732. Thus, the correct answer is D. 'It divides the interval size by 1.732.'

\*\*\*\*\*

Answer Option 5:

The confidence interval size is inversely proportional to the square root of the sample size. This means that increasing the sample size will decrease the confidence interval size.

To be more precise, if we triple the sample size, it means multiplying the square root of the sample size by  $\sqrt{3}$  approx 1.732\$. Therefore, tripling the sample size will divide the confidence interval size by  $\sqrt{3}$ \$.

Thus, the correct answer is  $\boxed{\text{D. It divides the interval size by 1.732.}}$ .

\*\*\*\*\*

Answer Option 6:

The confidence interval size is inversely proportional to the square root of the sample size.

Let's say the confidence interval size for a given sample size is  $\Delta x$ . When we triple the sample size, the new sample size is three times larger, and the new confidence interval size is  $\Delta x/\sqrt{3}$ .

Therefore, tripling the sample size divides the interval size by  $\sqrt{3}$ , which is approximately 1.732.

Hence, the correct answer is  $\boxed{\text{D. 'It divides the interval size by 1.732.'}}$ .

\*\*\*\*\*

Answer Option 7:

The confidence interval size is determined by the formula  $\text{interval size} = z \times \frac{\text{standard deviation}}{\sqrt{\text{sample size}}}$ , where  $z$  is the z-score corresponding to the desired level of confidence.

When we triple the sample size, the square root of the sample size in the denominator becomes  $\sqrt{3}$  times larger. As a result, the denominator becomes  $\frac{\text{standard deviation}}{\sqrt{3}}$ , which is equivalent to  $\frac{\text{standard deviation}}{\sqrt{\text{sample size}}}$  divided by  $\sqrt{3}$ . Therefore, the interval size is divided by  $\sqrt{3}$  when we triple the sample size.

In other words, tripling the sample size divides the interval size by  $\sqrt{3}$  approx 1.732\$.

Therefore, the correct answer is  $\boxed{\text{D. It divides the interval size by 1.732.}}$ .

\*\*\*\*\*

Answer Option 8:

The confidence interval size is determined by the standard error of the estimate, which is inversely proportional to the square root of the sample size.

Let's say the original sample size is  $n$ . Tripling the sample size would result in  $3n$ .

The standard error of the estimate is  $\frac{\sigma}{\sqrt{n}}$ , where  $\sigma$  is the population standard deviation.

For the new sample of size  $3n$ , the standard error of the estimate would be  $\frac{\sigma}{\sqrt{3n}}$ .

Comparing the two standard errors, we can see that  $\frac{\sigma}{\sqrt{3n}}$  is equal to  $\frac{\sigma}{\sqrt{n}} \cdot \frac{1}{\sqrt{3}}$ .

Therefore, tripling the sample size reduces the standard error of the estimate by a factor of  $\frac{1}{\sqrt{3}}$ , which means the confidence interval size is divided by  $\frac{1}{\sqrt{3}}$  or multiplied by  $\sqrt{3}$  approx 1.732\$.

Thus, tripling the sample size results in multiplying the confidence interval size by approximately  $\sqrt{3}$ .

Therefore, the correct answer is  $\boxed{\text{C. It multiplies the interval size by 1.732.}}$

\*\*\*\*\*

Answer Option 9:

The confidence interval size is determined by the standard error, which is inversely proportional to the square root of the sample size. When we triple the sample size, the square root of the sample size increases by a factor of  $\sqrt{3}$ . Since the standard error is inversely proportional to the square root of the sample size, it decreases by a factor of  $\frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$ . Therefore, the confidence interval size decreases by a factor of  $\frac{\sqrt{3}}{3} \approx 0.577$ .

In other words, tripling the sample size divides the interval size by approximately 0.577. The correct answer is  $\boxed{\text{D}}$ . It divides the interval size by 1.732. Final answer: D

\*\*\*\*\*



Question 57:

'Suppose the correlation between two variables is  $r = 0.19$ . What is the new correlation if 0.23 is added to all values of the x-variable, every value of the y-variable is doubled, and the two variables are interchanged?', A. '0.19', B. '0.42', C. '0.84', D. '-0.19'

=====

Answer Option 1:

Let's consider the effects of each transformation on the variables and correlation:

1. Adding 0.23 to all values of the x-variable: This does not change the correlation between the variables.
2. Doubling every value of the y-variable: This scales the y-variable by a factor of 2. Scaling the y-variable does not change the sign or magnitude of the correlation.
3. Interchanging the variables: This changes the sign of the correlation.

Since only the sign of the correlation changes, the new correlation will be  $r' = -0.19$ .

Therefore, the answer is  $\boxed{\text{D. } -0.19}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

Let's break down the steps of the transformation:

1. Adding 0.23 to all values of the x-variable: This does not affect the correlation coefficient. The correlation remains at  $r = 0.19$ .
2. Doubling every value of the y-variable: This scales the y-variable by a factor of 2. Since correlation is a measure of linear association, multiplying the y-variable by a constant does not change the correlation coefficient. The correlation remains at  $r = 0.19$ .
3. Interchanging the two variables: Interchanging the variables will change the sign of the correlation coefficient, but not its magnitude. So, the new correlation is -0.19.

Therefore, the new correlation after the transformation is  $\boxed{\text{D. } -0.19}$ . Final answer: D

\*\*\*\*\*

Answer Option 3:

Let  $X$  and  $Y$  be the original variables, and let  $X'$  and  $Y'$  be the transformed variables. We are given that  $\text{corr}(X, Y) = r = 0.19$ .

After adding 0.23 to all values of  $X$ , we have  $X' = X + 0.23$ .

After doubling all values of  $Y$ , we have  $Y' = 2Y$ .

Finally, interchanging the variables, we have  $\text{corr}(Y', X')$ .

Note that if  $X$  and  $Y$  are linearly related, then  $Y$  and  $X$  will also be linearly related. Therefore,  $\text{corr}(Y, X) = r$ , and the correlation coefficient will remain the same after interchanging the variables.

Thus,  $\text{corr}(Y', X') = \text{corr}(X', Y') = \text{corr}(X + 0.23, 2Y) = \text{corr}(X, Y) = r = \boxed{0.19}$ . Final answer: A

\*\*\*\*\*

Answer Option 4:

Let  $X$  and  $Y$  be the original variables, and let  $X'$  and  $Y'$  be the transformed variables. We are given that the correlation between  $X$  and  $Y$  is  $r = 0.19$ , and we want to find the correlation between  $X'$  and  $Y'$  after the transformation.

Adding 0.23 to all values of  $X$  gives us  $X' = X + 0.23$ . Doubling all values of  $Y$  gives us  $Y' = 2Y$ . Interchanging the variables gives us  $X'' = Y'$  and  $Y'' = X'$ .

To find the new correlation, we need to compute the correlation coefficient between  $X''$  and  $Y''$ . Using the formula for correlation, we have:

$$\text{corr}(X'', Y'') = \frac{\text{cov}(X'', Y'')}{\sqrt{\text{var}(X'') \cdot \text{var}(Y'')}}]$$

We know that  $\text{cov}(X'', Y'') = \text{cov}(Y', X') = \text{cov}(Y, X)$ , and since  $X$  and  $Y$  are independent, their covariance is 0. Additionally,  $\text{var}(X'') = \text{var}(Y') = 4\text{var}(Y)$ .

Thus, the correlation coefficient is:

$$\text{corr}(X'', Y'') = \frac{\text{cov}(X'', Y'')}{\sqrt{\text{var}(X'') \cdot \text{var}(Y'')}} = \frac{0}{\sqrt{4\text{var}(Y) \cdot 4\text{var}(Y)}} = \frac{0}{4\text{var}(Y)} = 0]$$

So the new correlation after the transformation is  $\boxed{0.00}$ .

Final Answer: None of the above (not listed as an option)

\*\*\*\*\*

Answer Option 5:

Adding a constant to the x-variable and doubling the y-variable does not change the correlation. However, interchanging the two variables changes the sign of the correlation. Therefore, the new correlation is

$\boxed{-0.19}$ . Final Answer: D

\*\*\*\*\*

Answer Option 6:

Let  $(x, y)$  be a pair of data points. After adding 0.23 to all values of the x-variable, every value of the y-variable is doubled, and the two variables are interchanged, the new pair of data points is  $(y', x') = (2y, x + 0.23)$ . We want to find the correlation between  $x'$  and  $y'$ .

First, let's find the means and standard deviations of  $x$  and  $y$ . Let  $\bar{x}$  and  $\bar{y}$  represent the means of  $x$  and  $y$ , respectively, and let  $s_x$  and  $s_y$  represent the standard deviations of  $x$  and  $y$ , respectively.

The correlation coefficient  $r$  is defined by

$$r = \frac{\sum{(x_i - \bar{x})(y_i - \bar{y})}}{s_x s_y}]$$

Let's consider the numerator of this expression. Using the new variables, we have

\begin{align\*}

$$\sum{(x_i - \bar{x})(y_i - \bar{y})} = \sum{(y'_i - \bar{y}')(x'_i - \bar{x}')}\backslash$$

$$= \sum{(2y_i - 2\bar{y})(x_i + 0.23 - \bar{x})}\backslash$$

$$= 2\sum{(x_i - \bar{x})(y_i - \bar{y})} + 0.46\sum{(x_i - \bar{x})} - 0.46\sum{(y_i - \bar{y})} + 0.46n,$$

\end{align\*}

where  $n$  is the number of data points.

Plugging this result back into the correlation coefficient formula, we get

\begin{align\*}

$$r = \frac{\sum{(x_i - \bar{x})(y_i - \bar{y})}}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}}$$

$$= \frac{2\sum{(x_i - \bar{x})(y_i - \bar{y})} + 0.46\sum{(x_i - \bar{x})} - 0.46\sum{(y_i - \bar{y})} + 0.46n}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}}$$

$$= 2\frac{\sum{(x_i - \bar{x})(y_i - \bar{y})}}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}} + \frac{0.46\sum{(x_i - \bar{x})}}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}} - \frac{0.46\sum{(y_i - \bar{y})}}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}} + \frac{0.46n}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}}$$

$$= 2r + \frac{0.46\sum{(x_i - \bar{x})}}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}} - \frac{0.46\sum{(y_i - \bar{y})}}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}} + \frac{0.46n}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}}$$

$$= 2r + \frac{0.46n}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}}.$$

\end{align\*}

Since  $r = 0.19$  and  $\frac{0.46n}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}} \neq 0$ , we have

$$\sqrt{r} = 2(0.19) + \frac{0.46n}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}} = 0.38 + \frac{0.46n}{\sqrt{\sum{(x_i - \bar{x})^2} \sum{(y_i - \bar{y})^2}}}.$$

Thus, the new correlation after the transformations is  $\boxed{0.38}$ . Final answer: A

\*\*\*\*\*

Answer Option 7:

Let  $(X, Y)$  represent the original variables and  $(X', Y')$  represent the transformed variables. We are given that the correlation between  $XY$  and  $Y$  is  $r = 0.19$ . We want to find the correlation between  $X'Y'$  and  $Y'$ .

First, we add 0.23 to all values of the  $XY$ -variable to obtain  $X' = X + 0.23$ .

Next, we double every value of the  $Y$ -variable to obtain  $Y' = 2Y$ .

Finally, we interchange the two variables to obtain  $(Y', X')$ .

Using the formula for calculating the correlation, the correlation between  $X'Y'$  and  $Y'$  is given by:

$$r' = \frac{\text{Cov}(X', Y')}{\sigma_{X'} \sigma_{Y'}}$$

Since we only added a constant to  $XY$ , the covariance remains the same:  $\text{Cov}(X', Y') = \text{Cov}(X, Y)$ . Doubling the  $Y$ -variable does not affect the covariance either.

The standard deviation of  $X'$  is equal to the standard deviation of  $XY$  since we only added a constant.

The standard deviation of  $Y'$  is twice the standard deviation of  $Y$  since we doubled every value.

Therefore, the correlation between  $X'Y'$  and  $Y'$  is:

$$r' = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_{2Y}} = \frac{\text{Cov}(X, Y)}{\sigma_X (2\sigma_Y)} = \frac{r}{2} = \frac{0.19}{2} = \boxed{0.095} \text{ (rounded to three decimal places).}$$

Final Answer: Approximately 0.095

\*\*\*\*\*

Answer Option 8:

Let's break down the steps one by one:

- Adding 0.23 to all values of the  $x$ -variable: This does not affect the correlation between the two variables.
- Doubling every value of the  $y$ -variable: This scales the  $y$ -variable but does not affect the correlation between the two variables.
- Interchanging the two variables: This changes the sign of the correlation. If the original correlation was positive, it becomes negative, and if the original correlation was negative, it becomes positive.

Since the original correlation was 0.19, and the final step changes the sign, the new correlation is  $\boxed{-0.19}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

Let  $X$  and  $Y$  be the original variables, and let  $X'$  and  $Y'$  be the transformed variables. We know that  $r_{XY} = 0.19$ .

After adding 0.23 to all values of  $X$ , every value of  $Y$  is doubled, and the two variables are interchanged, we have  $X' = Y$  and  $Y' = X + 0.23$ .

To find the new correlation, we need to find  $r_{X'Y'}$ . We can use the formula for correlation to calculate this:

$$r_{X'Y'} = \frac{\sum{(X'_i - \bar{X}')(Y'_i - \bar{Y}')})}{\sqrt{\sum{(X'_i - \bar{X}')^2} \sum{(Y'_i - \bar{Y}')^2}}}$$

Since we have a linear transformation, the means of the transformed variables are:

$$\bar{X}' = \bar{Y} \text{ and } \bar{Y}' = \bar{X} + 0.23$$

where  $\bar{X}$  and  $\bar{Y}$  are the means of the original variables.

The sum in the numerator is:

$$\sum{(X'_i - \bar{X}')(Y'_i - \bar{Y}')} = \sum{(Y_i - \bar{Y})(X_i + 0.23 - \bar{X} - 0.23)}$$

$$= 2\sum{(X_i - \bar{X})(Y_i - \bar{Y})} = 2 \sum{r_{XY}}$$

The sum in the denominator is:

$$\sqrt{\sum{(X'_i - \bar{X}')^2} \sum{(Y'_i - \bar{Y}')^2}} = \sqrt{\sum{(Y_i - \bar{Y})^2} \sum{(X_i + 0.23 - \bar{X} - 0.23)^2}}$$

$$\sqrt{\sum (y_i - \bar{y})^2 \cdot \sum 4(x_i - \bar{x})^2)}$$

$$\sqrt{2 \sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2)}$$

$$\sqrt{2 \sum (x_i - \bar{x})(y_i - \bar{y})^2)}$$

$$\sqrt{2 \sum (r_{xy})^2)}$$

$$\sqrt{2r_{xy}}$$

Substituting these results into the correlation formula, we get:

$$r_{xy} = \frac{2 \cdot r_{xy}}{2r_{xy}} = \frac{r_{xy}}{r_{xy}} = \frac{0.19}{0.19} = \frac{0.19}{0.19} = 1$$

Therefore, the new correlation is  $\boxed{1}$ . Final answer: None of the given choices.

\*\*\*\*\*

Question 58:

'Students in AP Statistics class were timed in seconds while solving a series of mathematical brainteasers. One student's time had a standardized score of  $z = 2.40$ . If the times are all changed to minutes, what will then be the student's standardized score?', A. ' $z = 0.04$ ', B. ' $z = 0.4$ ', C. ' $z = 1.80$ ', D. ' $z = 2.40$ '

=====

Answer Option 1:

To convert the time from seconds to minutes, we divide by 60. Since the standardized score is calculated by subtracting the mean and dividing by the standard deviation, this calculation remains the same.

Therefore, the student's standardized score will still be  $\boxed{\text{z} = 2.40}$ \$. Final answer: D

\*\*\*\*\*

Answer Option 2:

If the student's original standardized score is  $z = 2.40$ \$, it means that their original time was 2.40 standard deviations above the mean.

When we change the times to minutes, the mean will be multiplied by 60 and the standard deviation will also be multiplied by 60. This does not affect the student's position relative to the mean, so the student's standardized score will remain the same.

Therefore, the student's standardized score after changing the times to minutes will still be  $z = \boxed{2.40}$ \$. Final Answer: D

\*\*\*\*\*

Answer Option 3:

To convert the time from seconds to minutes, we divide the time by 60. Since the standardized score is a measure of how many standard deviations the time is from the mean, it is a relative measure that does not depend on the unit of measurement. Therefore, the student's standardized score will remain the same after converting the time to minutes. Therefore, the student's standardized score will still be

$\boxed{\text{z} = 2.40}$ \$. Final Answer: D

\*\*\*\*\*

Answer Option 4:

If the student's original standardized score is  $z = 2.40$ \$, it means their original time is 2.40 standard deviations above the mean. Since the times are changed to minutes, the new time will be 2.40 standard deviations above the mean in minutes as well. Therefore, the student's standardized score after changing the units will still be  $z = \boxed{2.40}$ \$. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The standardized score, or z-score, is calculated by subtracting the mean from an observation and dividing by the standard deviation.

When the times are changed to minutes, both the mean and the standard deviation will change. However, the relative position of the student's time compared to the rest of the data will remain the same, so the student's standardized score will remain the same.

Therefore, the student's standardized score will still be  $z = 2.40$ , so the answer is  $\boxed{\text{D. 'z} = 2.40\text{'}}$ \$.

\*\*\*\*\*

Answer Option 6:

To convert the time from seconds to minutes, we divide the time by 60.

If the original standardized score is  $z = 2.40$ \$, it means that the student's time is 2.40 standard deviations above the mean.

When we convert the time to minutes, the mean and standard deviation will also change. However, the student's position relative to the mean will remain the same.

Therefore, the student's standardized score after the time is converted to minutes will still be  $z = \boxed{2.40}$ \$. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The standardized score, or z-score, is a measure of how many standard deviations a data point is away from the mean. It is calculated as the difference between the data point and the mean, divided by the standard deviation.

If the original standardized score is  $z = 2.40$ , it means the data point is 2.40 standard deviations above the mean.

When the times are changed to minutes, the data point will also be divided by 60 (since there are 60 seconds in a minute).

Since dividing a value by a positive constant does not change the number of standard deviations it is away from the mean, the student's new standardized score will still be  $z = 2.40$ .

Therefore, the correct answer is  $\boxed{\text{D. 'z} = 2.40\text{'}}$ \$.

\*\*\*\*\*

Answer Option 8:

The standardized score, also known as the z-score, is a measure of how many standard deviations a data point is away from the mean of a distribution. The formula for calculating the z-score is given by:

$$z = \frac{x - \mu}{\sigma}$$

Where  $x$  is the data point,  $\mu$  is the mean of the distribution, and  $\sigma$  is the standard deviation of the distribution.

In this case, the student's time has a standardized score of  $z = 2.40$ . If we convert the time from seconds to minutes, we need to adjust both the data point and the standard deviation. Since time is being converted from seconds to minutes, the mean and the standard deviation both need to be divided by 60.

Therefore, the new standardized score after converting to minutes can be calculated as:

$$z_{\text{new}} = \frac{\left(\frac{x}{60}\right) - \left(\frac{\mu}{60}\right)}{\left(\frac{\sigma}{60}\right)} = \frac{x - \mu}{\sigma} = z = 2.40$$

Thus, the student's standardized score remains the same at  $z = 2.40$ . Therefore, the correct option is  $\boxed{\text{D. 'z} = 2.40\text{'}}$ \$. Final answer: D

\*\*\*\*\*

Answer Option 9:

A standardized score, or z-score, is a measure of how many standard deviations a data point is from the mean. We can calculate the z-score using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the data point,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case, the student's original time has a standardized score of  $z = 2.40$ . Let's call the original time  $x$ . We can write the equation as:

$$2.40 = \frac{(x - \mu)}{\sigma}$$

Now, let's convert the time from seconds to minutes. Since there are 60 seconds in a minute, the new time  $y$  (in minutes) is related to the original time  $x$  (in seconds) by the equation:

$$y = x / 60.$$

Substituting this into the equation above, we have:

$$2.40 = \frac{(y / 60 - \mu)}{\sigma}$$

We want to find the standardized score of the new time, which we'll call  $z'$ . Using the same formula, we have:

$$z' = \frac{(y - \mu')}{\sigma'}$$

where  $y$  is the new time,  $\mu'$  is the new mean, and  $\sigma'$  is the new standard deviation.

Comparing the equations, we can see that:

$$\mu' = \mu / 60 \text{ (since we divided by 60 to convert from seconds to minutes),}$$

and

$$\sigma' = \sigma / 60 \text{ (since we divided by 60 to convert from seconds to minutes).}$$

Substituting these values into the equation for  $z'$ , we get:

$$z' = \frac{(y - \mu / 60)}{(\sigma / 60)}$$

Simplifying, we have:

$$z' = \frac{(y - \mu)}{\sigma}$$

This is the same formula as the original z-score formula, so the standardized score of the new time is the same as the original standardized score,  $z = 2.40$ .

Therefore, the student's standardized score after converting the time to minutes is  $z = 2.40$ . The correct answer is: D.  $z = 2.40$ .

\*\*\*\*\*

Question 59:

'An experiment was designed to test the effects of three different types of paint on the durability of wooden toys. Because boys and girls tend to play differently with toys, a randomly selected group of children was divided into two groups by sex. Which of the following statements about this experiment is true?', A. 'There are three types of paint and two sex groups, giving a total of six treatment combinations in this experiment.', B. 'Type of paint is a blocking factor.', C. 'Sex is a blocking factor.', D. 'This is a completely randomized design.'

=====

Answer Option 1:

In this experiment, the independent variable is the type of paint, and the dependent variable is the durability of the wooden toys. The researcher divided the children into two groups by sex, which means that sex is a factor in the experiment, not a blocking factor. Blocking is a method used to control for the effects of a nuisance variable, but in this case, sex is not considered a nuisance variable.

Since the researcher is specifically testing the effects of three different types of paint on the durability of wooden toys, the statement in option A is correct. There are indeed three types of paint and two sex groups, resulting in a total of six treatment combinations.

Therefore, the correct statement about this experiment is: A) 'There are three types of paint and two sex groups, giving a total of six treatment combinations in this experiment.'

Final Answer: A

\*\*\*\*\*

Answer Option 2:

The statement that is true is: C. 'Sex is a blocking factor.'

In this experiment, the factor being tested is the type of paint, while the sex of the children is a blocking factor. A blocking factor is a factor that is controlled for in the experimental design to reduce the variability introduced by that factor. In this case, by dividing the children into two groups based on sex, we are controlling for any potential differences in toy durability between boys and girls.

Option A is not true because there are only three types of paint being tested, not six.

Option B is not true because the type of paint is the main factor being tested, not a blocking factor.

Option D is not true because the experiment has a specific design with blocking factors (sex), so it is not completely randomized.

Therefore, the correct answer is C. 'Sex is a blocking factor.'

\*\*\*\*\*

Answer Option 3:

The correct answer is C. 'Sex is a blocking factor.' In this experiment, the factor being tested is the type of paint. The researcher is interested in comparing the effects of different types of paint on the durability of wooden toys. To control for potential differences between boys and girls, the children were divided into two groups by sex. This means that the factor of sex is being used as a blocking factor to ensure that any differences observed in the durability of the toys can be attributed to the type of paint rather than the sex of the children. Final answer: C

\*\*\*\*\*

Answer Option 4:

The correct answer is C. 'Sex is a blocking factor.'

In this experiment, the factor being tested is the type of paint, while the factor being controlled for is the sex of the children. By dividing the children into two groups based on sex, we are effectively blocking the effect of sex on the durability of the wooden toys. This allows us to analyze the effect of paint type while taking into account any potential differences between boys and girls in how they play with toys.

Option A is incorrect because there are only three types of paint, not six treatment combinations.

Option B is incorrect because type of paint is the factor being tested, not a blocking factor.

Option D is incorrect because the experiment involves blocking the factor of sex, making it not a completely randomized design.

\*\*\*\*\*

Answer Option 5:

The correct answer is B. 'Type of paint is a blocking factor.'

In this experiment, the researcher is interested in studying the effects of different types of paint on the durability of wooden toys. However, since boys and girls tend to play differently with toys, the researcher wants to control for the potential influence of sex on the results. Therefore, the researcher divides the randomly selected group of children into two groups by sex.

By dividing the children into two sex groups, the researcher is effectively blocking the potential influence of sex on the results. This means that within each sex group, the three types of paint are randomly assigned to the wooden toys. This ensures that any observed differences in durability between the three types of paint can be attributed to the paint itself, rather than any potential differences between boys and girls.

Therefore, the statement 'Type of paint is a blocking factor' is true. Final Answer: B

\*\*\*\*\*

Answer Option 6:

In this experiment, the independent variable is the type of paint (with three levels: the three different types of paint), and the grouping variable is the sex of the children (with two levels: boys and girls).

Option A is incorrect because it states that there are six treatment combinations, which is not accurate. The treatment combinations are the different combinations of the type of paint and the sex of the children. Since there are three types of paint and two sex groups, there are a total of six treatment combinations.

Option B is incorrect because a blocking factor is a variable that is controlled and used to create homogeneous groups. In this experiment, the sex of the children is not a blocking factor because it is not used to create homogeneous groups for comparison. Instead, it is a grouping variable.

Option C is incorrect because the type of paint is not a blocking factor in this experiment. A blocking factor is a variable that is controlled and used to create homogeneous groups for comparison. In this experiment, the type of paint is the independent variable being tested, not a variable used for blocking.

Option D is correct. This is a completely randomized design because the children were randomly assigned to the different treatment groups (corresponding to the different types of paint) without any additional factors being used to create homogeneous groups. Final Answer: D

\*\*\*\*\*

Answer Option 7:

In this experiment, the independent variable is the type of paint, and the dependent variable is the durability of wooden toys. The experiment is designed to test the effects of the different types of paint on the toys.

A. The statement "There are three types of paint and two sex groups, giving a total of six treatment combinations in this experiment" is incorrect. The total number of treatment combinations in this experiment is three (three types of paint) because the sex of the children is not considered when assigning the paint types.

B. The statement "Type of paint is a blocking factor" is incorrect. A blocking factor is a variable that is controlled for in an experiment to reduce variability and potential confounding effects. In this experiment, the type of paint is the independent variable, not a blocking factor.

C. The statement "Sex is a blocking factor" is also incorrect. Similar to the previous statement, the sex of the children is not used as a blocking factor in this experiment.

D. The statement "This is a completely randomized design" is true. In a completely randomized design, the subjects (in this case, the children) are randomly assigned to treatment groups (in this case, the different types of paint). The random assignment helps to eliminate bias and ensures that the groups are comparable.

Therefore, the correct statement is that "This is a completely randomized design." The answer is  $\boxed{\text{D}}$

\*\*\*\*\*

Answer Option 8:

Based on the information given, the experiment is designed to test the effects of three different types of paint on the durability of wooden toys, with the children divided into two groups by sex.

A. 'There are three types of paint and two sex groups, giving a total of six treatment combinations in this experiment.' This statement is incorrect because the combination of paint and sex does not create new treatment combinations. The experiment only has three types of paint, regardless of the sex of the children. Therefore, there are not six treatment combinations.

B. 'Type of paint is a blocking factor.' This statement is incorrect because a blocking factor is a factor intentionally used to reduce the variability in the experiment. In this case, the experiment is not trying to reduce variability based on the type of paint.

C. 'Sex is a blocking factor.' This statement is incorrect because, based on the information given, the sex of the children is not used as a blocking factor. The purpose of dividing the children into two groups by sex is to account for potential differences in how boys and girls play with toys.

D. 'This is a completely randomized design.' This statement is correct. In a completely randomized design, the treatments (in this case, types of paint) are randomly assigned to the experimental units (the children), without any specific blocking or stratification. From the information given, it is clear that the assignment of paint types to children has been done randomly, making this a completely randomized design.

Therefore, the correct statement is: 'This is a completely randomized design.' The answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 9:

The statement "There are three types of paint and two sex groups, giving a total of six treatment combinations in this experiment" is not true because the number of treatment combinations is determined by the multiplication of the number of levels of each factor. In this case, there are three levels of paint and two levels of sex, so the total number of treatment combinations is  $3 \times 2 = 6$ . Therefore, statement A is not true.

The statement "Type of paint is a blocking factor" is not true because a blocking factor is a factor that is intentionally used to create homogeneous groups within each block. In this experiment, there are no blocks defined based on any factor, so statement B is not true.

The statement "Sex is a blocking factor" is also not true because, like the previous statement, a blocking factor is used to create homogeneous groups within each block. In this experiment, the sex factor is used to divide the children into two groups, but it does not create any homogeneous groups within each group. Therefore, statement C is not true.

The correct statement is "This is a completely randomized design." In a completely randomized design, the treatments (types of paint in this case) are randomly assigned to the experimental units (wooden toys) without any specific blocking or control of factors. The assignment of children into two sex groups does not affect the random assignment of paint types to the toys. Therefore, statement D is true.

The correct answer is  $\boxed{\text{D. This is a completely randomized design.}}$

\*\*\*\*\*

Question 60:

'Suppose there is a correlation of  $r = 0.9$  between number of hours per day students study and GPAs. Which of the following is a reasonable conclusion?', A. '90% of students who study receive high grades.', B. '90% of students who receive high grades study a lot.', C. '90% of the variation in GPAs can be explained by variation in number of study hours per day.', D. '81% of the variation in GPAs can be explained by variation in number of study hours per day.'

=====

Answer Option 1:

The correlation coefficient,  $r$ , measures the strength and direction of the linear relationship between two variables. A correlation of 0.9 indicates a strong positive linear relationship between the number of hours studied and GPAs. However, it does not provide information about the direction of causation nor does it imply a causal relationship.

Option A, '90% of students who study receive high grades,' is not a reasonable conclusion because the correlation coefficient does not provide information about the percentage of students who receive high grades.

Option B, '90% of students who receive high grades study a lot,' is not a reasonable conclusion either because the correlation coefficient does not provide information about the characteristics of students who receive high grades.

Option C, '90% of the variation in GPAs can be explained by variation in the number of study hours per day,' is not a reasonable conclusion. A correlation coefficient of 0.9 implies a strong linear relationship, but it does not provide information about the proportion of variation explained.

Option D, '81% of the variation in GPAs can be explained by variation in the number of study hours per day,' is not a reasonable conclusion either because it is not supported by the given correlation coefficient.

Therefore, none of the given options are reasonable conclusions based solely on the correlation coefficient of 0.9.

\*\*\*\*\*

Answer Option 2:

The statement " $r = 0.9$ " refers to the correlation coefficient between the number of hours per day students study and their GPAs. The correlation coefficient measures the strength and direction of the linear relationship between two variables. A correlation coefficient of 0.9 indicates a strong positive linear relationship between the number of study hours and GPAs.

Option A, "90% of students who study receive high grades," is not a reasonable conclusion. Correlation does not imply causation, so we cannot make definitive statements about the relationship between studying and receiving high grades based on the correlation coefficient alone.

Option B, "90% of students who receive high grades study a lot," is also not a reasonable conclusion. The correlation coefficient does not provide information about the proportion of students who study a lot and receive high grades.

Option C, "90% of the variation in GPAs can be explained by variation in the number of study hours per day," is not accurate. The correlation coefficient of 0.9 indicates that 81% of the variation in GPAs can be explained by the variation in the number of study hours per day.

Therefore, the correct answer is option D, "81% of the variation in GPAs can be explained by variation in the number of study hours per day." This conclusion is supported by the given correlation coefficient.

Final answer: D

\*\*\*\*\*

Answer Option 3:

The correlation coefficient is a measure of the linear relationship between two variables. A correlation of  $r = 0.9$  suggests a strong positive linear relationship between the number of hours students study and their GPAs. However, it does not imply any of the given conclusions.

Option A, which states that 90% of students who study receive high grades, cannot be concluded from the correlation coefficient alone. The correlation coefficient measures the strength and direction of the relationship, but it does not provide information about the proportions or percentages of students falling into particular categories.

Option B, which states that 90% of students who receive high grades study a lot, cannot be concluded either. The correlation coefficient does not provide information about causality or the direction of the relationship. It only measures the strength and direction of the linear relationship, so it is not appropriate to make conclusions about causation based solely on the correlation coefficient.

Option C, which states that 90% of the variation in GPAs can be explained by the variation in the number of study hours per day, is also not correct. The correlation coefficient does not provide information about the proportion of variation explained by the relationship. It only measures the strength and direction of the linear relationship between the two variables.

Option D is the most accurate conclusion. The square of the correlation coefficient ( $r^2$ ) represents the proportion of variation in one variable that can be explained by the other variable. In this case,  $r^2 = 0.9^2 = 0.81$ , so 81% of the variation in GPAs can be explained by the variation in the number of study hours per day. Therefore, the correct conclusion is (D. '81% of the variation in GPAs can be explained by variation in number of study hours per day.')

\*\*\*\*\*

Answer Option 4:

The correct conclusion is option C: '90% of the variation in GPAs can be explained by variation in number of study hours per day.'

A correlation coefficient of 0.9 indicates a strong positive linear relationship between the number of study hours and GPAs. However, it does not imply a causal relationship or provide information about the percentage of students who study or receive high grades.

Option A is incorrect because the correlation does not provide information about the percentage of students who study or receive high grades.

Option B is incorrect because the correlation does not provide information about the percentage of students who receive high grades based on their study hours.

Option D is incorrect because it incorrectly states the percentage of variation in GPAs that can be explained by variation in study hours. The correct percentage is 90%, not 81%.

Therefore, the reasonable conclusion is option C. Final Answer: C

\*\*\*\*\*

Answer Option 5:



The correlation of  $r = 0.9$  implies a strong positive relationship between the number of hours students study and their GPAs. However, we cannot make direct conclusions about the percentage of students who study and receive high grades from this information alone. Therefore, options A and B are not reasonable conclusions.

Regarding option C, a correlation of 0.9 does not mean that 90% of the variation in GPAs can be explained by the variation in the number of study hours per day. It only indicates a strong linear relationship between the two variables.

Option D is the most reasonable conclusion. A correlation of 0.9 means that 81% of the variation in GPAs can be explained by the variation in the number of study hours per day. This is because the square of the correlation coefficient ( $r^2$ ) represents the proportion of the variation in the dependent variable (GPAs) that is explained by the independent variable (study hours per day).

Therefore, the correct answer is  $\boxed{\text{D. '81\% of the variation in GPAs can be explained by variation in number of study hours per day.'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 6:

The correlation coefficient measures the strength and direction of the linear relationship between two variables. In this case, a correlation coefficient of 0.9 indicates a strong positive linear relationship between the number of hours per day students study and their GPAs.

However, we cannot conclude that 90% of students who study receive high grades (option A) because the correlation coefficient does not provide information about the proportion of students who receive high grades.

We also cannot conclude that 90% of students who receive high grades study a lot (option B) because the correlation coefficient does not provide information about the proportion of students who receive high grades.

Option C states that 90% of the variation in GPAs can be explained by the variation in the number of study hours per day. This is not a reasonable conclusion based solely on the correlation coefficient. The correlation coefficient does not provide information about the amount of variation that can be explained.

Option D states that 81% of the variation in GPAs can be explained by the variation in the number of study hours per day. This is also not a reasonable conclusion based solely on the correlation coefficient. The correlation coefficient does not provide information about the amount of variation that can be explained.

Therefore, none of the given options are reasonable conclusions based on the correlation coefficient of 0.9.

\*\*\*\*\*

Answer Option 7:

The correlation coefficient,  $r$ , measures the strength and direction of the linear relationship between two variables. A correlation of  $r = 0.9$  indicates a strong positive linear relationship between the number of hours per day students study and their GPAs. However, this does not imply a cause-and-effect relationship or make any conclusions about the percentage of students or variation.

Option A, "90% of students who study receive high grades," is not a reasonable conclusion because the correlation does not provide information about the percentage of students who receive high grades.

Option B, "90% of students who receive high grades study a lot," is not a reasonable conclusion because the correlation does not provide information about the percentage of students who receive high grades.

Option C, "90% of the variation in GPAs can be explained by variation in the number of study hours per day," is not a reasonable conclusion based solely on the correlation coefficient. The correlation coefficient measures the strength of the linear relationship, but it does not provide information about the percentage of variation explained.

Option D, "81% of the variation in GPAs can be explained by variation in the number of study hours per day," is also not a reasonable conclusion based solely on the correlation coefficient. The correlation coefficient of 0.9 implies that 81% of the variation in GPAs can be explained by variation in the number of study hours per day.

Thus, the correct answer is  $\boxed{\text{No reasonable conclusion can be drawn based solely on the correlation coefficient.}}$

\*\*\*\*\*

Answer Option 8:

The statement " $r = 0.9$  between number of hours per day students study and GPAs" means that there is a strong positive linear relationship between the number of hours students study and their GPAs.

However, correlation does not imply causation.

Option A, "90% of students who study receive high grades," is not a reasonable conclusion. The correlation coefficient measures the strength and direction of the linear relationship between two variables, but it does not determine the percentage of students who receive high grades.

Option B, "90% of students who receive high grades study a lot," is also not a reasonable conclusion. While there may be a positive correlation between study hours and GPAs, it does not imply that all students who receive high grades necessarily study a lot.

Option C, "90% of the variation in GPAs can be explained by variation in the number of study hours per day," is not correct. The correlation coefficient does not indicate the proportion of variation in one variable that can be explained by the other variable.

Option D, "81% of the variation in GPAs can be explained by variation in the number of study hours per day," is the most reasonable conclusion. The correlation coefficient squared ( $r^2$ ) represents the proportion of variation in one variable that can be explained by the other variable. In this case,  $r^2 = 0.9^2 = 0.81$ , so 81% of the variation in GPAs can be explained by the variation in the number of study hours per day. Therefore, the correct answer is D.

\*\*\*\*\*

Answer Option 9:

The correlation coefficient measures the strength and direction of the linear relationship between two variables. In this case, a correlation coefficient of  $r = 0.9$  indicates a strong positive linear relationship between the number of hours students study and their GPAs. However, it does not imply any causation or provide information about the percentages mentioned in the answer choices.

Option A, '90% of students who study receive high grades', cannot be concluded solely based on the correlation coefficient. It is possible that some students who study may not receive high grades and vice versa.

Option B, '90% of students who receive high grades study a lot', also cannot be concluded solely based on the correlation coefficient. It is possible that some students who receive high grades may not study a lot and vice versa.

Option C, '90% of the variation in GPAs can be explained by variation in the number of study hours per day', is not accurate. The correlation coefficient of 0.9 does not provide information about the percentage of variation explained.

Option D, '81% of the variation in GPAs can be explained by variation in the number of study hours per day', is also not accurate. The correlation coefficient of 0.9 does not provide information about the percentage of variation explained.

Therefore, none of the given options can be reasonably concluded based solely on the correlation coefficient of  $r = 0.9$ .

\*\*\*\*\*

Question 61:

Estrella wants to become a paramedic and takes a screening exam. Scores on the exam have been approximately normally distributed over the years it has been given. The exam is normed with a mean of 80 and a standard deviation of 9. Only those who score in the top 15% on the test are invited back for further evaluation. Estrella received a 90 on the test. What was her percentile rank on the test, and did she qualify for further evaluation?'. A. '13.35; she didn't qualify.'. B. '54.38; she didn't qualify.'. C. '86.65; she qualified.'. D. '84.38; she didn't qualify.'

=====

Answer Option 1:

To find Estrella's percentile rank, we need to find the percentage of scores that are equal to or lower than her score. To do this, we calculate the z-score corresponding to a score of 90 using the formula:

$$Z = \frac{x - \mu}{\sigma} = \frac{90 - 80}{9} = \frac{10}{9} \approx 1.11$$

We can then use a standard normal table or calculator to find the area to the left of the z-score 1.11.

Looking up this value in the table, we find that the area to the left of 1.11 is approximately 0.8665.

This area represents Estrella's percentile rank, or roughly 86.65%.

Since she scored in the top 15%, or  $100\% - 15\% = 85\%$ , Estrella did qualify for further evaluation.

Therefore, the answer is  $\boxed{\text{C. '86.65; she qualified.'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

To find Estrella's percentile rank on the test, we need to find the percentage of scores that are less than or equal to her score. First, we need to calculate the z-score of Estrella's score.

The z-score is calculated using the formula:

$$Z = \frac{x - \mu}{\sigma}$$

where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For Estrella,  $x = 90$ ,  $\mu = 80$ , and  $\sigma = 9$ . Substituting these values into the formula, we get:

$$Z = \frac{90 - 80}{9} = \frac{10}{9}$$

Using a standard normal distribution table or a calculator, we can find that the area to the left of a z-score of  $\frac{10}{9}$  is approximately 0.8665.

To find Estrella's percentile rank, we can multiply this value by 100:

$$0.8665 \times 100 \approx 86.65$$

Estrella's percentile rank on the test is approximately 86.65.

Since she scored higher than the top 15% of test-takers, Estrella qualifies for further evaluation.

Therefore, the answer is: Estrella's percentile rank is approximately 86.65, and she qualified for further evaluation. The correct answer is  $\boxed{\text{C. '86.65; she qualified.'}}$ .

\*\*\*\*\*

Answer Option 3:

To find Estrella's percentile rank, we need to find the percentage of scores that are below her score of 90. First, we calculate the z-score using the formula:

$$Z = \frac{x - \mu}{\sigma} = \frac{90 - 80}{9} = \frac{10}{9}$$

Using a standard normal distribution table, we find that the area to the left of a z-score of  $\frac{10}{9}$  is approximately 0.8159. Since we want to find the percentage below her score, we subtract this value from 1 to get the percentage above her score:  $1 - 0.8159 = 0.1841$ .

However, we are looking for the top 15% of scores, so we need to find the percentage below her score. Subtracting 0.1841 from 1 gives us the percentage below her score:  $1 - 0.1841 = 0.8159$ .

Estrella's percentile rank is 81.59%. Since her percentile rank is below the top 15%, she did not qualify for further evaluation. Therefore, the answer is  $\boxed{\text{D. '84.38; she didn't qualify.'}}$ .

Final Answer: D

\*\*\*\*\*

Answer Option 4:

To find Estrella's percentile rank on the test, we need to find the percentage of test-takers who scored lower than her. We can do this by standardizing her score using the z-score formula:

$$Z = \frac{x - \mu}{\sigma}$$

where  $x$  is the individual score,  $\mu$  is the mean of the distribution, and  $\sigma$  is the standard deviation of the distribution.

In this case, Estrella's score is 90, the mean is 80, and the standard deviation is 9. Plugging these values into the formula, we get:

$$Z = \frac{90 - 80}{9} = \frac{10}{9} \approx 1.11$$

To find the percentage of test-takers who scored lower than Estrella, we can look up the z-score in a standard normal distribution table or use a calculator. The z-score of 1.11 corresponds to a cumulative probability of approximately 0.8665, or 86.65%.

This means that Estrella scored higher than approximately 86.65% of test-takers. Since only the top 15% are invited for further evaluation, Estrella did qualify for further evaluation.

Therefore, Estrella's percentile rank on the test is approximately 86.65%, and she qualified for further evaluation. The correct answer is  $\boxed{\text{C. '86.65; she qualified.'}}$

\*\*\*\*\*

Answer Option 5:

To determine Estrella's percentile rank, we need to find the percentage of test-takers who scored lower than her. First, we convert her score to a z-score by using the formula  $Z = \frac{x - \mu}{\sigma}$ ,

where  $x$  is her score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we get  $Z = \frac{90 - 80}{9} = \frac{10}{9} \approx 1.11$ .

Using a standard normal distribution table or calculator, we can find that the area under the curve to the left of  $Z = 1.11$  is approximately 0.8665. This means that approximately 86.65% of test-takers scored lower than Estrella.

Since only the top 15% of test-takers are invited back for further evaluation, Estrella did qualify for further evaluation because her percentile rank is greater than 15%. Therefore, the correct answer is  $\boxed{\text{C. '86.65; she qualified.'}}$ .

\*\*\*\*\*

Answer Option 6:

To find Estrella's percentile rank on the test, we need to determine the percentage of test takers who scored lower than her. We can do this by calculating the Z-score of her score and then using a standard normal distribution table.

The Z-score is calculated by subtracting the mean from Estrella's score and dividing by the standard deviation:

$$Z = \frac{90 - 80}{9} = \frac{10}{9} \approx 1.11$$

Using a standard normal distribution table, we find that the area to the left of a Z-score of  $\frac{10}{9}$  is approximately 0.8625. This means that approximately 86.25% of test takers scored lower than Estrella.

Since Estrella's score is in the top 15%, she did qualify for further evaluation. Therefore, her percentile rank on the test is approximately  $100 - 86.25 = 13.75\%$ . Rounded to two decimal places, this is 13.35%.

Therefore, the correct answer is  $\boxed{\text{(A) 13.35; she didn't qualify.}}$ . Final answer: A

\*\*\*\*\*

Answer Option 7:

To find Estrella's percentile rank, we need to determine the percentage of people who scored lower than her on the exam. We can do this by calculating the z-score for her score of 90. The z-score is calculated using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the raw score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Plugging in the values, we get:

$$z = \frac{90 - 80}{9} = \frac{10}{9} \approx 1.11$$

Using a standard normal distribution table, we can find that the area to the left of  $z = 1.11$  is approximately 0.8665. This means that approximately 86.65% of the test-takers scored lower than Estrella.

Since only the top 15% are invited back for further evaluation, Estrella did indeed qualify for further evaluation. Therefore, the correct answer is  $\boxed{\text{C. '86.65; she qualified.'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 8:

To find Estrella's percentile rank, we need to determine the percentage of test-takers who scored lower than her.

We can first convert her score into a z-score using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $\mu$  is the mean and  $\sigma$  is the standard deviation. In this case, Estrella's z-score is:

$$z = \frac{90 - 80}{9} = \frac{10}{9} \approx 1.11$$

Using a standard normal distribution table or calculator, we can find that the area to the left of  $z = \frac{10}{9}$  is approximately 0.8625.

To find the percentile rank, we subtract this value from 1, and then multiply by 100:

$$\text{Percentile Rank} = (1 - 0.8625) \times 100 \approx 13.75$$

Therefore, Estrella's percentile rank on the test is approximately 13.75.

Since only the top 15% of test-takers are invited for further evaluation, Estrella did not qualify because her percentile rank is below 15%.

So the correct answer is  $\boxed{\text{(A. ) 13.35; she didn't qualify.}}$

\*\*\*\*\*

Answer Option 9:

To find Estrella's percentile rank, we need to determine the percentage of people who scored lower than her. We can use the z-score formula for this calculation.

The z-score is given by the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the individual score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For Estrella, with a score of 90, the z-score is:

$$z = \frac{90 - 80}{9} = \frac{10}{9} \approx 1.11$$

Using a standard normal distribution table or calculator, we find that the area to the left of  $z = \frac{10}{9}$  is approximately 0.8159.

This means that approximately 81.59% of the test-takers scored lower than Estrella.

To find Estrella's percentile rank, we subtract this percentage from 100%:

$$\text{Percentile rank} = 100\% - 81.59\% = 18.41\%$$

Therefore, Estrella's percentile rank on the test is approximately 18.41%.

Since only the top 15% are invited back for further evaluation, Estrella did not qualify for further evaluation.

Therefore, the correct answer is  $\boxed{\text{(B. '54.38; she didn't qualify.'}}$ .

\*\*\*\*\*

Question 62:

'To test whether extensive exercise lowers the resting heart rate, a study is performed by randomly selecting half of a group of volunteers to exercise 1 hour each morning, while the rest are instructed to perform no exercise. Is this study an experiment or an observational study?', A. 'An experiment with a control group and blinding', B. 'An experiment with blocking', C. 'An observational study with comparison and randomization', D. 'None of the above'

=====

Answer Option 1:

The study described is an experiment. In an experiment, the researcher manipulates one or more variables (in this case, exercise) and observes the effect on another variable (resting heart rate). The fact that the volunteers are randomly assigned to either the exercise group or the no exercise group indicates randomization, which is a characteristic of an experiment. Additionally, the use of a control group (the group that does not exercise) allows for comparison and helps establish a cause-and-effect relationship. Therefore, the study is an experiment with a control group and randomization, making the correct answer choice A. 'An experiment with a control group and blinding'. Final Answer: A

\*\*\*\*\*

Answer Option 2:

The study described is an experiment. It involves randomly selecting half of the group to exercise, while the other half serves as the control group. This random assignment of participants to different groups allows for a comparison between those who exercise and those who do not. Additionally, the fact that the participants are instructed to perform or not perform exercise shows that the researchers are actively manipulating the independent variable (exercise) to observe its effect on the dependent variable (resting heart rate). Therefore, the study is an experiment and the correct answer is A. 'An experiment with a control group and blinding'. Final Answer: A

\*\*\*\*\*

Answer Option 3:

The study described is an experiment. It involves randomly selecting half of a group of volunteers to exercise and instructing the other half to perform no exercise. By randomly assigning participants to either the exercise or no exercise group, the study incorporates randomization, which is a characteristic of experimental design. Additionally, the presence of a control group (the group instructed to perform no exercise) further supports the notion that this is an experiment. Therefore, the correct answer is  $\boxed{\text{A. An experiment with a control group and blinding}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 4:

The study described is an experiment. In an experiment, the researcher manipulates one or more variables (in this case, exercise) and observes the effect on another variable (resting heart rate). In this study, the researcher randomly assigned the participants into two groups: one that exercises and one that does not. This random assignment helps ensure that any differences in resting heart rate between the two groups can be attributed to the exercise and not to other factors. Additionally, the study involves a control group (the group that does not exercise) which allows for comparison and helps establish a baseline for determining the effect of exercise. Therefore, the study is an experiment with a control group and blinding. The correct answer is A. Final Answer: A

\*\*\*\*\*

Answer Option 5:

The study described is an experiment. It involves randomly selecting half of the group of volunteers to exercise, while the other half is instructed to perform no exercise. This random assignment of participants to the exercise group and the non-exercise group is a key characteristic of an experiment. Additionally, the study has a control group (the non-exercise group) which allows for a comparison between the effects of exercise and no exercise on the resting heart rate. Thus, the correct answer is A. 'An experiment with a control group and blinding'. Final answer: A

\*\*\*\*\*

Answer Option 6:

The study described is an experiment because the researchers randomly select half of the group to exercise and the other half to not exercise. This random assignment allows for a comparison between the two groups and allows the researchers to make causal inferences about the effect of exercise on resting heart rate. Additionally, the fact that the volunteers are instructed to either exercise or not exercise suggests that there is an element of control and comparison in the study. Therefore, the correct answer is  $\boxed{\text{A. 'An experiment with a control group and blinding'}}$ . Final answer: A

\*\*\*\*\*

Answer Option 7:

The study described is an experiment. It involves randomly assigning individuals to either an exercise group or a control group. This random assignment ensures that any differences observed in the resting heart rate between the two groups can be attributed to the exercise intervention. Additionally, the study includes a control group, which allows for comparison between the groups. The fact that the control group does not receive any exercise serves as a baseline for comparison. However, the study does not mention blinding or blocking, so the correct answer is  $\boxed{\text{A. 'An experiment with a control group and blinding'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 8:

The study described is an experiment.

An experiment involves the manipulation of variables by the researcher. In this case, the researcher is randomly assigning half of the volunteers to exercise 1 hour each morning, while the other half are instructed to perform no exercise. This manipulation of the exercise variable is a key characteristic of an experiment.

Additionally, the experiment includes a control group, which is the group of volunteers instructed to perform no exercise. This allows for comparison between the group that exercises and the group that does not, to assess the impact of exercise on resting heart rate.

Therefore, the correct answer is A. 'An experiment with a control group and blinding'. Final Answer: A

\*\*\*\*\*

Answer Option 9:

The study described is an experiment. In an experiment, the researchers actively manipulate the variables of interest and assign participants to different groups or conditions. In this case, the researchers randomly assigned half of the volunteers to the exercise group and the other half to the no exercise group. This random assignment helps to ensure that the groups are comparable, which is important for drawing causal conclusions. Additionally, the study has a control group (the no exercise group) which allows for comparison between the groups. The fact that blinding or blocking is not mentioned does not change the fact that it is an experiment. Therefore, the correct answer is: A. 'An experiment with a control group and blinding'. Final Answer: A

\*\*\*\*\*

Question 63:

'Both over-the-counter niacin and the prescription drug Lipitor are known to lower blood cholesterol levels. In one double-blind study Lipitor outperformed niacin. The 95% confidence interval estimate of the difference in mean cholesterol level lowering was (18, 41). Which of the following is a reasonable conclusion?', A. 'Niacin lowers cholesterol an average of 18 points, while Lipitor lowers cholesterol an average of 41 points.', B. 'There is a 0.95 probability that Lipitor will outperform niacin in lowering the cholesterol level of any given individual.', C. 'There is a 0.95 probability that Lipitor will outperform niacin by at least 23 points in lowering the cholesterol level of any given individual.', D. 'None of the above.'

=====

Answer Option 1:

The correct answer is D. 'None of the above.'

The given information is about the confidence interval estimate of the difference in mean cholesterol level lowering between Lipitor and niacin. It does not provide information about the individual average cholesterol level lowering for each drug or the probability of Lipitor outperforming niacin in lowering cholesterol for any given individual. Therefore, none of the statements A, B, or C can be concluded based on the information given. Final answer: D

\*\*\*\*\*

Answer Option 2:

The correct conclusion is D. 'None of the above.' The given confidence interval of (18, 41) does not provide information about the average cholesterol lowering effects of niacin and Lipitor. It only provides information about the likely range of the difference in mean cholesterol level lowering between the two drugs. Additionally, it does not provide information about the probability of one drug outperforming the other in individual cases. Therefore, we cannot draw any conclusions about the average cholesterol lowering or the probability of one drug outperforming the other based on the given confidence interval.

Final Answer: D

\*\*\*\*\*

Answer Option 3:

The correct conclusion is D. 'None of the above.'

The 95% confidence interval estimate (18, 41) means that we are 95% confident that the true difference in mean cholesterol level lowering lies between 18 and 41 points. It does not provide information about the actual average cholesterol lowering for niacin or Lipitor or the probabilities of Lipitor outperforming niacin by a certain amount. Therefore, all of the given conclusions are not reasonable based on the information provided. Final answer: D.

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'None of the above.'

The 95% confidence interval estimate (18, 41) means that we are 95% confident that the true difference in mean cholesterol level lowering between Lipitor and niacin falls within the interval (18, 41). It does not provide information about the average cholesterol level lowering for each drug or the probability of Lipitor outperforming niacin by a certain amount.

To draw a more reasonable conclusion, we would need additional information or a different type of statistical analysis.

\*\*\*\*\*

Answer Option 5:

The correct conclusion is D. 'None of the above.'

The given confidence interval (18, 41) represents the range of values within which the true difference in mean cholesterol level lowering between Lipitor and niacin is likely to fall, with 95% confidence. However, this does not provide information about the average cholesterol level lowering of either niacin or Lipitor, nor does it provide probabilities about individual performance or a specific difference in cholesterol level lowering.

Therefore, none of the given statements can be reasonably concluded based on the provided information. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The correct answer is D. 'None of the above.'

The given confidence interval (18, 41) does not provide information about the average cholesterol-lowering effects of niacin or Lipitor. It only provides information about the uncertainty in the estimated difference in mean cholesterol level lowering between the two treatments.

Option A states specific average cholesterol-lowering values for both niacin and Lipitor, which cannot be concluded from the confidence interval.

Option B states a probability, but this is incorrect. A confidence interval provides information about the amount of uncertainty in the estimate, not the probability of an event occurring.

Option C states a specific difference of at least 23 points, but this cannot be concluded from the confidence interval. It only provides a range of plausible values for the difference.

Therefore, the reasonable conclusion is 'None of the above.' Final answer: D

\*\*\*\*\*

Answer Option 7:

The correct conclusion is D. 'None of the above.'

The 95% confidence interval estimate of the difference in mean cholesterol level lowering is (18, 41). This means that we are 95% confident that the true difference in mean cholesterol level lowering between Lipitor and niacin falls between 18 and 41 points. It does not provide any information about the average cholesterol level lowering for niacin or Lipitor individually. Therefore, option A is incorrect.

Additionally, the confidence interval estimate does not imply anything about the probability that Lipitor will outperform niacin in lowering the cholesterol level for any given individual. Therefore, option B is incorrect.

Similarly, the confidence interval estimate does not imply anything about the probability that Lipitor will outperform niacin by at least 23 points in lowering the cholesterol level for any given individual.

Therefore, option C is incorrect.

Hence, the correct conclusion is D. 'None of the above.' Final Answer: D

\*\*\*\*\*

Answer Option 8:

The correct reasonable conclusion is D. 'None of the above.'

The 95% confidence interval estimate of the difference in mean cholesterol level lowering is (18, 41). This means that we are 95% confident that the true difference in mean cholesterol level lowering between Lipitor and niacin is somewhere between 18 and 41 points. It does not mean that niacin lowers cholesterol by an average of 18 points and Lipitor lowers cholesterol by an average of 41 points (option A).

Similarly, it does not mean that there is a 0.95 probability that Lipitor will outperform niacin in lowering the cholesterol level of any given individual (option B) or that Lipitor will outperform niacin by at least 23 points in lowering the cholesterol level of any given individual (option C).

The confidence interval provides an estimate of the range within which the true difference in mean cholesterol level lowering likely falls, but it does not give probabilities or average values for individual cases.

Therefore, none of the given options are valid reasonable conclusions. Final answer: D

\*\*\*\*\*

Answer Option 9:

The correct answer is D. 'None of the above.'

First, let's analyze the options:

A. 'Niacin lowers cholesterol an average of 18 points, while Lipitor lowers cholesterol an average of 41 points.' This option is not supported by the information given. The confidence interval only gives an estimate of the difference in mean cholesterol level lowering, not the absolute values for each drug.

B. 'There is a 0.95 probability that Lipitor will outperform niacin in lowering the cholesterol level of any given individual.' This option is incorrect. The confidence interval does not provide probabilities or guarantees about individual outcomes. It only provides an interval estimate for the difference in means based on the sample data.

C. 'There is a 0.95 probability that Lipitor will outperform niacin by at least 23 points in lowering the cholesterol level of any given individual.' This option is also incorrect. The confidence interval does not provide probabilities or guarantees about individual outcomes. It only provides an interval estimate for the difference in means based on the sample data.

Since none of the options accurately represents the information provided by the confidence interval, the correct answer is D. 'None of the above.' Final Answer: D

\*\*\*\*\*



Question 64:

'Fifty oranges of one variety were selected at random and weighed. A 95 percent confidence interval computed from the sample weights to estimate the mean weight of oranges of this variety is (7.58, 8.72) ounces. Which of the following statements is true with respect to the confidence level used?', A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.', B. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the sampled mean weight of the oranges.', C. 'Ninety-five percent of the oranges in this sample weigh between 7.58 and 8.72 ounces.', D. 'Ninety-five percent of the oranges of this variety weigh between 7.58 and 8.72 ounces.'

=====

Answer Option 1:

The correct answer is B. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the sampled mean weight of the oranges.'

A confidence interval is a range of values constructed from a sample that is likely to contain the true population parameter with a certain level of confidence. In this case, the confidence level used is 95 percent. This means that if we were to take multiple samples of the same size and construct confidence intervals for each sample, approximately 95 percent of those intervals would contain the true mean weight of the oranges of this variety.

Option A is incorrect because it states that 95 percent of the intervals constructed will contain the true mean weight of the oranges. This is not necessarily true as the intervals are constructed from samples and may or may not include the true mean.

Option C is incorrect because it refers to the specific sample of oranges used to construct the confidence interval, not the population as a whole.

Option D is incorrect because it also refers to the specific sample of oranges used to construct the confidence interval, not the population as a whole.

\*\*\*\*\*

Answer Option 2:

The correct statement is option A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.'

A confidence interval provides a range of plausible values for the population parameter (in this case, the mean weight of oranges of this variety) based on a sample. The 95 percent confidence interval (7.58, 8.72) means that if we were to take many samples of the same size and construct 95 percent confidence intervals for each sample, approximately 95 percent of these intervals would contain the true mean weight of the oranges. This statement accurately describes the interpretation of a confidence interval.

Option B is incorrect because the confidence interval is not based on the sampled mean weight of the oranges; it is based on the true mean weight of the oranges, which is unknown.

Option C is incorrect because the confidence interval does not provide information about individual oranges in the sample; it is focused on estimating the mean weight of oranges of this variety.

Option D is incorrect because the confidence interval is not about the individual oranges; it is about the mean weight of oranges of this variety.

\*\*\*\*\*

Answer Option 3:

The correct statement is A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.'

A 95 percent confidence interval means that if we were to take 100 different samples of the same size, the true mean weight of the oranges would fall within the interval for approximately 95 of those samples. It does not guarantee that any particular sample will have 95 percent of the oranges falling within the interval, as stated in statements C and D. It also does not guarantee that the sampled mean weight will fall within the interval for 95 percent of the samples, as stated in statement B. Final Answer: A

\*\*\*\*\*

Answer Option 4:

The correct statement is A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.'

A confidence interval is constructed to estimate the true population parameter, which in this case is the mean weight of oranges of the variety. The 95 percent confidence interval means that in repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight. This statement reflects the concept of confidence intervals and their interpretation. It does not mean that 95 percent of the oranges in the sample or 95 percent of the oranges of this variety weigh between 7.58 and 8.72 ounces. So options B, C, and D are incorrect. Final Answer: A

\*\*\*\*\*

Answer Option 5:

The correct statement is A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.'

A 95 percent confidence interval means that if we were to take many random samples of the same size and compute a confidence interval for each sample, approximately 95 percent of the intervals would contain the true mean weight of the oranges of this variety.

Statement B is incorrect because the confidence interval is not about the sampled mean weight, but about the true mean weight.

Statement C is incorrect because the confidence interval is not about the individual oranges in the sample, but about the mean weight.

Statement D is incorrect because the confidence interval provides a range estimate for the true mean weight, not for the weight of individual oranges.

\*\*\*\*\*

Answer Option 6:

The correct answer is A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.'

A confidence interval is constructed to estimate a population parameter, in this case, the mean weight of oranges of a certain variety. The statement in option A is the correct interpretation of a 95 percent confidence interval. It means that if we were to take multiple samples of the same size and construct a 95 percent confidence interval using each sample, approximately 95 percent of these intervals will contain the true population mean.

Option B is incorrect because the confidence interval is used to estimate the population mean, not the sample mean. Option C is incorrect because the confidence interval applies to the population, not the sample. Option D is incorrect because the confidence interval provides information about the population mean, not individual observations.

\*\*\*\*\*

Answer Option 7:

The correct statement with respect to the confidence level used is A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.'

A confidence interval is a range of values that is likely to contain the population parameter (in this case, the mean weight of oranges of this variety) with a certain level of confidence. In this problem, the 95 percent confidence interval (7.58, 8.72) means that if we were to take repeated samples of the same size and compute confidence intervals using the same method, we would expect that approximately 95 percent of those intervals would contain the true mean weight of the oranges of this variety.

Option B is incorrect because the confidence interval is not about the sampled mean weight of the oranges. Option C is incorrect because it refers to the specific sample, not the population. Option D is incorrect because it refers to the specific oranges in the sample, not the population mean weight. Final Answer: A

\*\*\*\*\*

Answer Option 8:

The correct statement is A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.'

A confidence interval is an estimate of a population parameter, in this case, the mean weight of oranges of a particular variety. The given confidence interval (7.58, 8.72) means that there is a 95% chance that the true mean weight of the oranges lies within this interval. In repeated sampling, 95% of the constructed confidence intervals will contain the true mean weight.

Option B is incorrect because the confidence interval is not about the sampled mean weight, but rather about the true mean weight.

Option C is incorrect because it is not correct to say that 95% of the oranges in the sample weigh between 7.58 and 8.72 ounces. The confidence interval is about the population mean, not individual oranges in the sample.

Option D is incorrect for the same reason as option C. The confidence interval is about the population mean, not individual oranges.

\*\*\*\*\*

Answer Option 9:

The correct statement is A. 'In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.'

A confidence interval is a range of values that is likely to contain the true value of the population parameter being estimated. In this case, the confidence interval (7.58, 8.72) is used to estimate the mean weight of oranges of this variety.

The confidence level, which is 95 percent in this case, refers to the long-term frequency of the intervals that contain the true value of the parameter. This means that if we were to take many samples of the same size and construct confidence intervals using the same method, we would expect 95 percent of those intervals to contain the true mean weight of the oranges.

Option B is incorrect because the confidence interval is not used to estimate the sampled mean weight of the oranges, but rather the true mean weight of the oranges. Option C is incorrect because it is not accurate to say that 95 percent of the oranges in this sample weigh between 7.58 and 8.72 ounces, as the confidence interval is used to estimate the population mean weight, not individual weights. Option D is incorrect because the confidence interval is not making a statement about the weights of individual oranges, but rather the true mean weight of oranges of this variety.

\*\*\*\*\*

Question 65:

There are two games involving flipping a fair coin. In the first game you win a prize if you can throw between 45% and 55% heads. In the second game you win if you can throw more than 80% heads. For each game would you rather flip the coin 30 times or 300 times?, A. '30 times for each game', B. '300 times for each game', C. '30 times for the first game and 300 times for the second', D. '300 times for the first game and 30 times for the second'

=====

Answer Option 1:

To determine which option is better, we need to consider the expected number of heads in each game.

In the first game, the expected number of heads in 30 coin flips is  $30 \cdot 0.5 = 15$ . Since we want between 45% and 55% heads, this means we want between 13.5 and 16.5 heads. Since 15 falls within this range, we would rather flip the coin 30 times in the first game.

In the second game, the expected number of heads in 300 coin flips is  $300 \cdot 0.5 = 150$ . Since we want more than 80% heads, this means we want more than 240 heads. Since 150 is less than 240, we would rather flip the coin 300 times in the second game.

Therefore, the answer is  $\boxed{\text{30 times for the first game and 300 times for the second}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

To determine which option is better for each game, we need to calculate the probability of winning in each scenario.

For the first game, we want to find the probability of throwing between 45% and 55% heads. Since the coin is fair, the probability of getting heads on a single flip is 0.5. We can use the binomial distribution to calculate the probability of getting a specific number of heads in a given number of flips.

Using 30 flips, the probability of getting between 45% and 55% heads can be calculated by summing the probabilities of getting 14, 15, ..., 16 heads:

$$P(\text{45\% to 55\% heads in 30 flips}) = \sum_{k=14}^{16} \binom{30}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{(30-k)}$$

Similarly, for 300 flips, the probability of getting between 45% and 55% heads can be calculated by summing the probabilities of getting 134, 135, ..., 165 heads:

$$P(\text{45\% to 55\% heads in 300 flips}) = \sum_{k=134}^{165} \binom{300}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{(300-k)}$$

For the second game, we want to find the probability of getting more than 80% heads. Again, we use the binomial distribution to calculate the probabilities.

Using 30 flips, the probability of getting more than 80% heads can be calculated by summing the probabilities of getting 25, 26, ..., 30 heads:

$$P(\text{more than 80\% heads in 30 flips}) = \sum_{k=25}^{30} \binom{30}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{(30-k)}$$

Similarly, for 300 flips, the probability of getting more than 80% heads can be calculated by summing the probabilities of getting 241, 242, ..., 300 heads:

$$P(\text{more than 80\% heads in 300 flips}) = \sum_{k=241}^{300} \binom{300}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{(300-k)}$$

After calculating these probabilities, we find that the probabilities for the first game are extremely small, regardless of the number of flips. On the other hand, the probabilities for the second game are much higher when flipping the coin 300 times compared to 30 times.

Therefore, we would rather flip the coin  $\boxed{\text{300 times for the second game and 30 times for the first game}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

In both games, the probability of getting a heads on any given flip is  $1/2$ , and the probability of getting a tails is also  $1/2$ .

In the first game, the probability of getting between 45% and 55% heads after 30 flips is given by the binomial distribution. This can be calculated using a computer or a calculator. Similarly, the probability of getting between 45% and 55% heads after 300 flips can also be calculated.

In the second game, the probability of getting more than 80% heads after 30 flips can also be calculated using the binomial distribution. Similarly, the probability of getting more than 80% heads after 300 flips can also be calculated.

Comparing the probabilities for each game, we see that as the number of flips increases, the probabilities become more accurate representations of the true probabilities. Therefore, in both games, we would prefer to flip the coin 300 times rather than 30 times.

Therefore, the answer is  $\boxed{\text{B. } 300 \text{ times for each game}}$ .

\*\*\*\*\*

Answer Option 4:

We can use the binomial distribution to analyze the probabilities of winning each game.

For the first game, flipping the coin 30 times or 300 times will have a similar effect. The probability of getting between 45% and 55% heads in 30 flips is the sum of the probabilities of getting 14, 15, ..., or 16 heads, which is approximately 0.914. The probability of getting between 45% and 55% heads in 300 flips is the sum of the probabilities of getting between 134 and 166 heads, which is approximately 0.907.

Therefore, it does not matter whether you flip the coin 30 times or 300 times for the first game.

For the second game, flipping the coin 300 times is clearly advantageous. The probability of getting more than 80% heads in 30 flips is the sum of the probabilities of getting 24, 25, ..., or 30 heads, which is approximately 0.001. The probability of getting more than 80% heads in 300 flips is the sum of the probabilities of getting more than 239 heads, which is significantly greater.

Therefore, the answer is  $\boxed{\text{300 times for the first game and 30 times for the second}}$ . Final answer: D

\*\*\*\*\*

Answer Option 5:

To determine which option is better for each game, we need to compare the probabilities of winning in each scenario.

In the first game, the probability of throwing between 45% and 55% heads in 30 coin flips can be approximated using the normal distribution. The mean is  $30 \cdot 0.5 = 15$ , and the standard deviation is  $\sqrt{30 \cdot 0.5 \cdot 0.5} = \sqrt{7.5} \approx 2.74$ . Using a z-score table, we can find the probability of falling within the given range. The probability turns out to be approximately 0.806.

Similarly, in the second game, the probability of throwing more than 80% heads in 30 coin flips can be approximated using the same method. The mean is  $30 \times 0.5 = 15$ , and the standard deviation is  $\sqrt{30 \times 0.5 \times 0.5} = \sqrt{7.5} \approx 2.74$ . Again, using a z-score table, we can find the probability of exceeding 80%. The probability turns out to be approximately 0.0003.

Now let's consider the scenario of flipping the coin 300 times for each game. In both cases, the mean is  $300 \times 0.5 = 150$ , and the standard deviation is  $\sqrt{300 \times 0.5 \times 0.5} = \sqrt{75} \approx 8.66$ . Using the same method as before, we find that the probability of throwing between 45% and 55% heads in 300 coin flips is approximately 0.954, while the probability of exceeding 80% heads is approximately 0.000.

Comparing the probabilities, we can conclude that for the first game, it is better to flip the coin 300 times, while for the second game, it is better to flip the coin 30 times. Therefore, the answer is

$\boxed{\text{D. } 300 \text{ times for the first game and 30 times for the second}}$ . Final answer: D.

\*\*\*\*\*

Answer Option 6:

In both games, we are interested in the probability of obtaining a certain number of heads out of the total number of flips. We can use the binomial distribution to calculate these probabilities.

For the first game, we want to find the probability of getting between 45% and 55% heads out of 30 flips. This is equivalent to finding the probability of getting between 13.5 and 16.5 heads. Calculating these probabilities for 30 flips using the binomial distribution is relatively straightforward.

For the second game, we want to find the probability of getting more than 80% heads out of 300 flips. This is equivalent to finding the probability of getting more than 240 heads. Calculating these probabilities for 300 flips using the binomial distribution becomes more difficult and time-consuming.

Therefore, it would be more efficient to flip the coin 30 times for the first game and 300 times for the second game. This way, we can easily calculate the probabilities needed for both games. Hence, the

answer is  $\boxed{\text{C. } 30 \text{ times for the first game and 300 times for the second}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 7:

For each game, we can use the binomial probability formula to calculate the probability of winning.

In the first game, we want to throw between 45% and 55% heads. If we flip the coin 30 times, the probability of winning can be calculated as the sum of the probabilities of getting 14, 15, ..., or 16 heads. This calculation involves calculating the probability of getting a specific number of heads in a specific order, which can be time-consuming. However, if we flip the coin 300 times, we can use the normal approximation to the binomial distribution, which is much faster and easier to calculate. The normal approximation can be applied when both the number of trials and the probability of success are large, which is the case when flipping a coin 300 times.

In the second game, we want to throw more than 80% heads. Again, if we flip the coin 30 times, we would need to calculate the probabilities of getting 25, 26, ..., or 30 heads, which can be time-consuming.

However, if we flip the coin 300 times, we can again use the normal approximation to the binomial distribution.

Therefore, the best strategy is to flip the coin 300 times for both games, so the answer is  $\boxed{\text{B. } 300 \text{ times for each game}}$ .

\*\*\*\*\*

Answer Option 8:

To determine which option is better for each game, we need to compare the expected number of heads obtained in each scenario.

For the first game, the probability of obtaining a head on a fair coin flip is \$0.5. The expected number of heads in 30 flips is  $30 \times 0.5 = 15$ , and the expected number of heads in 300 flips is  $300 \times 0.5 = 150$ . Since 45% to 55% of 30 is 13.5 to 16.5, and 45% to 55% of 300 is 135 to 165, flipping the coin 300 times would be preferable for the first game.

For the second game, the probability of obtaining a head on a fair coin flip is still \$0.5. The expected number of heads in 30 flips is  $30 \times 0.5 = 15$ , and the expected number of heads in 300 flips is  $300 \times 0.5 = 150$ . Since 80% of 30 is 24, and 80% of 300 is 240, flipping the coin 30 times would be preferable for the second game.

Therefore, the answer is  $\boxed{\text{C. } 30 \text{ times for the first game and 300 times for the second}}$ .

\*\*\*\*\*

Answer Option 9:

We can use the concept of the binomial distribution to analyze this problem.

For the first game, if we flip the coin 30 times, the mean number of heads is  $30 \times 0.5 = 15$ , and the standard deviation is  $\sqrt{30 \times 0.5 \times 0.5} \approx 2.74$ . The probability of getting between 45% and 55% heads is the probability of getting between 13.5 and 16.5 heads, which can be approximated using a normal distribution. The z-scores for 13.5 and 16.5 are approximately -0.91 and 0.91 respectively. Using a standard normal distribution table, the probability of getting between these two values is approximately 0.41.

For the second game, if we flip the coin 30 times, the mean number of heads is  $30 \times 0.5 = 15$ , and the standard deviation is  $\sqrt{30 \times 0.5 \times 0.5} \approx 2.74$ . The probability of getting more than 80% heads is the probability of getting more than 24 heads, which can again be approximated using a normal distribution. The z-score for 24 is approximately 3.27. Using a standard normal distribution table, the probability of getting more than 24 heads is approximately 0.0006.

For the first game, if we flip the coin 300 times, the mean number of heads is  $300 \times 0.5 = 150$ , and the standard deviation is  $\sqrt{300 \times 0.5 \times 0.5} \approx 8.66$ . The probability of getting between 45% and 55% heads is the probability of getting between 135 and 165 heads. Again, we can use a normal distribution approximation. The z-scores for 135 and 165 are approximately -1.73 and 1.73 respectively. Using a standard normal distribution table, the probability of getting between these two values is approximately 0.0801.

For the second game, if we flip the coin 300 times, the mean number of heads is  $300 \times 0.5 = 150$ , and the standard deviation is  $\sqrt{300 \times 0.5 \times 0.5} \approx 8.66$ . The probability of getting more than 80% heads is the probability of getting more than 240 heads, which can be approximated using a normal distribution. The z-score for 240 is approximately 9.01. Using a standard normal distribution table, the probability of getting more than 240 heads is approximately 0.

Comparing the probabilities, we can see that for the first game, it is better to flip the coin 30 times. For the second game, it is better to flip the coin 300 times. Therefore, the answer is  $\boxed{\text{C. } 30 \text{ times for the first game and 300 times for the second}}$ .

\*\*\*\*\*

Question 66:

'A company is interested in comparing the mean sales revenue per salesperson at two different locations. The manager takes a random sample of 10 salespeople from each location independently and records the sales revenue generated by each person during the last four weeks. He decides to use a t-test to compare the mean sales revenue at the two locations. Which of the following assumptions is necessary for the validity of the t-test?', A. 'The population standard deviations at both locations are equal.', B. 'The population standard deviations at both locations are not equal.', C. 'The population standard deviations at both locations are known.', D. 'The population of the sales records at each location is normally distributed.'

=====

Answer Option 1:

The correct assumption necessary for the validity of the t-test is D. 'The population of the sales records at each location is normally distributed.' The t-test assumes that the population follows a normal distribution in order to make inferences about the means of the two populations. Therefore, option D is the correct assumption required for the validity of the t-test. Final Answer: D

\*\*\*\*\*

Answer Option 2:

The correct assumption necessary for the validity of the t-test is D. 'The population of the sales records at each location is normally distributed.' The t-test assumes that the population from which the samples are taken is normally distributed. The other options, A, B, and C, relate to the assumption of the equality of population standard deviations, which is not necessary for the validity of the t-test. Therefore, the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The correct assumption necessary for the validity of the t-test is D. 'The population of the sales records at each location is normally distributed.' The t-test assumes that the data is normally distributed in order to make valid inferences about the population mean. Options A and B are not assumptions necessary for the t-test, as the t-test can still be used even if the population standard deviations are not known or not equal. Option C is also not necessary, as the t-test does not require knowledge of the population standard deviations. Therefore, the correct answer is D. Final Answer: D.

\*\*\*\*\*

Answer Option 4:

The correct assumption necessary for the validity of the t-test is D. 'The population of the sales records at each location is normally distributed.'

The t-test assumes that the data follow a normal distribution. This assumption is needed to ensure the validity of the t-test and for accurate inference about the means of the two populations. The other options, A, B, and C, are assumptions related to the equal or unequal variances between the two populations, which affect the choice of the specific t-test to be used (equal variance t-test or unequal variance t-test), but are not necessary assumptions for the validity of the t-test itself.

Therefore, the correct answer is D. 'The population of the sales records at each location is normally distributed.' Final answer: D.

\*\*\*\*\*

Answer Option 5:

The necessary assumption for the validity of the t-test is that the population of the sales records at each location is normally distributed. Therefore, the correct answer is  $\boxed{\text{D. 'The population of the sales records at each location is normally distributed.'}}$

\*\*\*\*\*

Answer Option 6:

The correct assumption necessary for the validity of the t-test is D. 'The population of the sales records at each location is normally distributed.' The t-test assumes that the data come from normally distributed populations. The other options, A, B, and C, are not necessary assumptions for the t-test. Final answer: D

\*\*\*\*\*

Answer Option 7:

The correct assumption necessary for the validity of the t-test is that the population of the sales records at each location is normally distributed. This assumption is necessary to ensure that the sampling distribution of the means is also normally distributed, which is a requirement for the t-test. Therefore, the answer is  $\boxed{\text{D. 'The population of the sales records at each location is normally distributed.'}}$  Final Answer: D

\*\*\*\*\*

Answer Option 8:

The necessary assumption for the validity of the t-test is that the population of the sales records at each location is normally distributed. This assumption allows us to use the t-distribution to make inferences about the means of the populations based on the sample means. Therefore, the correct answer is  $\boxed{\text{D. 'The population of the sales records at each location is normally distributed.'}}$

\*\*\*\*\*

Answer Option 9:

The correct assumption necessary for the validity of the t-test is option D: 'The population of the sales records at each location is normally distributed.' The t-test assumes that the populations from which the samples are drawn are normally distributed. This assumption is necessary to ensure that the sampling distribution of the mean is also normally distributed. The assumptions regarding the equality or inequality of population standard deviations (options A and B) and the knowledge of population standard deviations (option C) are not required for the validity of the t-test. Final Answer: D

\*\*\*\*\*

Question 67:

'Suppose you did 10 independent tests of the form  $H_0: \mu = 25$  versus  $H_a: \mu < 25$ , each at the  $\alpha = 0.05$  significance level. What is the probability of committing a Type I error and incorrectly rejecting a true  $H_0$  with at least one of the 10 tests?', A. '0.05', B. '0.40', C. '0.50', D. '0.60'

=====

Answer Option 1:

The probability of committing a Type I error in a single test at the  $\alpha = 0.05$  significance level is 0.05.

To find the probability of committing a Type I error in at least one of the 10 tests, we can find the probability of not committing a Type I error in all 10 tests and subtract it from 1.

The probability of not committing a Type I error in a single test is  $1 - 0.05 = 0.95$ .

Since the tests are independent, the probability of not committing a Type I error in all 10 tests is  $(0.95)^{10}$ .

Therefore, the probability of committing a Type I error in at least one of the 10 tests is  $1 - (0.95)^{10} = 0.4013$ .

Rounded to two decimal places, this is approximately 0.40.

Therefore, the correct answer is  $\boxed{\text{B. '0.40'}}$ .

\*\*\*\*\*

Answer Option 2:

The probability of committing a Type I error in a single test is the significance level, which is given as  $\alpha = 0.05$ .

The probability of not committing a Type I error in a single test is  $1 - \alpha = 1 - 0.05 = 0.95$ .

Since the tests are independent, the probability of not committing a Type I error in all 10 tests is  $(0.95)^{10}$ .

Therefore, the probability of committing a Type I error in at least one of the 10 tests is  $1 - (0.95)^{10} = 0.4013$ .

Rounded to two decimal places, the probability is approximately 0.40.

Thus, the correct answer is B. 0.40. Final Answer: B

\*\*\*\*\*

Answer Option 3:

The probability of committing a Type I error in a single test is equal to the significance level  $\alpha$ , which is given as 0.05.

In each test, the probability of not committing a Type I error (correctly accepting  $H_0$ ) is  $1 - \alpha = 1 - 0.05 = 0.95$ .

The probability of not committing a Type I error in all 10 tests is  $(0.95)^{10} = 0.5987$ .

Therefore, the probability of committing a Type I error in at least one of the 10 tests is  $1 - 0.5987 = 0.4013$ , which is approximately 0.40.

So, the answer is  $\boxed{\text{B. '0.40'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 4:

The probability of committing a Type I error in a single test, at the  $\alpha = 0.05$  significance level, is 0.05.

To find the probability of committing a Type I error in at least one of the 10 tests, we can find the complement of the event that no Type I errors are committed in any of the tests.

The probability of not committing a Type I error in a single test is  $1 - 0.05 = 0.95$ .

Since the 10 tests are independent, the probability of not committing a Type I error in any of the 10 tests is  $(0.95)^{10} \approx 0.5987$ .

Therefore, the probability of committing a Type I error in at least one of the 10 tests is  $1 - 0.5987 \approx 0.4013$ .

This is approximately equal to  $\boxed{\text{B. 0.40}}$ . Final answer: B

\*\*\*\*\*

Answer Option 5:

The probability of committing a Type I error (rejecting a true null hypothesis) in a single test at the  $\alpha = 0.05$  significance level is 0.05.

If we perform 10 independent tests, the probability of committing a Type I error in at least one of the tests can be calculated as the complement of the probability of not committing a Type I error in any of the tests.

The probability of not committing a Type I error in a single test is  $1 - 0.05 = 0.95$ .

Since the tests are independent, the probability of not committing a Type I error in any of the 10 tests is  $(0.95)^{10}$ .

Therefore, the probability of committing a Type I error in at least one of the tests is  $1 - (0.95)^{10} \approx 0.401263$ .

Rounded to two decimal places, this is approximately 0.40.

Thus, the answer is  $\boxed{\text{B. } 0.40}$ .

\*\*\*\*\*

Answer Option 6:

If we conduct each test at the  $\alpha = 0.05$  significance level, then the probability of committing a Type I error in a single test is 0.05.

The probability of not committing a Type I error (i.e., correctly failing to reject a true null hypothesis) in a single test is  $1 - 0.05 = 0.95$ .

Since the tests are independent, the probability of not committing a Type I error in any of the 10 tests is  $(0.95)^{10}$ .

Therefore, the probability of committing a Type I error in at least one of the 10 tests is  $1 - (0.95)^{10} \approx 0.4013$ .

Rounded to two decimal places, the answer is  $\boxed{\text{B. } 0.40}$ .

\*\*\*\*\*

Answer Option 7:

The probability of committing a Type I error in a single test at the  $\alpha = 0.05$  significance level is 0.05.

To find the probability of committing a Type I error in at least one of the 10 tests, we can use the complement rule. The probability of not committing a Type I error in a single test is  $1 - 0.05 = 0.95$ .

The probability of not committing a Type I error in any of the 10 tests is  $(0.95)^{10}$ .

Therefore, the probability of committing a Type I error in at least one of the 10 tests is  $1 - (0.95)^{10}$ .

Using a calculator, we find that  $1 - (0.95)^{10} \approx 0.401263$ .

So the probability of committing a Type I error in at least one of the 10 tests is approximately 0.401263, which rounded to two decimal places is 0.40.

Therefore, the answer is  $\boxed{\text{B. } 0.40}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

The probability of committing a Type I error for each individual test is  $\alpha = 0.05$ .

The probability of not committing a Type I error (correctly failing to reject a true null hypothesis) for each individual test is  $1 - \alpha = 0.95$ .

Since the tests are independent, the probability of not committing a Type I error for all 10 tests is  $(0.95)^{10}$ .

Therefore, the probability of committing a Type I error with at least one of the 10 tests is  $1 - (0.95)^{10} \approx 0.40$ .

Thus, the answer is  $\boxed{\text{B. } 0.40}$ .

\*\*\*\*\*

Answer Option 9:

The probability of committing a Type I error in a single test is the significance level, which is given as  $\alpha = 0.05$ .

The probability of not committing a Type I error in a single test is  $1 - \alpha = 1 - 0.05 = 0.95$ .

Since the tests are independent, the probability of not committing a Type I error in all 10 tests is  $(0.95)^{10} = 0.5987$ .

Therefore, the probability of committing a Type I error in at least one of the 10 tests is  $1 - 0.5987 = 0.4013$ .

This is approximately equal to 0.40, so the answer is  $\boxed{\text{B. } 0.40}$ . Final answer: B

\*\*\*\*\*

Question 68:

'Suppose that 60% of a particular electronic part last over 3 years, while 70% last less than 6 years. Assuming a normal distribution, what are the mean and standard deviation with regard to length of life of these parts?', A. ' $\mu = 3.677$ ,  $\sigma = 3.561$ ', B. ' $\mu = 3.977$ ,  $\sigma = 3.861$ ', C. ' $\mu = 4.177$ ,  $\sigma = 3.561$ ', D. ' $\mu = 4.377$ ,  $\sigma = 3.261$ '

=====

Answer Option 1:

Let's assume that the length of life of these parts follows a normal distribution. We know that 60% of the parts last over 3 years, which means that 40% of the parts last less than or equal to 3 years. Similarly, 70% of the parts last less than 6 years, which means that 30% of the parts last more than or equal to 6 years.

Using the standard normal distribution table or a calculator, we can find that the z-score corresponding to 40% is approximately -0.253 and the z-score corresponding to 30% is approximately 0.524.

Since the normal distribution is symmetric, the mean ( $\mu$ ) will be the average of the z-scores, which is  $(-0.253 + 0.524) / 2 = 0.1355$ .

To find the standard deviation ( $\sigma$ ), we need to find the difference between the z-score corresponding to 60% and the mean. The z-score corresponding to 60% is approximately -0.8416. Therefore,  $\sigma = 0.8416 - 0.1355 = 0.7061$ .

Converting the z-scores back to years, we need to multiply the z-scores by the standard deviation and add the mean. Thus, the mean is  $0.1355 * 3 + 3 = 3.4065$ , and the standard deviation is  $0.7061 * 3 = 2.1183$ .

Therefore, the mean and standard deviation are approximately  $\mu = 3.4065$  and  $\sigma = 2.1183$ , respectively. Rounded to three decimal places, the answer is ' $\mu = 3.407$ ,  $\sigma = 2.118$ ', which corresponds to the answer choice B. Final Answer: B

\*\*\*\*\*

Answer Option 2:

Let's denote the length of life of the electronic part by  $X$ . We are given that 60% of the parts last over 3 years, which means 40% of the parts last less than or equal to 3 years. Similarly, 70% of the parts last less than 6 years, which means 30% of the parts last more than or equal to 6 years.

We can use the standard normal distribution to find the mean and standard deviation. We know that 40% of the parts last less than or equal to 3 years, which corresponds to a z-score of  $Z = \frac{X - \mu}{\sigma} = -0.253$ . Using a standard normal distribution table, we can find that the corresponding z-score is approximately -0.253. Similarly, 30% of the parts last more than or equal to 6 years, which corresponds to a z-score of  $Z = \frac{X - \mu}{\sigma} = 0.524$ . Using the standard normal distribution table, we can find that the corresponding z-score is approximately 0.524.

From the z-scores, we can find the values of  $\mu$  and  $\sigma$ . We have:

$$-0.253 = \frac{3 - \mu}{\sigma}$$

$$0.524 = \frac{6 - \mu}{\sigma}$$

Solving these two equations, we can find that  $\mu = 4.377$  and  $\sigma = 3.261$ . Therefore, the answer is  $\boxed{\text{D) } \mu = 4.377, \sigma = 3.261}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

Let  $X$  be the length of life of the electronic part. We have the following information:

$$P(X > 3) = 0.6 \text{ (60\% last over 3 years)}$$

$$P(X < 6) = 0.7 \text{ (70\% last less than 6 years)}$$

We can use the standard normal distribution to find the corresponding z-scores for these probabilities.

For  $P(X > 3) = 0.6$ , we have:

$$P(X \leq 3) = 1 - P(X > 3) = 0.4$$

For  $P(X < 6) = 0.7$ , we have:

$$P(X \geq 6) = 1 - P(X < 6) = 0.3$$

Using the z-score formula,  $Z = \frac{X - \mu}{\sigma}$ , we can find the z-scores for the corresponding probabilities:

For  $P(X \leq 3) = 0.4$ :

$$Z_1 = \frac{3 - \mu}{\sigma}$$

For  $P(X \geq 6) = 0.3$ :

$$Z_2 = \frac{6 - \mu}{\sigma}$$

Since the standard normal distribution is symmetric, we have:

$$P(X \leq 3) = P(X \geq 6)$$

Therefore, we can set  $Z_1 = -Z_2$  and solve for  $\mu$  and  $\sigma$ .

$$\frac{3 - \mu}{\sigma} = -\frac{6 - \mu}{\sigma}$$

Solving this equation gives us  $\mu = 4.5$ .

To find  $\sigma$ , we can use the fact that the z-score for  $P(X \leq 3) = 0.4$  is -0.253 (found using a standard normal distribution table).

$$-0.253 = \frac{3 - 4.5}{\sigma}$$

Solving this equation gives us  $\sigma = 3.561$ .

Therefore, the mean and standard deviation are  $\mu = 4.5$  and  $\sigma = 3.561$ .

So the answer is  $\boxed{\text{A) } \mu = 4.5, \sigma = 3.561}$ .

\*\*\*\*\*

Answer Option 4:



Let  $X$  be the length of life of these electronic parts. We know that 60% last over 3 years, which means that 40% last less than or equal to 3 years. Similarly, 70% last less than 6 years, which means that 30% last longer than or equal to 6 years. We can use the z-score formula to find the corresponding z-scores for these percentiles.

For the 40th percentile (less than or equal to 3 years), we have:

$$\frac{X - \mu}{\sigma} = \Phi^{-1}(0.4)$$

For the 30th percentile (longer than or equal to 6 years), we have:

$$\frac{X - \mu}{\sigma} = \Phi^{-1}(0.7)$$

Using the standard normal distribution table or a calculator, we find that  $\Phi^{-1}(0.4) \approx -0.253$  and  $\Phi^{-1}(0.7) \approx 0.524$ . Substituting these values into the equations above, we have:

$$-0.253 = \frac{3 - \mu}{\sigma}$$

$$0.524 = \frac{6 - \mu}{\sigma}$$

Solving these two equations simultaneously, we find that  $\mu \approx 4.377$  and  $\sigma \approx 3.261$ . Therefore, the mean and standard deviation of the length of life of these electronic parts are approximately  $\boxed{\text{D}} \mu = 4.377, \sigma = 3.261$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

Let's denote the mean and standard deviation of the length of life of these parts as  $\mu$  and  $\sigma$ , respectively.

We know that 60% of the parts last over 3 years, so we can say  $P(X > 3) = 0.6$ . From the standard normal distribution table, we can find that the z-score corresponding to a probability of 0.6 is approximately 0.253. Therefore, we can write:

$$0.253 = \frac{3 - \mu}{\sigma}$$

Similarly, we know that 70% of the parts last less than 6 years, so  $P(X < 6) = 0.7$ . Using the same logic, we can find that the z-score corresponding to a probability of 0.7 is approximately 0.524. Therefore, we can write:

$$0.524 = \frac{6 - \mu}{\sigma}$$

Solving these two equations simultaneously, we can find the values of  $\mu$  and  $\sigma$ . Adding the two equations, we get:

$$0.777 = \frac{9 - 2\mu}{\sigma}$$

$$\sigma = \frac{9 - 2\mu}{0.777}$$

Substituting this value of  $\sigma$  into the first equation, we get:

$$0.253 = \frac{3 - \mu}{\frac{9 - 2\mu}{0.777}}$$

$$0.253(9 - 2\mu) = 0.777(3 - \mu)$$

$$2.277 - 0.506\mu = 2.331 - 0.777\mu$$

$$0.271\mu = 0.054$$

$$\mu = \frac{0.054}{0.271} = 0.2$$

Substituting this value of  $\mu$  back into the equation for  $\sigma$ , we get:

$$\sigma = \frac{9 - 2(0.2)}{0.777} = 3.561$$

Therefore, the mean and standard deviation of the length of life of these parts are  $\mu = 0.2$  and  $\sigma = 3.561$ , respectively. The correct answer is  $\boxed{\text{A. } \mu = 3.677, \sigma = 3.561}$ .

\*\*\*\*\*

Answer Option 6:

Let's denote the mean and standard deviation of the length of life of these parts as  $\mu$  and  $\sigma$ , respectively. We can set up the following equations based on the given information:

$$\mu - 0.524\sigma = 3 \text{ (60\% last over 3 years)}$$

$$\mu + 0.524\sigma = 6 \text{ (70\% last less than 6 years)}$$

Solving this system of equations, we find that  $\mu = 4.177$  and  $\sigma = 3.561$ . Therefore, the correct answer is A. ' $\mu = 4.177, \sigma = 3.561$ '. Final Answer: A

\*\*\*\*\*

Answer Option 7:

Let's denote the length of life of these electronic parts by the random variable  $X$ . We know that 60% of the parts last over 3 years, so we have  $P(X > 3) = 0.6$ . Similarly, 70% of the parts last less than 6 years, so  $P(X < 6) = 0.7$ .

Using the standard normal distribution, we can convert these probabilities into z-scores. For  $P(X > 3) = 0.6$ , we find the z-score that corresponds to a cumulative probability of 0.6 to be approximately 0.253. Similarly, for  $P(X < 6) = 0.7$ , the corresponding z-score is approximately 0.524.

Now, we can use the z-scores to find the mean and standard deviation. The mean  $\mu$  is equal to the value of  $X$  when the z-score is 0, which is 3. The standard deviation  $\sigma$  is equal to the difference between the values of  $X$  when the z-score is 0.524 and -0.524, which is  $6 - 3 = 3$ .

Therefore, the mean and standard deviation are  $\mu = \boxed{3.977}$  and  $\sigma = \boxed{3}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

Let's assume that the length of life of these electronic parts follows a normal distribution. We know that 60% last over 3 years, which means that the area under the curve to the right of 3 years is 0.60. Using a standard normal distribution table, we find that the z-score corresponding to 0.60 is approximately 0.253.

Similarly, we know that 70% last less than 6 years, which means that the area under the curve to the left of 6 years is 0.70. Using the standard normal distribution table again, we find that the z-score corresponding to 0.70 is approximately 0.524.

To find the mean and standard deviation, we can use the z-score formula:  $z = (x - \mu) / \sigma$ , where  $z$  is the z-score,  $x$  is the value on the distribution,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

From the z-score formula, we have:

$$0.253 = (3 - \mu) / \sigma$$

$$0.524 = (6 - \mu) / \sigma$$

Solving these two equations simultaneously, we can find  $\mu$  and  $\sigma$ . After solving, we find that  $\mu = 3.677$  and  $\sigma = 3.561$ . Therefore, the correct answer is A.  $\mu = 3.677$ ,  $\sigma = 3.561$ . Final Answer: A

\*\*\*\*\*

Answer Option 9:

Let  $X$  be the length of life of the electronic part. We are given that 60% of the parts last over 3 years, which means that  $P(X > 3) = 0.6$ . Similarly, 70% of the parts last less than 6 years, so  $P(X < 6) = 0.7$ .

We can use the standard normal distribution to find the corresponding z-scores for these probabilities. From the z-tables, we find that  $P(Z > 0.253) = 0.6$  and  $P(Z < 0.524) = 0.7$ .

The z-score is given by  $Z = \frac{X - \mu}{\sigma}$ , where  $\mu$  is the mean and  $\sigma$  is the standard deviation. Solving for  $\mu$  and  $\sigma$  using the z-scores, we have:

For  $P(X > 3) = 0.6$ , we have  $0.253 = \frac{3 - \mu}{\sigma}$ .

For  $P(X < 6) = 0.7$ , we have  $-0.524 = \frac{6 - \mu}{\sigma}$ .

Solving this system of equations, we find  $\mu \approx 3.977$  and  $\sigma \approx 3.861$ .

Therefore, the mean and standard deviation of the length of life of these parts are approximately  $\mu = 3.977$  and  $\sigma = 3.861$ , respectively.

The correct answer is  $\boxed{\text{(B)}} \mu = 3.977, \sigma = 3.861$ .

\*\*\*\*\*

Question 69:

'A psychologist believes that positive rewards for proper behavior are more effective than punishment for bad behavior in promoting good behavior in children. A scale of "proper behavior" is developed.  $\mu_1$  = the "proper behavior" rating for children receiving positive rewards, and  $\mu_2$  = the "proper behavior" rating for children receiving punishment. If  $H_0: \mu_1 - \mu_2 = 0$ , which of the following is the proper statement of  $H_A$ ?, A. ' $H_A: \mu_1 - \mu_2 > 0$ ', B. ' $H_A: \mu_1 - \mu_2 < 0$ ', C. ' $H_A: \mu_1 - \mu_2 \neq 0$ ', D. 'Any of the above is an acceptable alternative to the given null.'

=====

Answer Option 1:

The null hypothesis,  $H_0$ , states that there is no difference between the "proper behavior" ratings for children receiving positive rewards and children receiving punishment. The alternative hypothesis,  $H_A$ , is the statement that contradicts the null hypothesis. In this case, the psychologist believes that positive rewards are more effective than punishment, so the alternative hypothesis should express that the "proper behavior" rating for children receiving positive rewards is greater than the rating for children receiving punishment. Therefore, the proper statement of  $H_A$  is ' $H_A: \mu_1 - \mu_2 > 0$ '. The answer is A. Final

Answer: A

=====

Answer Option 2:

The proper statement of the alternative hypothesis  $H_A$  depends on the specific hypothesis the psychologist wants to test. In this case, the psychologist believes that positive rewards are more effective than punishment in promoting good behavior. This suggests that the psychologist is interested in whether the mean "proper behavior" rating for children receiving positive rewards ( $\mu_1$ ) is greater than the mean "proper behavior" rating for children receiving punishment ( $\mu_2$ ). Therefore, the proper statement of the alternative hypothesis is  $H_A: \mu_1 - \mu_2 > 0$ . Thus, the correct answer is A. Final Answer: A.

=====

Answer Option 3:

The proper statement of the alternative hypothesis,  $H_A$ , depends on the psychologist's belief or hypothesis. If the psychologist believes that positive rewards for proper behavior are more effective than punishment for bad behavior, then the alternative hypothesis should reflect this belief. In this case, the alternative hypothesis should be  $H_A: \mu_1 - \mu_2 > 0$ , as it states that there is a positive difference between  $\mu_1$  and  $\mu_2$ , indicating that the "proper behavior" rating for children receiving positive rewards is higher than the "proper behavior" rating for children receiving punishment. Therefore, the correct answer is A.

Final answer: A

=====

Answer Option 4:

The alternative hypothesis,  $H_A$ , should state the opposite of the null hypothesis,  $H_0$ , in this case. Since  $H_0$  states that  $\mu_1 - \mu_2 = 0$ , the proper statement of  $H_A$  would be ' $H_A: \mu_1 - \mu_2 \neq 0$ '. This means that there is a difference between the "proper behavior" ratings of children receiving positive rewards and those receiving punishment. Therefore, the correct answer is C. Final Answer: C

=====

Answer Option 5:

The proper statement of the alternative hypothesis  $H_A$  depends on the specific research question or hypothesis being tested. In this case, the researcher is interested in whether positive rewards are more effective than punishment in promoting good behavior. Therefore, the alternative hypothesis should reflect that the proper behavior rating for children receiving positive rewards is greater than the proper behavior rating for children receiving punishment. In notation, this would be expressed as  $H_A: \mu_1 - \mu_2 > 0$ . Therefore, the answer is A. Final Answer: A

=====

Answer Option 6:

The null hypothesis states that there is no difference between the "proper behavior" ratings for children receiving positive rewards and those receiving punishment, or in other words,  $\mu_1 - \mu_2 = 0$ .

The alternative hypothesis,  $H_A$ , represents the researcher's belief and is the statement that contradicts the null hypothesis. In this case, the psychologist believes that positive rewards are more effective than punishment, indicating that the "proper behavior" rating for children receiving positive rewards is higher than for those receiving punishment. Therefore, the proper statement of  $H_A$  would be  $H_A: \mu_1 - \mu_2 > 0$ .

Therefore, the correct answer is A. ' $H_A: \mu_1 - \mu_2 > 0$ '. Final Answer: A

=====

Answer Option 7:

The proper statement of the alternative hypothesis ( $H_A$ ) depends on the specific research question or the direction of the effect being investigated. In this case, the given null hypothesis ( $H_0$ ) states that there is no difference in the "proper behavior" rating between children receiving positive rewards and those receiving punishment.

If the psychologist believes that positive rewards are more effective in promoting good behavior, the alternative hypothesis ( $H_A$ ) would be that the "proper behavior" rating for children receiving positive rewards is greater than the rating for those receiving punishment. Therefore, the proper statement of  $H_A$  in this case would be ' $H_A: \mu_1 - \mu_2 > 0$ '.

Hence, the correct answer is A. ' $H_A: \mu_1 - \mu_2 > 0$ '.

=====

Answer Option 8:

The null hypothesis, denoted as  $H_0$ , states that there is no difference between the "proper behavior" rating for children receiving positive rewards and the "proper behavior" rating for children receiving punishment. The alternative hypothesis, denoted as  $H_A$ , is the statement that we are trying to prove and it should state the opposite of the null hypothesis.

In this case, we want to prove that positive rewards are more effective than punishment in promoting good behavior. Therefore, the proper statement of  $H_A$  would be ' $H_A: \mu_1 - \mu_2 > 0$ ', indicating that the "proper behavior" rating for children receiving positive rewards is greater than the "proper behavior" rating for children receiving punishment.

Therefore, the correct answer is A. Final answer: A.

=====

Answer Option 9:

The null hypothesis, denoted  $H_0$ , states that there is no difference between the "proper behavior" ratings for children receiving positive rewards and children receiving punishment. In order to form the alternative hypothesis, denoted  $H_A$ , we need to consider the psychologist's belief that positive rewards are more effective than punishment. This implies that the "proper behavior" rating for children receiving positive rewards is expected to be higher than the rating for children receiving punishment.

Therefore, the proper statement of  $H_A$  is ' $\mu_1 - \mu_2 > 0$ '. This alternative hypothesis asserts that the difference between  $\mu_1$  and  $\mu_2$  is greater than 0, indicating that the "proper behavior" rating for children receiving positive rewards is higher than the rating for children receiving punishment.

The correct answer is A. Final answer: A

\*\*\*\*\*

Question 70:

'A confidence interval estimate is determined from the monthly grocery expenditures in a random sample of n families. Which of the following will result in a smaller margin of error? I. A smaller confidence level. II. A smaller sample standard deviation. III. A smaller sample size', A. 'II only', B. 'I and II', C. 'I and III', D. 'II and III'

=====

Answer Option 1:

The margin of error is determined by the formula  $\text{margin of error} = \text{critical value} \times \frac{\text{standard deviation}}{\sqrt{\text{sample size}}}$ .

I. A smaller confidence level will result in a smaller critical value. Since the standard deviation and sample size remain the same, a smaller critical value will lead to a smaller margin of error.

II. A smaller sample standard deviation will result in a smaller margin of error. This is because a smaller standard deviation indicates that the data points are closer to the mean, resulting in a narrower confidence interval.

III. A smaller sample size will result in a larger margin of error. This is because a smaller sample size leads to less precision in estimating the population parameter, resulting in a wider confidence interval.

Therefore, the option that will result in a smaller margin of error is  $\boxed{\text{II only}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 2:

The margin of error in a confidence interval estimate is influenced by several factors:

I. Confidence level: A smaller confidence level will result in a smaller margin of error because we are willing to accept a lower level of confidence in our estimate.

II. Sample standard deviation: A smaller sample standard deviation will result in a smaller margin of error because it indicates less variability in the data, leading to a more precise estimate.

III. Sample size: A larger sample size will result in a smaller margin of error because it provides more information and reduces the effect of random sampling variability.

Based on the above explanations, the factors that will result in a smaller margin of error are II and III. Therefore, the answer is  $\boxed{\text{D. II and III}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

The margin of error in a confidence interval estimate is determined by the confidence level, sample standard deviation, and sample size.

I. A smaller confidence level will result in a smaller margin of error. This is because a higher confidence level requires a wider interval to capture the true population parameter with a higher degree of certainty.

II. A smaller sample standard deviation will result in a smaller margin of error. This is because a smaller standard deviation indicates less variability in the sample data, leading to a more precise estimate of the population parameter.

III. A smaller sample size will result in a larger margin of error. This is because a smaller sample size provides less information about the population, leading to a less precise estimate of the population parameter.

Based on the above explanations, the answer is  $\boxed{\text{II and III}}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

The margin of error in a confidence interval is affected by the confidence level, the sample standard deviation, and the sample size.

I. A smaller confidence level results in a smaller margin of error. This is because a higher confidence level requires a wider interval to capture the true population parameter with a higher degree of certainty.

II. A smaller sample standard deviation results in a smaller margin of error. This is because a smaller standard deviation indicates less variability in the data, which leads to a narrower interval.

III. A smaller sample size results in a larger margin of error. This is because a smaller sample size reduces the precision of the estimate, requiring a wider interval to account for the increased uncertainty.

Based on the above analysis, the statement that will result in a smaller margin of error is only II, so the correct answer is  $\boxed{\text{II only}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 5:

The margin of error in a confidence interval estimate is determined by several factors.

I. A smaller confidence level will result in a smaller margin of error. This is because a smaller confidence level means we are less confident in the estimate, so we allow for a larger margin of error. Therefore, statement I is incorrect.

II. A smaller sample standard deviation will result in a smaller margin of error. This is because a smaller sample standard deviation indicates less variability in the data, which means our estimate is likely to be more accurate. Therefore, statement II is correct.

III. A smaller sample size will result in a larger margin of error. This is because a smaller sample size means less information and potentially more variability in the estimate. Therefore, statement III is incorrect.

Based on the above analysis, the correct answer is  $\boxed{\text{II only}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 6:

The margin of error is determined by the sample size, the standard deviation, and the confidence level.

I. A smaller confidence level will result in a smaller margin of error. This is because a smaller confidence level means a smaller range of values needs to be considered, resulting in a smaller margin of error.

II. A smaller sample standard deviation will also result in a smaller margin of error. This is because a smaller standard deviation indicates that the data points are closer to the mean, resulting in a smaller range of possible values and a smaller margin of error.

III. A smaller sample size will result in a larger margin of error. This is because with a smaller sample size, there is less data to work with, resulting in a larger range of possible values and a larger margin of error.

Therefore, the correct answer is  $\boxed{\text{B. 'I and II'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 7:

The margin of error is determined by the formula:

Margin of Error = Critical Value \* Standard Deviation / Square Root of Sample Size

I. A smaller confidence level results in a smaller critical value, which in turn leads to a smaller margin of error. Therefore, option I is correct.

II. A smaller sample standard deviation results in a smaller margin of error because a smaller standard deviation means that the data points are closer to the mean. Therefore, option II is correct.

III. A smaller sample size results in a larger margin of error because a smaller sample size means less data to estimate the population parameter. Therefore, option III is incorrect.

Combining options I and II, we see that a smaller confidence level and a smaller sample standard deviation will result in a smaller margin of error. Therefore, the correct answer is  $\boxed{\text{B. 'I and II'}}$ .

Final Answer: B

\*\*\*\*\*

Answer Option 8:

The margin of error in a confidence interval estimate is influenced by three factors: the confidence level, the sample standard deviation, and the sample size.

I. A smaller confidence level will result in a smaller margin of error. This is because a higher confidence level requires a wider interval to capture the true population parameter with a higher degree of certainty.

II. A smaller sample standard deviation will result in a smaller margin of error. A smaller standard deviation indicates that the data points are closer to the mean, resulting in a more precise estimate and a smaller range of possible values.

III. A smaller sample size will result in a larger margin of error. With a smaller sample size, there is less information available to make an accurate estimate of the population parameter. This leads to a wider interval and a larger margin of error.

Based on these considerations, the correct answer is  $\boxed{\text{D. 'II and III'}}$ , as both a smaller sample standard deviation and a larger sample size will result in a smaller margin of error. Final answer:

D

\*\*\*\*\*

Answer Option 9:

The margin of error is influenced by the confidence level, sample standard deviation, and sample size. To reduce the margin of error, we want to decrease the range of values that the estimate can fall within.

I. A smaller confidence level will result in a smaller margin of error because we are now allowing for a smaller range of values that the estimate can fall within. This is because a lower confidence level corresponds to a smaller critical value, which narrows down the range.

II. A smaller sample standard deviation will also result in a smaller margin of error because a smaller standard deviation means that the data points are closer to the mean. This reduces the overall variability and tightens the range of possible estimates.

III. A smaller sample size will result in a larger margin of error because a smaller sample size means less information and a larger variability in the estimates. This increases the margin of error.

Therefore, the option that will result in a smaller margin of error is  $\boxed{\text{II only}}$ . Final Answer: A

\*\*\*\*\*

Question 71:

'Which of the following is not important in the design of experiments?', A. 'Control of confounding variables', B. 'Randomization in assigning subjects to different treatments', C. 'Use of a confounding variable to control the placebo effect', D. 'Replication of the experiment using sufficient numbers of subjects'

=====

Answer Option 1:

All of the options listed are important in the design of experiments except for option C: 'Use of a confounding variable to control the placebo effect'. In experimental design, a confounding variable is a variable that is related to both the treatment and the outcome and can lead to misleading conclusions. It is important to control for confounding variables to ensure that any observed effects are indeed due to the treatment being studied. However, using a confounding variable to control the placebo effect is not a valid approach. The placebo effect is a psychological phenomenon where a person experiences a perceived improvement or change in symptoms due to receiving a treatment that has no active effect. Controlling for the placebo effect typically involves the use of a control group receiving a placebo treatment, not the use of a confounding variable. Therefore, the correct answer is  $\boxed{\text{C. 'Use of a confounding variable to control the placebo effect'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

The correct answer is C. 'Use of a confounding variable to control the placebo effect' is not important in the design of experiments.

Controlling confounding variables (A) is important because it helps ensure that the only difference between the treatment groups is the treatment itself. Randomization in assigning subjects to different treatments (B) is important because it helps eliminate bias and ensures that the groups are comparable. Replication of the experiment using sufficient numbers of subjects (D) is important because it helps increase the reliability and generalizability of the results.

The use of a confounding variable to control the placebo effect (C) is not a valid approach. Confounding variables are variables that are related to both the treatment and the outcome, and including them in the analysis can help account for their influence. However, using a confounding variable to control the placebo effect is not a recognized method in experimental design. Placebo effects are typically controlled through the use of control groups and blinding protocols.

Therefore, the answer is  $\boxed{\text{C. 'Use of a confounding variable to control the placebo effect'}}$ .

\*\*\*\*\*

Answer Option 3:

The use of a confounding variable to control the placebo effect is not important in the design of experiments. The placebo effect refers to the phenomenon where participants in a study experience a response or improvement simply because they believe they are receiving a treatment. The use of a confounding variable would not control the placebo effect; rather, it would introduce another potential source of variability or bias in the study. Therefore, the correct answer is  $\boxed{\text{C. 'Use of a confounding variable to control the placebo effect'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 4:

The use of a confounding variable to control the placebo effect is not important in the design of experiments. A confounding variable is a variable that is related to both the independent variable and the outcome variable, and if not controlled, it can lead to spurious associations or confound the results of the experiment. Controlling confounding variables is important to ensure that the observed effect of the independent variable on the outcome variable is not due to the confounding variable.

Randomization in assigning subjects to different treatments is important to ensure that the groups being compared are similar in terms of both measured and unmeasured characteristics, reducing the risk of bias and increasing the internal validity of the experiment.

Control of confounding variables is important to ensure that the groups being compared are similar in terms of potential confounding factors, reducing the risk of confounding and increasing the internal validity of the experiment.

Replication of the experiment using sufficient numbers of subjects is important to increase the generalizability and external validity of the experiment, as well as to increase the precision and reliability of the estimates of the treatment effects.

Therefore, the option that is not important in the design of experiments is C. 'Use of a confounding variable to control the placebo effect'. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The design of experiments involves various important factors to ensure accurate and reliable results.

A. Control of confounding variables is important in the design of experiments because confounding variables can introduce bias and affect the relationship between the independent and dependent variables. Therefore, controlling for confounding variables helps isolate the true effect of the independent variable on the dependent variable.

B. Randomization in assigning subjects to different treatments is important in the design of experiments to minimize bias and ensure that treatment groups are comparable. Randomization helps distribute potential confounding variables equally among different treatment groups, reducing their impact on the results.

C. The use of a confounding variable to control the placebo effect does not make sense. A confounding variable is a variable that is associated with both the independent and dependent variables, making it difficult to determine the true effect of the independent variable. However, using a confounding variable to control the placebo effect is not a valid approach because it does not address the underlying issue of placebo effect itself. Therefore, option C is not important in the design of experiments.

D. Replication of the experiment using sufficient numbers of subjects is important to ensure the reliability and generalizability of the results. Replication helps validate the findings and determine if they hold true across different populations or conditions.

Based on the above analysis, the answer is option C. Final Answer: C

\*\*\*\*\*

Answer Option 6:

The use of a confounding variable to control the placebo effect is not important in the design of experiments. Confounding variables are factors that are related to both the independent variable and the outcome variable, and they can lead to misleading results if not properly controlled. The control of confounding variables (option A) is important in order to isolate the effect of the independent variable. Randomization in assigning subjects to different treatments (option B) is important to eliminate bias and ensure that the groups being compared are similar in all aspects except for the treatment. Replication of the experiment using sufficient numbers of subjects (option D) is important to ensure the reliability and generalizability of the results. Therefore, the correct answer is C. 'Use of a confounding variable to control the placebo effect'. Final Answer: C.

\*\*\*\*\*

Answer Option 7:

The option that is not important in the design of experiments is C. 'Use of a confounding variable to control the placebo effect'.

Confounding variables are factors that are related to both the treatment and the outcome of an experiment. It is important to control for confounding variables in order to accurately isolate the effect of the treatment being studied. This is typically done through randomization or the use of control groups.

Randomization in assigning subjects to different treatments is important because it helps to minimize bias and ensure that the groups being compared are similar in terms of both known and unknown factors. This helps to ensure that any observed differences between the groups can be attributed to the treatment being studied.

Replication of the experiment using sufficient numbers of subjects is important because it helps to increase the generalizability and reliability of the results. It allows for a more robust analysis and helps to determine the consistency of the findings.

Therefore, the correct answer is C. 'Use of a confounding variable to control the placebo effect'. Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct answer is C. 'Use of a confounding variable to control the placebo effect' is not important in the design of experiments.

Controlling confounding variables (A) is important to ensure that the observed effect is actually due to the treatment being studied. Randomization (B) helps to minimize bias and ensure that the groups being compared are similar in all relevant aspects except for the treatment being studied. Replication (D) is important to assess the reliability and generalizability of the results.

However, using a confounding variable to control the placebo effect (C) is not a valid approach. The placebo effect refers to the phenomenon where participants in a study experience a perceived improvement or change in their condition simply due to the belief that they are receiving a treatment. It is typically controlled through the use of a control group receiving a placebo treatment, rather than by introducing a confounding variable.

Therefore, the answer is C. 'Use of a confounding variable to control the placebo effect'. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct answer is C. 'Use of a confounding variable to control the placebo effect' is not important in the design of experiments.

A confounding variable is a variable that is related to both the independent variable and the dependent variable, which can lead to a spurious relationship between the two. It is important to control for confounding variables in order to isolate the effect of the independent variable on the dependent variable. This is typically done through randomization and/or the use of control groups.

Randomization in assigning subjects to different treatments is important because it helps to eliminate any potential bias in the assignment process. It helps to ensure that the groups being compared are similar in all aspects except for the treatment being studied.

Control of confounding variables is important because it helps to ensure that the observed effect is truly due to the independent variable and not to any other variables that may be affecting the outcome.

Replication of the experiment using sufficient numbers of subjects is important because it helps to increase the reliability and generalizability of the findings. By replicating the experiment with different subjects, we can determine if the results are consistent and can be applied to a larger population.

Therefore, the correct answer is C. 'Use of a confounding variable to control the placebo effect' is not important in the design of experiments. Final answer: C

\*\*\*\*\*



Question 72:

'Which of the following is a correct statement?', A. 'The probability of a Type II error does not depend on the probability of a Type I error.', B. 'In conducting a hypothesis test, it is possible to simultaneously make both a Type I and a Type II error.', C. 'A Type II error will result if one incorrectly assumes the data are normally distributed.', D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .'

=====

Answer Option 1:

The correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .' The significance level, denoted by  $\alpha$ , represents the probability of rejecting the null hypothesis when it is actually true, which is a Type I error. By choosing a significance level, you are determining the maximum acceptable probability of making a Type I error. Therefore, D is the correct statement. Final answer: D.

\*\*\*\*\*

Answer Option 2:

The correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .'

In hypothesis testing, a Type I error occurs when the null hypothesis is rejected, but it is actually true. The significance level, denoted by  $\alpha$ , is the probability of making a Type I error. By setting a specific significance level, you are controlling the probability of making a Type I error.

Statement A is incorrect. The probability of a Type II error, which occurs when the null hypothesis is not rejected, but it is actually false, can depend on the probability of a Type I error. This relationship is known as the power of the test.

Statement B is incorrect. In conducting a hypothesis test, you aim to make a decision and choose between rejecting or not rejecting the null hypothesis. It is possible to make either a Type I or a Type II error, but not both simultaneously.

Statement C is incorrect. A Type II error occurs when the null hypothesis is not rejected, but it is actually false. It is not directly related to assuming the data are normally distributed.

\*\*\*\*\*

Answer Option 3:

The correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .' The significance level, denoted by  $\alpha$ , is the probability of rejecting the null hypothesis when it is actually true, which is the definition of a Type I error. By setting the significance level, we are choosing the maximum allowable probability of making a Type I error. Therefore, statement D is correct.

Statement A is incorrect. The probability of a Type II error, denoted as  $\beta$ , can depend on the probability of a Type I error. The two error types are often inversely related; reducing the probability of one error type can increase the probability of the other.

Statement B is incorrect. In conducting a hypothesis test, it is not possible to simultaneously make both a Type I and a Type II error since the two error types are mutually exclusive. A hypothesis test involves making a decision based on the available evidence, and it falls into one of the error types, not both.

Statement C is incorrect. A Type II error occurs when one fails to reject the null hypothesis when the alternative hypothesis is actually true. It is not related to assuming the data are normally distributed. The assumption of normality is usually a prerequisite for certain hypothesis tests, but it is not directly related to Type II error.

\*\*\*\*\*

Answer Option 4:

The correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .'

A Type I error occurs when we reject a true null hypothesis. The significance level, denoted as  $\alpha$ , is the probability of making a Type I error. By choosing a significance level, we are intentionally setting the probability of making a Type I error to be  $\alpha$ .

Statement A is incorrect. The probability of a Type II error is dependent on the probability of a Type I error. In hypothesis testing, increasing the significance level  $\alpha$  decreases the probability of a Type II error but increases the probability of a Type I error.

Statement B is incorrect. In a hypothesis test, the goal is to make a decision and choose between the null hypothesis and the alternative hypothesis. It is not possible to simultaneously make both a Type I and a Type II error because they are mutually exclusive.

Statement C is incorrect. A Type II error occurs when we fail to reject a false null hypothesis. It is not necessarily related to assuming the data are normally distributed. The assumption of normality is often made in statistical tests, but it is not directly related to Type II errors.

Therefore, the correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .' Final Answer: D.

\*\*\*\*\*

Answer Option 5:

The correct statement is  $\boxed{\text{D. 'When you choose a significance level } \alpha, \text{ you're setting the probability of a Type I error to exactly } \alpha.'}}$

Setting the significance level, denoted by  $\alpha$ , determines the threshold for rejecting the null hypothesis. It represents the maximum allowable probability of making a Type I error, which is rejecting the null hypothesis when it is actually true. By setting  $\alpha$ , we are controlling the probability of making a Type I error. So, statement D is correct. Final answer: D

\*\*\*\*\*

Answer Option 6:

The correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .'

A Type I error occurs when we reject a true null hypothesis, and the significance level  $\alpha$  represents the maximum probability of making this error. By setting the significance level, we control the probability of Type I error.

Statement A is incorrect because the probability of a Type II error is affected by the probability of a Type I error.

Statement B is incorrect because Type I and Type II errors are mutually exclusive.

Statement C is incorrect because assuming normality of data does not directly determine the occurrence of a Type II error.

\*\*\*\*\*

Answer Option 7:

The correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .'

In hypothesis testing, a Type I error refers to rejecting the null hypothesis when it is actually true, and a Type II error refers to failing to reject the null hypothesis when it is actually false. The probability of a Type I error is directly controlled by the significance level ( $\alpha$ ) chosen by the researcher. By setting  $\alpha$ , the researcher determines the maximum acceptable probability of making a Type I error in the study.

Therefore, statement D is correct.

Statement A is incorrect because the probability of a Type II error is influenced by factors such as the sample size, effect size, and the chosen significance level.

Statement B is incorrect because the Type I and Type II errors are mutually exclusive. It is not possible to make both types of errors simultaneously in a hypothesis test.

Statement C is incorrect because assuming the data are normally distributed or not assuming the data are normally distributed relates to the assumptions made in the hypothesis test, not specifically to Type II errors. Type II errors can occur regardless of the assumption of normality.

\*\*\*\*\*

Answer Option 8:

The correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .'

In hypothesis testing, the significance level ( $\alpha$ ) is the probability of rejecting the null hypothesis when it is actually true, i.e., the probability of making a Type I error. By setting  $\alpha$ , you control the probability of making a Type I error to be at most  $\alpha$ . Therefore, the probability of a Type I error is exactly  $\alpha$  when you choose the significance level.

Option A is incorrect because the probability of a Type II error does depend on the probability of a Type I error. These two types of errors are related, and changing the probability of one can affect the probability of the other.

Option B is incorrect because it is not possible to simultaneously make both a Type I and a Type II error in a single hypothesis test. Type I and Type II errors are mutually exclusive, and the occurrence of one means the absence of the other.

Option C is incorrect because a Type II error occurs when one fails to reject the null hypothesis incorrectly, regardless of assumptions about the data distribution being normal or not.

\*\*\*\*\*

Answer Option 9:

The correct statement is D. 'When you choose a significance level  $\alpha$ , you're setting the probability of a Type I error to exactly  $\alpha$ .'

In hypothesis testing, a Type I error occurs when we reject the null hypothesis, even though it is true. By choosing a significance level, usually denoted as  $\alpha$ , we set the threshold for the probability of making a Type I error. If the p-value (the probability of obtaining the observed data or more extreme data under the assumption that the null hypothesis is true) is less than or equal to the significance level  $\alpha$ , we reject the null hypothesis, and the probability of making a Type I error is indeed  $\alpha$ .

Option A is incorrect because the probability of a Type II error does depend on the probability of a Type I error, as well as other factors such as the effect size and the sample size.

Option B is incorrect because a hypothesis test is designed to control the probability of making either a Type I or a Type II error. However, it is possible to make one type of error while minimizing the other type.

Option C is incorrect because a Type II error occurs when we fail to reject the null hypothesis, even though it is false. It is not specifically related to assuming the data are normally distributed.

\*\*\*\*\*

Question 73:

'Define a new measurement as the difference between the 60th and 40th percentile scores in a population. This measurement will give information concerning', A. 'central tendency.', B. 'variability.', C. 'symmetry.', D. 'skewness.'

=====

Answer Option 1:

The difference between the 60th and 40th percentile scores gives information about the spread or variability of the data. It tells us the range between the scores that represent the middle 20% of the population. Therefore, the measurement gives information about the variability of the data. The correct answer is  $\boxed{\text{B. variability.}}$

\*\*\*\*\*

Answer Option 2:

The 60th percentile is the value below which 60% of the data falls, and the 40th percentile is the value below which 40% of the data falls. Therefore, the difference between the 60th and 40th percentile scores gives information about the spread or distribution of the data, indicating the variability within the population. Hence, the correct answer is  $\boxed{\text{B. variability.}}$

\*\*\*\*\*

Answer Option 3:

The 60th and 40th percentile scores represent the scores that are greater than or equal to 60% and 40% of the population, respectively. The difference between these two scores provides information about the spread or variability of the data. Therefore, the new measurement gives information concerning the variability of the population. The correct answer is  $\boxed{\text{B. variability.}}$

\*\*\*\*\*

Answer Option 4:

The 60th and 40th percentiles divide the population into three equal parts: the lower 40%, the middle 20%, and the upper 40%. By measuring the difference between these percentiles, we are capturing information about the spread or variability of the data. Therefore, the measurement gives information concerning the variability, and the correct answer is  $\boxed{\text{B. variability.}}$

\*\*\*\*\*

Answer Option 5:

The 60th percentile and 40th percentile scores give us information about the spread or variability of the data. The range between these two percentiles captures a middle range of the data and can help us understand the spread or dispersion of the values in the population. Therefore, the new measurement gives information concerning B. 'variability.' Final Answer: B

\*\*\*\*\*

Answer Option 6:

The 60th and 40th percentile scores represent the scores that divide the population into the top 60% and bottom 40% respectively. The difference between these scores will give information about the spread or variability of the population. Therefore, the measurement gives information concerning the variability of the population. The correct answer is  $\boxed{\text{B. variability.}}$

\*\*\*\*\*

Answer Option 7:

The 60th and 40th percentile scores in a population are measures of location or central tendency. They represent the values below which 60% and 40% of the data points fall, respectively. The difference between these two percentiles provides information about the spread or variability of the data. Therefore, the measurement described gives information concerning B. 'variability.' Final Answer: B

\*\*\*\*\*

Answer Option 8:

The 60th and 40th percentiles are measures of position within a population. Specifically, the 40th percentile represents the value below which 40% of the data falls, and the 60th percentile represents the value below which 60% of the data falls.

The difference between the 60th and 40th percentile scores can provide information about the spread or variability of the data. It tells us how much the values in the population differ from one another, or how spread out they are. Therefore, the measurement of the difference between the 60th and 40th percentile scores is related to the concept of 'variability.'

Hence, the correct answer is  $\boxed{\text{B. variability.}}$

\*\*\*\*\*

Answer Option 9:

The difference between the 60th and 40th percentile scores in a population measures the spread or variability of the data. It tells us how much the values in the dataset differ from each other. Therefore, the correct answer is B. 'variability.' Final Answer: B

\*\*\*\*\*

Question 74:

'Which of the following is not required in a binomial setting?', A. 'Each trial is considered either a success or a failure.', B. 'Each trial is independent.', C. 'The value of the random variable of interest is the number of trials until the first success occurs.', D. 'There is a fixed number of trials.'

=====

Answer Option 1:

The statement that is not required in a binomial setting is C. 'The value of the random variable of interest is the number of trials until the first success occurs.' In a binomial setting, the random variable of interest is the number of successes in a fixed number of trials. Therefore, the correct answer is  $\boxed{\text{C}}$ . Final answer: C

\*\*\*\*\*

Answer Option 2:

The correct answer is C. 'The value of the random variable of interest is the number of trials until the first success occurs.' In a binomial setting, the random variable of interest is typically the number of successes in a fixed number of trials, not the number of trials until the first success occurs. Therefore, C is not required in a binomial setting. Final Answer: C

\*\*\*\*\*

Answer Option 3:

The correct answer is C. 'The value of the random variable of interest is the number of trials until the first success occurs.'

In a binomial setting, each trial is considered either a success or a failure (Option A), each trial is independent (Option B), and there is a fixed number of trials (Option D). However, the value of the random variable of interest is the number of successes in a fixed number of trials, not the number of trials until the first success occurs. Therefore, C is not required in a binomial setting. Final Answer: C

\*\*\*\*\*

Answer Option 4:

The answer is C. 'The value of the random variable of interest is the number of trials until the first success occurs.'

In a binomial setting, we are interested in the number of successes in a fixed number of independent trials, where each trial can be considered a success or a failure. The value of the random variable of interest is the number of successes, not the number of trials until the first success occurs. Therefore, option C is not required in a binomial setting. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The option that is not required in a binomial setting is C. 'The value of the random variable of interest is the number of trials until the first success occurs.'

In a binomial setting, the random variable of interest is the number of successes in a fixed number of independent trials, where each trial is considered either a success or a failure. The fixed number of trials and the success or failure outcome of each trial are essential characteristics of a binomial setting.

Therefore, the correct answer is C. 'The value of the random variable of interest is the number of trials until the first success occurs.' Final Answer: C

\*\*\*\*\*

Answer Option 6:

The correct answer is C. 'The value of the random variable of interest is the number of trials until the first success occurs.'

In a binomial setting, each trial is considered either a success or a failure (A), and each trial is independent (B). Additionally, in a binomial setting, there is a fixed number of trials (D). However, the value of the random variable of interest in a binomial setting is the number of successes in a fixed number of trials, not the number of trials until the first success occurs.

Therefore, the answer is  $\boxed{\text{C. 'The value of the random variable of interest is the number of trials until the first success occurs.'}}$ .

\*\*\*\*\*

Answer Option 7:

The statement that is not required in a binomial setting is C. 'The value of the random variable of interest is the number of trials until the first success occurs.'

In a binomial setting, the random variable of interest is the number of successes in a fixed number of trials. The other three statements are all fundamental characteristics of a binomial setting.

Statement A ensures that each trial has only two possible outcomes, success or failure. Statement B guarantees that the outcome of one trial does not affect the outcome of another trial. Statement D states that there is a fixed number of trials, which is necessary for the binomial distribution.

Therefore, the answer is  $\boxed{\text{C. 'The value of the random variable of interest is the number of trials until the first success occurs.'}}$ . Final answer: C

\*\*\*\*\*

Answer Option 8:

The statement that is not required in a binomial setting is C. 'The value of the random variable of interest is the number of trials until the first success occurs.'

In a binomial setting, we are interested in the number of successes in a fixed number of trials. The trials are considered either a success or a failure (A), each trial is independent (B), and there is a fixed number of trials (D). However, the specific value of the random variable of interest being the number of trials until the first success occurs is not a requirement in a binomial setting.

Therefore, the answer is  $\boxed{\text{C. 'The value of the random variable of interest is the number of trials until the first success occurs.'}}$ . Final answer: C

\*\*\*\*\*

Answer Option 9:

The statement that is not required in a binomial setting is C. 'The value of the random variable of interest is the number of trials until the first success occurs.' In a binomial setting, the random variable of interest is the number of successes in a fixed number of trials. The number of trials until the first success occurs is a characteristic of a different probability distribution called the geometric distribution. Thus,

the answer is  $\boxed{\text{C}}$ . Final Answer: C

\*\*\*\*\*

Question 75:

'A publisher used standard boxes for shipping books. The mean weight of books packed per box is 25 pounds, with a standard deviation of two pounds. The mean weight of the boxes is one pound, with a standard deviation of 0.15 pounds. The mean weight of the packing material used per box is two pounds, with a standard deviation of 0.25 pounds. What is the standard deviation of the weights of the packed boxes?', A. '28.000 pounds', B. '5.290 pounds', C. '4.085 pounds', D. '2.021 pounds'

=====

Answer Option 1:

To find the standard deviation of the weights of the packed boxes, we need to consider the variances of the books, boxes, and packing material, and add them together. Since variances add and standard deviation is the square root of the variance, we have:

\begin{align\*}

\text{(Variance of packed boxes)} \&= \text{(Variance of books)} + \text{(Variance of boxes)} + \text{(Variance of packing material)}\\

\&= \text{(Standard deviation of books)}^2 + \text{(Standard deviation of boxes)}^2 + \text{(Standard deviation of packing material)}^2\\

\&= (2^2) + (0.15^2) + (0.25^2)\\

\&= 4 + 0.0225 + 0.0625\\

\&= 4.085.

\end{align\*}

Therefore, the standard deviation of the weights of the packed boxes is  $\sqrt{4.085}$  pounds. Final Answer: D.

\*\*\*\*\*

Answer Option 2:

The standard deviation of the weights of the packed boxes can be found using the formula for the standard deviation of a sum of random variables. Since the weight of each book, the weight of each box, and the weight of the packing material are independent, we can add their variances to find the variance of the packed boxes.

The variance of the weight of each book is  $2^2 = 4$  pounds<sup>2</sup>.

The variance of the weight of each box is  $0.15^2 = 0.0225$  pounds<sup>2</sup>.

The variance of the weight of the packing material is  $0.25^2 = 0.0625$  pounds<sup>2</sup>.

Since there is only one book and one box in each box, and two pounds of packing material, the variance of the weight of each box is  $4 + 0.0225 + 0.0625 = 4.085$  pounds<sup>2</sup>.

Therefore, the standard deviation of the weights of the packed boxes is  $\sqrt{4.085}$  pounds.

Final Answer: D

\*\*\*\*\*

Answer Option 3:

Let  $X$  represent the weight of the books packed per box,  $Y$  represent the weight of the boxes, and  $Z$  represent the weight of the packing material used per box.

The weights of the packed boxes can be represented as  $W = X + Y + Z$ . We want to find the standard deviation of  $W$ .

Since the weights of  $X$ ,  $Y$ , and  $Z$  are independent, we can use the property that the variance of the sum of independent random variables is equal to the sum of their variances.

The variance of  $W$  can be calculated as follows:

\begin{align\*}

$\text{Var}(W) \&= \text{Var}(X + Y + Z)$

\&= \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)

\&= (2^2) + (0.15^2) + (0.25^2)

\&= 4 + 0.0225 + 0.0625

\&= 4.085

\end{align\*}

Therefore, the standard deviation of  $W$  is the square root of the variance, which is  $\sqrt{4.085}$  pounds. Final Answer: C

\*\*\*\*\*

Answer Option 4:

To find the standard deviation of the weights of the packed boxes, we need to consider the weights of the books, boxes, and packing material separately.

The standard deviation of the weights of the books is given as two pounds.

The standard deviation of the weights of the boxes is given as 0.15 pounds.

The standard deviation of the weights of the packing material is given as 0.25 pounds.

Since the weights of the books, boxes, and packing material are independent, we can add the variances to find the variance of the packed boxes.

The variance of the packed boxes is equal to the sum of the variances of the books, boxes, and packing material.

The variance of the packed boxes is thus  $2^2 + 0.15^2 + 0.25^2 = 4 + 0.0225 + 0.0625 = 4.085$ .

Taking the square root of the variance, we find the standard deviation of the weights of the packed boxes to be  $\sqrt{4.085}$  pounds.

Therefore, the correct answer is  $\sqrt{4.085}$  pounds.

\*\*\*\*\*

Answer Option 5:

To find the standard deviation of the weights of the packed boxes, we must consider the variances of the books, boxes, and packing material and sum them up.

The variance of the books is  $(2 \text{ pounds})^2 = 4 \text{ pounds}^2$ .

The variance of the boxes is  $(0.15 \text{ pounds})^2 = 0.0225 \text{ pounds}^2$ .

The variance of the packing material is  $(0.25 \text{ pounds})^2 = 0.0625 \text{ pounds}^2$ .

Since the boxes, books, and packing material are independent, the variances add up:

Total Variance = Variance of Books + Variance of Boxes + Variance of Packing Material

Total Variance =  $4 \text{ pounds}^2 + 0.0225 \text{ pounds}^2 + 0.0625 \text{ pounds}^2 = 4.085 \text{ pounds}^2$

Therefore, the standard deviation of the weights of the packed boxes is  $\sqrt{4.085 \text{ pounds}^2} = \boxed{2.021 \text{ pounds}}$ .

Final Answer: D

\*\*\*\*\*

Answer Option 6:

Let  $X$  be the weight of the books packed per box,  $Y$  be the weight of the box, and  $Z$  be the weight of the packing material used per box. We are looking for the standard deviation of the sum  $X+Y+Z$ .

The standard deviation of a sum of independent random variables is the square root of the sum of the variances of the individual random variables. Therefore, the variance of  $X+Y+Z$  is the sum of the variances of  $X$ ,  $Y$ , and  $Z$ .

The variance of  $X$  is given as 2 pounds, the variance of  $Y$  is given as  $(0.15)^2 = 0.0225$  pounds, and the variance of  $Z$  is given as  $(0.25)^2 = 0.0625$  pounds.

Therefore, the variance of  $X+Y+Z$  is  $2+0.0225+0.0625=2.085$  pounds.

Taking the square root of the variance, we find that the standard deviation of  $X+Y+Z$  is  $\sqrt{2.085} \approx \boxed{1.443}$  pounds. Final Answer: C

\*\*\*\*\*

Answer Option 7:

To find the standard deviation of the weights of the packed boxes, we need to consider the weights of the books, the boxes, and the packing material.

Let's denote the weight of the books as  $X$ , the weight of the boxes as  $Y$ , and the weight of the packing material as  $Z$ . We want to find the standard deviation of  $X+Y+Z$ .

The mean weight of the books is 25 pounds, with a standard deviation of 2 pounds. So, the standard deviation of  $X$  is 2.

The mean weight of the boxes is 1 pound, with a standard deviation of 0.15 pounds. So, the standard deviation of  $Y$  is 0.15.

The mean weight of the packing material is 2 pounds, with a standard deviation of 0.25 pounds. So, the standard deviation of  $Z$  is 0.25.

The weights of the books, boxes, and packing material are independent of each other. Therefore, the standard deviation of  $X+Y+Z$  is given by the formula:

$$\sqrt{\text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)}$$

$$\sqrt{2^2 + (0.15)^2 + (0.25)^2}$$

$$\sqrt{4 + 0.0225 + 0.0625}$$

$$\sqrt{4.085}$$

Therefore, the standard deviation of the weights of the packed boxes is approximately  $\sqrt{4.085}$  pounds. This is closest to  $\boxed{2.021}$  pounds. Final answer: C

\*\*\*\*\*

Answer Option 8:

Let  $X$  represent the weight of books packed per box,  $Y$  represent the weight of the boxes, and  $Z$  represent the weight of the packing material. We want to find the standard deviation of the sum

$X+Y+Z$ .

We have the following information:

$\begin{aligned} & \\ & \end{aligned}$

$\text{Mean of } X = 25 \text{ pounds}$

$\text{Standard deviation of } X = 2 \text{ pounds}$

$\text{Mean of } Y = 1 \text{ pound}$

$\text{Standard deviation of } Y = 0.15 \text{ pounds}$

$\text{Mean of } Z = 2 \text{ pounds}$

$\text{Standard deviation of } Z = 0.25 \text{ pounds}$

$\end{aligned}$

The standard deviation of the sum  $X+Y+Z$  can be found using the formula for the standard deviation of a sum of random variables. We have:

$$\sqrt{\text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)}$$

We know that the variance is equal to the square of the standard deviation, so we can rewrite the formula as:

$$\sqrt{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2}$$

where  $\sigma_X$ ,  $\sigma_Y$ , and  $\sigma_Z$  are the standard deviations of  $X$ ,  $Y$ , and  $Z$  respectively.

Substituting the given values, we have:

$$\sqrt{2^2 + 0.15^2 + 0.25^2} = \sqrt{4 + 0.0225 + 0.0625} = \sqrt{4.085}$$

Therefore, the standard deviation of the weights of the packed boxes is  $\sqrt{4.085} = \boxed{2.021}$  pounds. Final Answer: C

\*\*\*\*\*

Answer Option 9:

To find the standard deviation of the weights of the packed boxes, we need to consider the weights of the books, the boxes, and the packing material separately.

The standard deviation of the weights of the books is given as two pounds.

The standard deviation of the weights of the boxes is given as 0.15 pounds.

The standard deviation of the weights of the packing material is given as 0.25 pounds.

Since the weights of the books, boxes, and packing material are independent, we can use the property that the standard deviation of the sum of independent random variables is equal to the square root of the sum of their variances.

The variance of the weights of the packed boxes is then given by:

$$\sqrt{2^2 + 0.15^2 + 0.25^2} = 4 + 0.0225 + 0.0625 = 4.085$$

Taking the square root, we find that the standard deviation of the weights of the packed boxes is approximately  $\sqrt{4.085} \approx 2.021$ .

Therefore, the answer is  $\boxed{\text{D. } 2.021 \text{ pounds}}$ .

\*\*\*\*\*

Question 76:

'A group of scientists wanted to estimate the proportion of geese returning to the same site for the next breeding season. Suppose they decided to increase the sample size from 200 to 2,000. How will this affect the distribution of the sample proportion?', A. 'The distribution of the sample proportion will be more spread out.', B. 'The distribution of the sample proportion will be less spread out.', C. 'The spread of the distribution of the sample proportion will remain unaffected.', D. 'The distribution of the sample proportion will more closely resemble the binomial distribution.'

=====

Answer Option 1:

Increasing the sample size from 200 to 2,000 will have the effect of reducing the variability of the sample proportion. As the sample size increases, the standard deviation of the sample proportion decreases. This means that the distribution of the sample proportion will be less spread out. Therefore, the correct answer is  $\boxed{\text{B. The distribution of the sample proportion will be less spread out.}}$

\*\*\*\*\*

Answer Option 2:

Increasing the sample size from 200 to 2,000 will have the effect of reducing the spread of the distribution of the sample proportion. This is because as the sample size increases, the standard error of the sample proportion decreases, leading to a more precise estimate of the true population proportion. Therefore, the correct answer is B. 'The distribution of the sample proportion will be less spread out.' Final

Answer: B

\*\*\*\*\*

Answer Option 3:

When the sample size increases, the distribution of the sample proportion becomes less spread out. This is known as the Central Limit Theorem. Therefore, the correct answer is  $\boxed{\text{B. The distribution of the sample proportion will be less spread out.}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 4:

When the sample size increases, the distribution of the sample proportion becomes less spread out. This is because as the sample size increases, the sample proportion becomes a better estimate of the true proportion. The larger sample size reduces the variability of the sample proportion, resulting in a narrower distribution. Therefore, the correct answer is: B. 'The distribution of the sample proportion will be less spread out.'

\*\*\*\*\*

Answer Option 5:

When the sample size is increased from 200 to 2,000, the distribution of the sample proportion will be less spread out. This is because as the sample size increases, the standard error of the sample proportion decreases. The standard error is the measure of the spread of the distribution of the sample proportion. As the sample size increases, the standard error decreases, indicating that the distribution of the sample proportion becomes more concentrated around the true population proportion. Therefore, the correct answer is  $\boxed{\text{B. The distribution of the sample proportion will be less spread out.}}$

\*\*\*\*\*

Answer Option 6:

As the sample size increases from 200 to 2,000, the distribution of the sample proportion will be less spread out. This is because the sample size is getting larger, which reduces the random sampling variability. According to the Central Limit Theorem, as the sample size increases, the distribution of the sample proportion becomes more normal and concentrated around the true population proportion. Therefore, the answer is  $\boxed{\text{B. The distribution of the sample proportion will be less spread out.}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 7:

When the sample size increases from 200 to 2,000, the distribution of the sample proportion will be less spread out. This is because as the sample size increases, the standard deviation of the sample proportion decreases. The standard deviation is inversely proportional to the square root of the sample size. Therefore, as the sample size increases, the variability of the sample proportion decreases, resulting in a less spread out distribution. Thus, the correct answer is  $\boxed{\text{B. The distribution of the sample proportion will be less spread out.}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

Increasing the sample size from 200 to 2,000 will decrease the variability of the sample proportion. This means that the distribution of the sample proportion will be less spread out. Therefore, the correct answer is  $\boxed{\text{B. The distribution of the sample proportion will be less spread out.}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

As the sample size increases from 200 to 2,000, the distribution of the sample proportion will be less spread out. This is because as the sample size increases, the standard deviation of the sample proportion decreases. This is known as the Central Limit Theorem, which states that as the sample size increases, the distribution of the sample proportion becomes approximately normal with a smaller standard deviation. Therefore, the correct answer is  $\boxed{\text{B. The distribution of the sample proportion will be less spread out.}}$

\*\*\*\*\*



Question 77:

'A telephone survey of 400 registered voters showed that 256 had not yet made up their minds 1 month before the election. How sure can we be that between 60% and 68% of the electorate were still undecided at that time?', A. '2.4%', B. '8.0%', C. '64.0%', D. '90.4%'

=====

Answer Option 1:

To find the confidence interval for the proportion of voters who were undecided at that time, we can use the formula:

$$\text{CI} = \hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where  $\hat{p}$  is the sample proportion,  $z$  is the z-score corresponding to the desired confidence level, and  $n$  is the sample size. In this case,  $\hat{p} = \frac{256}{400} = 0.64$ , and  $n = 400$ .

To find the z-score corresponding to a 95% confidence level, we can use a standard normal distribution table. The z-score for 95% confidence corresponds to a cumulative probability of 0.975, which gives us a z-score of approximately 1.96.

Plugging in the values, we get:

$$\text{CI} = 0.64 \pm 1.96 \sqrt{\frac{0.64(1-0.64)}{400}}$$

Calculating this, we find that the confidence interval is approximately (0.60, 0.68).

Therefore, we can be 90.4% sure that between 60% and 68% of the electorate were still undecided at that time. The answer is  $\boxed{\text{D. 90.4\%}}$ .

\*\*\*\*\*

Answer Option 2:

To find the margin of error, we can use the formula

$$E = z \sqrt{\frac{p(1-p)}{n}}$$

where  $E$  represents the margin of error,  $z$  is the z-score corresponding to the desired confidence level,  $p$  is the proportion of voters who are undecided, and  $n$  is the sample size.

In this case, we are looking for a confidence level between 60% and 68%, which corresponds to a range of proportions between 0.60 and 0.68. We can use the worst-case scenario ( $p = 0.60$ ) to maximize the margin of error. Plugging in the values, we have

$$E = z \sqrt{\frac{0.60(1-0.60)}{400}}$$

Since we want to be as sure as possible, we can use a z-score corresponding to a 99% confidence level. This value is approximately 2.57. Plugging this in, we have

$$E = 2.57 \sqrt{\frac{0.60(1-0.60)}{400}} \approx 0.024$$

Therefore, we can be sure that the percentage of undecided voters is within 60% to 68% with a margin of error of 2.4%. The answer is  $\boxed{\text{A. '2.4\%'}}$ .

\*\*\*\*\*

Answer Option 3:

To find the confidence interval for the proportion of voters who were still undecided, we can use the formula:

$$\text{CI} = \hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where  $\hat{p}$  is the sample proportion,  $z$  is the z-score corresponding to the desired confidence level, and  $n$  is the sample size.

In this case, the sample proportion is  $\hat{p} = \frac{256}{400} = 0.64$ . We want to find the confidence interval for the proportion of undecided voters at a 95% confidence level, which corresponds to a z-score of approximately 1.96.

Plugging these values into the formula, we get:

$$\text{CI} = 0.64 \pm 1.96 \sqrt{\frac{0.64(1-0.64)}{400}}$$

Simplifying, we find:

$$\text{CI} = 0.64 \pm 1.96 \cdot 0.0248$$

This gives us the confidence interval of \$(0.591, 0.689)\$. Therefore, we can be 95% confident that between 59.1% and 68.9% of the electorate were still undecided at that time.

The interval range is  $0.689 - 0.591 = 0.098$ , which is 9.8%. The midpoint of the interval is  $(0.689 + 0.591)/2 = 0.64$ , which is 64%. Thus, the answer is  $\boxed{\text{C. '64.0\%'}}$ . Final answer: C

\*\*\*\*\*

Answer Option 4:

To find the range of the percentage of voters who were still undecided, we need to find the range of the proportion of voters who were still undecided.

The maximum possible proportion of undecided voters is when all 256 of the undecided voters are included in the survey. In this case, the proportion of undecided voters is  $\frac{256}{400} = 0.64$ , or 64%.

The minimum possible proportion of undecided voters is when none of the undecided voters are included in the survey. In this case, the proportion of undecided voters is 0%.

Therefore, we can be sure that between 0% and 64% of the electorate were still undecided at that time.

The answer choice closest to this range is  $\boxed{\text{C. '64.0\%'}}$ . Final answer: C

\*\*\*\*\*

Answer Option 5:

To find the margin of error, we can use the formula for a confidence interval for a proportion. The formula is given by:

$$\text{Margin of Error} = Z \sqrt{\frac{p(1-p)}{n}}$$

where  $Z$  is the z-score corresponding to the desired confidence level,  $p$  is the proportion of interest, and  $n$  is the sample size.

In this case, we want to find the margin of error for the proportion of voters who are undecided, which is  $p = \frac{256}{400} = 0.64$ . We also want to find the margin of error for a 95% confidence level, which corresponds to a z-score of  $Z = 1.96$ . Plugging in these values, we get:

$$\text{Margin of Error} = 1.96 \sqrt{\frac{0.64(1-0.64)}{400}} \approx 0.036$$

To find the range of proportions within the margin of error, we add and subtract the margin of error from the proportion of voters who are undecided. Therefore, the range of proportions is  $0.64 \pm 0.036$ , or  $(0.604, 0.676)$ .

Since the desired range is between 60% and 68%, which is within the range of proportions, we can be sure that between 60% and 68% of the electorate were still undecided at that time with a confidence level of 95%. The answer is  $\boxed{\text{D. } 90.4\%}$ .

\*\*\*\*\*

Answer Option 6:

We can construct a confidence interval for the proportion of voters who are undecided using the formula:

$$\hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where  $\hat{p}$  is the proportion of undecided voters in the sample,  $n$  is the sample size, and  $z$  is the critical value corresponding to the desired level of confidence.

In this case,  $\hat{p} = \frac{256}{400} = 0.64$  and  $n = 400$ . We want to find the confidence interval for the proportion of undecided voters with a level of confidence between 60% and 68%, so we need to find the critical values  $z_1$  and  $z_2$  corresponding to those confidence levels.

Using a standard normal distribution table, we find that  $z_1 \approx -0.84$  and  $z_2 \approx 0.84$  for a 60% confidence level, and  $z_1 \approx -1.00$  and  $z_2 \approx 1.00$  for a 68% confidence level.

Plugging in the values, we get:

$$0.64 \pm 0.84 \sqrt{\frac{0.64(1-0.64)}{400}}$$

$$0.64 \pm 0.84 \sqrt{\frac{0.2304}{400}}$$

$$0.64 \pm 0.84 \times 0.01512$$

$$0.64 \pm 0.01270$$

Therefore, the confidence interval is approximately  $(0.6273, 0.6527)$ .

Since this interval is entirely within the range of 60% to 68%, we can be sure that between 60% and 68% of the electorate were still undecided at that time.

Thus, the answer is  $\boxed{\text{D. } 90.4\%}$ .

\*\*\*\*\*

Answer Option 7:

We can use the formula for the margin of error in a proportion given by

$$ME = Z \sqrt{\frac{p(1-p)}{n}}$$

where  $Z$  is the z-score corresponding to the desired level of confidence,  $p$  is the proportion of interest, and  $n$  is the sample size.

In this case, we have  $p = \frac{256}{400} = 0.64$ , and we want to find the margin of error for a 95% confidence interval, which corresponds to a z-score of approximately 1.96.

Substituting the values into the formula, we find

$$ME = 1.96 \sqrt{\frac{0.64(1-0.64)}{400}} \approx 0.0391$$

To find the range of proportions within the margin of error, we add and subtract the margin of error from the sample proportion:

$$0.64 \pm 0.0391 = (0.6009, 0.6791)$$

Therefore, we can be 95% confident that the proportion of voters who were undecided 1 month before the election is between 60.09% and 67.91%.

The range of proportions within 60% and 68% is a subset of this confidence interval, so we can be even more confident that the true proportion falls within that range. Therefore, we can be at least 95% confident that between 60% and 68% of the electorate were still undecided at that time.

Hence, the answer is  $\boxed{\text{D. } 90.4\%}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

To find the confidence interval for the proportion of undecided voters, we can use the formula:

$$\text{CI} = \left( \hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) \times 100\%$$

where  $\hat{p}$  is the sample proportion,  $z$  is the z-score corresponding to the desired level of confidence, and  $n$  is the sample size.

In this case, the sample proportion of undecided voters is  $\hat{p} = \frac{256}{400} = 0.64$ . We want to find the confidence interval for the proportion of undecided voters at the 95% confidence level, which corresponds to a z-score of approximately 1.96.

Plugging in these values into the formula, we get:

$$\text{CI} = \left( 0.64 \pm 1.96 \sqrt{\frac{0.64(1-0.64)}{400}} \right) \times 100\%$$

Simplifying, we find:

$$\text{CI} = (0.64 \pm 0.024) \times 100\%$$

Therefore, the confidence interval for the proportion of undecided voters is  $(61.6\%, 66.4\%)$ . This means we can be 95% confident that between 61.6% and 66.4% of the electorate were still undecided at that time.

So the answer is  $\boxed{\text{B. } 8.0\%}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

To find the confidence interval for the proportion of voters who were undecided, we can use the formula:

$$\text{CI} = \hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where  $\hat{p}$  is the sample proportion,  $z$  is the z-score corresponding to the desired confidence level, and  $n$  is the sample size.

In this case,  $\hat{p} = \frac{256}{400} = 0.64$  and  $n = 400$ . We want to find the confidence interval for the proportion of undecided voters at a confidence level between 60% and 68%. To do this, we can find the confidence intervals for both 60% and 68% and see where they overlap.

For 60% confidence level,  $z \approx 0.253$  (corresponding to a two-tailed test). Plugging in the values, we get:

$$\text{CI}_{60\%} = 0.64 \pm 0.253 \cdot \sqrt{\frac{0.64(1-0.64)}{400}} \approx 0.64 \pm 0.077$$

For 68% confidence level,  $z \approx 0.994$  (corresponding to a two-tailed test). Plugging in the values, we get:

$$\text{CI}_{68\%} = 0.64 \pm 0.994 \cdot \sqrt{\frac{0.64(1-0.64)}{400}} \approx 0.64 \pm 0.241$$

Since the confidence intervals for 60% and 68% overlap, we can be sure that between 60% and 68% of the electorate were still undecided at that time. The correct answer is D. '90.4%'.

\*\*\*\*\*

Question 78:

'Players in the National Football League weigh, on average, about 248 pounds with a standard deviation of about 47 pounds. If four players are to be selected at random, the expected value of the random variable W, the total combined weight of the four players, is 992 pounds. The standard deviation of Wis approximately', A. '47 pounds', B. '67 pounds', C. '94 pounds', D. '141 pounds'

=====

Answer Option 1:

The standard deviation of a sum of independent random variables is the square root of the sum of the variances of the individual random variables. Since we are selecting four players at random, the variance of the total combined weight of the four players is  $4 \cdot (47^2)$ \$. Thus, the standard deviation of  $W$  is  $\sqrt{4 \cdot (47^2)} = 2 \cdot 47 = 94$  pounds\$. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The standard deviation of the sum of four independent random variables is equal to the square root of the sum of the variances of the individual random variables. Since the standard deviation of each player's weight is 47 pounds, the variance is  $47^2 = 2209$  pounds $^2$ \$. Therefore, the standard deviation of the sum of four players' weights is  $\sqrt{4 \cdot 2209} = 94$  pounds. The answer is  $\boxed{94}$  pounds\$. Final Answer: C

\*\*\*\*\*

Answer Option 3:

The expected value of the sum of random variables is the sum of the expected values of the individual random variables. Since the expected weight of one player is 248 pounds, the expected value of the total combined weight of four players is  $4 \cdot 248 = 992$  pounds, as given in the problem.

For independent random variables, the variance of the sum of the random variables is the sum of the variances of the individual random variables. Since the standard deviation is the square root of the variance, the standard deviation of the total combined weight of four players is  $\sqrt{4 \cdot (47^2)} = \sqrt{4 \cdot 2209} = \sqrt{8836}$ \$. Simplifying, we find that this is equal to  $\sqrt{8836}$  pounds, or approximately 94 pounds. Final Answer: C.

\*\*\*\*\*

Answer Option 4:

The expected value of the sum of random variables is equal to the sum of the expected values of the individual random variables. In this case, since the expected value of each player's weight is 248 pounds, the expected value of the total combined weight of four players is 4 times 248 pounds, which is 992 pounds.

The standard deviation of the sum of independent random variables is equal to the square root of the sum of the variances of the individual random variables. The variance of each player's weight is equal to the square of the standard deviation, which is  $(47)^2 = 2209$ \$. Therefore, the variance of the total combined weight of four players is 4 times the variance of the weight of a single player, which is  $4(2209) = 8836$ \$.

Taking the square root of 8836 gives us approximately 94 pounds. Therefore, the standard deviation of W is approximately  $\boxed{94}$  pounds\$.

\*\*\*\*\*

Answer Option 5:

The expected value of the sum of independent random variables is equal to the sum of the expected values of the individual random variables. In this case, since the expected value of a single player's weight is 248 pounds and there are four players, the expected value of the total combined weight is  $4 \cdot 248 = 992$  pounds.

The standard deviation of the sum of independent random variables is equal to the square root of the sum of the variances of the individual random variables. The variance of a single player's weight is equal to the square of the standard deviation, which is  $(47)^2 = 2209$  pounds squared. Since there are four players, the variance of the total combined weight is  $4 \cdot 2209 = 8836$  pounds squared.

Therefore, the standard deviation of the total combined weight is the square root of 8836, which is approximately  $\sqrt{8836} \approx 94$  pounds. Final Answer: C

\*\*\*\*\*

Answer Option 6:

The expected value of the sum of random variables is equal to the sum of the expected values. Since there are four players and the expected weight of one player is 248 pounds, the expected value of the total combined weight is 4 times 248, which is 992 pounds.

The standard deviation of the sum of independent random variables is equal to the square root of the sum of the variances. Since the standard deviation of one player is 47 pounds, the variance is 47 squared. The variance of the sum of four independent variables is 4 times the variance of one variable, which is 4 times 47 squared. Therefore, the standard deviation of the total combined weight is the square root of 4 times 47 squared, which is 4 times 47, or 188 pounds.

Therefore, the standard deviation of W is approximately  $\boxed{141}$  pounds\$. Final Answer: D

\*\*\*\*\*

Answer Option 7:

Let  $X$  be the weight of a single player. We know that  $E(X) = 248$  and  $SD(X) = 47$ .

Let  $W$  be the total combined weight of the four players. We want to find  $SD(W)$ . Since  $W$  is the sum of four independent random variables  $X_1, X_2, X_3, X_4$ , we have:

$$SD(W) = \sqrt{SD(X_1)^2 + SD(X_2)^2 + SD(X_3)^2 + SD(X_4)^2}.$$

Since all  $X_i$  are identically distributed, we have  $SD(X_1) = SD(X_2) = SD(X_3) = SD(X_4) = SD(X) = 47$ . Therefore,

$$SD(W) = \sqrt{47^2 + 47^2 + 47^2 + 47^2} = \sqrt{4 \cdot 47^2} = \boxed{94} \text{ pounds}.$$

Final Answer: C

\*\*\*\*\*

Answer Option 8:

The expected value of a sum of random variables is equal to the sum of the expected values of the random variables. In this case, the expected value of  $W$ , the total combined weight of the four players, is 992 pounds. Since there are four players and the expected weight of each player is 248 pounds, we have:

$$E[W] = E[X_1 + X_2 + X_3 + X_4] = E[X_1] + E[X_2] + E[X_3] + E[X_4] = 248 + 248 + 248 + 248 = 992$$

The standard deviation of a sum of independent random variables is equal to the square root of the sum of the variances of the random variables. Since there are four players and the standard deviation of each player is 47 pounds, we have:

$$SD[W] = \sqrt{\text{Var}[X_1] + \text{Var}[X_2] + \text{Var}[X_3] + \text{Var}[X_4]} = \sqrt{(47^2) + (47^2) + (47^2) + (47^2)} = \sqrt{4 \cdot (47^2)} = 47 \cdot \sqrt{4} = 47 \cdot 2 = 94$$

Therefore, the standard deviation of  $W$  is approximately 94 pounds, so the answer is  $\boxed{\text{(C) 94 pounds}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 9:

Since the expected value of the sum of random variables is equal to the sum of the expected values, we know that

$$E[W] = E[X_1 + X_2 + X_3 + X_4] = E[X_1] + E[X_2] + E[X_3] + E[X_4] = 4 \cdot 248 = 992.$$

For random variables  $X_1, X_2, X_3, X_4$  representing the weights of the four players, we have that  $E[X_i] = 248$  for each  $i$ .

The standard deviation of the sum of independent random variables is equal to the square root of the sum of their variances. Since the variance of  $X_i$  is equal to the square of the standard deviation of  $X_i$ , we can write the variance of the sum of the  $X_i$ 's as

$$\text{Var}(W) = \text{Var}(X_1 + X_2 + X_3 + X_4) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) + \text{Var}(X_4).$$

Since the standard deviation is the square root of the variance, we can calculate the standard deviation of  $W$  as

$$\sqrt{\text{Var}(W)} = \sqrt{\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) + \text{Var}(X_4)}.$$

Since the variance of  $X_i$  is equal to the square of the standard deviation of  $X_i$ , we have that  $\text{Var}(X_i) = (\text{SD}(X_i))^2$ . Therefore,

$$\sqrt{\text{Var}(W)} = \sqrt{(\text{SD}(X_1))^2 + (\text{SD}(X_2))^2 + (\text{SD}(X_3))^2 + (\text{SD}(X_4))^2}.$$

We are given that  $\text{SD}(X_i) = 47$  for each  $i$ . Plugging in these values, we get

$$\sqrt{\text{Var}(W)} = \sqrt{(47)^2 + (47)^2 + (47)^2 + (47)^2} = \sqrt{4 \cdot (47)^2} = 2 \cdot 47 = \boxed{94}.$$

Final Answer: C

\*\*\*\*\*

Question 79:

"A hypothesis test yields a P-value of 0.20." Which of the following best describes what is meant by this statement?', A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.', B. 'The probability of getting a finding as extreme as that obtained by chance alone from repeated random sampling is 0.20.', C. 'The probability is 0.20 that our finding is significant.', D. 'The probability of getting this finding is 0.20.'

=====

Answer Option 1:

The correct interpretation of a p-value is that it represents the probability of obtaining a result as extreme as, or more extreme than, the observed result, assuming that the null hypothesis is true. Therefore, the best description of the statement "A hypothesis test yields a P-value of 0.20" is option A: "The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20." Final answer: A

=====

Answer Option 2:

The correct answer is A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.' The p-value represents the probability of obtaining a test statistic as extreme as, or more extreme than, the observed value, assuming the null hypothesis is true. It is not the probability of the null hypothesis being true itself, which is why options B, C, and D are incorrect. Final Answer: A

=====

Answer Option 3:

The correct answer is A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.'

The p-value is a measure of the strength of evidence against the null hypothesis in a hypothesis test. It represents the probability of observing a test statistic as extreme as, or more extreme than, the one obtained from the sample data, assuming that the null hypothesis is true. In other words, it quantifies the likelihood of obtaining the observed result by chance alone. Therefore, option A accurately describes the meaning of the p-value. Final Answer: A

=====

Answer Option 4:

The correct answer is A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.'

The p-value is the probability of obtaining a test statistic as extreme as the one observed, or even more extreme, under the assumption that the null hypothesis is true. It measures the strength of evidence against the null hypothesis. A p-value of 0.20 indicates that there is a 20% chance of obtaining a result as extreme as the one observed if the null hypothesis is true. It does not provide information about the probability of the null hypothesis being true or the probability of the finding being significant. Final Answer: A

=====

Answer Option 5:

The correct answer is A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.'

A p-value is the probability of obtaining a test statistic as extreme as, or more extreme than, the observed data, assuming that the null hypothesis is true. It measures the strength of evidence against the null hypothesis. A p-value of 0.20 indicates that there is a 20% chance of obtaining a result as extreme as, or more extreme than, the observed data by chance alone, assuming that the null hypothesis is true.

This means that if the null hypothesis is true, there is a 20% probability of obtaining the observed data purely due to random chance. Final answer: A

=====

Answer Option 6:

The correct answer is A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.'

The p-value is the probability of observing a test statistic or a more extreme value if the null hypothesis is true. It measures the strength of evidence against the null hypothesis. A p-value of 0.20 means that there is a 20% chance of obtaining a result as extreme or more extreme than the one observed, purely due to chance, if the null hypothesis is true.

Option B describes the concept of a significance level, which is different from the p-value. Option C and D are incorrect because the p-value does not directly indicate the probability that the finding is significant or the probability of observing the specific finding.

=====

Answer Option 7:

The correct answer is A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.'

The P-value is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming that the null hypothesis is true. In other words, it measures the strength of evidence against the null hypothesis. A smaller P-value indicates stronger evidence against the null hypothesis, while a larger P-value suggests weaker evidence against the null hypothesis.

Therefore, the statement "The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20" best describes what is meant by the P-value. Final

Answer: A

=====

Answer Option 8:

The correct answer is A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.' The P-value is a measure of the strength of evidence against the null hypothesis. A P-value of 0.20 means that if the null hypothesis is true, the chances of obtaining a result as extreme as the one observed (or more extreme) is 0.20. It does not directly provide information about the probability of other statements described in the other answer choices. Final Answer: A

=====

Answer Option 9:

The correct answer is A. 'The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20.'

The p-value is defined as the probability of obtaining a test statistic as extreme as, or more extreme than, the observed test statistic, assuming that the null hypothesis is true. In other words, it quantifies the evidence against the null hypothesis.

Therefore, the statement "The probability of getting a finding at least as extreme as that obtained by chance alone if the null hypothesis is true is 0.20" accurately represents the meaning of the p-value. Final

Answer: A

\*\*\*\*\*

Question 80:

'Which of the following is not true with regard to contingency tables for chi-square tests for independence?', A. 'The categories are not numerical for either variable.', B. 'Observed frequencies should be whole numbers.', C. 'Expected frequencies should be whole numbers.', D. 'Expected frequencies in each cell should be at least 5, and to achieve this, one sometimes combines categories for one or the other or both of the variables.'

=====

Answer Option 1:

The correct answer is C. 'Expected frequencies should be whole numbers.' In chi-square tests for independence, expected frequencies are calculated based on the assumption of independence between the variables. These expected frequencies are not required to be whole numbers. The other options are true statements regarding contingency tables and chi-square tests for independence. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The correct answer is C. 'Expected frequencies should be whole numbers.' In chi-square tests for independence, the expected frequencies are calculated based on the assumption of independence between the variables. These expected frequencies are typically decimals and do not need to be whole numbers. Thus, the statement in option C is not true. Final Answer: C

\*\*\*\*\*

Answer Option 3:

The correct answer is C. 'Expected frequencies should be whole numbers.' In chi-square tests for independence, the expected frequencies are calculated based on the assumption of independence between the two variables. These expected frequencies are calculated using the observed frequencies and the total sample size, and they are generally not whole numbers. Therefore, option C is not true. Final

Answer: C

\*\*\*\*\*

Answer Option 4:

The correct answer is C. 'Expected frequencies should be whole numbers.' In chi-square tests for independence, the expected frequencies are calculated based on the observed frequencies and the assumption of independence between the variables. The expected frequencies can take on any real value, not just whole numbers. Therefore, option C is not true with regard to contingency tables for chi-square tests for independence. Final answer: C

\*\*\*\*\*

Answer Option 5:

The correct answer is C. 'Expected frequencies should be whole numbers.' In chi-square tests for independence, the observed frequencies are the actual counts of data in each cell of the contingency table. The expected frequencies are the values that would be expected if there was no association between the two variables being tested. These expected frequencies are calculated based on the row and column totals of the observed frequencies, and they are typically not whole numbers. Therefore, C is the statement that is not true with regard to contingency tables for chi-square tests for independence. Final

Answer: C

\*\*\*\*\*

Answer Option 6:

The correct answer is C. 'Expected frequencies should be whole numbers.'

In chi-square tests for independence, expected frequencies are calculated based on the assumption of independence between the variables. These expected frequencies can take any real value, not just whole numbers. The other options are true statements about contingency tables for chi-square tests for independence.

Option A is true because the categories for both variables in a contingency table are usually qualitative or categorical, not numerical.

Option B is true because observed frequencies represent actual counts, so they should be whole numbers.

Option D is true because one of the assumptions of the chi-square test for independence is that the expected frequencies in each cell should be at least 5. If any expected frequency is less than 5, categories may need to be combined to meet this assumption. Final Answer: C

\*\*\*\*\*

Answer Option 7:

The correct answer is C. 'Expected frequencies should be whole numbers.'

In a chi-square test for independence, the expected frequencies are calculated based on the assumption of independence between the two variables. These expected frequencies are not required to be whole numbers. They are typically decimal values that represent the expected counts under the assumption of independence.

The other options are true statements about contingency tables for chi-square tests for independence:

A. 'The categories are not numerical for either variable.' - This is true because contingency tables are used to analyze the relationship between two categorical variables, not numerical variables.

B. 'Observed frequencies should be whole numbers.' - This is true because the observed frequencies represent the actual counts of observations in each category, which are whole numbers.

D. 'Expected frequencies in each cell should be at least 5, and to achieve this, one sometimes combines categories for one or the other or both of the variables.' - This is true because having expected frequencies below 5 in any cell can lead to issues with the chi-square test, such as low expected cell counts. To address this, categories may be combined to ensure that the expected frequencies in each cell are at least 5.

\*\*\*\*\*

Answer Option 8:

The statement that is not true with regard to contingency tables for chi-square tests for independence is option C: 'Expected frequencies should be whole numbers.'

In chi-square tests for independence, the expected frequencies are calculated based on the assumption of independence between the two variables. These expected frequencies are calculated using the observed frequencies and the marginal totals of the contingency table. However, the expected frequencies are not required to be whole numbers. They can be decimals or fractions depending on the specific



values of the observed frequencies and the total sample size.

Therefore, the correct answer is C. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct answer is C. 'Expected frequencies should be whole numbers.'

In chi-square tests for independence, the observed frequencies represent the actual counts in each cell of the contingency table, while the expected frequencies are the counts that would be expected under the assumption that the variables are independent. It is not necessary for the expected frequencies to be whole numbers.

The other options are all true statements about contingency tables for chi-square tests for independence:

A. The categories are not numerical for either variable. Chi-square tests for independence are used to determine if there is a relationship between two categorical variables.

B. Observed frequencies should be whole numbers. This is because the observed frequencies represent the actual counts and cannot be fractional or negative.

D. Expected frequencies in each cell should be at least 5, and to achieve this, one sometimes combines categories for one or the other or both of the variables. This is a general guideline to ensure the validity of the chi-square test. If the expected frequencies are too small, the chi-square test may not be reliable, so categories may need to be combined to increase the expected frequencies.

\*\*\*\*\*

Question 81:

The observed times (in minutes) it takes a runner to complete a marathon are normally distributed. The z-score for his running time this week is  $-2$ . Which one of the following statements is a correct interpretation of his z-score? A. 'This week his time was two minutes lower than his time last week.', B. 'This week his time was two minutes lower than his best time ever.', C. 'This week his time was two minutes lower than his average time.', D. 'This week his time was two standard deviations lower than his average time.'

=====

Answer Option 1:

A z-score represents the number of standard deviations a data point is away from the mean. In this case, the runner's z-score is  $-2$ , which means his running time is two standard deviations below the mean. Therefore, the correct interpretation of his z-score is: 'This week his time was two standard deviations lower than his average time.' The answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 2:

The z-score represents the number of standard deviations a data point is away from the mean. In this case, since the z-score is  $-2$ , it means that the runner's running time this week is two standard deviations lower than the mean running time. Therefore, the correct interpretation is  $\boxed{\text{D. 'This week his time was two standard deviations lower than his average time.'}}$

\*\*\*\*\*

Answer Option 3:

The z-score represents the number of standard deviations a value is away from the mean. A z-score of  $-2$  means that the runner's time this week is 2 standard deviations below the mean. Therefore, the correct interpretation is that 'This week his time was two standard deviations lower than his average time.' Thus, the answer is  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

The z-score measures how many standard deviations an observation is away from the mean. In this case, the z-score is  $-2$ , which means that the runner's running time this week is 2 standard deviations below the mean. Therefore, the correct interpretation is that 'This week his time was two standard deviations lower than his average time.' Hence, the answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 5:

The correct interpretation of the z-score is D. 'This week his time was two standard deviations lower than his average time.'

A z-score measures how many standard deviations an individual data point is from the mean. In this case, the z-score of  $-2$  means that the runner's time this week is two standard deviations below the mean time. It does not provide any information about his time last week, his best time ever, or his average time. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The z-score is a measure of how many standard deviations an observation is away from the mean. In this case, a z-score of  $-2$  means that the runner's running time this week is 2 standard deviations below the mean.

Therefore, the correct interpretation is: 'This week his time was two standard deviations lower than his average time.' This corresponds to choice D. Final answer: D.

\*\*\*\*\*

Answer Option 7:

The z-score measures how many standard deviations an observation is from the mean of a distribution. In this case, the z-score is  $-2$ , which means that the runner's running time this week is 2 standard deviations below the mean. Therefore, the correct interpretation is that 'This week his time was two standard deviations lower than his average time.' This corresponds to answer choice D. Final answer: D

\*\*\*\*\*

Answer Option 8:

A z-score measures the number of standard deviations a data point is away from the mean. In this case, the z-score is  $-2$ , which means the runner's running time is two standard deviations below the mean running time. Therefore, the correct interpretation of his z-score is that 'This week his time was two standard deviations lower than his average time.' Thus, the answer is  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

The z-score is defined as the number of standard deviations an observation is away from the mean. A z-score of  $-2$  means that the observation is 2 standard deviations below the mean.

Therefore, the correct interpretation of his z-score is: 'This week his time was two standard deviations lower than his average time.'

So, the answer is  $\boxed{\text{D}}$  Final Answer: D

\*\*\*\*\*

Question 82:

The mean Law School Aptitude Test (LSAT) score for applicants to a particular law school is 650 with a standard deviation of 45. Suppose that only applicants with scores above 700 are considered. What percentage of the applicants considered have scores below 740? (Assume the scores are normally distributed.)', A. '13.3%', B. '17.1%', C. '82.9%', D. '86.7%'

=====

Answer Option 1:

To find the percentage of applicants with scores below 740, we need to find the area under the normal distribution curve to the left of 740.

First, we need to standardize the score of 740. Using the formula for standardization, we have

$$Z = \frac{x - \mu}{\sigma} = \frac{740 - 650}{45} = \frac{90}{45} = 2$$

Next, we find the area to the left of  $Z = 2$  using a standard normal distribution table or a calculator. The area to the left of 2 is approximately 0.9772.

However, we are only considering applicants with scores above 700. Since the normal distribution is symmetric, the percentage of applicants with scores below 740 is equal to half the percentage of applicants with scores above 700. Thus, the percentage of applicants considered with scores below 740 is

$$\frac{1 - 0.9772}{2} = 0.0114$$

Converting this to a percentage, we have 1.14%.

Therefore, the answer is  $\boxed{\text{A. '13.3\%'}}$ .

=====

Answer Option 2:

To solve this problem, we need to find the area under the normal distribution curve to the left of 740.

First, we need to calculate the z-score for a score of 740. The z-score is calculated as:

$$Z = \frac{x - \mu}{\sigma}$$

where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we get:

$$Z = \frac{740 - 650}{45} = \frac{90}{45} = 2$$

Next, we need to find the area under the normal distribution curve to the left of a z-score of 2. We can look this up in the standard normal distribution table or use a calculator to find that the area is approximately 0.9772.

Finally, we subtract this value from 1 to find the percentage of applicants considered that have scores below 740:

$$1 - 0.9772 = 0.0228$$

Converting this to a percentage, we get 2.28%. Therefore, the answer is approximately 2.28%, which is closest to  $\boxed{\text{A. '13.3\%'}}$ . Final Answer: A

=====

Answer Option 3:

To solve this problem, we need to standardize the scores using the z-score formula:  $Z = \frac{x - \mu}{\sigma}$ , where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

First, we standardize the score 740:

$$Z = \frac{740 - 650}{45} = \frac{90}{45} = 2$$

Next, we find the area to the left of  $Z = 2$ . Using a standard normal distribution table or a calculator, we find that the area to the left of  $Z = 2$  is approximately 0.9772.

However, we are interested in the percentage of applicants with scores below 740, which is the area to the left of 740. Since the normal distribution is symmetrical, the area to the right of  $Z = 2$  is also 0.9772. Therefore, the area to the left of 740 is  $1 - 0.9772 = 0.0228$ .

Finally, we convert this to a percentage by multiplying by 100:

$$0.0228 \times 100 = 2.28\%$$

So, the percentage of the applicants considered with scores below 740 is approximately  $\boxed{\text{A. '13.3\%'}}$ . Final answer: None of the given choices

=====

Answer Option 4:

To solve this problem, we need to use the concept of Z-scores. The Z-score measures how many standard deviations an observation is from the mean of a distribution. Given that the mean LSAT score is 650 and the standard deviation is 45, we can calculate the Z-score for an LSAT score of 740 as follows:

$$Z = \frac{740 - 650}{45}$$

$$= \frac{90}{45}$$

$$= 2$$

Using a Z-score table or calculator, we can find that the percentage of scores below a Z-score of 2 is approximately 97.7%. However, we need to take into account that only applicants with scores above 700 are considered. To calculate the percentage of applicants considered with scores below 740, we need to subtract the percentage of applicants with scores between 700 and 740.

To find the Z-score for an LSAT score of 700:

$$Z = \frac{700 - 650}{45}$$

$$= \frac{50}{45}$$

$$= 1.11$$

Using the Z-score table or calculator, we find that the percentage of scores below a Z-score of 1.11 is approximately 86.7%.

Therefore, the percentage of applicants considered with scores below 740 is  $97.7\% - 86.7\% = 11\%$ . Thus, the answer is approximately 11%, or in percentage form, '17.1%'. Therefore, the correct answer is B.

=====

Answer Option 5:

To find the percentage of applicants considered with scores below 740, we need to find the area under the normal distribution curve to the left of 740.

First, we calculate the z-score for 740 using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $\mu$  is the mean and  $\sigma$  is the standard deviation.

Plugging in the values, we have:

$$z = \frac{740 - 650}{45} = 2$$

Next, we use a standard normal distribution table or calculator to find the area to the left of  $z = 2$ .

The area to the left of  $z = 2$  is approximately 0.9772 or 97.72%.

However, we need to consider that only applicants with scores above 700 are considered.

To find the percentage of applicants considered with scores below 740, we subtract the percentage of applicants with scores below 700.

Using the same method, we find the z-score for 700:

$$z = \frac{700 - 650}{45} = 1.111$$

The area to the left of  $z = 1.111$  is approximately 0.8664 or 86.64%.

Subtracting 86.64% from 97.72%, we get:

$$97.72\% - 86.64\% = 11.08\%$$

Therefore, the percentage of the applicants considered with scores below 740 is approximately 11.1%. Final answer: None of the above.

\*\*\*\*\*

Answer Option 6:

To find the percentage of applicants with scores below 740, we need to find the area under the normal distribution curve to the left of 740.

First, we standardize the scores using the z-score formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For a score of 740:

$$z = \frac{740 - 650}{45} = \frac{90}{45} = 2$$

Using a standard normal distribution table or a calculator, we can find that the area to the left of  $z = 2$  is approximately 0.9772.

Therefore, the percentage of applicants with scores below 740 is approximately 97.72%. However, only applicants with scores above 700 are considered, so we need to subtract the percentage of applicants with scores between 700 and 740.

To find the z-score for a score of 700:

$$z = \frac{700 - 650}{45} = \frac{50}{45} = \frac{10}{9}$$

Using a standard normal distribution table or a calculator, we can find that the area to the left of  $z = 10/9$  is approximately 0.7123.

Subtracting this from the percentage we found earlier:

$$0.9772 - 0.7123 = 0.2649$$

So, the percentage of the applicants considered with scores below 740 is approximately 26.49%.

Therefore, the answer is approximately 26.49% or 17.1%. Final Answer: B

\*\*\*\*\*

Answer Option 7:

To find the percentage of applicants with scores below 740, we need to find the area under the normal distribution curve to the left of 740. First, we calculate the z-score for 740 using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Substituting in the values, we have:

$$z = \frac{740 - 650}{45} = \frac{90}{45} = 2$$

Next, we use a standard normal distribution table or a calculator to find the area to the left of  $z = 2$ . This area represents the percentage of applicants with scores below 740. From the table or calculator, we find that the area to the left of  $z = 2$  is approximately 0.9772.

However, we need to consider that only applicants with scores above 700 are considered. This means that the percentage of applicants with scores below 740 is the percentage of the total area to the left of 740, minus the percentage of the total area to the left of 700.

Using the same process as above, we find that the z-score for 700 is:

$$z = \frac{700 - 650}{45} = \frac{50}{45} = \frac{10}{9} \approx 1.11$$

The area to the left of  $z = 1.11$  is approximately 0.8665.

Therefore, the percentage of applicants considered with scores below 740 is approximately  $0.9772 - 0.8665 = 0.1107$ , or 11.07%.

Thus, the answer is approximately 11.07%. Final answer: None of the given answer choices

\*\*\*\*\*

Answer Option 8:

To find the percentage of applicants considered who have scores below 740, we need to find the area under the normal distribution curve to the left of 740.

First, we need to standardize the score of 740 using the z-score formula:

$$z = \frac{x - \mu}{\sigma} = \frac{740 - 650}{45} = \frac{90}{45} = 2$$

Using a standard normal distribution table (or a calculator), we can find that the area to the left of a z-score of 2 is approximately 0.9772.

Therefore, the percentage of applicants considered with scores below 740 is approximately 97.72%. However, we need to consider that only applicants with scores above 700 are considered.

The percentage of applicants considered with scores below 740 among those with scores above 700 can be calculated by subtracting the percentage of applicants with scores between 700 and 740 from the total.

To find the z-score for a score of 700:

$$z = \frac{x - \mu}{\sigma} = \frac{700 - 650}{45} = \frac{50}{45} = \frac{10}{9} \approx 1.11$$

Using the standard normal distribution table, we can find that the area to the left of a z-score of 1.11 is approximately 0.8665.

Therefore, the percentage of applicants with scores between 700 and 740 is approximately  $(0.9772 - 0.8665) \times 100 = 11.07\%$ .

Subtracting 11.07% from 97.72%, we find that the percentage of applicants considered with scores below 740 is approximately 86.65%, or  $\boxed{86.7\%}$ .

\*\*\*\*\*

Answer Option 9:

To find the percentage of applicants considered with scores below 740, we need to find the area under the normal distribution curve to the left of 740.

First, we calculate the z-score for 740 using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we are interested in,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we get:

$$z = \frac{740 - 650}{45} = \frac{90}{45} = 2$$

Next, we look up the corresponding area under the standard normal distribution curve for a z-score of 2. Using a standard normal table, we find that the area to the left of 2 is approximately 0.9772.

However, we need to take into account that only applicants with scores above 700 are considered. To find the percentage of applicants considered with scores below 740, we subtract the percentage of applicants with scores between 700 and 740 from the total percentage of applicants below 740.

To find the area between 700 and 740, we calculate the z-scores for both values:

For 700:

$$z = \frac{700 - 650}{45} = \frac{50}{45} \approx 1.111$$

For 740:

$$z = \frac{740 - 650}{45} = \frac{90}{45} = 2$$

Using the standard normal table again, we find that the area to the left of 1.111 is approximately 0.8664, and the area to the left of 2 is approximately 0.9772.

The percentage of applicants considered with scores between 700 and 740 is approximately  $(0.9772 - 0.8664) \times 100 = 11.08\%$ .

Finally, we subtract this percentage from the total area to the left of 740 to find the percentage of applicants considered with scores below 740:

$$0.9772 - 0.8664 = 0.1108$$

So the percentage of applicants considered with scores below 740 is approximately  $0.9772 - 0.1108 \approx 0.8664 \times 100\% = 86.64\%$ .

Therefore, the answer is  $\boxed{86.7\%}$ .

\*\*\*\*\*

Question 83:

'On the same test, Mary and Pam scored at the 64th and 56th percentiles, respectively. Which of the following is a true statement?', A. 'Mary scored eight more points than Pam.', B. 'Mary's score is 8% higher than Pam's.', C. 'Eight percent of those who took the test scored between Pam and Mary.', D. 'Thirty-six people scored higher than both Mary and Pam.'

=====

Answer Option 1:

The 64th percentile means that Mary scored better than 64% of the people who took the test. Similarly, the 56th percentile means that Pam scored better than 56% of the people who took the test. Since Mary's percentile is higher than Pam's, it implies that Mary's score is better than Pam's.

However, we cannot make a direct comparison between the two scores or determine the exact difference in points based solely on the given information. Therefore, we cannot conclude that Mary scored eight more points than Pam (option A) or that Mary's score is 8% higher than Pam's (option B).

Additionally, we cannot determine the number of people who scored between Pam and Mary, so we cannot conclude that 8% of those who took the test scored between them (option C).

Therefore, the only true statement among the given options is option D: 'Thirty-six people scored higher than both Mary and Pam.' Final answer: D.

\*\*\*\*\*

Answer Option 2:

The statement 'Mary scored eight more points than Pam' cannot be determined based on the given information about percentiles. Percentiles indicate the percentage of scores that are below a given score, but they do not provide information about the actual scores.

The statement 'Mary's score is 8% higher than Pam's' cannot be determined either. Percentiles indicate relative positions, not absolute differences in scores.

The statement 'Eight percent of those who took the test scored between Pam and Mary' is also not necessarily true. Percentiles indicate the relative positions of the scores, and the percentage of scores between Pam and Mary depends on the distribution of scores in that range. It is possible for the percentage of scores between Pam and Mary to be different from 8%.

The only statement that can be determined as true is 'Thirty-six people scored higher than both Mary and Pam.' Percentiles indicate the percentage of scores below a given score, so if Mary is at the 64th percentile and Pam is at the 56th percentile, it means that 64% of the scores are below Mary's score and 56% of the scores are below Pam's score. Therefore, at least 36% of the scores are above both Mary and Pam's scores, which means at least 36 people scored higher than both Mary and Pam. Therefore, the correct answer is  $\boxed{\text{(D)}}$ .

\*\*\*\*\*

Answer Option 3:

The 64th percentile means that Mary scored better than 64% of the people who took the test, and the 56th percentile means that Pam scored better than 56% of the people who took the test. Since the percentiles are based on rankings, we cannot determine the actual scores of Mary and Pam. Therefore, statement A and statement B cannot be determined to be true.

Statement C cannot be determined to be true either. It is possible that more than 8% of the people who took the test scored between Pam and Mary.

Statement D is the only statement that can be determined to be true. Since Mary scored higher than 64% of the people who took the test and Pam scored higher than 56% of the people who took the test, it follows that there are at least  $100 - 64 = 36$  people who scored higher than both Mary and Pam.

Therefore, the correct answer is  $\boxed{\text{(D) 'Thirty-six people scored higher than both Mary and Pam.'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

The percentile of a score represents the percentage of scores that are lower than that score. Therefore, if Mary scored at the 64th percentile, it means that 64% of test-takers scored lower than her, and if Pam scored at the 56th percentile, it means that 56% of test-takers scored lower than her.

Since Mary scored at a higher percentile than Pam, it means that Mary's score is higher than Pam's score. Therefore, option A is not true.

Option B states that Mary's score is 8% higher than Pam's score. However, we do not have any information about the actual scores of Mary and Pam, only their percentiles. Therefore, we cannot determine if option B is true or not.

Option C states that 8% of test-takers scored between Pam and Mary. Since 64% of test-takers scored below Mary and 56% of test-takers scored below Pam, it means that there is an overlap of 8% of test-takers between the scores of Pam and Mary. Therefore, option C is true.

Option D states that 36 people scored higher than both Mary and Pam. However, we do not have any information about the number of test-takers or the distribution of scores. Therefore, we cannot determine if option D is true or not.

Therefore, the only true statement is option C: 'Eight percent of those who took the test scored between Pam and Mary.' Final Answer: C

\*\*\*\*\*

Answer Option 5:

The statement "Mary scored eight more points than Pam" cannot be determined from the percentiles given. Percentiles only provide information on how a student's score compares to the rest of the test takers, not the actual scores themselves.

The statement "Mary's score is 8% higher than Pam's" also cannot be determined from the percentiles given. Percentiles do not represent percentage differences in scores.

The statement "Thirty-six people scored higher than both Mary and Pam" also cannot be determined from the percentiles given. Percentiles only provide information on how a student's score compares to the rest of the test takers, not the actual number of people who scored higher.

However, the statement "Eight percent of those who took the test scored between Pam and Mary" is true. If Mary is at the 64th percentile and Pam is at the 56th percentile, then 8% of the test takers scored between them. This statement is supported by the given information.

Therefore, the correct answer is  $\boxed{\text{(C. 'Eight percent of those who took the test scored between Pam and Mary.'})}$

\*\*\*\*\*

Answer Option 6:

The statement "Mary scored eight more points than Pam" cannot be determined from the given information. The percentiles only provide information about the relative rankings of their scores, not the actual scores themselves.

The statement "Mary's score is 8% higher than Pam's" also cannot be determined from the given information. The percentiles indicate the percentage of test-takers that scored lower than Mary and Pam, but they do not provide direct information about the difference in their scores.

The statement "Eight percent of those who took the test scored between Pam and Mary" is also not necessarily true. The percentiles only provide information about the rankings of their scores relative to other test-takers, not the actual number of test-takers.

Therefore, the only statement that can be true is "Thirty-six people scored higher than both Mary and Pam." This statement does not contradict the given information and is a reasonable conclusion based on the percentiles. Hence, the correct answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 7:

The 64th percentile means that 64% of the people scored below Mary, and the 56th percentile means that 56% of the people scored below Pam. Therefore, it is possible that Mary scored eight more points than Pam, but it is also possible that she scored less or the same amount. So statement A is not necessarily true.

Similarly, the percentiles do not tell us anything about the actual scores, so statement B is not necessarily true either.

Statement C is not necessarily true either. The percentiles tell us nothing about the number of people who scored between Pam and Mary. It is possible that 8% of the people scored between them, but it could be more or less than that.

Statement D is true based on the information given. Since Pam scored at a lower percentile than Mary, it means that there are at least 36 people who scored higher than both Mary and Pam.

Therefore, the correct answer is  $\boxed{\text{D}}$ . "Thirty-six people scored higher than both Mary and Pam." Final answer: D

\*\*\*\*\*

Answer Option 8:

The percentile score represents the percentage of scores that are lower than a given score. Since Mary scored at the 64th percentile and Pam at the 56th percentile, it means that 64% of the scores are lower than Mary's score and 56% of the scores are lower than Pam's score.

From this information, we cannot determine the actual scores of Mary and Pam, so we cannot conclude that Mary scored eight more points than Pam (option A) or that Mary's score is 8% higher than Pam's (option B).

Option C states that 8% of those who took the test scored between Pam and Mary. Since Mary scored higher than Pam, it is not possible for 8% of the scores to fall between them. Therefore, option C is not true.

Option D states that thirty-six people scored higher than both Mary and Pam. We do not have information about the number of people who took the test, so we cannot determine if this statement is true or not.

Therefore, none of the options can be determined to be a true statement. The correct answer is: None of the above.

\*\*\*\*\*

Answer Option 9:

The 64th percentile means that Mary scored better than 64% of the test takers, and the 56th percentile means that Pam scored better than 56% of the test takers.

Since Mary scored better than Pam, it is not possible for Mary to have scored eight more points than Pam. Thus, statement A is not true.

The percentiles do not directly correspond to the percentage difference in scores. For example, if the test had a total of 100 points, Mary could have scored 64 points while Pam scored 56 points. In this case, Mary's score is only 14.3% higher than Pam's. Thus, statement B is not necessarily true.

The fact that the percentiles are different implies that there are test takers whose scores fall between Mary and Pam. However, the statement does not provide enough information to determine whether this percentage is exactly 8%. Thus, statement C is not necessarily true.

Finally, we can conclude that since Mary scored better than 64% of the test takers and Pam scored better than 56% of the test takers, there must be at least 36 people who scored higher than both Mary and Pam. Thus, statement D is true.

Therefore, the correct answer is  $\boxed{\text{D}}$ . "Thirty-six people scored higher than both Mary and Pam."

\*\*\*\*\*

Question 84:

'A study looked at medical records of about 23,000 patients, mostly in Asia. They found that patients who drank at least three cups of green tea per day had a much lower incidence of depression. In an American newspaper article about this study, which of the following statements should not be made?', A. 'It is possible that people who drink green tea also tend to have healthier lifestyles than those who don't.', B. 'It is possible that people who drink green tea also tend to have a more positive outlook on life than those who don't.', C. 'Because this is observational data, the association between drinking green tea and a lower incidence of depression is not an indication of a causal relationship.', D. 'People who want to avoid depression should consider drinking green tea on a daily basis.'

=====

Answer Option 1:

The statement that should not be made is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.'

The reason is that the study only found an association between drinking green tea and a lower incidence of depression, not a causal relationship. This means that there may be other factors involved that contribute to the lower incidence of depression among green tea drinkers. Recommending green tea as a way to avoid depression based solely on this association would be misleading and potentially unsupported by the available evidence. Final Answer: D

\*\*\*\*\*

Answer Option 2:

The statement that should not be made is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.' This is because the study only found an association between drinking green tea and a lower incidence of depression, not a causal relationship. It is possible that there are other factors or variables that contribute to the lower incidence of depression among green tea drinkers. Therefore, it would be inappropriate to make a direct recommendation based solely on this observational study. Final answer: D

\*\*\*\*\*

Answer Option 3:

The correct answer is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.'

This statement cannot be made because although the study found an association between drinking green tea and a lower incidence of depression, it does not establish a causal relationship. The study is based on observational data, which means that other factors or variables not accounted for in the study could be contributing to the observed association. Therefore, it is not appropriate to make a direct recommendation for people to drink green tea to avoid depression based on this study alone. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The statement that should not be made is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.'

The study provides observational data, which can only establish an association or correlation between drinking green tea and a lower incidence of depression. It cannot prove a causal relationship. Therefore, it is not appropriate to make a direct recommendation for people to drink green tea to avoid depression based solely on this study. Other factors could be contributing to the observed association, and further research would be needed to establish a causal link. Therefore, option D should not be made in the newspaper article. Final answer: D

\*\*\*\*\*

Answer Option 5:

The statement that should not be made is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.'

This statement implies a direct causation between drinking green tea and avoiding depression, which cannot be inferred from the observational data presented in the study. While the study found an association between drinking green tea and a lower incidence of depression, it does not establish a cause-and-effect relationship. Other factors, such as healthier lifestyles or a more positive outlook, could be contributing to the observed association. Therefore, it would be inappropriate to make a direct recommendation based solely on the observational data.

The other statements (A, B, and C) acknowledge the limitations of the study and the possibility of confounding factors, which is a more accurate interpretation of the findings.

\*\*\*\*\*

Answer Option 6:

The statement that should not be made is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.'

This statement goes beyond the scope of the study's findings. While the study found an association between drinking green tea and a lower incidence of depression, it does not establish a causal relationship. Therefore, it is not appropriate to make a strong recommendation like "people should consider drinking green tea on a daily basis" based solely on this observational study. The other statements (A, B, C) are all valid interpretations of the study's findings and do not make strong claims beyond what the study suggests. Final answer: D

\*\*\*\*\*

Answer Option 7:

The correct statement that should not be made is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.'

While the study found an association between drinking green tea and a lower incidence of depression, it is important to note that this is observational data. Observational studies can only establish a correlation between variables, not a causal relationship. Therefore, it would be incorrect to make a direct recommendation based solely on this association.

Statements A and B are plausible because it is possible that individuals who drink green tea may also engage in healthier lifestyles or have a more positive outlook on life, which could contribute to lower rates of depression. Statement C is also correct because it acknowledges the limitations of observational data.

\*\*\*\*\*

Answer Option 8:

The statement that should not be made is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.' The reason is that the study only found an association between drinking green tea and a lower incidence of depression. It does not establish a causal relationship. Additionally, the study was observational in nature, which means that other factors could be contributing to the observed association. Therefore, it would be incorrect to make a direct recommendation based solely on this study. Final Answer: D



\*\*\*\*\*

Answer Option 9:

The statement that should not be made is D. 'People who want to avoid depression should consider drinking green tea on a daily basis.'

This statement goes beyond the scope of the study and the available evidence. While the study may have found an association between drinking green tea and a lower incidence of depression, it cannot establish a causal relationship. Additionally, there may be other factors or variables that contribute to the lower incidence of depression among green tea drinkers. Therefore, it is not appropriate to make a direct recommendation based solely on the observational data. Final Answer: D

\*\*\*\*\*

Question 85:

'A researcher planning a survey of school principals in a particular state has lists of the school principals employed in each of the 125 school districts. The procedure is to obtain a random sample of principals from each of the districts rather than grouping all the lists together and obtaining a sample from the entire group. Which of the following is a correct conclusion?', A. 'This is a simple random sample obtained in an easier and less costly manner than procedures involving sampling from the entire population of principals.', B. 'This is a cluster sample in which the population was divided into heterogeneous groups called clusters.', C. 'This is an example of systematic sampling, which gives a reasonable sample as long as the original order of the list is not related to the variables under consideration.', D. 'This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.'

=====

Answer Option 1:

The correct conclusion is D. 'This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.'

In this scenario, the researcher is dividing the population (school principals) into distinct groups (school districts) and obtaining a random sample from each group. This is known as stratified sampling. By doing so, the researcher can ensure that each group is represented in the sample, which allows for comparative analysis between different districts within the state. This method can provide more detailed and nuanced information compared to a simple random sample, where the entire population is treated as one group. Final Answer: D

\*\*\*\*\*

Answer Option 2:

The correct conclusion is D. This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.

In this case, the researcher has divided the population of school principals into 125 school districts and plans to obtain a random sample from each district. This approach is known as stratified sampling, where the population is divided into homogeneous groups or strata. By sampling from each district, the researcher ensures representation from all districts and can compare the responses and characteristics of principals across different districts. This comparative information would not be available with a simple random sample or other sampling methods. Final answer: D.

\*\*\*\*\*

Answer Option 3:

The correct conclusion is D. 'This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.'

In this sampling procedure, the population (school principals) is divided into different groups or strata (school districts). A random sample is then taken from each stratum, which ensures that principals from each district have a chance of being included in the survey.

This method allows for the comparison of responses and characteristics among different school districts, which a simple random sample would not provide. It allows for specific analysis of each stratum, and insights into any variations or patterns that may exist in different districts.

Therefore, Option D is the correct conclusion.

\*\*\*\*\*

Answer Option 4:

The correct conclusion is D. 'This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.'

In this scenario, the researcher is obtaining a random sample from each of the 125 school districts. By doing so, they are dividing the population (school principals) into homogeneous groups (school districts) and then sampling from each group. This is known as stratified sampling.

Stratified sampling is useful when there are subgroups within the population that may have different characteristics or proportions. By ensuring that each subgroup is represented in the sample, the researcher can obtain comparative information that a simple random sample wouldn't provide.

Options A, B, and C are incorrect because they do not accurately describe the sampling procedure described in the scenario.

\*\*\*\*\*

Answer Option 5:

The correct conclusion is D. 'This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.'

In this scenario, the population of school principals is divided into school districts, and a random sample is obtained from each district. This is an example of stratified sampling because the population is divided into homogeneous groups (school districts) and a sample is taken from each group. This method allows for comparisons and analysis between the different school districts, which a simple random sample wouldn't provide.

Option A is incorrect because it implies that this method is easier and less costly, which is not necessarily true. Option B is incorrect because cluster sampling involves randomly selecting entire clusters, whereas in this scenario, individual principals are selected from each district. Option C is incorrect because systematic sampling involves selecting every kth element from a list, which is not the case here.

\*\*\*\*\*

Answer Option 6:

The correct conclusion is D. 'This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.'

In this scenario, the researcher is dividing the population (school principals) into homogeneous groups (school districts) and obtaining a random sample from each group. This is known as stratified sampling. Stratified sampling is useful when there are important subgroups within the population that may have different characteristics or variations. By obtaining a sample from each district, the researcher can ensure that principals from each district are represented in the sample, allowing for comparative information between the districts.

Option A is incorrect because this sampling procedure does not involve sampling from the entire population of principals.

Option B is incorrect because in cluster sampling, the population is divided into heterogeneous groups, not homogeneous groups.

Option C is incorrect because systematic sampling involves selecting every nth individual from a list, which is different from the procedure described in the question.

\*\*\*\*\*

Answer Option 7:

The correct conclusion is B. This is a cluster sample in which the population was divided into heterogeneous groups called clusters. In this case, the clusters are the individual school districts. The researcher is obtaining a random sample from each of the districts, rather than sampling from the entire population of principals. This method is less costly and easier to implement compared to sampling from the entire population of principals. Final answer: B.

\*\*\*\*\*

Answer Option 8:

The correct conclusion is D. 'This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.'

In this case, the population (school principals) is divided into homogeneous groups (school districts) called strata. A random sample of principals is then obtained from each district. This allows for comparison and analysis of data at the district level, which a simple random sample from the entire population would not provide.

Option A is incorrect because it implies that this method is easier and less costly, which is not necessarily true. Option B is incorrect because cluster sampling involves dividing the population into heterogeneous groups, not homogeneous groups. Option C is incorrect because systematic sampling involves selecting individuals at regular intervals from an ordered list, which is not the case here. Final

Answer: D

\*\*\*\*\*

Answer Option 9:

The correct conclusion is D. 'This is a stratified sample, which may give comparative information that a simple random sample wouldn't give.'

In this scenario, the researcher is obtaining a random sample of principals from each of the 125 school districts. This is an example of stratified sampling, where the population (school principals) is divided into homogeneous groups (school districts) and a random sample is taken from each group.

Stratified sampling allows for the comparison of different groups within the population, in this case, the different school districts. This can provide valuable information and insights that a simple random sample from the entire population may not be able to capture.

Option A is incorrect because the sampling procedure described is not an easier or less costly method compared to sampling from the entire population.

Option B is incorrect because the sampling procedure described is not a cluster sample, where the population is divided into heterogeneous groups.

Option C is incorrect because systematic sampling involves selecting every nth element from a list, which is not the case in this scenario.

\*\*\*\*\*

Question 86:

'In an experiment, the control group should receive', A. 'treatment opposite that given the experimental group.', B. 'the same treatment given the experimental group without knowing they are receiving the treatment.', C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.', D. 'a procedure identical to that given the experimental group except for a random decision on receiving the treatment under examination.'

=====

Answer Option 1:

The control group is a group in an experiment that does not receive the treatment under examination. Its purpose is to provide a baseline for comparison with the experimental group. Therefore, the correct answer is C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.' Final answer: C.

\*\*\*\*\*

Answer Option 2:

The correct answer is C. The control group should be treated identically to the experimental group, except for not receiving the specific treatment under examination. This allows for comparison between the effects of the treatment and the absence of the treatment. The control group provides a baseline for comparison and helps to determine the effectiveness of the treatment being studied. Final answer: C.

\*\*\*\*\*

Answer Option 3:

The correct answer is C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.' The control group is used as a benchmark to compare the effects of the treatment being tested in the experimental group. Therefore, the control group should be treated the same as the experimental group, with the exception that they do not receive the specific treatment under examination. This allows researchers to determine the impact of the treatment by comparing the outcomes of the experimental group to those of the control group. Final answer: C

\*\*\*\*\*

Answer Option 4:

The correct answer is C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.'

In an experiment, the control group serves as a baseline group that does not receive the treatment being tested. This allows researchers to compare the effects of the treatment by comparing the experimental group (which receives the treatment) to the control group (which does not receive the treatment). To ensure a fair comparison, the control group should undergo all other procedures that the experimental group undergoes, except for receiving the treatment. This helps isolate the effects of the treatment being examined. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The correct answer is C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.'

The control group serves as a baseline for comparison in an experiment. It should be treated in the same way as the experimental group, except for the specific treatment being tested. This allows researchers to isolate the effects of the treatment by comparing the outcomes of the control group to those of the experimental group. The control group helps researchers determine whether the observed effects are due to the treatment or other factors. Therefore, the control group should receive the same procedure as the experimental group, but without the treatment under examination. Final answer: C.

\*\*\*\*\*

Answer Option 6:

The correct answer is C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.'

In an experiment, the control group is used as a baseline to compare the effect of the treatment given to the experimental group. To accurately measure the effectiveness of the treatment, it is important to have a group that is not exposed to the treatment. This control group should be treated in the same way as the experimental group, except for the specific treatment being tested. This allows researchers to isolate the effect of the treatment and determine if it has any impact compared to the baseline. Final Answer: C

\*\*\*\*\*

Answer Option 7:

The correct answer is C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.'

In an experiment, the control group serves as a baseline for comparison. It is used to assess the effect of the treatment being tested by providing a standard against which the experimental group can be evaluated. In order to isolate the effect of the treatment, the control group should not receive the treatment under examination. Instead, they should receive the same procedure and conditions as the experimental group, except for the treatment itself. This allows researchers to compare the outcomes between the two groups and determine the effectiveness of the treatment. Therefore, option C is the most accurate description of what the control group should receive.

\*\*\*\*\*

Answer Option 8:

The correct answer is C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.'

The control group serves as a baseline comparison for the experimental group. Therefore, the control group should be treated in the same way as the experimental group, with the only difference being the absence of the specific treatment under examination. This allows researchers to isolate and measure the effects of the treatment by comparing the outcomes of the experimental group to those of the control group. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct answer is C. 'a procedure identical to that given the experimental group except for receiving the treatment under examination.'

In an experiment, the control group is used as a baseline for comparison. It is important to have a control group that is treated identically to the experimental group, except for the variable being tested. This allows researchers to isolate the effects of the treatment and determine its true impact. By comparing the control group's results to those of the experimental group, researchers can determine whether the

treatment is effective or not. Therefore, the control group should receive the same procedure as the experimental group, except for the treatment under examination. Final answer: C.

\*\*\*\*\*

Question 87:

Tina's science fair project was to estimate the mean amount of chemicals in her city's water supply. At first, she had decided to use a random sample of 15 observations. But her teacher asked her to take 35 observations. The mean and standard deviation from 35 observations turned out to be approximately the same as those from 15 observations. Is there any advantage in using 35 observations instead of 15 observations? A. 'There is no advantage. Because the mean and the standard deviation are about the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations.'. B. 'There is no advantage. In fact, the 20 extra observations will increase the likelihood of error.'. C. 'There is no advantage. Because she took 35 observations instead of 15 observations, the confidence interval using 35 observations will be wider than that using 15 observations.'. D. 'There is some advantage. Because she took 35 observations instead of 15 observations, the confidence interval using 35 observations will be narrower than that using 15 observations.'

=====

Answer Option 1:

The larger sample size (35 observations) provides Tina with a more accurate estimate of the mean amount of chemicals in the water supply. With a larger sample size, the standard deviation becomes a more reliable measure of variability, leading to a narrower confidence interval. Therefore, the correct answer is D. 'There is some advantage. Because she took 35 observations instead of 15 observations, the confidence interval using 35 observations will be narrower than that using 15 observations.' Final Answer: D

\*\*\*\*\*

Answer Option 2:

The advantage of increasing the sample size is that it generally leads to a more accurate estimation of the population mean. In this case, since the mean and standard deviation from the 35 observations are approximately the same as those from the 15 observations, it suggests that the larger sample size of 35 provides a more reliable estimate of the population mean. Therefore, the correct answer is D. 'There is some advantage. Because she took 35 observations instead of 15 observations, the confidence interval using 35 observations will be narrower than that using 15 observations.' Final Answer: D

\*\*\*\*\*

Answer Option 3:

Based on the given information, since the mean and standard deviation from the 35 observations are approximately the same as those from the 15 observations, it can be concluded that there is no advantage in using 35 observations instead of 15 observations. The confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations. Therefore, the correct answer is  $\boxed{\text{A. There is no advantage. Because the mean and the standard deviation are about the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations.}}$

\*\*\*\*\*

Answer Option 4:

Since the mean and standard deviation from 35 observations are approximately the same as those from 15 observations, it suggests that the sample mean and standard deviation have stabilized and are not significantly changing with additional observations. Therefore, there is no advantage in using 35 observations instead of 15 observations. The confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations. Thus, the correct answer is  $\boxed{\text{A. There is no advantage. Because the mean and the standard deviation are about the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations.}}$ . Final answer: A

\*\*\*\*\*

Answer Option 5:

The advantage of using a larger sample size is that it generally leads to a more accurate estimate of the population mean. In this case, since the mean and standard deviation from the larger sample size of 35 observations are approximately the same as those from the smaller sample size of 15 observations, it suggests that the population mean is relatively stable and not likely to change significantly with additional observations.

Therefore, the correct answer is:

A. 'There is no advantage. Because the mean and the standard deviation are about the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations.'

\*\*\*\*\*

Answer Option 6:

Option A is the correct answer: 'There is no advantage. Because the mean and the standard deviation are about the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations.'

The mean and standard deviation are measures of central tendency and variability, respectively. In this case, since the mean and standard deviation from both the 15 and 35 observations are approximately the same, it suggests that the data points in the additional 20 observations do not significantly affect the overall mean and variability of the data. Therefore, there is no advantage in using 35 observations instead of 15 observations in terms of estimating the mean amount of chemicals in the water supply. The confidence interval, which is associated with the mean and standard deviation, should be approximately the same regardless of the sample size. Final Answer: A

\*\*\*\*\*

Answer Option 7:

The correct answer is A. 'There is no advantage. Because the mean and the standard deviation are about the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations.'

When the mean and standard deviation from 35 observations are approximately the same as those from 15 observations, it suggests that the additional 20 observations did not provide any new or valuable information. In this case, using 35 observations instead of 15 observations does not provide any advantage in terms of estimating the mean amount of chemicals in the water supply.

The confidence interval is a measure of the precision of the estimate. Since the mean and standard deviation are approximately the same, it is reasonable to expect that the confidence interval computed using 35 observations will be approximately the same as that computed using 15 observations. Therefore, there is no advantage in using 35 observations instead of 15 observations. Final Answer: A

\*\*\*\*\*

Answer Option 8:

The correct answer is A. 'There is no advantage. Because the mean and the standard deviation are about the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations.'

The mean and standard deviation are measures of the central tendency and variability of the data, respectively. In this case, since the mean and standard deviation from 35 observations are approximately the same as those from 15 observations, it suggests that the additional 20 observations did not provide any new or different information about the population.

Therefore, there is no advantage in using 35 observations instead of 15 observations. The confidence interval, which is a measure of the precision of the estimate, will be approximately the same using either sample size.

\*\*\*\*\*

Answer Option 9:

The answer is C, 'There is no advantage. Because she took 35 observations instead of 15 observations, the confidence interval using 35 observations will be wider than that using 15 observations.'

When estimating the mean, a larger sample size generally leads to a more accurate estimate and a narrower confidence interval. However, in this case, we are given that the mean and standard deviation from the 35 observations are approximately the same as those from the 15 observations. This suggests that the additional 20 observations did not provide any additional information or decrease the variability in the data.

Since the mean and standard deviation are the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations. Therefore, there is no advantage in using 35 observations instead of 15 observations. In fact, the confidence interval using 35 observations may be wider due to the increased sample size.

\*\*\*\*\*

Question 88:

'The total cholesterol level in a large population of people is strongly skewed right with a mean of 210 mg/dL and a standard deviation of 15 mg/dL. If random samples of size 16 are repeatedly drawn from this population, which of the following appropriately describes the sampling distribution of these sample means?', A. 'The shape is unknown with a mean of 210 and a standard deviation of 15.', B. 'The shape is somewhat skewed right with a mean of 210 and a standard deviation of 3.75.', C. 'The shape is approximately normal with a mean of 210 and a standard deviation of 15.', D. 'The shape is approximately normal with a mean of 210 and a standard deviation of 3.75.'

=====

Answer Option 1:

According to the Central Limit Theorem, for a large enough sample size, the sampling distribution of the sample means will be approximately normal, regardless of the shape of the original population distribution.

The mean of the sampling distribution of the sample means will be equal to the mean of the population, which is 210 mg/dL.

The standard deviation of the sampling distribution of the sample means, also known as the standard error, is equal to the standard deviation of the population divided by the square root of the sample size. In this case, the standard deviation of the population is 15 mg/dL and the sample size is 16. Therefore, the standard deviation of the sampling distribution of the sample means is  $15/\sqrt{16} = 3.75$  mg/dL.

Hence, the sampling distribution of these sample means can be appropriately described as approximately normal with a mean of 210 and a standard deviation of 3.75. Therefore, the correct option is D. 'The shape is approximately normal with a mean of 210 and a standard deviation of 3.75.' Final Answer: D

\*\*\*\*\*

Answer Option 2:

According to the Central Limit Theorem, for a large enough sample size, the sampling distribution of the sample means will be approximately normal, regardless of the shape of the population distribution.

Therefore, the correct answer is C. 'The shape is approximately normal with a mean of 210 and a standard deviation of 15.' Final Answer: C

\*\*\*\*\*

Answer Option 3:

According to the Central Limit Theorem, when random samples of size  $n$  are drawn from a population, the sampling distribution of the sample means will be approximately normal, regardless of the shape of the population distribution, as long as the sample size is sufficiently large (usually  $n \geq 30$ ).

In this case, since the sample size is 16 (which is less than 30), we cannot immediately conclude that the sampling distribution of the sample means will be exactly normal. However, the population distribution is strongly skewed right, and the Central Limit Theorem suggests that for larger sample sizes, the sampling distribution will become more approximately normal.

Therefore, the correct answer is  $\boxed{\text{The shape is approximately normal with a mean of 210 and a standard deviation of 15.}}$  Final Answer: C

\*\*\*\*\*

Answer Option 4:

The sampling distribution of sample means follows the Central Limit Theorem. For large sample sizes, regardless of the shape of the population distribution, the sampling distribution of sample means will be approximately normal. The mean of the sampling distribution of sample means is equal to the population mean, and the standard deviation is given by the population standard deviation divided by the square root of the sample size.

Therefore, the correct answer is: 'The shape is approximately normal with a mean of 210 and a standard deviation of  $\frac{15}{\sqrt{16}}=3.75$ .' Hence, the correct answer is  $\boxed{\text{D.}}$

\*\*\*\*\*

Answer Option 5:

According to the Central Limit Theorem, when random samples of sufficiently large size are drawn from a population, the sampling distribution of the sample means will be approximately normal, regardless of the shape of the population distribution. Additionally, the mean of the sampling distribution of the sample means will be equal to the mean of the population, and the standard deviation of the sampling distribution of the sample means will be equal to the standard deviation of the population divided by the square root of the sample size.

In this case, the sampling distribution of the sample means will have a mean of 210 mg/dL (same as the population mean) and a standard deviation of 15 mg/dL divided by the square root of 16 (sample size).

Simplifying, we get a standard deviation of  $15/4 = 3.75$  mg/dL.

Therefore, the correct answer is  $\boxed{\text{D. 'The shape is approximately normal with a mean of 210 and a standard deviation of 3.75.'}}$  Final answer: D

\*\*\*\*\*

Answer Option 6:

According to the Central Limit Theorem, when random samples of sufficient size are repeatedly drawn from any population, the sampling distribution of the sample means will be approximately normal, regardless of the shape of the original population.

The mean of the sampling distribution of the sample means is equal to the mean of the population, which is given as 210 mg/dL. The standard deviation of the sampling distribution of the sample means is equal to the standard deviation of the population, divided by the square root of the sample size. In this case, the standard deviation of the population is given as 15 mg/dL, and the sample size is 16.

Therefore, the standard deviation of the sampling distribution of the sample means is 15 mg/dL divided by the square root of 16, which is 3.75 mg/dL.

Thus, the appropriate description of the sampling distribution of these sample means is: 'The shape is approximately normal with a mean of 210 and a standard deviation of 3.75.' Therefore, the correct answer is  $\boxed{\text{D.}}$

\*\*\*\*\*

Answer Option 7:

The sampling distribution of sample means from a large population follows the Central Limit Theorem. According to the Central Limit Theorem, regardless of the shape of the population distribution, the sampling distribution of the sample means will approach a normal distribution as the sample size increases. Additionally, the mean of the sampling distribution will be equal to the mean of the population, and the standard deviation of the sampling distribution (also known as the standard error) will be equal to the population standard deviation divided by the square root of the sample size.



In this case, the sampling distribution of sample means will be approximately normal with a mean of 210 (same as the population mean) and a standard deviation of 15 divided by the square root of 16 (since the sample size is 16). Thus, the standard deviation of the sampling distribution is  $15/4 = 3.75$ .

Therefore, the appropriate description is: "The shape is approximately normal with a mean of 210 and a standard deviation of 3.75." Hence, the answer is  $\boxed{\text{(D)}}$ .

\*\*\*\*\*

Answer Option 8:

The sampling distribution of sample means follows the central limit theorem. According to the central limit theorem, regardless of the shape of the original population, as the sample size increases, the sampling distribution of sample means will approach a normal distribution.

In this case, the population is strongly skewed right. However, since the sample size is 16 (which is reasonably large) and the central limit theorem applies, the sampling distribution of sample means will be approximately normal.

The mean of the sampling distribution of sample means is equal to the mean of the population, which is 210 mg/dL.

The standard deviation of the sampling distribution of sample means is equal to the population standard deviation divided by the square root of the sample size. Hence, the standard deviation of the sampling distribution of sample means is  $15 \text{ mg/dL} / \sqrt{16} = 15 \text{ mg/dL} / 4 = 3.75 \text{ mg/dL}$ .

Therefore, the appropriate description of the sampling distribution of these sample means is: 'The shape is approximately normal with a mean of 210 and a standard deviation of 3.75,' which corresponds to answer choice D. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The sampling distribution of sample means follows the Central Limit Theorem, which states that for a large enough sample size, the sampling distribution will be approximately normal regardless of the shape of the population distribution.

In this case, the population distribution is strongly skewed right. However, since the sample size is 16 (which is considered fairly large), the sampling distribution of sample means will be approximately normal.

The mean of the sampling distribution will be equal to the mean of the population, which is 210 mg/dL.

The standard deviation of the sampling distribution, also known as the standard error, is calculated by dividing the standard deviation of the population by the square root of the sample size. In this case, the standard deviation of the population is 15 mg/dL, and the sample size is 16. Therefore, the standard deviation of the sampling distribution is  $15 / \sqrt{16} = 15 / 4 = 3.75 \text{ mg/dL}$ .

Thus, the appropriate description of the sampling distribution of sample means is: 'The shape is approximately normal with a mean of 210 and a standard deviation of 3.75.' This corresponds to option D. Final

Answer: D

\*\*\*\*\*

Question 89:

'An experiment is conducted in which the response variable is the average gain in participants' performance in the long jump. A two-sample t-test with a 5% level of significance will be used to analyze the results. If all else is kept the same, which of the following descriptions of a possible change in procedure is true?', A. 'Change from equal size treatment groups to very different size treatment groups would increase the power of the test.', B. 'Change from a 5% significance level to a 1% significance level would increase the power of the test.', C. 'Taking more careful measurements to reduce variability in the response would increase the power of the test.', D. 'Increasing the sample size would reduce the probability of a Type I error.'

=====

Answer Option 1:

The power of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. It is influenced by several factors, including the significance level, sample size, effect size, and variability in the data.

A. Changing from equal size treatment groups to very different size treatment groups would not necessarily increase the power of the test. The power is generally maximized when the sample sizes in the two groups are equal.

B. Changing from a 5% significance level to a 1% significance level would not increase the power of the test. A lower significance level reduces the probability of a Type I error (incorrectly rejecting the null hypothesis), but it does not directly impact the power of the test.

C. Taking more careful measurements to reduce variability in the response would increase the power of the test. When there is less variability in the data, the effect size becomes more apparent, making it easier to detect a significant difference between groups.

D. Increasing the sample size would increase the power of the test. A larger sample size provides more information and reduces the standard error, resulting in increased power to detect a true difference.

Therefore, the correct statement is: 'Taking more careful measurements to reduce variability in the response would increase the power of the test.' Option C is the correct answer. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The power of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. To increase the power of the test, we want to increase the probability of detecting a true difference in the population.

A. Changing from equal size treatment groups to very different size treatment groups would not increase the power of the test. In fact, having equal size treatment groups is preferred for a balanced and accurate comparison.

B. Changing from a 5% significance level to a 1% significance level would not increase the power of the test. Lowering the significance level decreases the probability of rejecting the null hypothesis, making it more difficult to detect a true difference.

C. Taking more careful measurements to reduce variability in the response would increase the power of the test. By reducing variability, the signal of the true difference becomes stronger, making it easier to detect.

D. Increasing the sample size would not reduce the probability of a Type I error. The probability of a Type I error is determined by the significance level chosen for the test, not the sample size. Increasing the sample size can, however, increase the power of the test by reducing sampling variability and increasing the precision of the estimate.

Therefore, the correct description of a possible change in procedure that would increase the power of the test is C. 'Taking more careful measurements to reduce variability in the response would increase the power of the test.' Final Answer: C

\*\*\*\*\*

Answer Option 3:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when it is false. In this case, the null hypothesis is that there is no difference in the average gain in participants' performance in the long jump between the treatment groups.

A. Changing from equal size treatment groups to very different size treatment groups would not increase the power of the test. The power of the test depends on the sample size, not the sizes of the treatment groups.

B. Changing from a 5% significance level to a 1% significance level would not increase the power of the test. A lower significance level would actually decrease the power of the test, as it would require stronger evidence to reject the null hypothesis.

C. Taking more careful measurements to reduce variability in the response would increase the power of the test. Reducing variability would make it easier to detect a significant difference between the treatment groups.

D. Increasing the sample size would increase the power of the test. A larger sample size provides more information and reduces the variability, making it easier to detect a significant difference.

Therefore, the correct description of a possible change in procedure that would increase the power of the test is: 'Increasing the sample size would reduce the probability of a Type I error.' The answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 4:

The power of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. In this case, the null hypothesis is that there is no difference in the average gain in performance between the treatment groups.

A. Changing from equal size treatment groups to very different size treatment groups would not necessarily increase the power of the test. In fact, having unequal group sizes can decrease the power of the test because it can introduce bias.

B. Changing from a 5% significance level to a 1% significance level would not increase the power of the test. A lower significance level reduces the probability of a Type I error (rejecting the null hypothesis when it is true), but it does not necessarily improve the ability to detect a true effect (increase power).

C. Taking more careful measurements to reduce variability in the response would increase the power of the test. By reducing variability, the signal (difference in average gain in performance) becomes more distinguishable from the noise (random variability), making it easier to detect a true effect.

D. Increasing the sample size would increase the power of the test. A larger sample size provides more information and reduces random variability, making it easier to detect a true effect.

Therefore, the correct description of a possible change in procedure that would increase the power of the test is D. 'Increasing the sample size would reduce the probability of a Type I error.' Final Answer: D  
\*\*\*\*\*

Answer Option 5:

The power of a statistical test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the alternative hypothesis is that there is a difference in the average gain in performance between the treatment groups.

A. Changing from equal size treatment groups to very different size treatment groups would not necessarily increase the power of the test. The power of the test depends on the sample size and the effect size, but not on the size of the treatment groups.

B. Changing from a 5% significance level to a 1% significance level would not increase the power of the test. The significance level determines the threshold for rejecting the null hypothesis, but it does not directly affect the power of the test.

C. Taking more careful measurements to reduce variability in the response would likely increase the power of the test. Reducing variability can make the effect size more apparent and increase the likelihood of detecting a significant difference.

D. Increasing the sample size would increase the power of the test. With a larger sample size, there is a greater chance of detecting a true difference and reducing the probability of a Type II error (failing to reject the null hypothesis when it is false).

Therefore, the correct answer is  $\boxed{\text{C. "Taking more careful measurements to reduce variability in the response would increase the power of the test."}}$   
\*\*\*\*\*

Answer Option 6:

The power of a statistical test is the probability of correctly rejecting the null hypothesis, given that the alternative hypothesis is true. Increasing the power of the test means increasing the probability of detecting a true effect.

A. Changing from equal size treatment groups to very different size treatment groups would not increase the power of the test. The power of the test depends on the sample size, not the size of the treatment groups.

B. Changing from a 5% significance level to a 1% significance level would decrease the power of the test. A lower significance level means that we have a higher threshold for rejecting the null hypothesis, making it harder to detect a true effect.

C. Taking more careful measurements to reduce variability in the response would increase the power of the test. Reducing variability in the response variable increases the signal-to-noise ratio, making it easier to detect a true effect.

D. Increasing the sample size would increase the power of the test. A larger sample size provides more information and reduces the sampling variability, making it easier to detect a true effect.

Therefore, the correct answer is  $\boxed{\text{C. "Taking more careful measurements to reduce variability in the response would increase the power of the test."}}$   
\*\*\*\*\*

Answer Option 7:

The power of a statistical test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the null hypothesis is that there is no difference in the average gain in performance between the treatment groups, and the alternative hypothesis is that there is a difference.

A. Changing from equal size treatment groups to very different size treatment groups would not necessarily increase the power of the test. The power of the test depends on the sample size, not the size of the treatment groups.

B. Changing from a 5% significance level to a 1% significance level would not increase the power of the test. A lower significance level actually decreases the power of the test because it makes it more difficult to reject the null hypothesis.

C. Taking more careful measurements to reduce variability in the response would increase the power of the test. By reducing variability, the effect size becomes larger relative to the noise, making it easier to detect a significant difference.

D. Increasing the sample size increases the power of the test. With a larger sample size, there is more information available to make a decision, and the test is more likely to correctly reject the null hypothesis when there is a true difference.

Therefore, the correct answer is  $\boxed{\text{D. Increasing the sample size would reduce the probability of a Type I error.}}$  Final Answer: D  
\*\*\*\*\*

Answer Option 8:

The power of a statistical test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. Increasing the power of the test means increasing the probability of detecting a true effect.

A. Changing from equal size treatment groups to very different size treatment groups would not increase the power of the test. In fact, having different size treatment groups can lead to biased results and decreased power.

B. Changing from a 5% significance level to a 1% significance level would decrease the power of the test. A lower significance level makes it harder to reject the null hypothesis, reducing the probability of detecting a true effect.

C. Taking more careful measurements to reduce variability in the response would increase the power of the test. Reducing variability increases the signal-to-noise ratio, making it easier to detect a true effect.

D. Increasing the sample size would reduce the probability of a Type I error. Type I error refers to rejecting the null hypothesis when it is actually true. Increasing the sample size reduces the chance of this error as it provides more evidence to support the conclusion.

Therefore, the correct answer is C. 'Taking more careful measurements to reduce variability in the response would increase the power of the test.' Final Answer: C

\*\*\*\*\*

Answer Option 9:

The power of a statistical test is the probability that the test correctly rejects the null hypothesis when the alternative hypothesis is true. In this case, the alternative hypothesis would be that there is a difference in the average gain in participants' performance in the long jump between the treatment groups.

A change from equal size treatment groups to very different size treatment groups would not necessarily increase the power of the test. In fact, having very different size treatment groups may introduce bias and affect the validity of the test results. Therefore, option A is not true.

Changing from a 5% significance level to a 1% significance level would not increase the power of the test. A lower significance level reduces the probability of rejecting the null hypothesis when it is true, which means it reduces the chance of detecting a true difference. Therefore, option B is not true.

Taking more careful measurements to reduce variability in the response would increase the power of the test. A smaller variability in the response variable increases the ability to detect a true difference between the treatment groups. Therefore, option C is true.

Increasing the sample size would not reduce the probability of a Type I error, which is the rejection of the null hypothesis when it is true. The probability of a Type I error is determined by the significance level, not the sample size. Therefore, option D is not true.

Thus, the correct statement is option C: 'Taking more careful measurements to reduce variability in the response would increase the power of the test.' Final Answer: C

\*\*\*\*\*

Question 90:

'An assembly line machine is supposed to turn out ball bearings with a diameter of 1.25 centimeters. Each morning the first 30 bearings produced are pulled and measured. If their mean diameter is under 1.23 centimeters or over 1.27 centimeters, the machinery is stopped and an engineer is called to make adjustments before production is resumed. The quality control procedure may be viewed as a hypothesis test with the null hypothesis  $H_0: \mu = 1.25$  and the alternative hypothesis  $H_a: \mu \neq 1.25$ . The engineer is asked to make adjustments when the null hypothesis is rejected. In test terminology, what would a Type II error result in?', A. 'A warranted halt in production to adjust the machinery', B. 'An unnecessary stoppage of the production process', C. 'Continued production of wrong size ball bearings', D. 'Continued production of proper size ball bearings'

=====

Answer Option 1:

A Type II error occurs when the null hypothesis is true, but we incorrectly fail to reject it. In this case, it means that the mean diameter of the ball bearings is actually different from 1.25 centimeters, but we fail to detect this and continue production without making any adjustments. Therefore, a Type II error would result in the production of ball bearings with the wrong size. Hence, the correct answer is

$\boxed{\text{C. Continued production of wrong size ball bearings}}$ .

\*\*\*\*\*

Answer Option 2:

A Type II error occurs when the null hypothesis is not rejected, even though it is false. In this case, the null hypothesis is that the mean diameter of the ball bearings is 1.25 centimeters. Therefore, a Type II error would result in the production process continuing, even though the ball bearings are not actually the proper size.

Therefore, the correct answer is C. 'Continued production of wrong size ball bearings'. Final answer: C.

\*\*\*\*\*

Answer Option 3:

A Type II error occurs when the null hypothesis is not rejected, even though it is false. In this case, the null hypothesis is that the mean diameter of the ball bearings is 1.25 centimeters. So, a Type II error would result in accepting the null hypothesis and continuing production, even though the mean diameter is actually different from 1.25 centimeters. Therefore, a Type II error would result in the continued production of ball bearings that are not the proper size. The correct answer is  $\boxed{\text{C. 'Continued production of wrong size ball bearings'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 4:

A Type II error in hypothesis testing occurs when we fail to reject the null hypothesis, even though the alternative hypothesis is true. In this case, it means that the engineer will fail to make adjustments to the machinery even though the mean diameter of the ball bearings is different from the desired value of 1.25 centimeters. Therefore, a Type II error would result in the continued production of ball bearings that are not the proper size. Hence, the correct answer is  $\boxed{\text{C. 'Continued production of wrong size ball bearings'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 5:

A Type II error occurs when the null hypothesis is not rejected, even though it is actually false. In this case, the null hypothesis is that the mean diameter of the ball bearings is 1.25 centimeters. If a Type II error occurs, it means that the machinery is not stopped and adjustments are not made, even though the mean diameter is actually different from 1.25 centimeters. Therefore, a Type II error would result in the continued production of ball bearings with the wrong size. The correct answer is  $\boxed{\text{C}} \setminus \text{Continued production of wrong size ball bearings}$ . Final Answer: C

\*\*\*\*\*

Answer Option 6:

In test terminology, a Type II error occurs when the null hypothesis is not rejected, but it is actually false. In this scenario, the null hypothesis is that the mean diameter is 1.25 centimeters. A Type II error would mean that the machinery is not stopped and adjusted, even though the mean diameter is actually under 1.23 centimeters or over 1.27 centimeters. Therefore, a Type II error would result in the continued production of wrong size ball bearings. The correct answer is  $\boxed{\text{C. 'Continued production of wrong size ball bearings'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 7:

A Type II error in hypothesis testing occurs when we fail to reject the null hypothesis when it is actually false. In this case, the null hypothesis is that the mean diameter of the ball bearings is 1.25 centimeters. If a Type II error occurs, it means that we failed to detect that the mean diameter is different from 1.25 centimeters, and thus we would continue production of ball bearings without making any adjustments to the machinery. Therefore, a Type II error in this context would result in the continued production of ball bearings of the wrong size. Hence, the correct answer is  $\boxed{\text{C. 'Continued production of wrong size ball bearings'}}$ . Final answer: C

\*\*\*\*\*

Answer Option 8:

A Type II error occurs when the null hypothesis is not rejected, even though it is false (i.e., the alternative hypothesis is true). In this case, the null hypothesis states that the mean diameter of the ball bearings is 1.25 centimeters. If a Type II error occurs, it means that the machinery is not stopped and no adjustments are made, even though the mean diameter is actually different from 1.25 centimeters. Therefore, a Type II error would result in the continued production of ball bearings with the wrong size. Hence, the correct answer is  $\boxed{\text{C. 'Continued production of wrong size ball bearings'}}$ .

\*\*\*\*\*

Answer Option 9:

A Type II error occurs when the null hypothesis is actually false, but we fail to reject it. In this case, the null hypothesis is that the mean diameter of the ball bearings is 1.25 centimeters. If there is a Type II error, it means that the mean diameter is actually different from 1.25 centimeters (i.e., the alternative hypothesis is true), but we fail to detect this and continue production without making adjustments. Therefore, a Type II error would result in the production of ball bearings with the wrong size, leading to the answer choice C. 'Continued production of wrong size ball bearings'. Final Answer: C

\*\*\*\*\*



Question 91:

'We are interested in the proportion  $p$  of people who are unemployed in a large city. Eight percent of a simple random sample of 500 people are unemployed. What is the midpoint for a 95% confidence interval estimate of  $p$ ?' A. '0.012', B. '0.025', C. '0.475', D. 'None of the above.'

=====

Answer Option 1:

The midpoint for a confidence interval estimate of a proportion is simply the point estimate of the proportion itself. In this case, we are given that 8% of a random sample of 500 people are unemployed.

Therefore, the midpoint for the 95% confidence interval estimate of the proportion of people who are unemployed is 8%.

Thus, the correct answer is D. 'None of the above.'

\*\*\*\*\*

Answer Option 2:

The midpoint for a 95% confidence interval estimate of a proportion can be found by taking the sample proportion and adding/subtracting the margin of error. The margin of error is equal to the critical value multiplied by the standard error.

The critical value for a 95% confidence interval is approximately 1.96 (assuming a large sample size). The standard error can be calculated as the square root of  $(p(1-p))/n$ , where  $p$  is the sample proportion and  $n$  is the sample size.

In this case, the sample proportion is 0.08 and the sample size is 500. So the standard error is  $\sqrt{(0.08(1-0.08))/500} = 0.012$ .

The margin of error is then  $1.96 * 0.012 = 0.02352$ .

The midpoint for the confidence interval is the sample proportion (0.08) plus/minus the margin of error (0.02352). So the midpoint is  $0.08 \pm 0.02352 = [0.05648, 0.10352]$ .

Therefore, the correct answer is 'None of the above'. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The midpoint for a 95% confidence interval estimate of a proportion can be calculated using the formula:

Midpoint = Sample proportion  $\pm$  Margin of error

Since we are given that 8% of a simple random sample of 500 people are unemployed, the sample proportion is 0.08.

To calculate the margin of error, we use the formula:

Margin of error = Critical value \* Standard error

The critical value for a 95% confidence interval for a proportion is approximately 1.96.

The standard error can be calculated using the formula:

Standard error =  $\sqrt{(p \cdot q) / n}$

where  $p$  is the sample proportion,  $q$  is  $1 - p$ , and  $n$  is the sample size.

Using the given information, we have:

$p = 0.08$

$q = 1 - p = 0.92$

$n = 500$

Plugging these values into the formula, we get:

Standard error =  $\sqrt{(0.08 * 0.92) / 500} = 0.013$

Finally, we can calculate the midpoint:

Midpoint =  $0.08 \pm (1.96 * 0.013) = 0.08 \pm 0.025$

Therefore, the midpoint for a 95% confidence interval estimate of  $p$  is approximately  $0.08 \pm 0.025$ .

None of the available answer choices match this value, so the correct answer is D. None of the above.

\*\*\*\*\*

Answer Option 4:

The midpoint for a 95% confidence interval estimate of a proportion is equal to the point estimate, which in this case is the sample proportion. The sample proportion is given as 8% or 0.08. Therefore, the

midpoint for the confidence interval is 0.08. None of the given options A, B, or C correspond to the correct midpoint, so the answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The midpoint of a confidence interval is the point estimate, which is the sample statistic. In this case, the sample statistic is the proportion of the sample that is unemployed, which is 0.08 (or 8%). Therefore,

the midpoint for a 95% confidence interval estimate of  $p$  is 0.08. None of the given answer choices (A, B, C) match this, so the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 6:

A 95% confidence interval estimate for a proportion is given by

$$\left[ p \pm 1.96 \sqrt{\frac{p(1-p)}{n}} \right]$$

where  $p$  is the sample proportion and  $n$  is the sample size. In this case, we have  $p = 0.08$  and  $n = 500$ . Plugging these values into the formula, we get

$$\left[ 0.08 \pm 1.96 \sqrt{\frac{0.08(1-0.08)}{500}} \right]$$

Calculating this, we find that the midpoint of the confidence interval is approximately 0.08, which means the answer is  $\boxed{\text{None of the above}}$ . The midpoint is not given by any of the answer choices. Final answer: D

\*\*\*\*\*

Answer Option 7:

The midpoint for a 95% confidence interval estimate of a proportion is equal to the point estimate plus or minus the margin of error.

The point estimate is the sample proportion, which is 8% or 0.08.

The margin of error can be calculated using the formula: margin of error =  $z \cdot \sqrt{p \cdot (1-p)/n}$ , where  $z$  is the  $z$ -value corresponding to the desired level of confidence (for a 95% confidence interval,  $z = 1.96$ ),  $p$  is the sample proportion, and  $n$  is the sample size.

Using the given information, we have: margin of error =  $1.96 \cdot \sqrt{(0.08 \cdot (1-0.08))/500} = 0.025$

Therefore, the midpoint for the 95% confidence interval estimate of  $p$  is  $0.08 \pm 0.025 = 0.055$ .

None of the answer choices provided match this result, so the correct answer is D. None of the above.

\*\*\*\*\*

Answer Option 8:

The midpoint for a 95% confidence interval estimate can be found by taking the sample proportion, 0.08, and adding/subtracting the margin of error.

The margin of error is given by the formula:

$$\text{margin of error} = z \cdot \sqrt{\frac{p(1-p)}{n}}$$

where  $z$  is the  $z$ -score corresponding to the desired confidence level (in this case, 95%),  $p$  is the sample proportion, and  $n$  is the sample size.

Since the sample size is 500, we can calculate the margin of error using the formula above. The  $z$ -score for a 95% confidence level is approximately 1.96.

$$\text{margin of error} = 1.96 \cdot \sqrt{\frac{0.08 \cdot 0.92}{500}} \approx 0.024$$

The midpoint of the confidence interval is given by the sample proportion plus or minus the margin of error, so the midpoint is approximately  $0.08 \pm 0.024 = [0.056, 0.104]$ .

Therefore, the correct answer is D. 'None of the above.'

\*\*\*\*\*

Answer Option 9:

Since we have a simple random sample, we can use the formula for a confidence interval for a proportion:  $\hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ , where  $\hat{p}$  is the proportion of unemployed in the sample,  $n$  is the sample size, and  $z$  is the critical value from the standard normal distribution for the desired level of confidence.

In this case,  $\hat{p} = 0.08$ ,  $n = 500$ , and we want a 95% confidence interval, so the critical value  $z$  is approximately 1.96.

$$\text{margin of error} = z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \cdot \sqrt{\frac{0.08(1-0.08)}{500}} \approx 0.0254$$

Therefore, the confidence interval estimate for the proportion of people unemployed is  $0.08 \pm 0.0254$ .

The midpoint of the confidence interval is the average of the upper and lower bounds, which is  $0.08$ .

Thus, the answer is  $\boxed{\text{None of the above}}$ . Final Answer: D

\*\*\*\*\*



Question 92:

'An insurance company wishes to study the number of years drivers in a large city go between automobile accidents. They plan to obtain and analyze the data from a sample of drivers. Which of the following is a true statement?', A. 'A reasonable time-and-cost-saving procedure would be to use systematic sampling on an available list of all AAA (Automobile Association of America) members in the city.', B. 'A reasonable time-and-cost-saving procedure would be to randomly choose families and include all drivers in each of these families in the sample.', C. 'To determine the mean number of years between accidents, randomness in choosing a sample of drivers is not important as long as the sample size is very large.', D. 'None of the above are true statements.'

=====

Answer Option 1:

The correct answer is D. 'None of the above are true statements.'

- A. Using systematic sampling on an available list of all AAA members would introduce bias because it would only include members of AAA, which may not be representative of all drivers in the city.
- B. Randomly choosing families and including all drivers in each of these families would also introduce bias because it would oversample certain families and potentially exclude other drivers who are not part of the selected families.
- C. Randomness in choosing a sample of drivers is important to ensure that the sample is representative of the population of drivers. If the sample is not chosen randomly, it may not accurately reflect the characteristics and behaviors of all drivers in the city.

Therefore, none of the statements are true. Final Answer: D

\*\*\*\*\*

Answer Option 2:

The correct answer is D. 'None of the above are true statements.'

- A. Systematic sampling on an available list of all AAA members in the city would introduce bias, as it would only include AAA members and may not be representative of the entire population of drivers in the city.
- B. Randomly choosing families and including all drivers in each of these families in the sample may introduce bias, as it may overrepresent certain families and underrepresent others. Additionally, the sample may not be representative of the entire population of drivers in the city.
- C. Randomness in choosing a sample of drivers is important in order to obtain an unbiased estimate of the mean number of years between accidents. A random sample helps ensure that the sample is representative of the entire population, and randomness is necessary for statistical inference.

Therefore, the correct answer is D. 'None of the above are true statements.' Final answer: D

\*\*\*\*\*

Answer Option 3:

The correct answer is D. 'None of the above are true statements.'

- A. Systematic sampling on an available list of all AAA members would introduce a bias, as it only includes AAA members and excludes drivers who are not members of AAA. This would not provide a representative sample of all drivers in the city.
- B. Randomly choosing families and including all drivers in each of these families would also introduce a bias, as it only includes drivers from selected families. This may not represent the entire population of drivers in the city.
- C. Randomness in choosing a sample of drivers is important when determining the mean number of years between accidents. Random sampling helps to ensure that the sample is representative of the population and reduces the risk of bias.

Therefore, the correct answer is D. 'None of the above are true statements.' Final Answer: D

\*\*\*\*\*

Answer Option 4:

In order to obtain accurate and unbiased results, it is crucial to use appropriate sampling methods.

Option A suggests using systematic sampling on an available list of all AAA members in the city. This method may introduce bias since it only includes AAA members and may not be representative of all drivers in the city. Therefore, option A is not a true statement.

Option B suggests randomly choosing families and including all drivers in each of these families in the sample. This method may introduce bias as well because it only includes drivers from selected families and may not be representative of the entire population of drivers in the city. Therefore, option B is not a true statement.

Option C states that randomness in choosing a sample of drivers is not important as long as the sample size is very large. This is not a true statement. Randomness in selecting a sample is essential to ensure that the sample is representative of the entire population of drivers. Even with a large sample size, if the sampling method is not random, the results may still be biased.

Based on the analysis above, the correct answer is  $\boxed{\text{D. None of the above are true statements.}}$  Final Answer: D

\*\*\*\*\*

Answer Option 5:

The most appropriate answer is D. 'None of the above are true statements.'

- A. Using systematic sampling on an available list of all AAA members in the city would introduce bias because it would only include members of the AAA, not all drivers in the city.
- B. Randomly choosing families and including all drivers in each of these families in the sample would introduce bias because it would oversample certain groups, such as large families.
- C. Randomness in choosing a sample of drivers is important to ensure that the sample is representative of the population of drivers in the city. Without randomness, the sample may not accurately reflect the characteristics of all drivers in the city.

Therefore, none of the statements are true. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The correct statement is D. 'None of the above are true statements.'

Option A suggests using systematic sampling on an available list of all AAA members. However, this would introduce bias since it would only include AAA members and not all drivers in the city.

Option B suggests randomly choosing families and including all drivers in each family. While this method might save time and cost, it may not provide a representative sample of all drivers in the city.

Option C suggests that randomness in choosing a sample of drivers is not important as long as the sample size is very large. However, randomness is crucial in obtaining an unbiased and representative sample. A large sample size alone does not guarantee representativeness.

Therefore, none of the given options are true statements. The correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct answer is D. 'None of the above are true statements.'

A. Systematic sampling on an available list of all AAA members would introduce bias because not all drivers in the large city are members of AAA. The sample would not be representative of the entire population of drivers in the city.

B. Randomly choosing families and including all drivers in each family in the sample would also introduce bias because it assumes that all drivers within a family have the same number of years between accidents. This assumption may not hold true, and the sample would not be representative of the entire population of drivers in the city.

C. Randomness in choosing a sample of drivers is important to ensure that the sample is representative of the entire population of drivers in the city. Without randomness, the sample may be skewed and not reflect the true mean number of years between accidents for all drivers in the city.

Therefore, none of the given statements are true. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The most appropriate answer choice is D. 'None of the above are true statements.'

In order to obtain an unbiased and representative sample of drivers in a large city, it is important to use random sampling techniques.

Option A suggests using systematic sampling on an available list of all AAA members in the city. This would introduce a bias as it would only include members of the AAA, excluding those who are not members.

Option B suggests randomly choosing families and including all drivers in each of these families in the sample. This would introduce a bias as well, as it would not include drivers who are not part of the chosen families.

Option C states that randomness in choosing a sample of drivers is not important as long as the sample size is very large. However, randomness is crucial in obtaining an unbiased sample that represents the larger population. A large sample size alone does not guarantee representativeness.

Therefore, the correct answer is D. 'None of the above are true statements.'

\*\*\*\*\*

Answer Option 9:

Option B is a reasonable time-and-cost-saving procedure. By randomly choosing families and including all drivers in each of these families in the sample, the insurance company can obtain a representative sample of drivers in the city while minimizing the time and cost required to collect the data.

Option A is not a suitable procedure because using systematic sampling on an available list of all AAA members in the city would introduce bias into the sample. The sample would not be representative of all drivers in the city as it would only include AAA members.

Option C is also incorrect. Randomness in choosing a sample of drivers is important to ensure that the sample is representative of the population of drivers in the city. Without randomness, the sample may be biased and not accurately reflect the mean number of years between accidents for all drivers in the city.

Thus, the correct answer is  $\boxed{\text{B. 'A reasonable time-and-cost-saving procedure would be to randomly choose families and include all drivers in each of these families in the sample.'}}$

\*\*\*\*\*

Question 93:

'Which of the following is the central limit theorem?', A. 'No matter how the population is distributed, as the sample size increases, the mean of the sample means becomes closer to the mean of the population.', B. 'No matter how the population is distributed, as the sample size increases, the standard deviation of the sample means becomes closer to the standard deviation of the population divided by the square root of the sample size.', C. 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.', D. 'The central limit theorem refers to something else.'

=====

Answer Option 1:

The correct statement for the central limit theorem is C. 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.' This theorem states that regardless of the shape of the population distribution, as the sample size increases, the sampling distribution of the sample mean approaches a normal distribution. This is a fundamental concept in statistics and has wide-ranging applications in hypothesis testing and confidence intervals. Therefore, the correct answer is C. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The correct statement of the central limit theorem is C. 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.' This theorem states that even if the population is not normally distributed, as the sample size increases, the distribution of the sample mean approaches a normal distribution. The other statements A and B are not true for all population distributions. Thus, the answer is  $\boxed{\text{C}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 3:

The correct statement of the central limit theorem is option C. 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.' This theorem is a fundamental concept in statistics that states that regardless of the shape of the population distribution, as the sample size increases, the sampling distribution of the sample mean approaches a normal distribution. This is a crucial result that allows for the use of many statistical techniques. Therefore, the correct answer is C. Final answer: C

\*\*\*\*\*

Answer Option 4:

The correct statement of the central limit theorem is C. 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.' Final answer: C

\*\*\*\*\*

Answer Option 5:

The correct statement of the central limit theorem is C. 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.' This theorem is a fundamental result in statistics that allows us to make inferences about a population based on a sample. Final Answer: C

\*\*\*\*\*

Answer Option 6:

The correct statement of the central limit theorem is:

C. 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.'

This theorem is a fundamental concept in statistics and describes the behavior of sample means as the sample size increases. Final Answer: C

\*\*\*\*\*

Answer Option 7:

The correct statement for the central limit theorem is option C: 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.' This is an important result in statistics, which states that regardless of the shape of the population distribution, the distribution of sample means will approach a normal distribution as the sample size increases. This is a fundamental concept in statistical inference. Final answer: C

\*\*\*\*\*

Answer Option 8:

The central limit theorem states that if we have a population with any distribution, and we take sufficiently large random samples from that population with replacement, then the sampling distribution of the sample mean will be approximately normally distributed. Therefore, the correct statement of the central limit theorem is:

C. 'If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution.'

Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct statement of the central limit theorem is option C: "If the population is normally distributed, then as the sample size increases, the sampling distribution of the sample mean becomes closer to a normal distribution." This theorem is a fundamental result in statistics that states that regardless of the distribution of the population, as the sample size increases, the distribution of sample means approaches a normal distribution. Option A describes the law of large numbers, option B is not a correct statement of the central limit theorem, and option D is incorrect as the central limit theorem refers specifically to the convergence of the sampling distribution of the sample mean to a normal distribution. Therefore, the answer is C.

\*\*\*\*\*

Question 94:

'A manufacturer of ready-bake cake mixes is interested in designing an experiment to test the effects of four different temperature levels (300, 325, 350, and 375F), two different types of pans (glass and metal), and three different types of ovens (gas, electric, and microwave) on the texture of its cakes, in all combinations. Which of the following below is the best description of the design of the necessary experiment?', A. 'A completely randomized design with nine treatment groups', B. 'A completely randomized design with 24 treatment groups', C. 'A randomized block design, blocked on temperature, with six treatment groups', D. 'A randomized block design, blocked on type of pan, with 12 treatment groups'

=====

Answer Option 1:

To test the effects of the temperature levels, types of pans, and types of ovens on the texture of the cakes, we need to consider all possible combinations of these factors. There are four temperature levels, two types of pans, and three types of ovens, resulting in a total of  $4 \times 2 \times 3 = 24$  treatment groups.

Therefore, the best description of the design of the necessary experiment is option B: 'A completely randomized design with 24 treatment groups'. Final Answer: B

\*\*\*\*\*

Answer Option 2:

In this experiment, there are 4 temperature levels, 2 types of pans, and 3 types of ovens. To test all possible combinations, we need to multiply the number of levels for each factor together. This gives us  $4 \times 2 \times 3 = 24$  treatment groups.

Since the experiment involves testing all combinations of treatments, it is a completely randomized design. Therefore, the best description of the design of the necessary experiment is:

B. 'A completely randomized design with 24 treatment groups'.

\*\*\*\*\*

Answer Option 3:

To test the effects of temperature, type of pan, and type of oven on the texture of the cakes, we need to consider all possible combinations of these factors.

There are 4 temperature levels, 2 types of pans, and 3 types of ovens. Therefore, the total number of treatment groups is  $4 \times 2 \times 3 = 24$ .

Since all treatment groups are considered and there is no specific blocking or grouping of the treatments, the best description of the design of the necessary experiment is:

B. 'A completely randomized design with 24 treatment groups'.

\*\*\*\*\*

Answer Option 4:

To design the necessary experiment, we need to consider the number of levels for each factor and the number of factors involved.

The experiment involves four temperature levels, two types of pans, and three types of ovens.

Since we have four temperature levels, two types of pans, and three types of ovens, the total number of treatment groups will be  $4 \times 2 \times 3 = 24$ .

Therefore, the best description of the design of the necessary experiment is "A completely randomized design with 24 treatment groups."

Hence, the correct answer is  $\boxed{\text{B. 'A completely randomized design with 24 treatment groups'}}$ .

\*\*\*\*\*

Answer Option 5:

To test the effects of temperature, type of pan, and type of oven on the texture of the cakes, we need to consider all possible combinations of these factors. We have 4 temperature levels, 2 types of pans, and 3 types of ovens, so the total number of treatment groups is  $4 \times 2 \times 3 = 24$ .

Since the experiment involves testing all possible combinations of the factors, the best description of the design is a completely randomized design with 24 treatment groups. Therefore, the correct answer is

$\boxed{\text{B. 'A completely randomized design with 24 treatment groups'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 6:

The experiment is testing the effects of four temperature levels, two types of pans, and three types of ovens. To account for all combinations, we need to test each temperature with each type of pan and each type of oven. This means that we have 4 temperature levels, 2 types of pans, and 3 types of ovens, resulting in a total of  $4 \times 2 \times 3 = 24$  treatment groups.

Since there is no blocking or grouping of treatments based on any specific factor, this design can be described as a completely randomized design. Therefore, the best description of the design is 'A

completely randomized design with 24 treatment groups'. Thus, the correct answer is  $\boxed{\text{B. 'A completely randomized design with 24 treatment groups'}}$ . Final answer: B

\*\*\*\*\*

Answer Option 7:

The best description of the design of the necessary experiment is option B. 'A completely randomized design with 24 treatment groups'. This is because there are 4 temperature levels, 2 types of pans, and 3 types of ovens, resulting in a total of  $4 \times 2 \times 3 = 24$  treatment groups. In a completely randomized design, each treatment group is assigned randomly to the different combinations of temperature, pan type, and oven type. Final answer: B

\*\*\*\*\*

Answer Option 8:

To test the effects of the different temperature levels, types of pans, and types of ovens, we need to consider the combination of all these factors. There are 4 temperature levels, 2 types of pans, and 3 types of ovens, which gives us a total of  $4 \times 2 \times 3 = 24$  treatment groups.

Therefore, the best description of the design of the necessary experiment is 'A completely randomized design with 24 treatment groups', which corresponds to option  $\boxed{\text{B}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

The experiment involves testing the effects of four temperature levels, two types of pans, and three types of ovens. To account for all possible combinations, we need to multiply the number of levels for each factor.

There are 4 temperature levels, 2 types of pans, and 3 types of ovens. Therefore, the total number of treatment groups is  $4 \times 2 \times 3 = 24$ .

Since the treatments are assigned randomly, we have a completely randomized design. Hence, the correct description of the experiment is:

B. 'A completely randomized design with 24 treatment groups'

\*\*\*\*\*

Question 95:

'A study is made to determine whether more hours of academic studying leads to higher point scoring by basketball players. In surveying 50 basketball players, it is noted that the 25 who claim to study the most hours have a higher point average than the 25 who study less. Based on this study, the coach begins requiring the players to spend more time studying. Which of the following is a correct statement?', A. 'While this study may have its faults, it still does prove causation.', B. 'There could well be a confounding variable responsible for the seeming relationship.', C. 'While this is a controlled experiment, the conclusion of the coach is not justified.', D. 'To get the athletes to study more, it would be more meaningful to have them put in more practice time on the court to boost their point averages, as higher point averages seem to be associated with more study time.'

=====

Answer Option 1:

The correct statement is B. 'There could well be a confounding variable responsible for the seeming relationship.'

The study only shows a correlation between studying more hours and having a higher point average. It does not establish a causal relationship. There could be other factors, such as natural ability or dedication to the sport, that are influencing both the amount of studying and the point average. Therefore, it is important to consider other variables that may be responsible for the observed relationship before concluding that increasing studying hours will directly lead to higher point averages. Final Answer: B

\*\*\*\*\*

Answer Option 2:

The correct statement is B. 'There could well be a confounding variable responsible for the seeming relationship.'

While the study observed a correlation between studying hours and point averages, there could be other factors that contribute to this relationship. For example, basketball players who are more dedicated and motivated might be more likely to both study more hours and put more effort into their basketball performance, leading to higher point averages.

Without controlling for other variables, it is difficult to establish a causal relationship between studying hours and point averages. Therefore, it is important to consider the possibility of confounding variables that could explain the observed relationship before making any conclusions or implementing changes based on the study.

\*\*\*\*\*

Answer Option 3:

The correct statement is: B. 'There could well be a confounding variable responsible for the seeming relationship.'

The observed correlation between studying more hours and having a higher point average does not necessarily imply a causal relationship. There could be other factors, known as confounding variables, that are responsible for the observed relationship. For example, it is possible that the players who study more hours also have better time management skills or are more motivated, which could contribute to their higher point averages.

Additionally, the study is based on observational data, not a controlled experiment. Therefore, it is difficult to establish a cause-and-effect relationship between studying hours and point averages.

The other options are not supported by the information provided or are not relevant to the study. Final Answer: B

\*\*\*\*\*

Answer Option 4:

The correct statement is B. 'There could well be a confounding variable responsible for the seeming relationship.'

The study only shows a correlation between studying hours and point scoring. It does not establish a causal relationship. There could be other factors at play, such as natural talent or dedication, which lead to both higher point averages and more studying hours. Additionally, the study does not account for the possibility of outliers, measurement errors, or other variables that could affect the results.

Without further evidence, it is not justified to conclude that requiring players to spend more time studying will directly result in higher point averages. Option A implies that the study proves causation, which is not supported by the given information. Option C states that the conclusion of the coach is not justified, which is true based on the limitations of the study. Option D suggests an alternative approach that could be more meaningful, but it does not directly address the issue of causation or confounding variables. Thus, the correct statement is B. Final Answer: B

\*\*\*\*\*

Answer Option 5:

The correct statement is B. 'There could well be a confounding variable responsible for the seeming relationship.'

The study only shows a correlation between studying more hours and having a higher point average. However, it does not establish a cause-and-effect relationship or rule out the possibility of other factors influencing the results. There could be other variables, such as natural talent, motivation, or prior academic abilities, that are responsible for both the increased study hours and the higher point average.

Therefore, it is premature to conclude that studying more hours directly leads to higher point scoring. Final Answer: B

\*\*\*\*\*

Answer Option 6:

The correct statement is B. 'There could well be a confounding variable responsible for the seeming relationship.'

The study only shows a correlation between studying more hours and having a higher point average. It does not establish a cause-and-effect relationship. There may be other factors that contribute to the higher point averages of the players who study more, such as natural talent, motivation, or access to better resources. It is possible that these factors, rather than studying more, are the true cause of the higher point averages. Therefore, there could be a confounding variable responsible for the observed relationship. Final Answer: B

\*\*\*\*\*

Answer Option 7:

The correct statement is B. 'There could well be a confounding variable responsible for the seeming relationship.'

The study only shows a correlation between studying more hours and having a higher point average. However, there could be other factors that are responsible for the observed relationship. For example, it is possible that the players who study more hours are naturally more talented or motivated, which could be the true reason for their higher point averages.

Without further evidence, it is not valid to conclude that studying more hours causes higher point averages. Therefore, it is important to consider other factors (confounding variables) that could be influencing the results. Final Answer: B

\*\*\*\*\*

Answer Option 8:

The correct statement is B. 'There could well be a confounding variable responsible for the seeming relationship.'

The study shows a correlation between studying more hours and having a higher point average. However, correlation does not imply causation. There may be other factors at play that are responsible for the higher point averages among the players who study more hours. For example, the players who study more may have better time management skills or a stronger work ethic, which could also contribute to their higher point averages.

Without controlling for all other possible variables, it is not possible to conclude that studying more hours directly causes higher point averages. Therefore, the coach's conclusion that requiring players to spend more time studying will lead to higher point averages is not justified.

Option D is also incorrect because it suggests that practicing more on the court is associated with higher point averages, which is not supported by the information given in the study.

\*\*\*\*\*

Answer Option 9:

The correct statement is B. 'There could well be a confounding variable responsible for the seeming relationship.'

Although it is observed that the 25 players who claim to study the most hours have a higher point average than the 25 players who study less, there could be other factors at play. For example, it is possible that the players who study more hours are generally more dedicated or have better study habits, which could also contribute to their higher point averages. Without controlling for all possible confounding variables, we cannot conclude that studying more hours directly leads to higher point scoring.

The other answer choices are incorrect. Option A states that the study proves causation, which is not supported by the given information. Option C suggests that while it is a controlled experiment, the coach's conclusion is not justified, but we do not have enough information to make that conclusion. Option D suggests that the athletes should put in more practice time on the court instead of studying more, but this recommendation is not supported by the given information. Final Answer: B

\*\*\*\*\*

Question 96:

'Biologists around the world have increased efforts at conservation. Monitoring wildlife populations is important so that appropriate management efforts can be implemented, but such monitoring is often difficult. One study found a cheap and easy way to estimate the number of nesting sites of terns (a type of seabird) by monitoring the number of calls heard per minute. More calls happen when there are more birds, as one would expect. In fact, it turned out that the number of calls explained 71% of the variation in the abundance of nests between breeding sites. Which of the following statements is correct about the correlation between the number of calls and the abundance of nests?', A. 'The correlation coefficient is -0.71.', B. 'The correlation coefficient is 0.71.', C. 'The correlation coefficient is -0.84.', D. 'The correlation coefficient is 0.84.'

=====

Answer Option 1:

The statement says that the number of calls explains 71% of the variation in the abundance of nests. This means that there is a positive correlation between the number of calls and the abundance of nests. The correlation coefficient is a measure of the strength and direction of the linear relationship between two variables. Since there is a positive correlation, the correlation coefficient should be positive. Therefore, the correct statement is: 'The correlation coefficient is 0.71.' Hence, the answer is  $\boxed{\text{B}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 2:

The statement "the number of calls explained 71% of the variation in the abundance of nests" indicates that the correlation coefficient between the number of calls and the abundance of nests is  $\sqrt{0.71} \approx 0.84$ . Therefore, the correct statement is 'The correlation coefficient is  $\boxed{0.84}$ .' Final Answer: D

\*\*\*\*\*

Answer Option 3:

Since the number of calls explains 71% of the variation in the abundance of nests, we know that the correlation coefficient, which measures the strength and direction of the linear relationship between two variables, is the square root of 71%, or approximately  $\sqrt{0.71} \approx 0.84$ . Therefore, the correct statement is 'The correlation coefficient is 0.84.' The answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 4:

Since the number of calls explains 71% of the variation in the abundance of nests, we know that there is a positive correlation between the number of calls and the abundance of nests. A positive correlation means that as the number of calls increases, the abundance of nests also increases. Therefore, the correct statement is 'The correlation coefficient is  $\boxed{\text{B. } 0.71}$ .' Final Answer: B

\*\*\*\*\*

Answer Option 5:

The statement that is correct about the correlation between the number of calls and the abundance of nests is option B, 'The correlation coefficient is 0.71.' The fact that the number of calls explains 71% of the variation in the abundance of nests indicates a strong positive correlation between the two variables. The correlation coefficient represents the strength and direction of the linear relationship between two variables, and a positive value of 0.71 suggests a strong positive correlation. Final Answer: B.

\*\*\*\*\*

Answer Option 6:

The statement "the number of calls explained 71% of the variation in the abundance of nests between breeding sites" means that the coefficient of determination,  $R^2$ , is equal to 0.71.

The coefficient of determination is the square of the correlation coefficient, so we have  $R^2 = 0.71$ , which means  $R = \sqrt{0.71} \approx 0.8416$ .

Since the correlation coefficient is positive (more calls correspond to more nests), the correct statement is: 'The correlation coefficient is 0.71.' Therefore, the answer is  $\boxed{\text{B}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 7:

The statement that is correct about the correlation between the number of calls and the abundance of nests is B. 'The correlation coefficient is 0.71.'

The given information states that the number of calls explains 71% of the variation in the abundance of nests. This indicates a positive correlation between the number of calls and the abundance of nests.

Since the correlation coefficient measures the strength and direction of the linear relationship between two variables, a positive correlation coefficient of 0.71 indicates a moderately strong positive correlation between the number of calls and the abundance of nests. Final answer: B.

\*\*\*\*\*

Answer Option 8:

The statement says that the number of calls explains 71% of the variation in the abundance of nests. This means that there is a positive correlation between the number of calls and the abundance of nests.

Therefore, the correct statement about the correlation coefficient is  $\boxed{\text{B}}$  The correlation coefficient is 0.71. Final Answer: B

\*\*\*\*\*

Answer Option 9:

The statement says that the number of calls explains 71% of the variation in the abundance of nests between breeding sites. This means that there is a positive correlation between the number of calls and the abundance of nests. The correct statement about the correlation between the number of calls and the abundance of nests is therefore 'The correlation coefficient is  $\boxed{\text{B. The correlation coefficient is 0.71.}}$ '. Final Answer: B

\*\*\*\*\*



Question 97:

'A company has 1000 employees evenly distributed throughout five assembly plants. A sample of 30 employees is to be chosen as follows. Each of the five managers will be asked to place the 200 time cards of their respective employees in a bag, shake them up, and randomly draw out six names. The six names from each plant will be put together to make up the sample. Will this method result in a simple random sample of the 1000 employees?', A. 'Yes, because every employee has the same chance of being selected.', B. 'Yes, because every plant is equally represented.', C. 'Yes, because this is an example of stratified sampling, which is a special case of simple random sampling.', D. 'No, because not every group of 30 employees has the same chance of being selected.'

=====

Answer Option 1:

The answer is D. 'No, because not every group of 30 employees has the same chance of being selected.'

In this method, each manager is selecting 6 employees randomly from their respective plant. This means that the chance of being selected for an employee varies depending on which plant they belong to. Employees from larger plants have a higher chance of being selected compared to employees from smaller plants. Therefore, this method does not result in a simple random sample of the 1000 employees.

Final Answer: D

\*\*\*\*\*

Answer Option 2:

The method described does not result in a simple random sample of the 1000 employees. In this method, each manager selects 6 employees from their respective plant, resulting in a sample of 30 employees. However, not every group of 30 employees has the same chance of being selected. Since the employees are evenly distributed throughout the five assembly plants, the chance of being selected depends on which plant an employee belongs to. Therefore, the correct answer is  $\boxed{\text{D. 'No, because not every group of 30 employees has the same chance of being selected.'}}$

\*\*\*\*\*

Answer Option 3:

The answer is  $\boxed{\text{D}}$  'No, because not every group of 30 employees has the same chance of being selected.'

In this sampling method, each manager selects six employees from their respective plant. Since the number of employees in each plant is equal, each employee within a plant has the same chance of being selected by their manager. However, this does not guarantee that each group of 30 employees from the entire company has the same chance of being selected. Certain groups of employees may have a higher chance of being selected if their plant's manager has a higher probability of selecting them. Therefore, this method does not result in a simple random sample of the 1000 employees. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'No, because not every group of 30 employees has the same chance of being selected.'

In this sampling method, each manager selects 6 employees from their respective plant. This means that employees in larger plants have a higher chance of being selected compared to employees in smaller plants. For example, if all 6 employees are selected from one plant, then no employees from the other plants will be included in the sample.

In a simple random sample, each employee should have an equal chance of being selected, regardless of the plant they are in. However, in this method, the selection process is not random or equal for every employee. Therefore, this method does not result in a simple random sample of the 1000 employees.

\*\*\*\*\*

Answer Option 5:

The method described does not result in a simple random sample of the 1000 employees. In this method, each manager selects 6 employees from their own plant, resulting in a sample where employees from different plants are not equally represented. Therefore, the correct answer is  $\boxed{\text{D. 'No, because not every group of 30 employees has the same chance of being selected.'}}$

\*\*\*\*\*

Answer Option 6:

The correct answer is D. 'No, because not every group of 30 employees has the same chance of being selected.'

In this method, every employee does not have the same chance of being selected. Employees within the same plant have a higher chance of being selected together as a group, compared to employees from different plants. Additionally, there is no guarantee that every possible group of 30 employees has an equal chance of being selected. Therefore, this method does not result in a simple random sample of the 1000 employees. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct answer is D. No, because not every group of 30 employees has the same chance of being selected.

In this sampling method, the probability of each individual employee being selected is not equal. Employees from plants with fewer employees have a higher chance of being selected compared to employees from plants with more employees. This violates the requirement of a simple random sample, where every individual should have an equal chance of being selected. Final answer: D

\*\*\*\*\*

Answer Option 8:

The method described does not result in a simple random sample of the 1000 employees. To have a simple random sample, each individual employee should have an equal chance of being selected.

However, in this method, the probability of being selected depends on which plant the employee belongs to. Therefore, the correct answer is  $\boxed{\text{D. 'No, because not every group of 30 employees has the same chance of being selected.'}}$

\*\*\*\*\*

Answer Option 9:

The method described in the problem does not result in a simple random sample of the 1000 employees. This is because not every group of 30 employees has the same chance of being selected. Each manager only selects from their own plant, so the employees within each plant have a higher chance of being chosen compared to employees from other plants. Therefore, the correct answer is

$\boxed{\text{D. 'No, because not every group of 30 employees has the same chance of being selected.'}}$

\*\*\*\*\*

Question 98:

'A doctor hopes that a new surgery technique will shorten the recovery time compared to the standard technique. To test this, he designed an experiment in which patients who required this type of surgery were randomly assigned to the standard technique or the new technique. Then the mean recovery time for each treatment group was compared. Assuming conditions for inference were met, which analysis should be used?', A. 'A t-test for a mean.', B. 'A t-test for a difference in means.', C. 'A z-test for a mean.', D. 'A z-test for a difference in means.'

=====

Answer Option 1:

To compare the mean recovery time between the standard technique and the new technique, we should use 'A t-test for a difference in means.' The t-test is appropriate when comparing the means of two groups, which is exactly what we want to do in this scenario. Final answer: B

\*\*\*\*\*

Answer Option 2:

Since the doctor wants to compare the mean recovery time between the standard technique and the new technique, the appropriate analysis to use would be a t-test for a difference in means. This test is used when comparing the means of two independent groups, which is exactly what the doctor is doing in this experiment. Therefore, the correct answer is  $\boxed{\text{B. 'A t-test for a difference in means.'}}$

\*\*\*\*\*

Answer Option 3:

Since the doctor wants to compare the mean recovery time between the standard technique and the new technique, the appropriate analysis to use is a t-test for a difference in means. This test is used to determine if there is a statistically significant difference between the means of two independent groups. Therefore, the correct answer is  $\boxed{\text{B. 'A t-test for a difference in means.'}}$  Final Answer: B

\*\*\*\*\*

Answer Option 4:

To compare the mean recovery time for the standard technique and the new technique, we should use a t-test for a difference in means. This is because we are comparing the means of two independent groups (standard technique group and new technique group). The conditions for inference include random assignment of patients to treatments and the assumption of approximately normal distribution of recovery times in each treatment group. Therefore, the correct analysis to be used is option B, 'A t-test for a difference in means.' Final Answer: B

\*\*\*\*\*

Answer Option 5:

To compare the mean recovery time between the standard technique and the new technique, a t-test for a difference in means should be used. This test is appropriate when comparing the means of two independent groups. The t-test for a mean would be used if we were comparing the mean recovery time within a single group (e.g., comparing the mean recovery time before and after a treatment in the same group). Therefore, the correct answer is  $\boxed{\text{B. A t-test for a difference in means}}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 6:

In this scenario, the doctor wants to compare the mean recovery time between the standard technique and the new technique. Therefore, the appropriate analysis to use would be a t-test for a difference in means.

A t-test for a mean is used when comparing a single sample mean to a known population mean or a hypothesized value.

A z-test for a mean is used when comparing a single sample mean to a known population mean, but the population standard deviation is known.

A z-test for a difference in means is used when comparing two independent sample means, and the population standard deviations are known.

Since the doctor is comparing the mean recovery time between two treatment groups, the appropriate analysis is a t-test for a difference in means, making option B, 'A t-test for a difference in means', the correct choice. Final Answer: B

\*\*\*\*\*

Answer Option 7:

To compare the mean recovery time between the standard technique and the new technique, we should use a hypothesis test for a difference in means. This can be done using a t-test for a difference in means when the sample sizes are small (typically less than 30) or when the population standard deviations are unknown. If the sample sizes are large (typically greater than 30) and the population standard deviations are known or assumed to be equal, a z-test for a difference in means can be used.

Since the experiment randomly assigns patients to the different techniques and we are comparing the mean recovery time between the two groups, we should use a t-test for a difference in means. Therefore, the correct analysis to be used is option B. 'A t-test for a difference in means.' Final Answer: B

\*\*\*\*\*

Answer Option 8:

Since the doctor is comparing the mean recovery time between two treatment groups (standard technique and new technique), the appropriate analysis to be used is 'A t-test for a difference in means.' This test is used when comparing the means of two independent groups to determine if there is a statistically significant difference between them. Final answer: B

\*\*\*\*\*

Answer Option 9:

The appropriate analysis in this scenario would be a t-test for a difference in means, which is option B. The t-test is used when comparing the means of two groups, in this case, the standard technique group and the new technique group. The t-test takes into consideration the variability within each group and determines whether the difference in means between the groups is statistically significant. Since we are comparing the mean recovery time between the two techniques, a t-test for a difference in means is the most suitable choice. Final Answer: B

\*\*\*\*\*



Question 99:

'A researcher plans a study to examine long-term confidence in the U.S. economy among the adult population. She obtains a simple random sample of 30 adults as they leave a Wall Street office building one weekday afternoon. All but two of the adults agree to participate in the survey. Which of the following conclusions is correct?', A. 'Proper use of chance as evidenced by the simple random sample makes this a well-designed survey.', B. 'The high response rate makes this a well-designed survey.', C. 'Selection bias makes this a poorly designed survey.', D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

=====

Answer Option 1:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.' In this scenario, the researcher obtained a simple random sample, which is a well-designed sampling method. However, the response rate was not high since only 28 out of 30 adults agreed to participate. This means that the participants are self-selecting, and those who chose to respond might have stronger opinions or more interest in the topic compared to those who did not respond. Therefore, the survey may not accurately represent the opinions of the entire adult population regarding long-term confidence in the U.S. economy. Final answer: D.

\*\*\*\*\*

Answer Option 2:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

In this scenario, the researcher obtains a simple random sample of 30 adults as they leave a Wall Street office building. However, all but two of the adults agree to participate in the survey. This means that the survey results are based on voluntary responses, which can introduce bias. People who choose to respond may have stronger opinions about the topic of the survey compared to those who choose not to respond. Therefore, the survey may not accurately represent the views of the entire adult population and cannot be considered a well-designed survey. Final answer: D

\*\*\*\*\*

Answer Option 3:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

The reason is that the researcher obtained a simple random sample of 30 adults, but the response rate was not high, as only 28 out of 30 adults agreed to participate. This means that the participants who responded to the survey were self-selected, and those who decided to participate may have had stronger opinions about the topic of the survey. This can lead to biased results and does not represent the entire population accurately. Therefore, a voluntary response study like this gives too much emphasis to persons with strong opinions, making it a poorly designed survey. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

In this study, the researcher obtained a simple random sample of adults leaving a Wall Street office building. However, only a portion of the adults agreed to participate in the survey. This creates a voluntary response study because individuals self-select whether or not to participate. Voluntary response studies can introduce bias because individuals with strong opinions or a vested interest are more likely to participate, while those with no interest or weak opinions are less likely to participate. Therefore, the conclusions drawn from this study may not be representative of the entire adult population's confidence in the U.S. economy. Final answer: D

\*\*\*\*\*

Answer Option 5:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

In this survey, the researcher obtained a simple random sample of adults leaving a Wall Street office building. However, the response rate was not high, as only two adults out of the 30 selected declined to participate. This means that the survey is a voluntary response study, where individuals have the choice of whether or not to participate.

In voluntary response studies, individuals with stronger opinions or a vested interest in the topic are more likely to respond, while individuals with less interest or weaker opinions are less likely to participate.

This can introduce bias and lead to results that do not accurately represent the opinions of the entire population.

Therefore, the survey is poorly designed due to the potential selection bias introduced by the voluntary response aspect. The correct conclusion is D.

\*\*\*\*\*

Answer Option 6:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

In this survey, the researcher obtained a simple random sample of adults leaving a Wall Street office building. However, the fact that all but two of the adults agreed to participate means that the sample is not truly random, but rather a voluntary response sample.

Voluntary response studies tend to attract individuals with strong opinions on the topic being surveyed. In this case, the survey is about long-term confidence in the U.S. economy. It is likely that the participants who agreed to take the survey have a particular interest or strong opinions about the economy, which can introduce bias into the results.

Therefore, the survey design is not ideal due to the voluntary response nature of the sample, and the correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.' Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

In this scenario, the researcher obtained a simple random sample of 30 adults from a Wall Street office building. However, the fact that all but two of the adults agreed to participate in the survey introduces a voluntary response bias.

Voluntary response bias occurs when individuals choose whether or not to participate in the survey, and this can lead to a biased sample as those with strong opinions, either positive or negative, are more likely to volunteer their response.

Therefore, the conclusion that this is a poorly designed survey due to voluntary response bias is the correct choice. Final answer: C

\*\*\*\*\*

Answer Option 8:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

The survey in question has a response rate of 28 out of 30, which is quite high. However, the fact that the sample was obtained from adults leaving a Wall Street office building on a weekday afternoon suggests that the sample may not be representative of the entire adult population.

Additionally, the survey is voluntary, meaning that individuals have the choice to participate or not. This can introduce bias, as those with strong opinions or a vested interest in the topic are more likely to respond. This can lead to an overrepresentation of certain viewpoints, making the results less generalizable.

Therefore, the survey suffers from selection bias and voluntary response bias, which makes it a poorly designed survey. The correct conclusion is C. 'Selection bias makes this a poorly designed survey.'

Final answer: C

\*\*\*\*\*

Answer Option 9:

The correct conclusion is D. 'A voluntary response study like this gives too much emphasis to persons with strong opinions.'

In this scenario, the researcher obtained a simple random sample of 30 adults. However, the response rate was not high as only two adults declined to participate. This means that the participants in the survey are those who voluntarily chose to respond. This type of study design is known as a voluntary response study.

Voluntary response studies tend to attract individuals who have strong opinions or are particularly interested in the topic being studied. Therefore, the results obtained from this survey would not be representative of the entire adult population, as it would be biased towards individuals who are more likely to respond.

It is important in survey research to aim for a high response rate and minimize selection bias in order to obtain more accurate and generalizable results. In this case, the low non-response rate and voluntary nature of the study make it a poorly designed survey. So, the correct conclusion is D.

\*\*\*\*\*

Question 100:

To study the effect of music on piecework output at a clothing manufacturer, two experimental treatments are planned: day-long classical music for one group versus day-long light rock music for another. Which one of the following groups would serve best as a control for this study?, A. 'A third group for which no music is played', B. 'A third group that randomly hears either classical or light rock music each day', C. 'A third group that hears day-long R & B music', D. 'A third group that hears classical music every morning and light rock every afternoon'

=====

Answer Option 1:

In order to effectively study the effect of music on piecework output, it is crucial to have a control group that does not receive any music. This enables the researchers to compare the performance of the groups listening to classical music and light rock music to a baseline group that is unaffected by music. Therefore, the best control group for this study would be option A, 'A third group for which no music is played'. Final Answer: A

\*\*\*\*\*

Answer Option 2:

In a scientific study, a control group is used to provide a baseline for comparison. The purpose of the control group is to observe and measure the effect of the experimental treatment by comparing it to a group that does not receive the treatment.

In this case, the experimental treatments are day-long classical music and day-long light rock music. To determine the effect of music on piecework output, the best control group would be one that does not receive any music. This allows for a direct comparison between the groups that receive the experimental treatments and the group that does not receive any music.

Therefore, the best choice for a control group in this study would be option A: 'A third group for which no music is played'. Final answer: A

\*\*\*\*\*

Answer Option 3:

In a scientific study, a control group is used as a baseline to compare the effects of the experimental treatments. The purpose of a control group is to provide a reference point that represents the normal or typical conditions, allowing researchers to determine whether the experimental treatments have any significant impact.

In this case, the control group should not be exposed to any of the experimental treatments, as it is meant to represent the absence of music or the normal working conditions without any specific type of music playing. Therefore, the best choice for the control group would be option A, 'A third group for which no music is played'. This group would provide a baseline for comparison against the groups exposed to classical music and light rock music.

The other options either introduce some form of music (options B, C, and D) or involve mixing different types of music (option D), which would not provide a suitable control to isolate the effects of classical music and light rock music. Therefore, option A is the most appropriate choice for the control group. Final answer: A

\*\*\*\*\*

Answer Option 4:

In a controlled study, it is important to have a group that serves as a control, meaning a group that does not receive the treatment being studied. In this case, the treatment being studied is the effect of music on piecework output.

Option A, 'A third group for which no music is played', would be the best choice for a control group. By not playing any music for this group, we can compare their piecework output to the groups that listen to classical music and light rock music. This allows us to isolate the effect of music on piecework output by comparing the control group to the experimental groups.

Option B, 'A third group that randomly hears either classical or light rock music each day', would not serve as a good control because this group is still being exposed to some form of music. We would not be able to determine if any observed differences in piecework output are due to the type of music or simply the presence of music.

Option C, 'A third group that hears day-long R & B music', would also not serve as a good control because this group is listening to a different genre of music. We would not be able to isolate the effect of classical music or light rock music on piecework output if we introduce a different genre as the control.

Option D, 'A third group that hears classical music every morning and light rock every afternoon', would not serve as a good control because this group is still being exposed to both types of music. Similar to option B, we would not be able to isolate the effect of each type of music on piecework output.

Therefore, the best choice for a control group would be option A, 'A third group for which no music is played'. Final Answer: A

\*\*\*\*\*

Answer Option 5:

In a scientific experiment, a control group is used as a baseline to compare the effects of the experimental treatments. The purpose of a control group is to isolate the effect of the independent variable being studied. In this case, the independent variable is the type of music - classical music or light rock music.

Option A, 'A third group for which no music is played', would serve as the best control group because it allows researchers to compare the effects of music (classical and light rock) with the absence of music. This control group would help determine if the music itself has an effect on piecework output.

Option B, 'A third group that randomly hears either classical or light rock music each day', would not serve as a good control group because it does not isolate the effect of the type of music being played. The random assignment of music to this group would make it difficult to determine if any observed effects are due to the type of music or simply the presence of any music.

Option C, 'A third group that hears day-long R & B music', is not a good control group because it introduces a different type of music that is not being studied. The researchers are specifically interested in comparing the effects of classical music and light rock music, so introducing a different genre would not provide a valid comparison.

Option D, 'A third group that hears classical music every morning and light rock every afternoon', could potentially be used as a control group if the researchers are specifically interested in comparing the effects of day-long music exposure versus split exposure (morning and afternoon). However, if the researchers are solely interested in comparing the effects of classical music and light rock music, this option would not be a suitable control group.

Therefore, the best control group for this study is option A, 'A third group for which no music is played'. Final answer: A

\*\*\*\*\*

Answer Option 6:

In order to study the effect of music on piecework output, it is important to have a control group that does not receive any music. This allows us to compare the effects of classical music and light rock music on piecework output without any confounding factors. Therefore, the best control group for this study would be option A. 'A third group for which no music is played'. Final Answer: A

\*\*\*\*\*

Answer Option 7:

In order to determine the effect of music on piecework output, it is important to have a control group that does not receive any music. This allows for a baseline comparison to measure the impact of the different types of music on the workers' productivity. Therefore, the best choice for a control group in this study is option A, 'A third group for which no music is played.' This group will provide a reference point to evaluate the relative effects of classical and light rock music on piecework output. Final Answer: A

\*\*\*\*\*

Answer Option 8:

In an experiment, the control group serves as a baseline comparison for the experimental groups. It allows us to compare the effects of the treatment being tested (in this case, the effect of music on piecework output) against the absence of that treatment.

Option A, 'A third group for which no music is played', would serve as the best control group in this study. This group would not be exposed to any music, allowing us to compare the effects of classical music and light rock music on piecework output to no music at all. By eliminating the variable of music entirely, we can isolate and evaluate the impact of music on productivity.

Therefore, the correct answer is A. 'A third group for which no music is played'. Final answer: A

\*\*\*\*\*

Answer Option 9:

In a scientific study, the control group serves as a baseline for comparison. It is important to have a group that does not receive the treatment being studied in order to assess the true effect of the treatment.

Option A, 'A third group for which no music is played', would serve as the best control group in this study. By not playing any music for this group, we can compare their piecework output to the groups that listen to classical or light rock music. This allows us to isolate the effect of music on productivity.

Option B, 'A third group that randomly hears either classical or light rock music each day', introduces a confounding variable. The group that randomly hears either type of music may have different levels of productivity due to the inconsistency in the music they listen to.

Option C, 'A third group that hears day-long R & B music', introduces a different genre of music. This would not provide a fair comparison to the groups listening to classical or light rock music.

Option D, 'A third group that hears classical music every morning and light rock every afternoon', introduces both types of music to the group. This does not allow us to isolate the effect of a specific type of music on productivity.

Therefore, the best control group for this study would be option A, 'A third group for which no music is played'. Thus, the answer is A.

\*\*\*\*\*



Question 101:

During the years 1886 through 2000 there were an average of 8.7 tropical cyclones per year, of which an average of 5.1 became hurricanes. Assuming that the probability of any cyclone becoming a hurricane is independent of what happens to any other cyclone, if there are five cyclones in one year, what is the probability that at least three become hurricanes?, A. '0.313', B. '0.345', C. '0.586', D. '0.658'

=====

Answer Option 1:

We can model this problem using a binomial distribution. Let  $p$  be the probability that a cyclone becomes a hurricane, and let  $n$  be the number of cyclones in one year. The probability that exactly  $k$  cyclones become hurricanes is given by the expression  $\binom{n}{k} p^k (1-p)^{n-k}$ . In this case,  $n=5$  and  $p=\frac{5.1}{8.7}$ . We want to find the probability that at least three cyclones become hurricanes, which is equivalent to finding the probability that three, four, or five cyclones become hurricanes. Therefore, the desired probability is

$$\binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p)^1 + \binom{5}{5} p^5 (1-p)^0.$$

Evaluating this expression, we find that the probability is approximately  $\boxed{0.586}$ , which corresponds to answer choice C.

\*\*\*\*\*

Answer Option 2:

We can model this situation using the binomial distribution. Let  $X$  be the number of cyclones that become hurricanes out of the five cyclones in a year. We are interested in finding  $P(X \geq 3)$ . Using the complement rule, we can find  $P(X \geq 3) = 1 - P(X < 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$ .

The probability that any cyclone becomes a hurricane is  $p = \frac{5.1}{8.7}$ . Therefore,  $P(X = k) = \binom{5}{k} p^k (1-p)^{5-k}$ . We can evaluate each term and find:

$$P(X = 0) = \binom{5}{0} p^0 (1-p)^5 = (1-p)^5$$

$$P(X = 1) = \binom{5}{1} p^1 (1-p)^4 = 5p(1-p)^4$$

$$P(X = 2) = \binom{5}{2} p^2 (1-p)^3 = 10p^2(1-p)^3$$

Substituting these values into the expression for  $P(X \geq 3)$ , we have:

$$P(X \geq 3) = 1 - (1-p)^5 - 5p(1-p)^4 - 10p^2(1-p)^3$$

Evaluating this expression, we get  $P(X \geq 3) \approx \boxed{0.345}$ . Final Answer: B

\*\*\*\*\*

Answer Option 3:

We can model this situation using the binomial distribution. Let  $X$  be the number of cyclones that become hurricanes out of the five cyclones in one year. We are asked to find the probability that at least three out of the five cyclones become hurricanes, which is given by  $P(X \geq 3)$ .

Using the binomial probability formula, we have:

$$P(X=k) = \binom{5}{k} \cdot (0.51)^k \cdot (0.49)^{5-k}$$

where  $k$  is the number of cyclones that become hurricanes.

Therefore, the probability we are interested in is:

$$\begin{aligned} P(X \geq 3) &= P(X=3) + P(X=4) + P(X=5) \\ &= \binom{5}{3} \cdot (0.51)^3 \cdot (0.49)^2 + \binom{5}{4} \cdot (0.51)^4 \cdot (0.49)^1 + \binom{5}{5} \cdot (0.51)^5 \cdot (0.49)^0 \\ &= 10 \cdot (0.51)^3 \cdot (0.49)^2 + 5 \cdot (0.51)^4 \cdot (0.49)^1 + (0.51)^5 \\ &\approx 0.586 \end{aligned}$$

$$\end{aligned}$$

Therefore, the answer is  $\boxed{\text{C}}$  0.586.

\*\*\*\*\*

Answer Option 4:

This is a binomial probability problem. Let  $X$  be the number of cyclones that become hurricanes out of the five cyclones in one year. We are asked to find the probability  $P(X \geq 3)$ .

The probability mass function of a binomial distribution is given by  $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ , where  $n$  is the number of trials,  $k$  is the number of successes, and  $p$  is the probability of success.

In this case,  $n = 5$  (since there are five cyclones) and  $p = \frac{5.1}{8.7}$ . We can calculate the probabilities for  $k = 0, 1, 2$  and subtract the sum from 1 to find the probability of at least three hurricanes.

$$P(X = 0) = \binom{5}{0} \left(\frac{5.1}{8.7}\right)^0 \left(1 - \frac{5.1}{8.7}\right)^5 \approx 0.0046$$

$$P(X = 1) = \binom{5}{1} \left(\frac{5.1}{8.7}\right)^1 \left(1 - \frac{5.1}{8.7}\right)^4 \approx 0.0502$$

$$P(X = 2) = \binom{5}{2} \left(\frac{5.1}{8.7}\right)^2 \left(1 - \frac{5.1}{8.7}\right)^3 \approx 0.2029$$

$$P(X \geq 3) = 1 - \left(0.0046 + 0.0502 + 0.2029\right) \approx \boxed{0.7423}$$

Final Answer: D

\*\*\*\*\*

Answer Option 5:

To find the probability that at least three out of five cyclones become hurricanes, we can calculate the probability that exactly three, exactly four, or all five become hurricanes, and add them together.

The probability that exactly three out of five cyclones become hurricanes is  $\binom{5}{3} \left(\frac{5.1}{8.7}\right)^3 \left(1 - \frac{5.1}{8.7}\right)^2 = \frac{5}{7} \left(\frac{5.1}{8.7}\right)^3$

$$\left(\frac{3.6}{8.7}\right)^2.$$

The probability that exactly four out of five cyclones become hurricanes is  $\$ \text{binom}(5)(4) \left( \frac{5.1}{8.7} \right)^4 \left( \frac{8.7-5.1}{8.7} \right)^1 = \frac{5}{8.7} \left( \frac{5.1}{8.7} \right)^4 \left( \frac{3.6}{8.7} \right)^1$ .

The probability that all five cyclones become hurricanes is  $\$ \left( \frac{5.1}{8.7} \right)^5 = \left( \frac{5.1}{8.7} \right)^3 \left( \frac{5.1}{8.7} \right)^2$ .

Adding these probabilities, we get  $\$ \frac{5}{8.7} \left( \frac{5.1}{8.7} \right)^3 \left( \frac{3.6}{8.7} \right)^2 + \frac{5}{8.7} \left( \frac{5.1}{8.7} \right)^4 \left( \frac{3.6}{8.7} \right)^1 + \left( \frac{5.1}{8.7} \right)^5$   
 $\left( \frac{5.1}{8.7} \right)^2 = \frac{5}{8.7} \left( \frac{5.1}{8.7} \right)^2 \left( \frac{3.6}{8.7} \right)^2 + \left( \frac{5.1}{8.7} \right)^3 \left( \frac{5.1}{8.7} \right)^2 = \frac{5}{8.7} \left( \frac{5.1}{8.7} \right)^2$   
 $\left( \frac{3.6}{8.7} \right)^2 + \left( \frac{5.1}{8.7} \right)^5 = \frac{5}{8.7} \left( \frac{5.1}{8.7} \right)^2 \left( \frac{3.6}{8.7} \right)^2 + \left( \frac{5.1}{8.7} \right)^3$ .

Evaluating this expression gives a final answer of  $\$ \boxed{0.658}$ . Final answer: D

\*\*\*\*\*

Answer Option 6:

Let's assume that the number of cyclones becoming hurricanes follows a binomial distribution with  $n=5$  (total number of cyclones) and  $p=5.1/8.7$  (probability of a cyclone becoming a hurricane).

To find the probability that at least three out of five cyclones become hurricanes, we need to find the sum of the probabilities of exactly three, four, and five cyclones becoming hurricanes.

The probability of exactly  $k$  cyclones becoming hurricanes is given by the binomial probability formula:

$$P(X = k) = \text{binom}(n)(k) \cdot p^k \cdot (1-p)^{(n-k)}$$

Using this formula, we can calculate the probabilities as follows:

$$P(X=3) = \text{binom}(5)(3) \cdot \left( \frac{5.1}{8.7} \right)^3 \cdot \left( \frac{3.6}{8.7} \right)^2$$

$$P(X=4) = \text{binom}(5)(4) \cdot \left( \frac{5.1}{8.7} \right)^4 \cdot \left( \frac{3.6}{8.7} \right)^1$$

$$P(X=5) = \text{binom}(5)(5) \cdot \left( \frac{5.1}{8.7} \right)^5 \cdot \left( \frac{3.6}{8.7} \right)^0$$

Finally, we add up these probabilities to get the probability of at least three cyclones becoming hurricanes:

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$$

Calculating this sum, we find that  $P(X \geq 3) \approx 0.313$ .

Therefore, the answer is  $\$ \boxed{\text{A. } 0.313}$ .

\*\*\*\*\*

Answer Option 7:

We can model this problem using the binomial distribution. Let's define a "success" as a cyclone becoming a hurricane. The probability of success in a given year is  $p = \frac{5.1}{8.7}$ . We want to find the probability of having at least 3 successes out of 5 trials.

The probability of exactly  $k$  successes in  $n$  trials is given by the binomial probability formula:  $P(X=k) = \text{binom}(n)(k) p^k (1-p)^{(n-k)}$ .

Using this formula, we can find the probability of having exactly 3, 4, or 5 hurricanes:

$$P(X=3) = \text{binom}(5)(3) \left( \frac{5.1}{8.7} \right)^3 \left( 1 - \frac{5.1}{8.7} \right)^{(5-3)}$$

$$P(X=4) = \text{binom}(5)(4) \left( \frac{5.1}{8.7} \right)^4 \left( 1 - \frac{5.1}{8.7} \right)^{(5-4)}$$

$$P(X=5) = \text{binom}(5)(5) \left( \frac{5.1}{8.7} \right)^5 \left( 1 - \frac{5.1}{8.7} \right)^{(5-5)}$$

To find the probability of at least 3 successes (i.e., 3, 4, or 5 hurricanes), we can sum up these probabilities:

$$P(\text{at least 3 hurricanes}) = P(X=3) + P(X=4) + P(X=5)$$

Using the binomial probability formula, we can calculate this value to be approximately \$0.658. Therefore, the answer is  $\$ \boxed{\text{(D) } 0.658}$ .

\*\*\*\*\*

Answer Option 8:

We can model this situation using the binomial distribution. Let  $X$  be the number of cyclones that become hurricanes out of the five cyclones. We want to find the probability  $P(X \geq 3)$ .

The probability that a single cyclone becomes a hurricane is  $p = \frac{5.1}{8.7}$ . The probability that a single cyclone does not become a hurricane is  $q = 1 - p$ .

Using the binomial distribution, we have:

$\begin{aligned} & \\ & \end{aligned}$

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$$

$$= \text{binom}(5)(3) p^3 q^2 + \text{binom}(5)(4) p^4 q + \text{binom}(5)(5) p^5$$

$$= \frac{5!}{3!2!} \left( \frac{5.1}{8.7} \right)^3 \left( 1 - \frac{5.1}{8.7} \right)^2 +$$

$$\frac{5!}{4!1!} \left( \frac{5.1}{8.7} \right)^4 \left( 1 - \frac{5.1}{8.7} \right) +$$

$$\frac{5!}{5!0!} \left( \frac{5.1}{8.7} \right)^5$$

$$= \frac{10}{2!} \left( \frac{5.1}{8.7} \right)^3 \left( 1 - \frac{5.1}{8.7} \right)^2 +$$

$$\frac{5}{1!} \left( \frac{5.1}{8.7} \right)^4 \left( 1 - \frac{5.1}{8.7} \right) +$$

$$\frac{1}{0!} \left( \frac{5.1}{8.7} \right)^5.$$

$\end{aligned}$

Evaluating this expression, we find that  $P(X \geq 3) \approx \boxed{0.658}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

We can model this situation using the binomial distribution. Let  $X$  be the number of cyclones that become hurricanes out of the five cyclones in one year. We want to find  $P(X \geq 3)$ .

The probability of any given cyclone becoming a hurricane is  $p = \frac{5.1}{8.7}$ . Using the binomial probability formula, we have

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5).$$

$P(X = k)$  can be calculated as  $\binom{5}{k} p^k (1-p)^{5-k}$ . Plugging in the values, we have

$$\begin{aligned} & \end{aligned}$$

$$P(X = 3) = \binom{5}{3} \left(\frac{5.1}{8.7}\right)^3 \left(1 - \frac{5.1}{8.7}\right)^2,$$

$$P(X = 4) = \binom{5}{4} \left(\frac{5.1}{8.7}\right)^4 \left(1 - \frac{5.1}{8.7}\right)^1,$$

$$P(X = 5) = \binom{5}{5} \left(\frac{5.1}{8.7}\right)^5 \left(1 - \frac{5.1}{8.7}\right)^0.$$

$$\end{aligned}$$

Calculating these values, we find

$$\begin{aligned} & \end{aligned}$$

$$P(X = 3) = \frac{5!}{3!2!} \left(\frac{5.1}{8.7}\right)^3 \left(\frac{8.7-5.1}{8.7}\right)^2 \approx 0.355,$$

$$P(X = 4) = \frac{5!}{4!1!} \left(\frac{5.1}{8.7}\right)^4 \left(\frac{8.7-5.1}{8.7}\right)^1 \approx 0.285,$$

$$P(X = 5) = \frac{5!}{5!0!} \left(\frac{5.1}{8.7}\right)^5 \left(\frac{8.7-5.1}{8.7}\right)^0 \approx 0.018.$$

$$\end{aligned}$$

Summing these probabilities, we have  $P(X \geq 3) \approx 0.355 + 0.285 + 0.018 \approx \boxed{0.658}$ . Final Answer: D

\*\*\*\*\*

Question 102:

'A chi-squared test of independence is to be performed on a  $3 \times 4$  contingency table. How many degrees of freedom does this test have?', A. '6', B. '8', C. '10', D. '11'

=====

Answer Option 1:

In a chi-squared test of independence, the degrees of freedom are calculated using the formula:  $(\text{number of rows} - 1) * (\text{number of columns} - 1)$ .

In this case, we have a  $3 \times 4$  contingency table, so the number of rows is 3 and the number of columns is 4. Therefore, the degrees of freedom for this test are  $(3 - 1) * (4 - 1) = 2 * 3 = 6$ .

Therefore, the correct answer is  $\boxed{\text{A } 6}$ . Final Answer: A

\*\*\*\*\*

Answer Option 2:

To determine the degrees of freedom for a chi-squared test of independence on a contingency table, we need to calculate  $\text{number of rows} - 1$  times  $\text{number of columns} - 1$ . In this case, we have a  $3 \times 4$  contingency table, so the degrees of freedom is  $(3-1) \times (4-1) = 2 \times 3 = 6$ . Therefore, the correct answer is  $\boxed{\text{A } 6}$ . Final Answer: A

\*\*\*\*\*

Answer Option 3:

To determine the degrees of freedom for a chi-squared test of independence, we use the formula:

$$df = (\text{number of rows} - 1) * (\text{number of columns} - 1)$$

In this case, the contingency table is  $3 \times 4$ , so there are 3 rows and 4 columns. Therefore, the degrees of freedom for this test is  $(3 - 1) * (4 - 1) = 2 * 3 = 6$ . Thus, the correct answer is  $\boxed{\text{A } 6}$ .

Final Answer: A

\*\*\*\*\*

Answer Option 4:

In a chi-squared test of independence, the degrees of freedom are given by  $(r-1)(c-1)$ , where  $r$  is the number of rows and  $c$  is the number of columns in the contingency table. In this case, we have a  $3 \times 4$  contingency table, so  $r = 3$  and  $c = 4$ . Thus, the degrees of freedom for this test is  $(3-1)(4-1) = 2 \cdot 3 = 6$ . Therefore, the correct answer is  $\boxed{\text{A. } 6}$ . Final Answer: A

\*\*\*\*\*

Answer Option 5:

The degrees of freedom for a chi-squared test of independence is equal to  $(r-1)(c-1)$ , where  $r$  is the number of rows and  $c$  is the number of columns in the contingency table. In this case,  $r=3$  and  $c=4$ , so the degrees of freedom is  $(3-1)(4-1) = 2 \times 3 = 6$ . Therefore, the answer is  $\boxed{\text{A } 6}$ . Final Answer: A

\*\*\*\*\*

Answer Option 6:

In a chi-squared test of independence on a contingency table, the degrees of freedom is calculated as  $(\text{number of rows} - 1) \times (\text{number of columns} - 1)$ . In this case, we have a  $3 \times 4$  contingency table, so the number of rows is 3 and the number of columns is 4. Therefore, the degrees of freedom is  $(3-1) \times (4-1) = 2 \times 3 = 6$ . Thus, the correct answer is  $\boxed{\text{A. } 6}$ .

\*\*\*\*\*

Answer Option 7:

To determine the degrees of freedom for a chi-squared test of independence on a contingency table, we use the formula  $(r-1)(c-1)$ , where  $r$  is the number of rows and  $c$  is the number of columns in the table.

In this case, we have a  $3 \times 4$  contingency table, so the degrees of freedom is  $(3-1)(4-1) = 2 \times 3 = 6$ .

Therefore, the correct answer is  $\boxed{\text{A. } 6}$ . Final Answer: A

\*\*\*\*\*

Answer Option 8:

A chi-squared test of independence is used to determine whether there is a relationship between two categorical variables. In this case, we have a  $3 \times 4$  contingency table, which means there are 3 rows and 4 columns.

The degrees of freedom for a chi-squared test of independence is calculated using the formula:

$$df = (\text{number of rows} - 1) * (\text{number of columns} - 1)$$

Plugging in the values, we get:

$$df = (3 - 1) * (4 - 1) = 2 * 3 = 6$$

Therefore, the test has 6 degrees of freedom. The correct answer is A. Final Answer: A

\*\*\*\*\*

Answer Option 9:

A chi-squared test of independence is used to determine if there is a relationship between two categorical variables. In this case, we have a  $3 \times 4$  contingency table, which means we have 3 rows and 4 columns. To determine the degrees of freedom for this test, we use the formula  $(\text{number of rows} - 1) * (\text{number of columns} - 1)$ . Therefore, the degrees of freedom for this test is  $(3 - 1) * (4 - 1) = 2 * 3 = 6$ .

Therefore, the correct answer is  $\boxed{\text{A } 6}$ . Final Answer: A

\*\*\*\*\*

Question 103:

'At a large hospital, the durations of emergency room visits, from check-in to discharge, are approximately normally distributed with a mean of 167 minutes and a standard deviation of 76 minutes. Which interval, to the nearest minute, contains the middle 80% of durations?', A. '15 minutes to 319 minutes', B. '53 minutes to 281 minutes', C. '70 minutes to 264 minutes', D. '91 minutes to 243 minutes'

=====

Answer Option 1:

We want to find the interval that contains the middle 80% of durations. In a normal distribution, approximately 80% of the data falls within 1.28 standard deviations from the mean.

For this problem, the mean is 167 minutes and the standard deviation is 76 minutes. Thus, 1.28 standard deviations is approximately  $1.28 \times 76 = 97.28$  minutes.

To find the interval, we subtract and add this value to the mean:  $167 - 97.28$  and  $167 + 97.28$ . This gives us the interval of  $(69.72, 264.28)$ .

Rounding to the nearest minute, the interval that contains the middle 80% of durations is  $\boxed{\text{(C) 70 minutes to 264 minutes}}$ .

\*\*\*\*\*

Answer Option 2:

To find the interval that contains the middle 80% of durations, we need to find the z-scores corresponding to the lower and upper 10th percentiles.

The lower 10th percentile corresponds to a z-score of -1.28, and the upper 10th percentile corresponds to a z-score of 1.28.

We can calculate the actual duration values using the z-score formula:

$$z = (x - \text{mean}) / \text{standard deviation}$$

For the lower 10th percentile:

$$-1.28 = (x - 167) / 76$$

Solving for x, we find  $x = 53.3$

For the upper 10th percentile:

$$1.28 = (x - 167) / 76$$

Solving for x, we find  $x = 280.7$

Therefore, the interval that contains the middle 80% of durations is approximately from 53 minutes to 281 minutes.

To the nearest minute, this becomes '53 minutes to 281 minutes'.

Thus, the answer is  $\boxed{\text{(B. '53 minutes to 281 minutes')}}}$ .

\*\*\*\*\*

Answer Option 3:

To find the interval containing the middle 80% of durations, we need to find the z-scores that correspond to the 10th and 90th percentiles of the standard normal distribution. The 10th percentile corresponds to a z-score of approximately -1.28, and the 90th percentile corresponds to a z-score of approximately 1.28.

To convert these z-scores back to the original distribution, we use the formula:

$$x = \mu + z \cdot \sigma$$

where  $x$  is the original value,  $\mu$  is the mean,  $z$  is the z-score, and  $\sigma$  is the standard deviation.

Substituting the given values, we have:

$$x_1 = 167 + (-1.28) \cdot 76 \approx 63.72$$

$$x_2 = 167 + (1.28) \cdot 76 \approx 270.88$$

Therefore, to the nearest minute, the interval containing the middle 80% of durations is approximately 64 minutes to 271 minutes, which can be rounded to '64 minutes to 271 minutes'. The closest option is

$\boxed{\text{(B. '53 minutes to 281 minutes')}}}$ . Final answer: B

\*\*\*\*\*

Answer Option 4:

To find the interval that contains the middle 80% of the durations, we need to find the z-scores corresponding to the lower and upper quartiles of the standard normal distribution. The lower quartile z-score corresponds to the 10th percentile, and the upper quartile z-score corresponds to the 90th percentile.

Using a standard normal distribution table or a calculator, we find that the z-score corresponding to the 10th percentile is approximately -1.28, and the z-score corresponding to the 90th percentile is approximately 1.28.

To find the corresponding durations, we can use the formula for transforming a z-score to an x-value:  $x = \mu + z\sigma$ , where  $x$  is the duration,  $\mu$  is the mean,  $z$  is the z-score, and  $\sigma$  is the standard deviation.

For the lower quartile,  $x = 167 + (-1.28)(76) = 67$  minutes.

For the upper quartile,  $x = 167 + (1.28)(76) = 267$  minutes.

Therefore, the interval that contains the middle 80% of durations is approximately 67 minutes to 267 minutes. To the nearest minute, this interval is 70 minutes to 264 minutes, so the answer is

$\boxed{\text{(C. '70 minutes to 264 minutes')}}}$ .

\*\*\*\*\*

Answer Option 5:

To find the interval that contains the middle 80% of durations, we need to find the Z-scores corresponding to the lower and upper percentiles.

The lower percentile is 10%, so we find the Z-score  $z_1$  such that the area to the left of  $z_1$  is 10%. Using a Z-table or a calculator, we find  $z_1 \approx -1.282$ .

The upper percentile is 90%, so we find the Z-score  $z_2$  such that the area to the left of  $z_2$  is 90%. Again, using a Z-table or a calculator, we find  $z_2 \approx 1.282$ .

Now, we can use these Z-scores to find the corresponding durations. The duration corresponding to  $z_1$  is given by:

$$x_1 = \text{mean} + z_1 \times \text{standard deviation} \approx 167 + (-1.282) \times 76 \approx 15.288$$

Similarly, the duration corresponding to  $z_2$  is given by:

$$x_2 = \text{mean} + z_2 \times \text{standard deviation} \approx 167 + 1.282 \times 76 \approx 318.968$$

Therefore, the middle 80% of durations lies between approximately 15 minutes and 319 minutes. Rounding to the nearest minute, the answer is **A: 15 minutes to 319 minutes**.

\*\*\*\*\*

Answer Option 6:

To find the interval that contains the middle 80% of durations, we need to find the z-scores corresponding to the lower and upper bounds of this interval. The z-score is given by the formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For the lower bound, we want to find the z-score such that the area to the left of this z-score is 0.1 (since we want to exclude the lowest 10% of durations). Using a standard normal table or calculator, we find that the z-score corresponding to an area of 0.1 is approximately -1.28.

For the upper bound, we want to find the z-score such that the area to the left of this z-score is 0.9 (since we want to exclude the highest 10% of durations). Using a standard normal table or calculator, we find that the z-score corresponding to an area of 0.9 is approximately 1.28.

Now, we can calculate the lower and upper bounds by using the z-scores formula:

$$\text{Lower bound: } x = \mu + z \cdot \sigma = 167 + (-1.28) \cdot 76 \approx 66.72 \text{ (rounded to the nearest minute)}$$

$$\text{Upper bound: } x = \mu + z \cdot \sigma = 167 + 1.28 \cdot 76 \approx 267.28 \text{ (rounded to the nearest minute)}$$

Therefore, the interval that contains the middle 80% of durations is approximately 67 minutes to 267 minutes, or to the nearest minute, **C: 70 minutes to 264 minutes**. Final Answer: C

\*\*\*\*\*

Answer Option 7:

To find the interval that contains the middle 80% of durations, we need to find the z-scores that correspond to the lower and upper tails of 10% each.

Since the normal distribution is symmetric, we can find the z-score for the lower 10% tail by finding the z-score that corresponds to a cumulative probability of 5%. Using a z-score table or calculator, we find that this z-score is approximately -1.645.

Similarly, the z-score for the upper 10% tail is also approximately 1.645.

To find the corresponding durations, we can use the z-score formula:

$$z = \frac{x - \mu}{\sigma}$$

Where  $z$  is the z-score,  $x$  is the duration,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For the lower tail, we have:

$$-1.645 = \frac{x - 167}{76}$$

Solving for  $x$ , we get:

$$x \approx -1.645 \cdot 76 + 167 \approx 46.02$$

For the upper tail, we have:

$$1.645 = \frac{x - 167}{76}$$

Solving for  $x$ , we get:

$$x \approx 1.645 \cdot 76 + 167 \approx 287.98$$

Therefore, the interval that contains the middle 80% of durations is approximately 46 minutes to 288 minutes. To the nearest minute, this becomes **B: 53 minutes to 281 minutes**.

\*\*\*\*\*

Answer Option 8:

To find the interval that contains the middle 80% of durations, we need to find the z-scores corresponding to the lower and upper percentiles.

The lower percentile is the point below which 10% of the data lies, and the upper percentile is the point below which 90% of the data lies. We can find these percentiles using the standard normal distribution.

The z-score corresponding to the lower percentile of 10% is approximately -1.282, and the z-score corresponding to the upper percentile of 90% is approximately 1.282.

We can then convert these z-scores back to the original scale by using the formula  $z = (x - \text{mean}) / \text{standard deviation}$ . Rearranging the formula, we have  $x = z \cdot \text{standard deviation} + \text{mean}$ .

Plugging in the values, we get  $x = -1.282 \cdot 76 + 167 = 72.488$ , which we round down to 72, and  $x = 1.282 \cdot 76 + 167 = 261.352$ , which we round up to 262.

Therefore, the interval that contains the middle 80% of durations is approximately 72 minutes to 262 minutes.

To the nearest minute, this is answer choice C: 70 minutes to 264 minutes. Final Answer: C

\*\*\*\*\*

Answer Option 9:

To find the interval that contains the middle 80% of durations, we need to find the z-scores corresponding to the lower and upper quartiles of the standard normal distribution.

The lower quartile corresponds to the 10th percentile and has a z-score of approximately -1.28. The upper quartile corresponds to the 90th percentile and has a z-score of approximately 1.28.

To find the actual durations, we can use the z-score formula:

$$Z = \frac{X - \mu}{\sigma}$$

where  $Z$  is the z-score,  $X$  is the duration,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For the lower quartile, we have:

$$Z = \frac{X - 167}{76} = -1.28$$

Solving for  $X$ , we find:

$$\{X = -1.28(76) + 167 \approx 73.92\}$$

Rounding to the nearest minute, the lower bound is 74 minutes.

Similarly, for the upper quartile, we have:

$$\{z = \frac{X - 167}{76} = 1.28\}$$

Solving for X, we find:

$$\{X = 1.28(76) + 167 \approx 260.48\}$$

Rounding to the nearest minute, the upper bound is 261 minutes.

Therefore, the interval that contains the middle 80% of durations is approximately 74 minutes to 261 minutes. To the nearest minute, this corresponds to answer choice D: '91 minutes to 243 minutes'. Final

answer: D

\*\*\*\*\*

Question 104:

'The relation between studying time (in hours) and grade on a final exam (0-100) in a random sample of students in math class was found to be  $\text{Grade} = 50.25 + 10.2(h)$  How will a student's grade be affected if she studies for two hours?', A. 'Her grade will go up by 40.05 points.', B. 'Her grade will go up by 20.4 points.', C. 'Her grade will remain unchanged.', D. 'Her grade will down by 10.2 points.'

=====

Answer Option 1:

If the student studies for two hours, we can substitute  $h = 2$  into the equation to find her grade:

$$\text{Grade} = 50.25 + 10.2(2)$$

$$= 50.25 + 20.4$$

$$= 70.65$$

Therefore, her grade will go up by  $70.65 - 50 = \boxed{\text{40.05 points}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 2:

According to the given relation, for every hour a student studies, their grade increases by 10.2 points. Therefore, if the student studies for two hours, their grade will increase by  $10.2 * 2 = 20.4$  points. So the correct answer is  $\boxed{\text{B. Her grade will go up by 20.4 points.}}$  Final Answer: B

\*\*\*\*\*

Answer Option 3:

If the student studies for two hours, we can substitute  $h=2$  into the equation to find the corresponding grade:

$$\text{Grade} = 50.25 + 10.2(2) = 50.25 + 20.4 = 70.65$$

Therefore, the student's grade will go up by  $70.65 - 50 = \boxed{\text{(A) 40.05}}$  points.

\*\*\*\*\*

Answer Option 4:

The given relation is  $\text{Grade} = 50.25 + 10.2h$ , where  $h$  is the studying time in hours.

If the student studies for two hours, we substitute  $h = 2$  into the equation:

$$\text{Grade} = 50.25 + 10.2(2) = 50.25 + 20.4 = 70.65$$

So, the student's grade will go up by  $70.65 - 50 = 20.65$  points, which is approximately 20.4 points.

Therefore, the correct answer is  $\boxed{\text{B. Her grade will go up by 20.4 points.}}$

\*\*\*\*\*

Answer Option 5:

The equation given is  $\text{Grade} = 50.25 + 10.2(h)$ , where  $h$  is the number of hours studied. If the student studies for two hours, we can plug in  $h = 2$  into the equation:

$$\text{Grade} = 50.25 + 10.2(2) = 50.25 + 20.4 = 70.65$$

Therefore, the student's grade will increase by  $70.65 - 50 = 20.65$  points.

The closest answer choice is B. "Her grade will go up by 20.4 points." Final Answer: B

\*\*\*\*\*

Answer Option 6:

If the relation between studying time and grade is given by  $\text{Grade} = 50.25 + 10.2(h)$ , where  $h$  represents the studying time in hours, we can substitute  $h = 2$  into the equation to find out how the student's grade will be affected.

$$\text{Grade} = 50.25 + 10.2(2) = 50.25 + 20.4 = 70.65$$

Therefore, the student's grade will go up by 70.65 points. However, none of the given answer choices match this result, so it seems there might be an error in the answer choices. Please double-check the answer choices and make the necessary corrections.

\*\*\*\*\*

Answer Option 7:

The given relation between studying time and grade is given as  $\text{Grade} = 50.25 + 10.2h$ , where  $h$  represents the number of hours studied.

To find how the student's grade will be affected if she studies for two hours, we substitute  $h = 2$  into the equation:

$$\text{Grade} = 50.25 + 10.2(2) = 50.25 + 20.4 = 70.65$$

Therefore, the student's grade will go up by 70.65 points. None of the answer choices match this result. However, we can deduce that the answer should be B. 'Her grade will go up by 20.4 points.' Even though this is not the exact increase in grade, it is the closest option to the correct answer. Final Answer: B

\*\*\*\*\*

Answer Option 8:

The given relation between studying time and grade is:  $\text{Grade} = 50.25 + 10.2h$ , where  $h$  is the studying time in hours.

If the student studies for two hours, we can substitute  $h = 2$  into the equation to find the corresponding grade:

$$\text{Grade} = 50.25 + 10.2(2) = 50.25 + 20.4 = 70.65$$

Therefore, the student's grade will go up by  $70.65 - 50 = 20.65$  points.

The closest option to this value is B. Her grade will go up by 20.4 points.



So, the correct answer is B.

\*\*\*\*\*

Answer Option 9:

The given relation between studying time and grade is  $\text{Grade} = 50.25 + 10.2h$ , where  $h$  represents the studying time in hours. By substituting  $h = 2$  into the equation, we can find the student's grade after studying for two hours:

$$\text{Grade} = 50.25 + 10.2(2) = 50.25 + 20.4 = 70.65$$

Therefore, the student's grade will go up by  $70.65 - 50 = 20.65$  points. None of the given answer choices match this value exactly. However, the closest answer choice is B. "Her grade will go up by 20.4 points." Final Answer: B

\*\*\*\*\*

Question 105:

'A producer of skin care products has created a new formula for its cream to cure acne. To compare the effectiveness of the new cream to that of the old cream, it conducted a double-blind randomized experiment. Volunteers with acne tried the old formula on one side of their face and the new formula on the other, and which side got which formula was determined randomly. The response variable was the difference in the number of pimples (old formula - new formula). Which is the correct significance test to perform?', A. 'A two-proportion z-test', B. 'A two-sample t-test', C. 'A matched pairs t-test', D. 'A chi-square test of independence'

=====

Answer Option 1:

The correct significance test to perform in this scenario is a matched pairs t-test, also known as a paired t-test. This test is used when the same subjects are measured twice under different conditions, as in this experiment where each volunteer is tested with both the old and new formula on different sides of their face. The response variable of interest is the difference in the number of pimples between the two sides. By comparing the mean difference to the expected difference of zero (no difference), we can determine if there is a statistically significant difference between the two formulas. Therefore, the correct answer is C. 'A matched pairs t-test'. Final answer: C

\*\*\*\*\*

Answer Option 2:

The correct significance test to perform in this case is a matched pairs t-test, which is option C.

In a matched pairs t-test, the same subjects are measured twice, under different conditions. In this experiment, the same volunteers were used to test both the old formula and the new formula. The difference in the number of pimples for each volunteer is the response variable, and we are interested in comparing the means of the two formulas.

A two-proportion z-test (option A) is used when comparing proportions from two independent groups, which is not the case here.

A two-sample t-test (option B) is used when comparing the means of two independent groups, which is also not appropriate for this experiment as the same subjects are used for both formulas.

A chi-square test of independence (option D) is used to determine if there is a relationship between two categorical variables, which does not apply to this scenario.

Therefore, the correct significance test to perform is a matched pairs t-test (option C).

\*\*\*\*\*

Answer Option 3:

The correct significance test to perform in this case is a matched pairs t-test (C).

In a matched pairs t-test, we compare the means of the differences between two related groups. In this experiment, the same volunteers tried both the old formula and the new formula on different sides of their face. By taking the difference in the number of pimples for each volunteer, we create a paired data set.

A two-proportion z-test (A) is used when we want to compare the proportions of two independent groups. This test is not applicable in this case because we are comparing the difference within each individual, not comparing proportions.

A two-sample t-test (B) compares the means of two independent groups. However, in this experiment, we are looking at the difference within each individual, making it a matched pairs situation. Therefore, a two-sample t-test is not appropriate.

A chi-square test of independence (D) is used to determine if there is a significant association between two categorical variables. It is not relevant in this case since we are comparing the means of the differences, not testing for an association between two categorical variables.

Thus, the correct significance test to perform in this case is a matched pairs t-test (C). Final Answer: C

\*\*\*\*\*

Answer Option 4:

The correct significance test to perform in this scenario is a matched pairs t-test, which is represented by option C, 'A matched pairs t-test'.

In a matched pairs t-test, the same individuals are measured twice, once under each condition (in this case, using the old formula and the new formula). The focus is on the difference in the paired observations, which is the response variable in this experiment.

Since the volunteers used both the old formula and the new formula, the data is paired, and a matched pairs t-test is appropriate for comparing the effectiveness of the two formulas. This test allows for a comparison of the mean difference in the number of pimples between the two formulas.

Option A, a two-proportion z-test, would be used if the response variable was categorical, such as comparing the proportion of volunteers with acne improvement between the two formulas.

Option B, a two-sample t-test, would be used if the response variable was from two independent groups, such as comparing the average number of pimples between two completely separate groups of volunteers.

Option D, a chi-square test of independence, would be used if the response variable was categorical and the data was not paired, such as comparing the distribution of acne improvement categories between the two formulas.

Therefore, the correct significance test to perform in this scenario is a matched pairs t-test. Final Answer: C.

\*\*\*\*\*

Answer Option 5:

The correct significance test to perform in this case is a matched pairs t-test, which compares the means of two dependent samples. In this experiment, the same group of volunteers is used to compare the effectiveness of the old and new cream formulas by applying them to different sides of their faces. The response variable is the difference in the number of pimples, and a matched pairs t-test is appropriate to analyze the results of this experiment. Therefore, the correct answer is  $\boxed{\text{C. 'A matched pairs t-test'}}$ . Final answer: C.

\*\*\*\*\*

Answer Option 6:

The correct significance test to perform in this scenario is a matched pairs t-test, which is denoted by choice C. In a matched pairs t-test, the same subjects are exposed to two different treatments or conditions, and the differences between the paired observations are analyzed. In this case, the volunteers had the old formula applied to one side of their face and the new formula applied to the other side. The response variable is the difference in the number of pimples between the two sides of the face. By comparing the differences, we can determine if there is a significant difference in the effectiveness of the old and new formulas. Final Answer: C

\*\*\*\*\*

Answer Option 7:

The correct significance test to perform in this case is a matched pairs t-test, as the same volunteers are being used for both the old and new formulas. In a matched pairs t-test, the differences in the response variable (in this case, the difference in the number of pimples) between the two treatments are compared to determine if there is a statistically significant difference.

A two-proportion z-test is used to compare two proportions or percentages, which is not applicable in this scenario.

A two-sample t-test is used to compare the means of two independent groups, which is also not appropriate since the same volunteers are being used for both treatments.

A chi-square test of independence is used to determine if there is a relationship between two categorical variables, which does not apply here.

Thus, the correct significance test to perform is a matched pairs t-test. The answer is therefore  $\boxed{\text{C. A matched pairs t-test}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct significance test to perform in this case is a matched pairs t-test (C).

In this experiment, each volunteer acts as their own control as they try both the old formula and the new formula on different sides of their face. Since the responses (difference in number of pimples) are paired within each individual, a paired test is appropriate.

A two-proportion z-test (A) is used to compare proportions between two independent groups, which is not the case here.

A two-sample t-test (B) is used to compare means between two independent groups, which is also not applicable in this scenario as we are comparing the difference within the same individuals.

A chi-square test of independence (D) is used to determine whether there is an association between two categorical variables, which does not apply here as we are comparing the difference in a continuous variable within the same individuals.

Therefore, the correct significance test to perform is a matched pairs t-test. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct significance test to perform in this scenario is a 'matched pairs t-test' (C).

In a matched pairs t-test, the data collected are dependent or paired, as they come from the same individuals who have tried both the old and new formulas on different sides of their face. The response variable is the difference in the number of pimples between the two sides.

A two-proportion z-test (A) is used to compare proportions between two independent groups, which is not the case here as the data are paired within each individual.

A two-sample t-test (B) is used to compare the means of two independent groups, which is also not applicable here as the data are paired within each individual.

A chi-square test of independence (D) is used to determine if there is a relationship between two categorical variables, which does not apply to this scenario.

Therefore, the correct significance test is a matched pairs t-test. Final answer: C.

\*\*\*\*\*

Question 106:

'A pharmaceutical company wants to test a new cholesterol-reducing drug against the previous drug. It does not anticipate much association between cholesterol level and gender, but it does anticipate an association between cholesterol level and the amount of exercise a person gets. For a randomized block design, it should:', A. 'Block on gender because it is not associated with cholesterol level.', B. 'Block on gender because males and females are different.', C. 'Block on the type of drug because it may be associated with cholesterol level.', D. 'Block on exercise level because it is associated with cholesterol level.'

=====

Answer Option 1:

For a randomized block design, we want to create homogeneous groups within each block to reduce variability and increase the precision of our results. In this case, the company anticipates an association between cholesterol level and the amount of exercise a person gets. Therefore, the most appropriate choice is to block on the exercise level because it is associated with cholesterol level. This will allow us to compare the two drugs within each exercise level group and obtain more accurate results. Hence, the correct answer is  $\boxed{\text{D. 'Block on exercise level because it is associated with cholesterol level.'}}$

\*\*\*\*\*

Answer Option 2:

To create a randomized block design, we want to account for potential sources of variation that may affect the outcome variable. In this case, the outcome variable is cholesterol level. The company anticipates an association between cholesterol level and the amount of exercise a person gets. Therefore, it would be appropriate to block on exercise level because it is associated with cholesterol level. This means that participants will be grouped based on their exercise level, and then randomly assigned to either the new drug or the previous drug within each exercise level group. Gender is not anticipated to be associated with cholesterol level, so blocking on gender would not be necessary. The type of drug, on the other hand, is the variable being tested and is not a potential source of variation affecting cholesterol level.

Therefore, the company should block on exercise level because it is associated with cholesterol level. The correct answer is D. 'Block on exercise level because it is associated with cholesterol level.'

\*\*\*\*\*

Answer Option 3:

For a randomized block design, we want to control for any potential confounding variables that may affect the outcome variable. In this case, we are interested in the effect of the new cholesterol-reducing drug compared to the previous drug on cholesterol levels. Since the pharmaceutical company does not anticipate much association between cholesterol level and gender, blocking on gender would not be necessary in this case.

However, the company does anticipate an association between cholesterol level and the amount of exercise a person gets. Therefore, it would be appropriate to block on the level of exercise in order to control for this potential confounding variable. This would ensure that the comparison between the two drugs is not influenced by differences in exercise levels among the participants.

Therefore, the correct answer is: D. 'Block on exercise level because it is associated with cholesterol level.'

\*\*\*\*\*

Answer Option 4:

In a randomized block design, we want to control for variables that are associated with the response variable. In this case, we want to control for the variable that is associated with cholesterol level, which is the amount of exercise a person gets. Therefore, we should block on exercise level. The correct answer is  $\boxed{\text{D. Block on exercise level because it is associated with cholesterol level.}}$  Final

Answer: D

\*\*\*\*\*

Answer Option 5:

In a randomized block design, we want to control for any potential confounding variables that may affect the outcome. In this case, the pharmaceutical company anticipates an association between cholesterol level and the amount of exercise a person gets. Therefore, to control for this potential confounding variable, it should block on the exercise level. The correct answer is thus  $\boxed{\text{D. 'Block on exercise level because it is associated with cholesterol level.'}}$

\*\*\*\*\*

Answer Option 6:

For a randomized block design, we want to control for any potential confounding variables that may be associated with the response variable. In this case, the company anticipates an association between cholesterol level and the amount of exercise a person gets. Therefore, it would be appropriate to block on the exercise level because it is associated with cholesterol level. The correct answer is therefore  $\boxed{\text{D. 'Block on exercise level because it is associated with cholesterol level.'}}$

\*\*\*\*\*

Answer Option 7:

In a randomized block design, we want to control for a factor that we believe may have an association with the response variable (cholesterol level) in order to reduce variability and improve the precision of our estimates.

Based on the information given, the company anticipates an association between cholesterol level and the amount of exercise a person gets. Therefore, it would be appropriate to block on the exercise level because it is associated with the cholesterol level.

The correct answer is therefore: D. 'Block on exercise level because it is associated with cholesterol level.'

\*\*\*\*\*

Answer Option 8:

In a randomized block design, we want to control for potential confounding variables or sources of variation that may affect the outcome variable. In this case, the pharmaceutical company anticipates an association between cholesterol level and the amount of exercise a person gets. Therefore, the company should block on exercise level because it is associated with cholesterol level. This will allow for a

more accurate comparison between the new drug and the previous drug, as it will help control for the potential influence of exercise on cholesterol levels. Hence, the correct answer is  $\boxed{\text{D. 'Block on exercise level because it is associated with cholesterol level.'}}$

\*\*\*\*\*

Answer Option 9:

In a randomized block design, the goal is to control for potential confounding variables that may affect the outcome variable. In this case, the company anticipates an association between cholesterol level and the amount of exercise a person gets. Therefore, it would be appropriate to block on the exercise level because it is associated with cholesterol level. This would help to ensure that the two groups being compared (new drug vs. previous drug) have similar distributions of exercise levels, allowing for a more accurate assessment of the drug's effectiveness. Therefore, the correct answer is  $\boxed{\text{D. 'Block on exercise level because it is associated with cholesterol level.'}}$

\*\*\*\*\*

Question 107:

'An experiment is to be conducted to determine whether taking fish oil capsules or garlic capsules has more of an effect on cholesterol levels. In past studies it was noted that daily exercise intensity (low, moderate, high) is associated with cholesterol level, but average sleep length (< 5, 5 - 8, > 8 hours) is not associated with cholesterol level. This experiment should be done', A. 'by blocking on exercise intensity', B. 'by blocking on sleep length', C. 'by blocking on cholesterol level', D. 'by blocking on capsule type'

=====

Answer Option 1:

Based on the information given, we know that daily exercise intensity is associated with cholesterol levels, while average sleep length is not associated with cholesterol levels. Therefore, it would be most appropriate to block the experiment on exercise intensity to account for this potential confounding factor. This will allow us to compare the effects of fish oil capsules and garlic capsules on cholesterol levels within each exercise intensity group, ensuring that any observed differences are not due to differences in exercise intensity. Therefore, the experiment should be done by blocking on exercise intensity, so the answer is  $\boxed{\text{A. 'by blocking on exercise intensity'}}$ .

\*\*\*\*\*

Answer Option 2:

To determine whether taking fish oil capsules or garlic capsules has more of an effect on cholesterol levels, it would be most appropriate to block on cholesterol level. Blocking on cholesterol level means dividing the subjects into groups based on their initial cholesterol levels, and then randomly assigning them to either the fish oil or garlic capsule group within each cholesterol level group. This helps to ensure that any observed differences in cholesterol levels between the two groups can be attributed to the capsules and not to differences in initial cholesterol levels. Therefore, the experiment should be done by blocking on cholesterol level, and the correct answer is  $\boxed{\text{C. 'by blocking on cholesterol level'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 3:

To determine whether taking fish oil capsules or garlic capsules has more of an effect on cholesterol levels, it would be appropriate to block on capsule type (D). Blocking is a technique used to control potential confounding variables in an experiment. By blocking on capsule type, we can ensure that each group receives both fish oil and garlic capsules, allowing for a direct comparison of the effects of the two types of capsules on cholesterol levels. Blocking on exercise intensity or sleep length would not be relevant in this case, as these variables are not associated with cholesterol level according to past studies. Blocking on cholesterol level would not be applicable either, as the purpose of the experiment is to determine the effect of the capsules on cholesterol levels. Final answer: D

\*\*\*\*\*

Answer Option 4:

To determine whether taking fish oil capsules or garlic capsules has more of an effect on cholesterol levels, the experiment should be done by blocking on capsule type. Blocking is a technique used in experiments to reduce the variability caused by certain factors that are known or suspected to affect the outcome. In this case, since the goal is to compare the effects of fish oil and garlic capsules on cholesterol levels, it is important to control for other factors that may confound the results.

Blocking on capsule type means that participants would be divided into groups based on the type of capsule they are taking (fish oil or garlic). This ensures that each group has a balanced representation of participants with different characteristics, such as exercise intensity and sleep length. By blocking on capsule type, any potential effects of exercise intensity or sleep length on cholesterol levels would be evenly distributed across the two groups, allowing for a more accurate comparison of the effects of the different capsule types.

Therefore, the correct answer is D. 'by blocking on capsule type'. Final answer: D

\*\*\*\*\*

Answer Option 5:

The experiment should be done by blocking on capsule type, which means that the participants should be divided into groups based on whether they are taking fish oil capsules or garlic capsules. This is because the goal of the experiment is to determine whether taking fish oil or garlic capsules has more of an effect on cholesterol levels. By blocking on capsule type, we can ensure that each group has an equal number of participants taking fish oil and garlic capsules. This will help eliminate any confounding variables that may arise from individual differences in cholesterol levels. Blocking on exercise intensity or sleep length would not necessarily address the main question of interest, which is the effect of the capsules on cholesterol levels. Blocking on cholesterol level would not be appropriate since the cholesterol level is the outcome variable being measured. Therefore, the correct answer is  $\boxed{\text{D. 'by blocking on capsule type'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 6:

Blocking is a technique used in experiments to control for variables that may affect the outcome. In this case, the experiment aims to determine whether taking fish oil capsules or garlic capsules has more of an effect on cholesterol levels. It is already known that daily exercise intensity is associated with cholesterol levels, but average sleep length is not associated with cholesterol levels.

To properly control for the variable of exercise intensity, the experiment should be done 'by blocking on exercise intensity', which means that participants should be grouped based on their exercise intensity level (low, moderate, high) and then randomly assigned to either the fish oil or garlic capsule group within each exercise intensity group. This allows for a more accurate comparison between the effects of the two capsule types on cholesterol levels within each exercise intensity level.

Blocking on sleep length or cholesterol level would not be necessary in this case as sleep length is not associated with cholesterol levels, and the experiment is specifically designed to compare the effects of the two capsule types on cholesterol levels.

Therefore, the correct answer is A. 'by blocking on exercise intensity'. Final Answer: A

\*\*\*\*\*

Answer Option 7:

Blocking is a technique used in experimental design to control for variables that may confound the results. In this case, the experiment aims to determine the effect of taking fish oil capsules or garlic capsules on cholesterol levels. Given that past studies have shown that exercise intensity is associated with cholesterol level, it would be appropriate to block on exercise intensity.

Blocking on exercise intensity means that within each level of exercise intensity (low, moderate, high), participants would be randomly assigned to either the fish oil or garlic capsule group. This ensures that the effects of exercise intensity on cholesterol levels are balanced between the two groups and allows for a more accurate comparison of the effects of the capsules.

Therefore, the experiment should be done by blocking on exercise intensity, which corresponds to option A. Final Answer: A

\*\*\*\*\*

Answer Option 8:

To determine whether taking fish oil capsules or garlic capsules has more of an effect on cholesterol levels, it would be ideal to control for any potential confounding variables that may affect cholesterol levels.

In this case, the past studies have shown that daily exercise intensity is associated with cholesterol levels, but average sleep length is not associated with cholesterol levels.

To control for the potential confounding effect of exercise intensity, it would be best to block on exercise intensity. This means that participants should be divided into groups based on their exercise intensity level (low, moderate, high), and then within each group, participants should be randomly assigned to either the fish oil or garlic capsule group. This will ensure that the effect of exercise intensity on cholesterol levels is balanced between the two capsule groups, allowing for a more accurate comparison of the effects of fish oil and garlic capsules on cholesterol levels.

Therefore, the experiment should be done by blocking on exercise intensity, and the correct answer is  $\boxed{\text{A. "by blocking on exercise intensity"}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 9:

Blocking is a technique used in experimental design to control for potential confounding variables and reduce variability. In this case, the experiment aims to determine the effect of fish oil capsules versus garlic capsules on cholesterol levels. It is noted that daily exercise intensity is associated with cholesterol levels, but average sleep length is not associated.

To control for the potential confounding variable of exercise intensity, the experiment should be done by blocking on exercise intensity. This means that participants should be grouped based on their exercise intensity level (low, moderate, high), and within each group, the fish oil and garlic capsules should be randomly assigned.

This way, any differences in cholesterol levels observed between the fish oil and garlic groups can be more confidently attributed to the type of capsule, as participants within each exercise intensity group are likely to have similar cholesterol level associations due to their exercise intensity level.

Therefore, the correct answer is  $\boxed{\text{A. "by blocking on exercise intensity"}}$ . Final Answer: A

\*\*\*\*\*

Question 108:

'A test for heartworm in dogs shows a positive result in 96% of dogs that actually have heartworm, and shows a negative result in 98% of dogs with no heartworm. If heartworm actually occurs in 10% of dogs, what is the probability that a randomly selected dog that tested positive for heartworm actually has heartworm?', A. '11%', B. '18%', C. '84%', D. '88%'

=====

Answer Option 1:

Let's consider a group of 1000 dogs for simplicity. Out of these 1000 dogs, 100 have heartworm (10% of 1000). The test shows a positive result in 96% of dogs with heartworm, so it correctly identifies 96 out of these 100 heartworm-positive dogs.

Out of the remaining 900 dogs without heartworm, the test shows a negative result in 98% of them, so it correctly identifies 882 out of these 900 dogs without heartworm.

Therefore, out of the total of 978 dogs that tested positive for heartworm, 96 actually have heartworm.

The probability that a randomly selected dog that tested positive for heartworm actually has heartworm is  $\frac{96}{978} \approx 9.8\%$ .

Final Answer: None of the above

\*\*\*\*\*

Answer Option 2:

Let's say we have a group of 1000 dogs.

Out of these 1000 dogs, 10% or 100 dogs actually have heartworm.

Out of these 100 dogs with heartworm, 96% or 96 dogs will test positive for heartworm.

Out of the remaining 900 dogs without heartworm, 2% or 18 dogs will test positive for heartworm.

Therefore, out of the total dogs that tested positive for heartworm ( $96 + 18 = 114$ ), the percentage of dogs with heartworm is  $96/114 = 84\%$ .

So, the probability that a randomly selected dog that tested positive for heartworm actually has heartworm is 84%.

Therefore, the answer is  $\boxed{\text{(C) } 84\%}$ . Final Answer: C

\*\*\*\*\*

Answer Option 3:

Let's assume there are 1000 dogs in total. Out of these, 100 dogs actually have heartworm (10% of 1000) and 900 dogs do not have heartworm (90% of 1000).

Out of the 100 dogs that have heartworm, the test correctly identifies 96% (96 dogs) as positive.

Out of the 900 dogs that do not have heartworm, the test incorrectly identifies 2% (18 dogs) as positive.

Therefore, out of the total of  $96 + 18 = 114$  dogs that tested positive, only 96 dogs actually have heartworm.

The probability that a randomly selected dog that tested positive for heartworm actually has heartworm is therefore  $96/114 = 8/9 = 88\%$ .

Therefore, the answer is  $\boxed{\text{(D) } 88\%}$

\*\*\*\*\*

Answer Option 4:

Let's assume we have 1000 dogs. Out of these 1000 dogs, 100 dogs actually have heartworm and 900 dogs do not have heartworm.

Out of the 100 dogs with heartworm, the test correctly shows positive results for 96% of them, which is 96 dogs.

Out of the 900 dogs without heartworm, the test incorrectly shows positive results for 4% of them, which is 36 dogs.

Therefore, out of the total of 132 dogs that tested positive, only 96 actually have heartworm.

The probability that a randomly selected dog that tested positive for heartworm actually has heartworm is  $\frac{96}{132} = \frac{8}{11} = 72.\overline{72}\%$ .

Therefore, the answer is approximately  $\boxed{\text{(C) } 84\%}$ .

\*\*\*\*\*

Answer Option 5:

Let's assume we have 100 dogs. 10% of them, or 10 dogs, have heartworm, and 90% of them, or 90 dogs, do not have heartworm.

Of the 10 dogs with heartworm, the test correctly shows a positive result for 96% of them, which is 9.6 dogs.

Of the 90 dogs without heartworm, the test incorrectly shows a positive result for 2% of them, which is 1.8 dogs.

Therefore, out of the 9.6 dogs that tested positive for heartworm, only  $9.6 - 1.8 = 7.8$  dogs actually have heartworm.

The probability that a randomly selected dog that tested positive for heartworm actually has heartworm is  $7.8/9.6 = 0.8125$ , or 81.25%.

Therefore, the answer is approximately 81.25%, which is closest to option C, '84%'. Final Answer: C

\*\*\*\*\*

Answer Option 6:

Let's assume that we have 1000 dogs for convenience. Out of these 1000 dogs, 10% (100 dogs) actually have heartworm. The test correctly identifies 96% of these dogs as positive, so 96 dogs with heartworm will test positive. Out of the remaining 900 dogs without heartworm, 98% (882 dogs) will correctly test negative. However, there will be 18 dogs without heartworm that will test positive falsely.

Therefore, out of the total of  $96 + 18 = 114$  dogs that tested positive, only 96 of them actually have heartworm. Thus, the probability that a randomly selected dog that tested positive for heartworm actually has heartworm is  $96/114 = 4/3 = 0.88$ , or 88%.

Therefore, the answer is  $\boxed{\text{(D) } 88\%}$ .

\*\*\*\*\*

Answer Option 7:



Let's assume there are 1000 dogs in total. Out of these, 100 dogs actually have heartworm (10% of 1000). The test shows a positive result in 96% of these dogs, which means it correctly identifies 96 dogs as having heartworm.

Out of the remaining 900 dogs without heartworm, the test shows a positive result in 4% of them, which means it incorrectly identifies 36 dogs as having heartworm.

Therefore, out of the 132 (96 + 36) dogs that tested positive for heartworm, only 96 actually have heartworm.

The probability that a randomly selected dog that tested positive for heartworm actually has heartworm is  $\frac{96}{132} = \frac{8}{11}$ . This is approximately 72.73%, or as a percentage, 72.73%.

Therefore, the answer is not listed among the options.

\*\*\*\*\*

Answer Option 8:

Let's assume that we test 1000 dogs. Out of these 1000 dogs, 10% or 100 dogs have heartworm, and 90% or 900 dogs do not have heartworm.

Out of the 100 dogs with heartworm, the test correctly shows a positive result for 96% of them, which is 96 dogs.

Out of the 900 dogs without heartworm, the test incorrectly shows a positive result for 2% of them, which is 18 dogs.

Therefore, out of the 114 dogs that tested positive for heartworm, only 96 of them actually have heartworm.

The probability that a randomly selected dog that tested positive for heartworm actually has heartworm is  $\frac{96}{114} = \frac{8}{9}$  approx 88.89%.

Therefore, the answer is  $\boxed{\text{(D) '88\%'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

Let's assume we have 1000 dogs to make the calculation easier. Out of these 1000 dogs, 10% or 100 dogs actually have heartworm, and 90% or 900 dogs do not have heartworm.

The test shows a positive result in 96% of dogs that actually have heartworm, so out of the 100 dogs with heartworm, 96 of them will test positive.

The test shows a negative result in 98% of dogs with no heartworm, so out of the 900 dogs without heartworm, 882 of them will test negative.

Therefore, out of the 978 dogs that tested positive for heartworm (96 with heartworm and 882 without heartworm), only 96 of them actually have heartworm.

The probability that a randomly selected dog that tested positive for heartworm actually has heartworm is  $\frac{96}{978}$  approx 0.098 or 9.8%.

Therefore, the answer is  $\boxed{\text{(A. '11\%)}}$ . Final Answer: A

\*\*\*\*\*

Question 109:

'Suppose X and Y are random variables with  $\mu_x = 32$ ,  $\sigma_x = 5$ ,  $\mu_y = 44$ , and  $\sigma_y = 12$ . Given that X and Y are independent, what are the mean and standard deviation of the random variable  $X + Y$ ?', A. ' $\mu_{x+y} = 76$ ,  $\sigma_{x+y} = 8.5$ ', B. ' $\mu_{x+y} = 76$ ,  $\sigma_{x+y} = 13$ ', C. ' $\mu_{x+y} = 76$ ,  $\sigma_{x+y} = 17$ ', D. ' $\mu_{x+y} = 38$ ,  $\sigma_{x+y} = 17$ '

=====

Answer Option 1:

The mean of the sum of two independent random variables is equal to the sum of their individual means. Therefore, the mean of  $X + Y$  is  $\mu_x + \mu_y = 32 + 44 = 76$ .

The variance of the sum of two independent random variables is equal to the sum of their individual variances. Therefore, the variance of  $X + Y$  is  $\sigma_x^2 + \sigma_y^2 = 5^2 + 12^2 = 25 + 144 = 169$ .

The standard deviation is the square root of the variance. Therefore, the standard deviation of  $X + Y$  is  $\sqrt{169} = 13$ .

Therefore, the mean and standard deviation of the random variable  $X + Y$  are  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ . The answer is B. Final answer: B

\*\*\*\*\*

Answer Option 2:

If X and Y are independent random variables, then the mean of  $X + Y$  is the sum of the means of X and Y, and the standard deviation of  $X + Y$  is the square root of the sum of the variances of X and Y.

Therefore, the mean of  $X + Y$  is  $32 + 44 = 76$ . The variance of  $X + Y$  is  $(5^2) + (12^2) = 169$ , so the standard deviation of  $X + Y$  is  $\sqrt{169} = 13$ . Thus, the mean and standard deviation of  $X + Y$  are  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ , respectively. Therefore, the correct answer is B.

\*\*\*\*\*

Answer Option 3:

The mean of the sum of two independent random variables is simply the sum of their means. Therefore, we have  $\mu_{x+y} = \mu_x + \mu_y = 32 + 44 = 76$ .

The standard deviation of the sum of two independent random variables is the square root of the sum of their variances. Therefore, we have  $\sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2} = \sqrt{5^2 + 12^2} = \sqrt{25 + 144} = \sqrt{169} = 13$ .

Thus, the mean and standard deviation of the random variable  $X + Y$  are  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ , respectively. The correct answer is B. Final answer: B

\*\*\*\*\*

Answer Option 4:

If X and Y are independent random variables, then the mean of the sum of X and Y is the sum of the means of X and Y, and the standard deviation of the sum of X and Y is the square root of the sum of the variances of X and Y.

Therefore,  $\mu_{x+y} = \mu_x + \mu_y = 32 + 44 = 76$ .

To find the standard deviation, we need to find the variances of X and Y. Since X and Y are independent, the variance of the sum of X and Y is the sum of the variances of X and Y.

$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) = \sigma_x^2 + \sigma_y^2 = 5^2 + 12^2 = 25 + 144 = 169$ .

Therefore,  $\sigma_{x+y} = \sqrt{\text{Var}(X+Y)} = \sqrt{169} = 13$ .

The mean and standard deviation of the random variable  $X + Y$  are therefore  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ .

The correct answer is B. Final answer: B.

\*\*\*\*\*

Answer Option 5:

If X and Y are independent random variables, then the mean of  $X + Y$  is the sum of the means of X and Y, and the variance of  $X + Y$  is the sum of the variances of X and Y. Therefore, the mean of  $X + Y$  is  $\mu_x + \mu_y = 32 + 44 = 76$ .

To find the standard deviation of  $X + Y$ , we need to find the variance first. The variance of  $X + Y$  is the sum of the variances of X and Y, since X and Y are independent. Therefore, the variance of  $X + Y$  is  $\sigma_x^2 + \sigma_y^2 = 5^2 + 12^2 = 25 + 144 = 169$ .

The standard deviation of  $X + Y$  is the square root of the variance, so the standard deviation of  $X + Y$  is  $\sqrt{169} = 13$ .

Therefore, the mean and standard deviation of the random variable  $X + Y$  are  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ , respectively. Thus, the correct answer is  $\boxed{\text{B}}$ . Final

Answer: B

\*\*\*\*\*

Answer Option 6:

The mean of the sum of two independent random variables is the sum of their individual means. Therefore, the mean of  $X + Y$  is  $\mu_x + \mu_y = 32 + 44 = 76$ .

The variance of the sum of two independent random variables is the sum of their individual variances. Therefore, the variance of  $X + Y$  is  $\sigma_x^2 + \sigma_y^2 = 5^2 + 12^2 = 25 + 144 = 169$ .

The standard deviation is the square root of the variance. Therefore, the standard deviation of  $X + Y$  is  $\sqrt{169} = 13$ .

Thus, the mean and standard deviation of  $X + Y$  are  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ . Therefore, the correct answer is  $\boxed{\text{B}}$ .

\*\*\*\*\*

Answer Option 7:

If X and Y are independent random variables, then the mean of  $X + Y$  is the sum of their individual means, and the variance of  $X + Y$  is the sum of their individual variances. The standard deviation is the square root of the variance.

The mean of  $X + Y$  is  $\mu_x + \mu_y = 32 + 44 = 76$ .

The variance of  $X + Y$  is  $\sigma_x^2 + \sigma_y^2 = 5^2 + 12^2 = 25 + 144 = 169$ .

The standard deviation of  $X + Y$  is  $\sqrt{169} = 13$ .

Therefore, the mean and standard deviation of the random variable  $X + Y$  are  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ .

The correct answer is B.  $\mu_{x+y} = 76$ ,  $\sigma_{x+y} = 13$ . Final answer: B.

\*\*\*\*\*

Answer Option 8:

The mean of the sum of two independent random variables is simply the sum of their individual means. Therefore, the mean of  $X + Y$  is  $\mu_x + \mu_y = 32 + 44 = 76$ .

The variance of the sum of two independent random variables is the sum of their individual variances. Therefore, the variance of  $X + Y$  is  $\sigma_x^2 + \sigma_y^2 = 5^2 + 12^2 = 25 + 144 = 169$ .

The standard deviation is the square root of the variance, so the standard deviation of  $X + Y$  is  $\sqrt{169} = 13$ .

Therefore, the mean and standard deviation of  $X + Y$  are  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ , respectively. The correct answer is B.

\*\*\*\*\*

Answer Option 9:

If  $X$  and  $Y$  are independent random variables, then the mean of  $X + Y$  is equal to the sum of their individual means:  $\mu_x + \mu_y = 32 + 44 = 76$ .

To find the standard deviation of  $X + Y$ , we use the property that the variance of the sum of independent random variables is equal to the sum of their individual variances. The variance of  $X + Y$  is equal to  $\sigma_x^2 + \sigma_y^2 = 5^2 + 12^2 = 25 + 144 = 169$ . Taking the square root of the variance gives us the standard deviation:  $\sigma_{x+y} = \sqrt{169} = 13$ .

Therefore, the mean and standard deviation of the random variable  $X + Y$  are  $\mu_{x+y} = 76$  and  $\sigma_{x+y} = 13$ . The correct answer is B. Final Answer: B

\*\*\*\*\*

Question 110:

'A high school sent a survey to a randomly selected sample of 75 of last year's graduating class. 27 of those selected did not return the survey. The best plan of action would be to:', A. 'Use the surveys that were returned and change the sample size to 48 for the analysis.', B. 'Use the surveys that were returned and leave the sample size at 75 for the analysis.', C. 'Randomly select 27 additional class members and send the survey to them.', D. 'Follow up with those that did not return the survey to encourage them to respond.'

=====

Answer Option 1:

The best plan of action would be to minimize non-response bias by following up with those who did not return the survey to encourage them to respond. This will help ensure a higher response rate and a more representative sample of the graduating class. Therefore, the correct answer is  $\boxed{\text{D. 'Follow up with those that did not return the survey to encourage them to respond.'}}$

\*\*\*\*\*

Answer Option 2:

The best plan of action would be to follow up with those who did not return the survey to encourage them to respond. This is because the sample that did not return the survey introduces non-response bias, which can affect the representativeness of the sample. By encouraging non-respondents to complete the survey, we increase the likelihood of obtaining a more accurate representation of the entire graduating class. Therefore, the correct answer is  $\boxed{\text{D. 'Follow up with those that did not return the survey to encourage them to respond.'}}$

\*\*\*\*\*

Answer Option 3:

To ensure the best representation of the entire graduating class, it is important to minimize non-response bias. In this case, 27 out of the selected sample of 75 did not return the survey. Therefore, the best plan of action would be to follow up with those who did not return the survey to encourage them to respond. This will potentially increase the response rate and reduce non-response bias. Therefore, the correct answer is  $\boxed{\text{D. 'Follow up with those that did not return the survey to encourage them to respond.'}}$

\*\*\*\*\*

Answer Option 4:

The best plan of action would be to follow up with those that did not return the survey to encourage them to respond. This is because the non-response rate is relatively high at 27/75, which may introduce bias into the analysis if only the returned surveys are used. By following up with those that did not respond, the school can potentially increase the response rate and obtain a more representative sample for analysis. Therefore, the correct answer is  $\boxed{\text{D. Follow up with those that did not return the survey to encourage them to respond.'}}$  Final Answer: D

\*\*\*\*\*

Answer Option 5:

In order to ensure the validity and representativeness of the survey results, it is important to have a high response rate. Since 27 out of 75 selected did not return the survey, the response rate is  $\frac{75-27}{75} = \frac{48}{75} = \frac{16}{25}$  or 64%. This response rate is relatively low, and therefore, it is best to encourage those who did not return the survey to respond.

The correct choice is  $\boxed{\text{D. 'Follow up with those that did not return the survey to encourage them to respond.'}}$  Final Answer: D

\*\*\*\*\*

Answer Option 6:

The best plan of action would be to follow up with those who did not return the survey to encourage them to respond. This will help increase the response rate and provide a more representative sample of the graduating class. Therefore, the correct answer is  $\boxed{\text{D. 'Follow up with those that did not return the survey to encourage them to respond.'}}$

\*\*\*\*\*

Answer Option 7:

The best plan of action would be to follow up with those who did not return the survey to encourage them to respond. This is because the survey was sent to a randomly selected sample, and the fact that 27 did not return it introduces non-response bias. By following up and encouraging them to respond, we can minimize this bias and improve the representativeness of the sample. Therefore, the answer is  $\boxed{\text{D. 'Follow up with those that did not return the survey to encourage them to respond.'}}$

\*\*\*\*\*

Answer Option 8:

The best plan of action in this situation would be to follow up with those who did not return the survey to encourage them to respond. This is because the response rate was low (only 48 out of 75 surveys were returned), and it is important to have a representative sample for accurate analysis. By following up with the non-respondents, we can increase the response rate and improve the representativeness of the sample. Therefore, the correct answer is  $\boxed{\text{D. 'Follow up with those that did not return the survey to encourage them to respond.'}}$  Final answer: D

\*\*\*\*\*

Answer Option 9:

The best plan of action would be to follow up with those who did not return the survey to encourage them to respond. This is because the response rate of the survey is only  $\frac{48}{75} = \frac{64}{100} = 64\%$ . It is important to have a high response rate in order to obtain accurate and representative data. By following up with those who did not respond, there is a higher chance of obtaining a more representative sample. Therefore, the answer is  $\boxed{\text{D. 'Follow up with those that did not return the survey to encourage them to respond.'}}$ .

\*\*\*\*\*

Question 111:

'Mr. Bee's statistics class had a standard deviation of 11.2 on a standardized test, while Mr. Em's class had a standard deviation of 5.6 on the same test. Which of the following is the most reasonable conclusion concerning the two classes' performance on the test?', A. 'Mr. Bee's class is less heterogeneous than Mr. Em's.', B. 'Mr. Em's class is more homogeneous than Mr. Bee's.', C. 'Mr. Bee's class performed twice as well as Mr. Em's.', D. 'Mr. Em's class did not do as well as Mr. Bee's.'

=====

Answer Option 1:

The standard deviation measures the dispersion or spread of data. A larger standard deviation indicates greater variability in the data, while a smaller standard deviation indicates less variability.

In this case, Mr. Bee's class has a standard deviation of 11.2, which is greater than Mr. Em's class with a standard deviation of 5.6. Therefore, Mr. Bee's class is more heterogeneous, meaning the performance of students in Mr. Bee's class is more varied compared to Mr. Em's class.

Therefore, the most reasonable conclusion is: A. 'Mr. Bee's class is less homogeneous than Mr. Em's.'

\*\*\*\*\*

Answer Option 2:

The standard deviation measures the spread or variability of data. A higher standard deviation indicates a wider spread of scores, while a lower standard deviation indicates a narrower spread of scores. In this case, Mr. Bee's class had a standard deviation of 11.2, which is twice as large as Mr. Em's class's standard deviation of 5.6. Therefore, we can conclude that Mr. Bee's class is more heterogeneous, or has a wider spread of scores, compared to Mr. Em's class. Thus, the most reasonable conclusion is that option A: 'Mr. Bee's class is less heterogeneous than Mr. Em's.' Final Answer: A

\*\*\*\*\*

Answer Option 3:

The standard deviation measures the spread or variability of a data set. A larger standard deviation indicates a greater spread and therefore a more heterogeneous group. Since Mr. Bee's class has a larger standard deviation of 11.2 compared to Mr. Em's class with a standard deviation of 5.6, the most reasonable conclusion is that Mr. Bee's class is less homogeneous than Mr. Em's class. Thus, the correct answer is  $\boxed{\text{A. 'Mr. Bee's class is less heterogeneous than Mr. Em's.'}}$

\*\*\*\*\*

Answer Option 4:

The standard deviation measures the spread or variability of a data set. A larger standard deviation indicates a greater spread of data, while a smaller standard deviation indicates less spread.

In this case, Mr. Bee's class has a standard deviation of 11.2, which is larger than Mr. Em's class with a standard deviation of 5.6. Therefore, we can reasonably conclude that Mr. Bee's class is more heterogeneous, or less homogeneous, than Mr. Em's class.

Thus, the most reasonable conclusion is: A. 'Mr. Bee's class is less heterogeneous than Mr. Em's.' Final answer: A

\*\*\*\*\*

Answer Option 5:

The standard deviation measures the spread or variability of data. A larger standard deviation indicates a greater spread of data points, while a smaller standard deviation indicates a lesser spread of data points.

Given that Mr. Bee's class had a standard deviation of 11.2 and Mr. Em's class had a standard deviation of 5.6, we can conclude that Mr. Bee's class is more heterogeneous than Mr. Em's class. This means that the performance of students in Mr. Bee's class varied more widely compared to the performance of students in Mr. Em's class.

Thus, the most reasonable conclusion is that option A is correct: 'Mr. Bee's class is less heterogeneous than Mr. Em's.' Final Answer: A

\*\*\*\*\*

Answer Option 6:

Standard deviation measures the spread or dispersion of data. A higher standard deviation indicates more variability in the data, while a lower standard deviation indicates less variability.

In this case, Mr. Bee's class has a standard deviation of 11.2, which is higher than Mr. Em's class with a standard deviation of 5.6. Therefore, it is reasonable to conclude that Mr. Bee's class is more heterogeneous or has more variability in their performance on the standardized test compared to Mr. Em's class.

Hence, the most reasonable conclusion is that 'Mr. Bee's class is less homogeneous than Mr. Em's.' This corresponds to answer choice A. Final Answer: A

\*\*\*\*\*

Answer Option 7:

The standard deviation measures the spread or dispersion of data. A larger standard deviation indicates greater variability, while a smaller standard deviation indicates less variability.

Here, Mr. Bee's class has a standard deviation of 11.2, which is larger than Mr. Em's class's standard deviation of 5.6. This means that Mr. Bee's class has a greater spread in their performance on the test compared to Mr. Em's class.

Therefore, the most reasonable conclusion is that Mr. Bee's class is more heterogeneous (has a greater range of performance) than Mr. Em's class.

The correct answer is  $\boxed{\text{A. 'Mr. Bee's class is less heterogeneous than Mr. Em's.'}}$  Final Answer: A

\*\*\*\*\*

Answer Option 8:

The standard deviation measures the dispersion or spread of data. A larger standard deviation indicates greater variability in the data, while a smaller standard deviation indicates less variability.

Here, Mr. Bee's class has a larger standard deviation of 11.2, while Mr. Em's class has a smaller standard deviation of 5.6. Therefore, we can conclude that Mr. Bee's class is more heterogeneous (has greater variability) than Mr. Em's class. Thus, the most reasonable conclusion is that 'Mr. Bee's class is less homogeneous than Mr. Em's.' This corresponds to option  $\boxed{\text{A}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 9:

The standard deviation measures the spread or variability of the data. A higher standard deviation indicates a greater spread of scores, while a lower standard deviation indicates a smaller spread of scores. In this case, Mr. Bee's class has a higher standard deviation of 11.2, while Mr. Em's class has a lower standard deviation of 5.6. This means that Mr. Bee's class is more spread out in terms of their scores, indicating a greater variability in performance. On the other hand, Mr. Em's class is more tightly clustered around the mean, indicating a smaller variability in performance. Therefore, the most reasonable conclusion is that Mr. Bee's class is less homogeneous than Mr. Em's. Thus, the correct answer is  $\boxed{\text{A. 'Mr. Bee's class is less heterogeneous than Mr. Em's.'}}$

\*\*\*\*\*

Question 112:

'When a patient complains to the doctor about a certain set of symptoms, the doctor diagnoses the patient with Condition A 15% of the time. If a patient with these symptoms is diagnosed with Condition A, he or she is diagnosed with Condition B 70% of the time. A patient with these symptoms that is not diagnosed with Condition A is diagnosed with Condition B 10% of the time. What is the probability that a patient with this set of symptoms will be diagnosed with at least one of these conditions?', A. '0.235', B. '0.25', C. '0.765', D. '0.85'

=====

Answer Option 1:

Let  $A$  be the event that the patient is diagnosed with Condition A, and  $B$  be the event that the patient is diagnosed with Condition B. We want to find  $P(A \cup B)$ , the probability that the patient is diagnosed with at least one of the conditions.

We can use the law of total probability to calculate this probability. We have:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B|A)P(A) + P(B|A^c)P(A^c) \\ &= 0.15 + 0.70(0.15) + 0.10(0.85) \\ &= 0.15 + 0.105 + 0.085 \\ &= 0.34 \end{aligned}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= P(A) + P(B|A)P(A) + P(B|A^c)P(A^c)$$

$$= 0.15 + 0.70(0.15) + 0.10(0.85)$$

$$= 0.15 + 0.105 + 0.085$$

$$= 0.34$$

$$\end{aligned}$$

Therefore, the probability that a patient with these symptoms will be diagnosed with at least one of these conditions is  $\boxed{0.34}$ . Final answer: None of the above.

\*\*\*\*\*

Answer Option 2:

Let's break down the problem into two cases: being diagnosed with Condition A and being diagnosed without Condition A.

Case 1: Being diagnosed with Condition A

The probability of being diagnosed with Condition A is 15%. In this case, the patient is diagnosed with Condition B 70% of the time. Therefore, the probability of being diagnosed with both Condition A and Condition B is  $15\% \times 70\% = 10.5\%$ .

Case 2: Being diagnosed without Condition A

The probability of not being diagnosed with Condition A is  $100\% - 15\% = 85\%$ . In this case, the patient is diagnosed with Condition B 10% of the time. Therefore, the probability of being diagnosed with only Condition B is  $85\% \times 10\% = 8.5\%$ .

To find the probability of being diagnosed with at least one of these conditions, we add the probabilities from both cases:

$$10.5\% + 8.5\% = 19\%$$

Therefore, the probability that a patient with these symptoms will be diagnosed with at least one of these conditions is 19%, which is equivalent to 0.19. None of the given answer choices match this result.

\*\*\*\*\*

Answer Option 3:

Let's denote the event that a patient is diagnosed with Condition A as  $A$ , and the event that a patient is diagnosed with Condition B as  $B$ . We want to find the probability of  $P(A \cup B)$ , which is the probability that a patient is diagnosed with at least one of these conditions.

Using the law of total probability, we can write:

$$P(A \cup B) = P(A) + P(B \mid A^c) \cdot P(A^c)$$

We are given that  $P(A) = 0.15$  and  $P(B \mid A^c) = 0.10$ .

To find  $P(A^c)$ , we can use the complement rule:

$$P(A^c) = 1 - P(A) = 1 - 0.15 = 0.85$$

Substituting the values we have into the equation, we get:

$$P(A \cup B) = 0.15 + 0.10 \cdot 0.85 = 0.15 + 0.085 = \boxed{0.235}$$

Therefore, the probability that a patient with these symptoms will be diagnosed with at least one of these conditions is 0.235. Final Answer: A

\*\*\*\*\*

Answer Option 4:

Let's calculate the probability that a patient with these symptoms is diagnosed with both Condition A and Condition B. This is given by the product of the probability of being diagnosed with Condition A and the probability of being diagnosed with Condition B given that the patient has Condition A. This is  $0.15 \times 0.70 = 0.105$ .

Now, let's calculate the probability that a patient with these symptoms is diagnosed with neither Condition A nor Condition B. This is given by the complement of the probability of being diagnosed with at least one of the conditions. Therefore, the probability of being diagnosed with at least one of the conditions is  $1 - \text{(probability of neither)}$ .

To find the probability of neither, we need to calculate the probability of not being diagnosed with Condition A and the probability of not being diagnosed with Condition B given that the patient does not have Condition A. These are  $0.85$  and  $0.90$  respectively. Therefore, the probability of neither is  $0.85 \times 0.90 = 0.765$ .

Finally, the probability of being diagnosed with at least one of the conditions is  $1 - 0.765 = 0.235$ . Therefore, the answer is  $\boxed{\text{(A) } 0.235}$ . Final Answer: A

\*\*\*\*\*

Answer Option 5:

Let  $A$  represent the event that a patient is diagnosed with Condition A, and let  $B$  represent the event that a patient is diagnosed with Condition B. We want to find  $P(A \cup B)$ , the probability that a patient is diagnosed with at least one of these conditions.

Using the law of total probability, we can break down this probability into two cases: when a patient is diagnosed with Condition A and when a patient is not diagnosed with Condition A.

Case 1: Patient is diagnosed with Condition A

The probability of this case is  $P(A) = 0.15$ . In this case, the patient is also diagnosed with Condition B with a probability of  $P(B|A) = 0.70$ .

Case 2: Patient is not diagnosed with Condition A

The probability of this case is  $P(\neg A) = 1 - P(A) = 0.85$ . In this case, the patient is diagnosed with Condition B with a probability of  $P(B|\neg A) = 0.10$ .

Using the law of total probability, we can combine these two cases:

$$P(A \cup B) = P(A)P(B|A) + P(\neg A)P(B|\neg A)$$

$$= (0.15)(0.70) + (0.85)(0.10)$$

$$= 0.105 + 0.085$$

$$= 0.19$$

Therefore, the probability that a patient with these symptoms will be diagnosed with at least one of these conditions is  $\boxed{0.19}$ . Final Answer: None of the above.

\*\*\*\*\*

Answer Option 6:

Let  $A$  be the event that a patient is diagnosed with Condition A, and  $B$  be the event that a patient is diagnosed with Condition B. We are trying to find  $P(A \cup B)$ , the probability that a patient is diagnosed with at least one of these conditions.

We can use the Law of Total Probability to calculate this. We have:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Given in the problem,  $P(A) = 0.15$ .

To calculate  $P(B)$ , we need to consider two cases:

- 1) The patient is diagnosed with Condition A and then with Condition B.
- 2) The patient is not diagnosed with Condition A and then diagnosed with Condition B.

For the first case, we have  $P(A \cap B) = P(A) \cdot P(B|A) = 0.15 \cdot 0.70 = 0.105$ .

For the second case, we have  $P(B) = P(\neg A) \cdot P(B|\neg A) = (1 - P(A)) \cdot P(B|\neg A) = 0.85 \cdot 0.10 = 0.085$ .

Substituting these values into the equation, we get:

$$P(A \cup B) = 0.15 + 0.085 - 0.105 = 0.13$$

Therefore, the probability that a patient with these symptoms will be diagnosed with at least one of these conditions is  $\boxed{0.13}$ . Final Answer: A

\*\*\*\*\*

Answer Option 7:

Let  $A$  be the event of being diagnosed with Condition A, and  $B$  be the event of being diagnosed with Condition B. We are looking for  $P(A \cup B)$ , the probability of being diagnosed with at least one of the conditions.

Using the law of total probability, we can express  $P(A \cup B)$  as:

$$P(A \cup B) = P(A \cap B^c) + P(B \cap A^c) + P(A \cap B)$$

where  $B^c$  denotes the complement of  $B$ , and  $A^c$  denotes the complement of  $A$ .

From the problem, we know that:

$$P(A) = 0.15$$

$$P(B|A) = 0.70$$

$$P(B|A^c) = 0.10$$

We can rewrite the equation as:

$$P(A \cup B) = P(A) + P(B|A^c) \cdot P(A^c) + P(A \cap B)$$

To find  $P(A \cap B)$ , we can use the fact that  $P(A \cap B) = P(A) \cdot P(B|A) = 0.15 \cdot 0.70 = 0.105$ .

$P(A^c)$  can be calculated as  $1 - P(A) = 1 - 0.15 = 0.85$ .

Substituting these values, we have:

$$P(A \cup B) = 0.15 + 0.10 \cdot 0.85 + 0.105 = 0.235$$

Therefore, the probability that a patient with these symptoms will be diagnosed with at least one of these conditions is  $\boxed{0.235}$ . Final answer: A

\*\*\*\*\*

Answer Option 8:

Let  $A$  be the event that the patient is diagnosed with Condition A, and let  $B$  be the event that the patient is diagnosed with Condition B. We want to find the probability  $P(A \cup B)$ .

Using the law of total probability, we can write:

$$P(A \cup B) = P(A) + P(B|A) \cdot P(A^c) + P(B|A^c) \cdot P(A^c)$$

where  $A^c$  denotes the complement of event  $A$ .

We are given that  $P(A) = 0.15$ ,  $P(B|A) = 0.70$ ,  $P(B|A^c) = 0.10$ , and  $P(A^c) = 1 - P(A) = 0.85$ .



Substituting these values into the equation, we have:

$$P(A \cup B) = 0.15 + 0.70 \cdot 0.85 + 0.10 \cdot 0.85 = 0.15 + 0.595 + 0.085 = 0.83$$

Therefore, the probability that a patient with these symptoms will be diagnosed with at least one of these conditions is  $\boxed{0.83}$ . Final answer: None of the above.

\*\*\*\*\*

Answer Option 9:

Let  $P(A)$  be the probability of being diagnosed with Condition A and  $P(B)$  be the probability of being diagnosed with Condition B. We want to find  $P(A \cup B)$ , the probability of being diagnosed with at least one of these conditions.

We know that  $P(A) = 0.15$ .

If a patient is diagnosed with Condition A, the probability of also being diagnosed with Condition B is  $0.70$ . So, the probability of being diagnosed with both Condition A and Condition B is  $P(A \cap B) = P(A) \cdot P(B|A) = 0.15 \cdot 0.70 = 0.105$ .

If a patient is not diagnosed with Condition A, the probability of being diagnosed with Condition B is  $0.10$ . So, the probability of being diagnosed with only Condition B is  $P(B) - P(A \cap B) = P(B) - 0.105$ .

Therefore, the probability of being diagnosed with at least one of these conditions is:

$$P(A \cup B) = P(A) + (P(B) - P(A \cap B)) = 0.15 + (P(B) - 0.105).$$

Since we don't have information about  $P(B)$ , we cannot determine the exact value of  $P(A \cup B)$ .

Thus, the correct answer is: None of the given choices (no answer choice matches the possible values of  $P(A \cup B)$ ).

\*\*\*\*\*

Question 113:

'A simple random sample is defined by', A. 'the method of selection.', B. 'examination of the outcome.', C. 'both of the above.', D. 'how representative the sample is of the population.'

=====

Answer Option 1:

A simple random sample is defined by the method of selection. Therefore, the correct answer is A. Final Answer: A

\*\*\*\*\*

Answer Option 2:

A simple random sample is defined by the method of selection. Therefore, the correct answer is A. 'the method of selection.' Final answer: A

\*\*\*\*\*

Answer Option 3:

A simple random sample is defined by the method of selection. This means that each member of the population has an equal chance of being selected to be a part of the sample. Therefore, the correct answer is A. 'the method of selection.' Final Answer: A

\*\*\*\*\*

Answer Option 4:

A simple random sample is defined by the method of selection. The method of selection for a simple random sample involves randomly selecting individuals from a population in such a way that each individual has an equal chance of being selected. Therefore, the answer is  $\boxed{\text{A. 'the method of selection.'}}$ .

\*\*\*\*\*

Answer Option 5:

A simple random sample is defined by the method of selection. Therefore, the answer is  $\boxed{\text{A. 'the method of selection.'}}$

\*\*\*\*\*

Answer Option 6:

A simple random sample is defined by the method of selection, which means that each member of the population has an equal chance of being chosen. Therefore, the correct answer is A. 'the method of selection.' Final answer: A

\*\*\*\*\*

Answer Option 7:

A simple random sample is defined by the method of selection. This means that each member of the population has an equal probability of being selected for the sample. Therefore, the correct answer is A. 'the method of selection.' Final Answer: A

\*\*\*\*\*

Answer Option 8:

A simple random sample is defined by the method of selection. This means that each member of the population has an equal chance of being selected to be part of the sample. Therefore, the correct answer is A. 'the method of selection.' Final Answer: A

\*\*\*\*\*

Answer Option 9:

A simple random sample is defined by 'the method of selection.' The method of selection is the process used to choose individuals from the population to be included in the sample. Therefore, the correct answer is  $\boxed{\text{A. 'the method of selection.'}}$

\*\*\*\*\*

Question 114:

'In one study on the effect that eating meat products has on weight level, an SRS of 500 subjects who admitted to eating meat at least once a day had their weights compared with those of an independent SRS of 500 people who claimed to be vegetarians. In a second study, an SRS of 500 subjects were served at least one meat meal per day for 6 months, while an independent SRS of 500 others were chosen to receive a strictly vegetarian diet for 6 months, with weights compared after 6 months.', A. 'The first study is a controlled experiment, while the second is an observational study.', B. 'The first study is an observational study, while the second is a controlled experiment.', C. 'Both studies are controlled experiments.', D. 'Both studies are observational studies.'

=====

Answer Option 1:

The first study compares the weights of two independent groups of people - those who eat meat at least once a day and those who claim to be vegetarians. These groups were not randomly assigned and the researchers did not control the diets of the participants. Therefore, the first study is an observational study.

The second study randomly assigns participants to either a meat-based diet or a vegetarian diet and compares their weights after 6 months. The researchers have control over the diets of the participants and can manipulate the independent variable (type of diet). Therefore, the second study is a controlled experiment.

Therefore, the correct answer is  $\boxed{\text{B. 'The first study is an observational study, while the second is a controlled experiment.'}}$

\*\*\*\*\*

Answer Option 2:

A controlled experiment is one in which the researchers actively manipulate the variables of interest. In the first study, the researchers did not assign the subjects to their diet groups; instead, they observed and compared the weights of two pre-existing groups (those who eat meat and those who are vegetarian). Therefore, the first study is an observational study, and the correct answer is B.

In the second study, the researchers actively assigned subjects to either a meat or vegetarian diet for 6 months. They then compared the weights of the two groups after the 6-month period. This study involves active manipulation of the diet variable and comparison of outcomes, making it a controlled experiment. However, since the question asks about each individual study, the correct answer is still B.

Final Answer: B

\*\*\*\*\*

Answer Option 3:

A controlled experiment is one in which the researcher actively manipulates the independent variable and controls other factors to determine their effect on the dependent variable. An observational study, on the other hand, is one in which the researcher observes and measures variables without intervening or manipulating them.

Based on the information given, the first study compared the weights of two groups - those who eat meat and those who are vegetarians. The researchers did not actively manipulate the diet of the participants; they simply observed and compared the weights of the two groups. Therefore, the first study is an observational study.

In contrast, the second study actively assigned participants to either a meat diet or a vegetarian diet for 6 months. The researchers manipulated the participants' diet and then compared their weights after the 6-month period. Therefore, the second study is a controlled experiment.

Hence, the correct answer is  $\boxed{\text{B. 'The first study is an observational study, while the second is a controlled experiment.'}}$

\*\*\*\*\*

Answer Option 4:

A controlled experiment is one in which the researcher actively manipulates the independent variable and randomly assigns participants to different treatment conditions. In the first study, the researcher did not assign participants to different diets or manipulate their eating habits. Instead, they observed and compared the weights of two independent groups of people (meat-eaters and vegetarians). Therefore, the first study is an observational study.

In the second study, the researcher actively manipulated the diet of the participants by assigning them to either a meat-based or a vegetarian diet. The researcher also compared the weights of the two groups after 6 months. This study involved intervention and control groups and random assignment, demonstrating characteristics of a controlled experiment.

Therefore, the first study is an observational study and the second study is a controlled experiment. The correct answer is  $\boxed{\text{B. 'The first study is an observational study, while the second is a controlled experiment.'}}$

\*\*\*\*\*

Answer Option 5:

A controlled experiment is one in which the researcher actively manipulates the variables being studied, while an observational study is one in which the researcher does not actively manipulate the variables and simply observes the subjects.

In the first study, the researcher compared the weight levels of two independent groups - those who eat meat and those who are vegetarians. The researcher did not actively manipulate the subjects' diets but rather observed the existing differences between the two groups. Therefore, the first study is an observational study.

In the second study, the researcher actively manipulated the subjects' diets by assigning them to either a meat or vegetarian diet for 6 months. The researcher then compared the weight levels of the two groups after 6 months. This active manipulation of the variables makes the second study a controlled experiment.

Therefore, the correct answer is  $\boxed{\text{B. 'The first study is an observational study, while the second is a controlled experiment.'}}$

\*\*\*\*\*

Answer Option 6:

The first study, where the weights of subjects who eat meat at least once a day are compared with those who claim to be vegetarians, is an observational study. The researchers are not manipulating the diets of the subjects, but instead are observing and comparing the weights of the two groups.

The second study, where one group is served at least one meat meal per day and the other group receives a strictly vegetarian diet, is a controlled experiment. The researchers are manipulating the diets of the subjects by assigning them to different groups and then comparing the weights after 6 months.

Therefore, the correct answer is  $\boxed{\text{B. 'The first study is an observational study, while the second is a controlled experiment.'}}$

\*\*\*\*\*

Answer Option 7:

In the first study, the researchers compare the weights of two groups of people - those who eat meat daily and those who are vegetarians. The researchers did not make any intervention or control the groups' diets. Therefore, the first study is an observational study.

In the second study, the researchers assign one group to eat meat daily for 6 months and another group to follow a strictly vegetarian diet for the same period. The researchers control the diet of each group for a specific time period. Therefore, the second study is a controlled experiment.

Therefore, the correct answer is  $\boxed{\text{B. "The first study is an observational study, while the second is a controlled experiment."}}$

\*\*\*\*\*

Answer Option 8:

The first study compares the weights of two independent groups - those who eat meat and those who are vegetarians. The researchers did not have control over who was in which group, as participants self-reported their meat-eating habits. Therefore, the first study is an observational study.

The second study randomly assigns participants to either the meat-eating group or the vegetarian group and compares their weights after 6 months. The researchers have control over the intervention (meat or vegetarian diet) and can measure the effect on weight. Therefore, the second study is a controlled experiment.

Based on this analysis, we can conclude that the correct answer is  $\boxed{\text{B. "The first study is an observational study, while the second is a controlled experiment."}}$  Final answer: B

\*\*\*\*\*

Answer Option 9:

In a controlled experiment, the researcher actively manipulates the independent variable (in this case, the type of diet) and randomly assigns participants to different groups. The researcher then measures the effect of the independent variable on the dependent variable (in this case, weight).

In the first study, the researchers compared the weights of two independent groups: those who ate meat at least once a day and those who claimed to be vegetarians. The researchers did not actively manipulate the type of diet or randomly assign participants to the groups. Therefore, the first study is an observational study.

In the second study, the researchers actively manipulated the diet by serving one group at least one meat meal per day and the other group a strictly vegetarian diet. The researchers then measured the effect of the diet on weight after 6 months. This study involved active manipulation of the independent variable and randomly assigned participants to different groups. Therefore, the second study is a controlled experiment.

Based on these characteristics, the correct answer is  $\boxed{\text{B. "The first study is an observational study, while the second is a controlled experiment."}}$

\*\*\*\*\*

Question 115:

'Which of these is a correct description of the term?', A. 'A factor is a response variable.', B. 'Replication means the experiment should be repeated several times.', C. 'Levels are the same as treatments.', D. 'Experimental units are the same as subjects.'

=====

Answer Option 1:

Among the given options, the correct description is option B: 'Replication means the experiment should be repeated several times.' Replication refers to the repetition of an experiment or study to increase the reliability and validity of the results. This involves conducting multiple trials or replicates under the same conditions to account for variability and assess the consistency of the findings. Final Answer: B

\*\*\*\*\*

Answer Option 2:

The correct description of the term is D. 'Experimental units are the same as subjects.' Experimental units refer to the objects or individuals on which the experiment is conducted. Subjects, on the other hand, are the individuals or participants in a study. Since experimental units can refer to objects or individuals, it is a more general term than subjects. Thus, the correct description is D. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The correct description of the term is B. 'Replication means the experiment should be repeated several times.' Replication refers to the process of conducting the experiment multiple times with different experimental units in order to account for variability and increase the reliability of the results. Final Answer: B

\*\*\*\*\*

Answer Option 4:

The correct description of the term is B. 'Replication means the experiment should be repeated several times.' In experimental design, replication refers to the process of conducting the experiment multiple times with different subjects or units in order to obtain more reliable and statistically valid results. This helps to decrease the impact of random variation and provides an opportunity to assess the consistency and reproducibility of the experiment. Final Answer: B

\*\*\*\*\*

Answer Option 5:

The correct description of the term is option D. 'Experimental units are the same as subjects.' Experimental units refer to the individuals or objects on which the experiment is conducted. They can be people, animals, plants, or any other entities that the experiment is being performed on. Subjects, on the other hand, typically refer to human individuals involved in the study. Therefore, experimental units and subjects are often used interchangeably in the context of human-based experiments. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The correct description of the term is D. 'Experimental units are the same as subjects.' Experimental units refer to the entities or objects on which the experiment is conducted or from which data is collected. Subjects typically refer to individuals or participants in a study. While experimental units can be subjects in certain experiments, they can also refer to other entities such as animals, plants, or inanimate objects. Therefore, experimental units and subjects are not always the same. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct description of the term is D. 'Experimental units are the same as subjects.' In an experiment, the experimental units refer to the individuals or objects on which the treatment or intervention is applied. These units can be people (referred to as subjects) or other entities such as animals, plants, or inanimate objects. Therefore, experimental units and subjects are synonymous in the context of an experiment. Final Answer: D

\*\*\*\*\*

Answer Option 8:

None of the options A, B, C, and D accurately describe the term.

A. 'A factor is a response variable.' - This is incorrect. A factor is an independent variable or a variable that is deliberately manipulated or controlled in an experiment.

B. 'Replication means the experiment should be repeated several times.' - This is partially correct. Replication does involve repeating the experiment multiple times to obtain more reliable and accurate results. However, it is not the only reason for replication. Replication also helps to assess the variability of the results and to establish statistical significance.

C. 'Levels are the same as treatments.' - This is incorrect. Levels refer to different values or settings of a factor or independent variable. Treatments, on the other hand, are the specific combinations of factor levels that are assigned to the experimental units.

D. 'Experimental units are the same as subjects.' - This is incorrect. Experimental units are the objects or individuals on which the experiment is conducted. Subjects, on the other hand, are a specific type of experimental units that are typically used in human studies.

Therefore, none of the options correctly describe the term.

\*\*\*\*\*

Answer Option 9:

The correct description of the term is D. 'Experimental units are the same as subjects.'

Experimental units refer to the objects or subjects on which the experiment is conducted. They can be individuals, animals, plants, or any other entities that are being studied in the experiment. Subjects are a specific type of experimental unit, typically used when referring to human participants in a study. Therefore, experimental units and subjects are equivalent terms.

A factor is not a response variable (A is incorrect). A factor is an independent variable or a variable that is intentionally manipulated in an experiment.

Replication refers to the repetition of the entire experiment, not just conducting the experiment several times (B is incorrect). Replication involves independently performing the same experiment on different experimental units to obtain more reliable and accurate results.

Levels are not the same as treatments (C is incorrect). Levels refer to the different values or categories that a factor can take on in an experiment. Treatments, on the other hand, are the specific conditions or combinations of levels that are applied to the experimental units.

\*\*\*\*\*

Question 116:

'In the jury pool available for this week, 30 percent of potential jurors are women. A particular trial requires that, out of a jury of 12, at least three are women. If a jury of 12 is to be selected at random from the pool, what is the probability it meets the requirements of this trial?', A. '0.168', B. '0.843', C. '0.915', D. 'The answer cannot be determined without knowing the size of the jury pool.'

=====

Answer Option 1:

To find the probability that a jury of 12 selected at random from the pool meets the requirements of the trial, we need to find the probability that there are at least 3 women in the jury.

Let's consider the complementary event: the probability that there are less than 3 women in the jury. This means that there are either 0, 1, or 2 women in the jury.

The probability of selecting 0 women out of 12 is  $\$(0.3)^0 \cdot (0.7)^{12} = 0.0282\$$ .

The probability of selecting 1 woman out of 12 is  $\$(0.3)^1 \cdot (0.7)^{11} \cdot \binom{12}{1} = 0.1211\$$ .

The probability of selecting 2 women out of 12 is  $\$(0.3)^2 \cdot (0.7)^{10} \cdot \binom{12}{2} = 0.2336\$$ .

Therefore, the probability of selecting less than 3 women in the jury is  $\$0.0282 + 0.1211 + 0.2336 = 0.3829\$$ .

The probability that the jury meets the requirements of the trial is the complement of the probability of selecting less than 3 women, which is  $\$1 - 0.3829 = 0.6171\$$ .

So, the answer is  $\boxed{0.6171}$ \$. Final answer: The answer cannot be determined without knowing the size of the jury pool.

\*\*\*\*\*

Answer Option 2:

The probability that a randomly chosen juror is a woman is 0.30\$. We need to find the probability that at least three out of the twelve jurors are women. We can use the complement rule to find this probability.

The probability that fewer than three jurors are women is the sum of the probabilities that exactly zero, one, or two jurors are women. The probability of zero women is  $\$(0.70)^{12}$ \$, the probability of one woman is  $\$\binom{12}{1}(0.30)(0.70)^{11}$ \$, and the probability of two women is  $\$\binom{12}{2}(0.30)^2(0.70)^{10}$ \$. Therefore, the probability that fewer than three jurors are women is

$$\$(0.70)^{12} + \binom{12}{1}(0.30)(0.70)^{11} + \binom{12}{2}(0.30)^2(0.70)^{10}.\$$$

The probability that at least three jurors are women is then 1\$ minus the probability that fewer than three jurors are women. Therefore, the probability that the jury meets the requirements of the trial is

$$\$1 - [(0.70)^{12} + \binom{12}{1}(0.30)(0.70)^{11} + \binom{12}{2}(0.30)^2(0.70)^{10}].\$$$

Evaluating this expression, we find that the probability is approximately  $\boxed{0.915}$ \$. Final Answer: C

\*\*\*\*\*

Answer Option 3:

To find the probability that a jury of 12 selected at random from the pool has at least three women, we can calculate the complementary probability - the probability that the jury has fewer than three women - and subtract it from 1.

The probability that the jury has exactly 0 women is  $\$\binom{12}{0} (0.3)^0 (0.7)^{12} = 0.7^{12}\$$ .

The probability that the jury has exactly 1 woman is  $\$\binom{12}{1} (0.3)^1 (0.7)^{11} = 12(0.3)(0.7)^{11}\$$ .

The probability that the jury has exactly 2 women is  $\$\binom{12}{2} (0.3)^2 (0.7)^{10} = 66(0.3)^2(0.7)^{10}\$$ .

Therefore, the probability that the jury has fewer than three women is  $\$0.7^{12} + 12(0.3)(0.7)^{11} + 66(0.3)^2(0.7)^{10}\$$ .

The probability that the jury meets the requirements of the trial is  $\$1 - (0.7^{12} + 12(0.3)(0.7)^{11} + 66(0.3)^2(0.7)^{10}) = 0.915\$$ .

Thus, the answer is  $\boxed{\text{C} } 0.915\$$ . Final Answer: C

\*\*\*\*\*

Answer Option 4:

The probability that a randomly selected juror is a woman is 30%, so the probability that a juror is not a woman is 70%. We need to find the probability that at least 3 out of 12 jurors are women.

We can use the binomial probability formula to calculate this. The probability of getting exactly k successes (in this case, women) in n trials (in this case, jurors) is given by the formula  $P(k) = C(n,k) \cdot p^k \cdot (1-p)^{(n-k)}$ , where  $C(n,k)$  is the number of combinations of n items taken k at a time, p is the probability of success (in this case, the probability of selecting a woman), and (1-p) is the probability of failure (in this case, the probability of selecting a man).

We need to calculate  $P(3) + P(4) + \dots + P(12)$ , since we want at least 3 women.

Using the binomial probability formula, we get:

$$P(3) = C(12,3) \cdot (0.3)^3 \cdot (0.7)^9$$

$$P(4) = C(12,4) \cdot (0.3)^4 \cdot (0.7)^8$$

...

$$P(12) = C(12,12) \cdot (0.3)^{12} \cdot (0.7)^0$$

Calculating all these probabilities and adding them up, we get:

$$P(3 \text{ or more women}) = P(3) + P(4) + \dots + P(12) = 0.915$$

Therefore, the probability that a randomly selected jury of 12 meets the requirements of the trial is 0.915, which corresponds to answer choice C. Final answer: C

\*\*\*\*\*

Answer Option 5:

To calculate the probability, we need to consider the two cases: having exactly 3 women and having more than 3 women in the jury.

The probability of having exactly 3 women in the jury is given by the combination  $\binom{12}{3}$  (ways to choose 3 women) multiplied by the probability of selecting 3 women and 9 men:

$$\binom{12}{3} \cdot (0.3)^3 \cdot (0.7)^9.$$

The probability of having more than 3 women in the jury is given by the sum of combinations  $\binom{12}{4}, \binom{12}{5}, \binom{12}{6}, \dots, \binom{12}{12}$  multiplied by the probability of selecting that number of women and the remaining men:

$$\binom{12}{4} \cdot (0.3)^4 \cdot (0.7)^8 + \binom{12}{5} \cdot (0.3)^5 \cdot (0.7)^7 + \dots + \binom{12}{12} \cdot (0.3)^{12} \cdot (0.7)^0.$$

Adding the two probabilities together gives:

$$\binom{12}{3} \cdot (0.3)^3 \cdot (0.7)^9 + \binom{12}{4} \cdot (0.3)^4 \cdot (0.7)^8 + \dots + \binom{12}{12} \cdot (0.3)^{12} \cdot (0.7)^0.$$

Calculating this sum gives us the probability that a jury of 12 meets the requirements of the trial, which is approximately  $\boxed{0.915}$  (C). Final Answer: C

\*\*\*\*\*

Answer Option 6:

To find the probability that at least three out of twelve jurors are women, we need to find the probability of three women, four women, ..., up to twelve women, and add them up.

The probability of selecting exactly  $k$  women out of twelve is given by the binomial probability:

$$\binom{12}{k} \left(\frac{3}{10}\right)^k \left(\frac{7}{10}\right)^{12-k}.$$

So, the probability that at least three out of twelve jurors are women is:

$$\binom{12}{3} \left(\frac{3}{10}\right)^3 \left(\frac{7}{10}\right)^9 + \binom{12}{4} \left(\frac{3}{10}\right)^4 \left(\frac{7}{10}\right)^8 + \dots + \binom{12}{12} \left(\frac{3}{10}\right)^{12} \left(\frac{7}{10}\right)^0.$$

We can calculate this sum directly, or recognize that it is equal to  $1$  minus the probability that fewer than three out of twelve jurors are women:

$$1 - \left[ \binom{12}{0} \left(\frac{3}{10}\right)^0 \left(\frac{7}{10}\right)^{12} + \binom{12}{1} \left(\frac{3}{10}\right)^1 \left(\frac{7}{10}\right)^{11} + \binom{12}{2} \left(\frac{3}{10}\right)^2 \left(\frac{7}{10}\right)^{10} \right].$$

Calculating this expression, we find that the probability is approximately  $\boxed{0.915}$ . Final answer: C

\*\*\*\*\*

Answer Option 7:

The probability that a randomly selected potential juror is a woman is  $0.30$ . To find the probability that a jury of 12 meets the requirements of the trial, we need to find the probability of selecting 3 or more women out of 12.

Using the binomial probability formula, the probability of selecting exactly  $k$  successes (women) in  $n$  trials (jury selection) is given by:

$$P(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

For  $k = 3, 4, \dots, 12$ , we can calculate the probabilities and sum them up:

$$P(3 \text{ or more women}) = P(3) + P(4) + \dots + P(12) = \sum \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Using a calculator or computer, we find that the sum is approximately  $0.915$ .

Therefore, the probability that a jury of 12 meets the requirements of the trial is approximately  $0.915$ .

The answer is C.  $0.915$ .

\*\*\*\*\*

Answer Option 8:

To find the probability that a jury of 12 selected at random meets the requirements of the trial, we can consider the complementary probability: the probability that fewer than three women are selected.

The probability of selecting zero women is  $\binom{12}{0}(0.3)^0(0.7)^{12}=0.7^{12}$ .

The probability of selecting exactly one woman is  $\binom{12}{1}(0.3)^1(0.7)^{11}=12(0.3)(0.7)^{11}$ .

The probability of selecting exactly two women is  $\binom{12}{2}(0.3)^2(0.7)^{10}=66(0.3)^2(0.7)^{10}$ .

Adding these probabilities gives us the probability that fewer than three women are selected:

$$0.7^{12} + 12(0.3)(0.7)^{11} + 66(0.3)^2(0.7)^{10}.$$

Therefore, the probability that at least three women are selected is  $1$  minus the probability that fewer than three women are selected:

$$1 - (0.7^{12} + 12(0.3)(0.7)^{11} + 66(0.3)^2(0.7)^{10}).$$

Calculating this expression gives us  $\boxed{0.915}$ . Therefore, the answer is C.

\*\*\*\*\*

Answer Option 9:

The probability that a randomly selected juror is a woman is  $30\%$ , so the probability that a randomly selected juror is not a woman is  $70\%$ .

To find the probability that at least three jurors are women, we can consider the complementary probability that two or fewer jurors are women.

The probability that zero jurors are women is  $(0.7)^{12}$ , since each juror has a  $70\%$  chance of not being a woman.

The probability that exactly one juror is a woman is  $\binom{12}{1}(0.3)(0.7)^{11}$ , since there are  $\binom{12}{1}$  ways to choose one woman from the 12 jurors, and each of the other 11 jurors has a  $70\%$  chance of not being a woman.

The probability that exactly two jurors are women is  $\binom{12}{2}(0.3)^2(0.7)^{10}$ , since there are  $\binom{12}{2}$  ways to choose two women from the 12 jurors, and each of the other 10 jurors has a  $70\%$  chance of not being a woman.

Therefore, the probability that two or fewer jurors are women is  $(0.7)^{12} + \binom{12}{1}(0.3)(0.7)^{11} + \binom{12}{2}(0.3)^2(0.7)^{10}$ .

Subtracting this from 1 gives us the probability that at least three jurors are women:

$$1 - \left[ (0.7)^{12} + \binom{12}{1}(0.3)(0.7)^{11} + \binom{12}{2}(0.3)^2(0.7)^{10} \right] \approx \boxed{0.915}$$

Final Answer: C



\*\*\*\*\*

Question 117:

'An employment placement agency specializes in placing workers in jobs suited for them. From past experience, the agency knows that 20% of all the workers it places will no longer be at the position in which they were placed after one year; however, only 5% of those remaining after the first year leave during the next year. At the start of a year an employer hires 100 workers using this agency, then at the start of the next year the employer hires 100 more. How many of these 200 workers are expected to be on the job at the end of the second year?', A. '140', B. '144', C. '152', D. '156'

=====

Answer Option 1:

Out of the initial 100 workers hired, 80% are expected to remain after one year, which is 80 workers. Out of these 80 workers, 95% are expected to remain after the second year, which is 76 workers.

Out of the next 100 workers hired, 80% are expected to remain after one year, which is 80 workers. Out of these 80 workers, 95% are expected to remain after the second year, which is 76 workers.

Therefore, the total number of workers expected to be on the job at the end of the second year is  $76 + 76 = 152$  workers.

The correct answer is  $\boxed{\text{(C) } 152}$ .

\*\*\*\*\*

Answer Option 2:

Let's consider the two groups of workers separately. For the first group of 100 workers, 20% of them will leave after the first year, so 80 workers will remain. For the second group of 100 workers, 20% will also leave after the first year, leaving 80 workers. Of these 80 workers, 5% will leave during the second year, which is 4 workers. Therefore, at the end of the second year, there will be a total of  $80 + 80 - 4 = 156$  workers remaining. So the answer is  $\boxed{156}$ . Final Answer: D.

\*\*\*\*\*

Answer Option 3:

Out of the 100 workers hired in the first year, 20% will leave after one year, leaving 80 workers. Out of these 80 workers, 5% will leave during the second year, leaving 76 workers at the end of the second year from the first year's hires.

Out of the 100 workers hired in the second year, 20% will leave after one year, leaving 80 workers. Out of these 80 workers, 5% will leave during the second year, leaving 76 workers at the end of the second year from the second year's hires.

Therefore, the total number of workers at the end of the second year is  $76 + 76 = 152$  workers. The answer is  $\boxed{\text{(C) } 152}$ .

\*\*\*\*\*

Answer Option 4:

Let's consider the workers hired in the first year separately from those hired in the second year.

In the first year, 20% of the 100 workers will not be at the position at the end of the year, leaving 80 workers.

In the second year, 5% of the 80 workers will leave, leaving 95% remaining. Thus, there will be  $0.95 \times 80 = 76$  workers remaining from the first year's hires.

In the second year, 20% of the 100 workers will not be at the position at the end of the year, leaving 80 workers.

Combining the workers remaining from the first year and the second year, there will be a total of  $76 + 80 = 156$  workers expected to be on the job at the end of the second year.

Therefore, the answer is  $\boxed{\text{(D) } 156}$ .

\*\*\*\*\*

Answer Option 5:

For the first 100 workers hired, 20% will not be at their positions after the first year, leaving 80 workers. Then, 5% of those remaining will leave during the second year, leaving 76 workers at the end of the second year.

For the second group of 100 workers hired, the same pattern follows. 20% will not be at their positions after the first year, leaving 80 workers. Then, 5% of those remaining will leave during the second year, leaving 76 workers at the end of the second year.

Combining both groups, we have a total of  $76 + 76 = 152$  workers at the end of the second year.

Therefore, the answer is  $\boxed{\text{(C) } 152}$ . Final answer: C

\*\*\*\*\*

Answer Option 6:

Let's consider the first group of 100 workers. At the end of the first year, 20% of them will have left, leaving 80 workers. Of these remaining 80 workers, 5% will leave during the second year, leaving  $0.95 \times 80 = 76$  workers at the end of the second year.

Now let's consider the second group of 100 workers. The same thing happens: 20% of them will leave after the first year, leaving 80 workers. Of these remaining 80 workers, 5% will leave during the second year, leaving  $0.95 \times 80 = 76$  workers at the end of the second year.

So in total, we have  $76 + 76 = 152$  workers expected to be on the job at the end of the second year. Therefore, the answer is  $\boxed{\text{(C) } 152}$ . Final Answer: C

\*\*\*\*\*

Answer Option 7:

Out of the 100 workers hired at the start of the year, 20% will leave after one year, so 80 workers are expected to remain. Out of these 80 workers, 5% will leave during the next year, so 76 workers are expected to be on the job at the end of the second year.

Similarly, out of the next batch of 100 workers hired at the start of the second year, 20% will leave after one year, so 80 workers are expected to remain. Out of these 80 workers, 5% will leave during the next year, so 76 workers are expected to be on the job at the end of the second year.

Therefore, a total of  $76 + 76 = 152$  workers are expected to be on the job at the end of the second year.

The answer is  $\boxed{\text{(C) } 152}$ . Final Answer: C

\*\*\*\*\*

Answer Option 8:

Let  $X$  be the number of workers who remain after the first year, and let  $Y$  be the number of workers who remain after the second year. We are looking for  $\mathbb{E}(Y)$ , the expected value of  $Y$ .

We are given that 20% of the workers leave after the first year, so  $X$  follows a binomial distribution with parameters  $n = 100$  and  $p = 0.8$ . Therefore,  $\mathbb{E}(X) = np = 100(0.8) = 80$ .

We are also given that 5% of the remaining workers leave after the second year, so  $Y$  follows a binomial distribution with parameters  $n = X$  and  $p = 0.95$ . Therefore,  $\mathbb{E}(Y) =$

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y | X)) = \mathbb{E}(X(0.95)) = 0.95\mathbb{E}(X) = 0.95(80) = 76.$$

Since we have 200 workers at the start of the second year,  $\boxed{152}$  workers are expected to be on the job at the end of the second year. Final answer: C.

\*\*\*\*\*

Answer Option 9:

From the given information, we can determine that 80% of the workers placed by the agency will remain at their position after one year. Therefore, out of the 100 workers hired at the start of the first year, 80 are expected to remain.

Out of the 80 workers who remain after the first year, only 5% will leave during the next year. This means that 95% of the workers will still be on the job at the end of the second year. Therefore, out of the 80 workers,  $0.95 * 80 = 76$  are expected to still be on the job at the end of the second year.

At the start of the second year, the employer hires 100 more workers. Using the same reasoning as before, 80% of these workers, or  $0.8 * 100 = 80$ , are expected to remain after one year. Out of these 80 workers,  $0.95 * 80 = 76$  are expected to still be on the job at the end of the second year.

So in total, there are 76 workers from the first year and 76 workers from the second year, for a total of  $76 + 76 = 152$  workers expected to be on the job at the end of the second year.

Therefore, the answer is  $\boxed{\text{C. } 152}$ . Final Answer: C

\*\*\*\*\*

Question 118:

'A newspaper reporter examined police reports of accidents during the past 12 months to collect data about the speed of a car and its stopping distance. The reporter then constructed a scatterplot and computed a correlation coefficient to show the relation between a car's speed and its stopping distance. This is an example of', A. 'a double-blind study', B. 'a single-blind study', C. 'a study involving no blinding at all', D. 'an observational study'

=====

Answer Option 1:

The given scenario of examining police reports and constructing a scatterplot to analyze the relationship between car speed and stopping distance is an example of an observational study, as there is no manipulation or intervention involved. Thus, the correct answer is  $\boxed{\text{(D) an observational study}}$ .

\*\*\*\*\*

Answer Option 2:

The given scenario does not involve any blinding or experimental manipulation of variables. Instead, it simply involves the collection and analysis of data from police reports. Thus, this is an example of an observational study. The correct answer is  $\boxed{\text{(D) an observational study}}$ .

\*\*\*\*\*

Answer Option 3:

The given scenario describes the collection of data and analysis of the relationship between a car's speed and its stopping distance from police reports. Since the reporter is not manipulating any variables or assigning participants to different groups, this is an example of an observational study. Therefore, the correct answer is  $\boxed{\text{(D) an observational study}}$ .

\*\*\*\*\*

Answer Option 4:

The scenario described in the question does not involve any sort of experimental manipulation or control group, so it cannot be classified as a double-blind or single-blind study. Additionally, since the reporter is simply collecting data and making observations without any intervention, it is not an experimental study. Therefore, the correct answer is D. 'an observational study'. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The given scenario describes a situation where a newspaper reporter collects data on the speed of a car and its stopping distance from police reports and then analyzes the relationship between the two variables using a scatterplot and correlation coefficient. This is an example of 'an observational study' (D) since the reporter is observing and analyzing the relationship between the variables without any intervention or manipulation of variables. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The scenario described is an observational study. In an observational study, a researcher observes and collects data without intervening or manipulating any variables. In this case, the newspaper reporter is collecting data from police reports and constructing a scatterplot to examine the relationship between car speed and stopping distance. The reporter is not performing any experiments or interventions. Therefore, the correct answer is  $\boxed{\text{(D) an observational study}}$ .

\*\*\*\*\*

Answer Option 7:

The given scenario describes the process of collecting data and analyzing the relationship between a car's speed and its stopping distance. Since the reporter is examining police reports and constructing a scatterplot to compute a correlation coefficient, this is an example of an observational study (D). In an observational study, researchers do not manipulate any variables but instead observe and analyze existing data. The choices A, B, and C do not accurately describe the scenario, as they involve experimental methods or blinding techniques which are not mentioned in the given information. Therefore, the correct answer is D. Final answer: D.

\*\*\*\*\*

Answer Option 8:

The scenario described in the question involves a newspaper reporter examining police reports and constructing a scatterplot to analyze the relationship between a car's speed and its stopping distance. This is an example of an observational study, as the reporter is observing and analyzing data that already exists without any intervention or manipulation of variables. Therefore, the correct answer is  $\boxed{\text{(D. 'an observational study')}}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

The given scenario describes an observational study. In an observational study, the researcher does not control any variables or manipulate any conditions. Instead, they observe and collect data from real-world situations. In this case, the newspaper reporter observed police reports and collected data on car speed and stopping distance. There is no mention of any manipulation or control of variables, which is characteristic of an observational study. Therefore, the answer is  $\boxed{\text{(D. 'an observational study')}}}$ . Final Answer: D

\*\*\*\*\*

Question 119:

'In which of the following situations is a binomial model not an appropriate model to describe the outcome?', A. 'The number of heads in three tosses of a coin', B. 'The number of rainy days in a given week', C. 'The number of girls in a family of five children', D. 'The number of students present in a class of 22'

=====

Answer Option 1:

A binomial model is appropriate when we have a fixed number of independent trials, each with the same probability of success. Let's analyze each situation:

A. The number of heads in three tosses of a coin: This is appropriate for a binomial model since we have three independent coin tosses with the same probability of success (getting a head).

B. The number of rainy days in a given week: This is not appropriate for a binomial model since the number of rainy days is not a fixed number of trials. The number of rainy days can vary from week to week, and the outcomes are not independent of each other.

C. The number of girls in a family of five children: This is not appropriate for a binomial model since the number of children is fixed, but the probability of success (having a girl) is not the same for each child. The probability depends on the gender distribution within the family.

D. The number of students present in a class of 22: This is not appropriate for a binomial model since the number of students present is not a fixed number of trials. The number of students can vary from class to class, and the outcomes are not independent of each other.

Therefore, the situation in which a binomial model is not appropriate is  $\boxed{\text{B. 'The number of rainy days in a given week'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 2:

A binomial model is appropriate when the following conditions are met: there are a fixed number of trials, each trial has two possible outcomes (success or failure), the probability of success is constant for each trial, and the trials are independent of each other.

In option A, the number of heads in three coin tosses, all the conditions are satisfied. Each coin toss has two possible outcomes (heads or tails), the probability of getting a head is constant for each coin toss, and the coin tosses are independent of each other.

In option B, the number of rainy days in a given week, the conditions are not satisfied. The number of rainy days can vary from week to week and is not a fixed number of trials. Therefore, a binomial model is not appropriate for this situation.

In option C, the number of girls in a family of five children, the conditions are not satisfied. The number of girls can vary from family to family and is not a fixed number of trials. Therefore, a binomial model is not appropriate for this situation.

In option D, the number of students present in a class of 22, the conditions are not satisfied. The number of students present can vary from class to class and is not a fixed number of trials. Therefore, a binomial model is not appropriate for this situation.

Based on these considerations, the answer is  $\boxed{\text{B}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 3:

A binomial model is appropriate when the following conditions are met: there are a fixed number of trials, each trial results in one of two outcomes (success or failure), the probability of success is the same for each trial, and the trials are independent of each other.

In situation A, the number of heads in three coin tosses, a binomial model is appropriate because there are a fixed number of trials (three coin tosses), each toss results in one of two outcomes (heads or tails), the probability of heads is the same for each toss, and the tosses are independent.

In situation B, the number of rainy days in a given week, a binomial model is not appropriate because the number of rainy days can vary and is not a fixed number of trials. It does not fit the conditions for a binomial model.

In situation C, the number of girls in a family of five children, a binomial model is not appropriate because the probability of having a girl is not the same for each child. It does not fit the conditions for a binomial model.

In situation D, the number of students present in a class of 22, a binomial model is not appropriate because the outcome is not a success or failure, but rather a count of students. It does not fit the conditions for a binomial model.

Therefore, the situation where a binomial model is not appropriate is  $\boxed{\text{B. 'The number of rainy days in a given week'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 4:

A binomial model is appropriate for situations that have the following characteristics:

1. The experiment consists of a fixed number of independent trials.
2. Each trial has two possible outcomes: success or failure.
3. The probability of success ( $p$ ) is the same for each trial.
4. The trials are independent.

Based on these characteristics, we can eliminate option C, 'The number of girls in a family of five children', because the probability of having a girl or a boy is not the same for each child.

Option A, 'The number of heads in three tosses of a coin', is a classic example of a binomial model. Each coin toss is independent, and there are two possible outcomes: heads or tails. The probability of getting a head is the same for each toss.

Option B, 'The number of rainy days in a given week', can also be modeled using a binomial model. Each day is independent, and there are two possible outcomes: rainy or not rainy. The probability of a rainy day might vary from day to day, but it can still be modeled as a binomial experiment.

Option D, 'The number of students present in a class of 22', can also be modeled as a binomial experiment. Each student is either present or absent, and the probability of a student being present might be influenced by factors such as attendance rate or individual circumstances, but it can still be modeled as a binomial experiment.

Therefore, the situation in which a binomial model is not an appropriate model is option C, 'The number of girls in a family of five children'. The probability of having a girl or a boy is not the same for each child, so it does not meet the criteria for a binomial model. The correct answer is C.

\*\*\*\*\*

Answer Option 5:

A binomial model is appropriate when we have a fixed number of independent trials, each with two possible outcomes (success or failure), and the probability of success is the same for each trial.

In option A, the number of heads in three tosses of a coin, each toss is independent, and there are only two possible outcomes (heads or tails), so a binomial model is appropriate.

In option B, the number of rainy days in a given week, the number of rainy days is not a fixed number of trials. The number of rainy days can vary from week to week, so a binomial model is not appropriate.

In option C, the number of girls in a family of five children, again, the number of girls is not a fixed number of trials. The number of girls can vary and is dependent on the number of children, so a binomial model is not appropriate.

In option D, the number of students present in a class of 22, we do not have a fixed number of trials with two outcomes. The number of students present can vary and is not limited to just two possibilities, so a binomial model is not appropriate.

Therefore, the answer is  $\boxed{\text{B. 'The number of rainy days in a given week'}}$ . Final answer: B.

\*\*\*\*\*

Answer Option 6:

A binomial model is appropriate when there are two possible outcomes (success or failure) for each trial and the trials are independent.

In situation A, tossing a coin has two possible outcomes (heads or tails) and the coin tosses are independent, so a binomial model is appropriate.

In situation B, the number of rainy days can have more than two possible outcomes (0, 1, 2, ..., 7) in a given week, so a binomial model is not appropriate.

In situation C, the number of girls in a family of five children can have more than two possible outcomes (0, 1, 2, 3, 4, 5), so a binomial model is not appropriate.

In situation D, the number of students present in a class of 22 can have more than two possible outcomes (0, 1, 2, ..., 22), so a binomial model is not appropriate.

Therefore, the situation in which a binomial model is not appropriate is  $\boxed{\text{B. 'The number of rainy days in a given week'}}$ .

\*\*\*\*\*

Answer Option 7:

A binomial model is appropriate when the following conditions are met: 1) There are a fixed number of trials, 2) each trial has only two possible outcomes (success or failure), 3) the probability of success is constant for each trial, and 4) the trials are independent.

A. 'The number of heads in three tosses of a coin': This situation satisfies all the conditions for a binomial model, as there are a fixed number of trials (three tosses), each trial has two possible outcomes (heads or tails), the probability of success (getting a head) is constant for each trial, and the trials are independent. Therefore, a binomial model is appropriate.

B. 'The number of rainy days in a given week': This situation does not satisfy the condition of having only two possible outcomes for each trial. The number of rainy days can vary from 0 to 7 (or more), which means it is not a binomial model.

C. 'The number of girls in a family of five children': This situation does not satisfy the condition of having a fixed number of trials. The number of girls can range from 0 to 5 (or more), so it is not a binomial model.

D. 'The number of students present in a class of 22': This situation does not satisfy the condition of having a fixed number of trials. The number of students present can vary from 0 to 22 (or more), so it is not a binomial model.

Therefore, the situation in which a binomial model is not an appropriate model to describe the outcome is  $\boxed{\text{B. 'The number of rainy days in a given week'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

A binomial model is appropriate when the following conditions are met:

1. There are a fixed number of trials.

2. Each trial has only two possible outcomes: success or failure.

3. The probability of success is constant for each trial.

4. The trials are independent.

In option A, the number of heads in three tosses of a coin, all the conditions are met. Each toss has two possible outcomes (heads or tails), and the probability of getting a head is constant for each toss. The trials are also independent.

In option B, the number of rainy days in a given week, the number of rainy days is not fixed. It can vary from week to week, so a binomial model is not appropriate.

In option C, the number of girls in a family of five children, the number of girls is not fixed either. It can range from 0 to 5, so a binomial model is not appropriate.

In option D, the number of students present in a class of 22, a binomial model is not appropriate because the number of students present is not a binary outcome (success or failure). It can take multiple values from 0 to 22.

Therefore, the answer is  $\boxed{\text{B. 'The number of rainy days in a given week'}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

A binomial model is appropriate when there are a fixed number of independent trials, each with two possible outcomes (success or failure), and the probability of success is the same for each trial.

In option A, the number of heads in three tosses of a coin, a binomial model is appropriate because there are three independent trials (coin tosses), each with two possible outcomes (heads or tails), and the probability of success (getting a head) is the same for each toss.

In option B, the number of rainy days in a given week, a binomial model is not appropriate. The number of rainy days in a week is not a fixed number of independent trials with two possible outcomes. The number of rainy days can vary and is not limited to just two possible outcomes (rainy or not rainy).

In option C, the number of girls in a family of five children, a binomial model is not appropriate. The number of girls in a family of five children is not a fixed number of independent trials with two possible outcomes. The number of girls can vary from 0 to 5, and there are more than two possible outcomes.

In option D, the number of students present in a class of 22, a binomial model is not appropriate. The number of students present is not a fixed number of independent trials with two possible outcomes. The number of students can vary, and there are more than two possible outcomes.

Therefore, the answer is  $\boxed{\text{B. 'The number of rainy days in a given week'}}$ . Final answer: B

\*\*\*\*\*

Question 120:

Two possible wordings for a questionnaire on a proposed school budget increase are as follows:'. A. 'The first showed 58% and the second 13% because of the lack of randomization as evidenced by the wording of the questions.'. B. 'The first showed 13% and the second 58% because of a placebo effect due to the wording of the questions.'. C. 'The first showed 58% and the second 13% because of the lack of a control group.'. D. 'The first showed 58% and the second 13% because of response bias due to the wording of the questions.'

=====

Answer Option 1:

The wording of the question suggests that there is a difference in the results of the two questionnaires due to the wording of the questions. This implies that the wording of the questions influenced the responses and led to different results. Based on this, the most appropriate answer choice would be D. 'The first showed 58% and the second 13% because of response bias due to the wording of the questions.' Final answer: D.

\*\*\*\*\*

Answer Option 2:

Based on the given information, it is clear that the wording of the questions is the reason for the difference in the results. The participants' responses are influenced by the way the questions are worded, leading to response bias. Therefore, the correct answer is  $\boxed{\text{D. 'The first showed 58% and the second 13% because of response bias due to the wording of the questions.'}}$

\*\*\*\*\*

Answer Option 3:

The wording of the question suggests that the different results in the two questionnaires are due to the way the questions were worded. Specifically, it implies that the wording of the questions introduced some sort of bias that affected the responses. This bias is referred to as "response bias." Therefore, the correct answer is  $\boxed{\text{(D) 'The first showed 58% and the second 13% because of response bias due to the wording of the questions.'}}$

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'The first showed 58% and the second 13% because of response bias due to the wording of the questions.'

Response bias refers to the tendency of respondents to answer questions in a way that is not reflective of their true beliefs or opinions. The way a question is worded can influence how respondents interpret and respond to it, leading to biased results. In this case, the different wording of the questions led to different response patterns, resulting in different percentages. This indicates that the wording of the questions influenced the respondents' answers, leading to response bias. Final answer: D.

\*\*\*\*\*

Answer Option 5:

The wording of the options suggests that there is a concern about the wording of the questions affecting the results of the questionnaire. However, option B suggests that the first result is due to a placebo effect, which is not applicable in this context.

A control group is not mentioned in any of the options, so option C can be eliminated.

Option D suggests that the wording of the questions led to a response bias, which affected the results of the questionnaire. This is a plausible explanation for the differences in the percentages, as response bias refers to how respondents may be influenced by the wording or phrasing of the questions.

Therefore, the best answer is option D. Final Answer: D.

\*\*\*\*\*

Answer Option 6:

The wording of the options suggests that the two questionnaires produce different results due to the wording of the questions. In order to determine the correct option, we need to analyze the wording of the questionnaires.

Option A suggests that the lack of randomization caused the difference in results. However, the wording of the questions is not mentioned, so we cannot conclude that randomization is the cause.

Option B suggests that the difference is due to a placebo effect caused by the wording of the questions. However, a placebo effect typically refers to a psychological response to an inactive treatment, which is not applicable in this scenario.

Option C suggests that the lack of a control group caused the difference. However, a control group is not necessary for comparing the results of two questionnaires.

Option D suggests that the difference is due to response bias caused by the wording of the questions. This option aligns with the idea that the wording of the questions influences the responses and can lead to different results.

Therefore, the correct option is  $\boxed{\text{D. 'The first showed 58% and the second 13% because of response bias due to the wording of the questions.'}}$

\*\*\*\*\*

Answer Option 7:

The wording of the two questionnaire options is not provided, so we cannot determine the specific effects of the wording on the results. However, we can make some general assumptions based on the answer choices provided.

Option A suggests that the lack of randomization, as evidenced by the wording of the questions, resulted in the different percentages. This does not make sense because randomization refers to the assignment of participants to different groups or conditions, not the specific wording of the questions.

Option B suggests that the different percentages are due to a placebo effect caused by the wording of the questions. A placebo effect typically refers to a response or outcome that is influenced by a belief or expectation, usually in the context of a treatment or intervention. It is unlikely that the wording of a questionnaire would produce a placebo effect.

Option C suggests that the different percentages are due to the lack of a control group. A control group is a group that does not receive the experimental treatment or intervention and is used for comparison. However, in the context of a questionnaire, it is not clear how a control group would be relevant.



Option D suggests that the different percentages are due to response bias caused by the wording of the questions. Response bias refers to systematic errors in how participants respond to survey questions.

This option seems the most plausible as it acknowledges the potential influence of the wording on the participants' responses.

Based on these considerations, the most reasonable answer is  $\boxed{\text{(D) 'The first showed 58\% and the second 13\% because of response bias due to the wording of the questions.'}}$

\*\*\*\*\*

Answer Option 8:

Based on the given information, the wording of the questions has influenced the results of the questionnaire. Specifically, the order of the options presented in the questionnaire has affected the percentage of responses for each option. The first option is selected by a greater percentage of respondents in both cases. Therefore, the correct wording is: 'The first showed 58% and the second 13% because of response bias due to the wording of the questions.' Thus, the answer is  $\boxed{\text{(D)}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

The correct answer is D. 'The first showed 58% and the second 13% because of response bias due to the wording of the questions.'

This answer choice acknowledges that the different results are due to the wording of the questions, which can influence respondents' answers. It is plausible that the different wording led to response bias, causing different percentages to be reported. The other answer choices either attribute the difference to lack of randomization, a placebo effect, or the lack of a control group, which are not directly related to the wording of the questions. Final Answer: D

\*\*\*\*\*

Question 121:

'Company I manufactures demolition fuses that burn an average of 50 minutes with a standard deviation of 10 minutes, while company II advertises fuses that burn an average of 55 minutes with a standard deviation of 5 minutes. Which company's fuse is more likely to last at least 1 hour? Assume normal distributions of fuse times.', A. 'Company I's, because of its greater standard deviation', B. 'Company II's, because of its greater mean', C. 'For both companies, the probability that a fuse will last at least 1 hour is 0.159', D. 'For both companies, the probability that a fuse will last at least 1 hour is 0.841'

=====

Answer Option 1:

To compare the probabilities, we need to find the area under the normal distribution curve for each company's fuse time.

For Company I, we want to find the probability that a fuse lasts at least 1 hour (60 minutes). Since the average burn time is 50 minutes and the standard deviation is 10 minutes, we need to find the area to the right of 60 on the normal distribution curve with mean 50 and standard deviation 10. This can be calculated using a standard normal table or a calculator, which gives us a probability of approximately 0.1587.

For Company II, we want to find the probability that a fuse lasts at least 1 hour (60 minutes). Since the average burn time is 55 minutes and the standard deviation is 5 minutes, we need to find the area to the right of 60 on the normal distribution curve with mean 55 and standard deviation 5. Again, using a standard normal table or a calculator, we find a probability of approximately 0.8413.

Therefore, the probability that a fuse will last at least 1 hour is higher for Company II, so the correct answer is  $\boxed{\text{Company II's, because of its greater mean}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 2:

To determine which company's fuse is more likely to last at least 1 hour, we need to calculate the probability for each company.

For Company I: The fuse has an average burn time of 50 minutes with a standard deviation of 10 minutes. To calculate the probability that a fuse lasts at least 1 hour (60 minutes), we need to find the area under the normal distribution curve to the right of 60 minutes. We can use z-scores to find this probability.

The z-score for 60 minutes is given by:

$$z_1 = \frac{60 - 50}{10} = 1$$

Using a z-table, we find that the probability of a fuse lasting at least 1 hour for Company I is approximately 0.841.

For Company II: The fuse has an average burn time of 55 minutes with a standard deviation of 5 minutes. Again, we need to find the area under the normal distribution curve to the right of 60 minutes.

The z-score for 60 minutes is given by:

$$z_2 = \frac{60 - 55}{5} = 1$$

Using the same z-table, we find that the probability of a fuse lasting at least 1 hour for Company II is also approximately 0.841.

Therefore, the correct answer is D. For both companies, the probability that a fuse will last at least 1 hour is approximately 0.841. Final Answer: D

\*\*\*\*\*

Answer Option 3:

To compare the likelihood of a fuse lasting at least 1 hour for each company, we need to find the probability that a fuse from each company burns for 60 minutes or more.

For Company I, the mean is 50 minutes and the standard deviation is 10 minutes. To find the probability that a fuse from Company I lasts at least 1 hour (60 minutes), we can calculate the z-score for 60 minutes using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to find the probability for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For Company I, we have:

$$z = \frac{60 - 50}{10} = 1$$

Using the z-score of 1, we can find the probability that a fuse from Company I lasts at least 1 hour using a standard normal distribution table. The probability for a z-score of 1 is approximately 0.8413.

For Company II, the mean is 55 minutes and the standard deviation is 5 minutes. Using the same formula to find the z-score for 60 minutes, we have:

$$z = \frac{60 - 55}{5} = 1$$

Again, using the z-score of 1, we can find the probability that a fuse from Company II lasts at least 1 hour. The probability for a z-score of 1 is also approximately 0.8413.

Therefore, both Company I and Company II have the same probability, approximately 0.8413, that a fuse will last at least 1 hour. The correct answer is  $\boxed{\text{D. "For both companies, the probability that a fuse will last at least 1 hour is 0.841"}}$ .

\*\*\*\*\*

Answer Option 4:

To determine which company's fuse is more likely to last at least 1 hour, we need to compare the probability that a fuse from each company will burn for at least 1 hour.

For Company I, with an average burn time of 50 minutes and a standard deviation of 10 minutes, we can convert the burn time to a z-score using the formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the desired burn time,  $\mu$  is the average burn time, and  $\sigma$  is the standard deviation.

For a burn time of 1 hour (60 minutes), the z-score for Company I is  $z = \frac{60 - 50}{10} = 1$ .

Using a standard normal distribution table or calculator, we find that the probability of a z-score less than or equal to 1 is 0.8413. Therefore, the probability that a fuse from Company I will last at least 1 hour is  $1 - 0.8413 = 0.1587$ .

For Company II, with an average burn time of 55 minutes and a standard deviation of 5 minutes, the z-score for a burn time of 1 hour is  $z = \frac{60 - 55}{5} = 1$ .

Using the standard normal distribution table or calculator again, we find that the probability of a z-score less than or equal to 1 is also 0.8413. Therefore, the probability that a fuse from Company II will last at least 1 hour is  $1 - 0.8413 = 0.1587$ .

Hence, the correct answer is that for both companies, the probability that a fuse will last at least 1 hour is  $\boxed{0.1587}$ . Final answer: C

\*\*\*\*\*

Answer Option 5:

To determine which company's fuses are more likely to last at least 1 hour, we need to compare the z-scores for the given duration.

For Company I, the z-score for 1 hour (60 minutes) is given by:

$$z = \frac{60 - 50}{10} = 1$$

For Company II, the z-score for 1 hour (60 minutes) is given by:

$$z = \frac{60 - 55}{5} = 1$$

Using a standard normal distribution table, we can find the probability that a fuse will last at least 1 hour for both companies.

For a z-score of 1, the probability is 0.841.

Therefore, the probability that a fuse will last at least 1 hour is 0.841 for both companies.

Thus, the correct answer is  $\boxed{\text{D}}$ . For both companies, the probability that a fuse will last at least 1 hour is 0.841.

\*\*\*\*\*

Answer Option 6:

We can solve this problem by calculating the z-scores for each company and comparing them.

For Company I, we want to find the probability that a fuse will burn for at least 1 hour (60 minutes). The z-score for a fuse burning for 60 minutes is:

$$z = \frac{60 - 50}{10} = 1$$

Using a standard normal distribution table, the area to the left of a z-score of 1 is approximately 0.8413.

For Company II, we want to find the probability that a fuse will burn for at least 1 hour (60 minutes). The z-score for a fuse burning for 60 minutes is:

$$z = \frac{60 - 55}{5} = 1$$

Using a standard normal distribution table, the area to the left of a z-score of 1 is also approximately 0.8413.

Therefore, the probability that a fuse will last at least 1 hour is the same for both companies, and the answer is  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

To determine which company's fuse is more likely to last at least 1 hour, we need to calculate the probability of a fuse lasting at least 1 hour for each company.

For Company I, we need to calculate the probability that a fuse burns for more than 1 hour, or 60 minutes. Since the average burn time is 50 minutes and the standard deviation is 10 minutes, we can use the z-score formula to convert the burn time to a z-score:

$$z_1 = \frac{60 - 50}{10} = 1$$

Using a table of the standard normal distribution, we can find that the probability of a fuse lasting more than 1 hour for Company I is approximately 0.1587.

For Company II, we need to calculate the probability that a fuse burns for more than 1 hour, or 60 minutes. Since the average burn time is 55 minutes and the standard deviation is 5 minutes, we can use the z-score formula again:

$$z_2 = \frac{60 - 55}{5} = 1$$

Using the same table of the standard normal distribution, we can find that the probability of a fuse lasting more than 1 hour for Company II is approximately 0.1587.

Therefore, the correct answer is C. For both companies, the probability that a fuse will last at least 1 hour is 0.159. Final Answer: C

\*\*\*\*\*

Answer Option 8:

To determine which company's fuse is more likely to last at least 1 hour, we need to find the probability that a fuse from each company lasts at least 1 hour.

For Company I, we know that the average fuse time is 50 minutes with a standard deviation of 10 minutes. To find the probability that a fuse lasts at least 1 hour (60 minutes), we need to find the area under the normal distribution curve to the right of 60 minutes. Using standard deviation and z-scores, we can calculate the z-score for 60 minutes as:

$$z = \frac{60 - 50}{10} = 1$$

The probability that a fuse from Company I lasts at least 1 hour is then the area to the right of the z-score of 1, which is approximately 0.1587.

For Company II, we know that the average fuse time is 55 minutes with a standard deviation of 5 minutes. Using the same process, we calculate the z-score for 60 minutes as:

$$z = \frac{60 - 55}{5} = 1$$

The probability that a fuse from Company II lasts at least 1 hour is also approximately 0.1587.

Therefore, the correct answer is  $\boxed{\text{C}}$ . For both companies, the probability that a fuse will last at least 1 hour is 0.159. Final Answer: C

\*\*\*\*\*

Answer Option 9:

To determine which company's fuse is more likely to last at least 1 hour, we need to calculate the probability that a fuse from each company lasts at least 1 hour.

For Company I, we know that the average burn time is 50 minutes with a standard deviation of 10 minutes. To find the probability that a fuse lasts at least 1 hour (60 minutes), we can calculate the z-score:

$$z = \frac{60 - 50}{10} = 1$$

Using a standard normal distribution table, we find that the probability of a fuse from Company I lasting at least 1 hour is approximately 0.841.

For Company II, we know that the average burn time is 55 minutes with a standard deviation of 5 minutes. Again, we calculate the z-score:

$$z = \frac{60 - 55}{5} = 1$$

Using the same standard normal distribution table, we find that the probability of a fuse from Company II lasting at least 1 hour is also approximately 0.841.

Therefore, the correct answer is  $\boxed{\text{D}}$ . For both companies, the probability that a fuse will last at least 1 hour is 0.841. Final answer: D

\*\*\*\*\*

Question 122:

'In a random sample of Toyota car owners, 83 out of 112 said they were satisfied with the Toyota front-wheel drive, while in a similar survey of Subaru owners, 76 out of 81 said they were satisfied with the Subaru four-wheel drive. A 90% confidence interval estimate for the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is reported to be  $-0.197 \pm 0.081$ . Which is a proper conclusion?', A. 'The interval is invalid because probabilities cannot be negative.', B. 'The interval is invalid because it does not contain zero.', C. 'Subaru owners are approximately 19.7% more satisfied with their drive systems than are Toyota owners.', D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.'

=====

Answer Option 1:

A 90% confidence interval estimate for the difference in proportions is reported to be  $-0.197 \pm 0.081$ . This means that the true difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is estimated to be between -0.278 and -0.116 with 90% confidence. Therefore, the proper conclusion is: 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.' So the correct answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 2:

The proper conclusion is D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.'

A confidence interval estimates a range of plausible values for a population parameter, in this case, the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems. The given confidence interval of  $-0.197 \pm 0.081$  provides an estimate of the plausible range for this difference.

Since the confidence interval does not contain zero, we can conclude that there is a significant difference in the proportions of satisfied car owners between Toyota and Subaru. The negative difference (-0.197) indicates that a higher proportion of Subaru owners are satisfied with their drive systems compared to Toyota owners. The confidence level of 90% indicates our level of certainty in the estimated range.

\*\*\*\*\*

Answer Option 3:

The correct conclusion is D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.'

This conclusion is supported by the given confidence interval estimate of  $-0.197 \pm 0.081$ . The confidence interval spans from -0.278 to -0.116, indicating that the true difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is likely to fall within this range with 90% confidence.

\*\*\*\*\*

Answer Option 4:

The proper conclusion is D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.'

The confidence interval estimate for the difference in proportions,  $-0.197 \pm 0.081$ , tells us that we are 90% confident that the true difference in proportions between satisfied Toyota and Subaru owners is between -0.278 and -0.116. This means that the difference in proportions is negative, indicating that Subaru owners are more satisfied with their drive systems than Toyota owners. The interval does contain zero, so it is a valid interval. Therefore, option D is the proper conclusion. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The proper conclusion is D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.'

A negative value for the difference in proportions indicates that Subaru owners are more satisfied with their drive systems compared to Toyota owners. The confidence interval provides a range of plausible values for the true difference in proportions, and since the interval does not contain zero, we can conclude that there is a significant difference between the proportions of satisfied owners for the two car brands. The interval is valid because it falls within the range of possible values for the difference in proportions. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The proper conclusion is D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.'

This is because the confidence interval given ( $-0.197 \pm 0.081$ ) provides an estimate of the range within which the true difference in proportions lies with 90% confidence. In this case, the interval spans from -0.278 to -0.116, indicating that the Subaru owners are more satisfied with their drive systems than Toyota owners. The interval contains only negative values, which suggests that Subaru owners are indeed more satisfied than Toyota owners.

\*\*\*\*\*

Answer Option 7:

A 90% confidence interval estimate for the difference in proportions is reported to be  $-0.197 \pm 0.081$ . This means that the true difference in proportions between Toyota and Subaru car owners who are

satisfied with their drive systems lies within the range from -0.278 to -0.116 with 90% confidence. Since the interval does not contain zero, we can conclude that there is a significant difference in satisfaction between Toyota and Subaru car owners regarding their drive systems. Therefore, the proper conclusion is: D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.' Final Answer: D

\*\*\*\*\*

Answer Option 8:

The proper conclusion is D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.'

The confidence interval estimate given in the question is  $-0.197 \pm 0.081$ . This means that we are 90% confident that the true difference in proportions between satisfied Toyota and Subaru car owners is between  $-0.197 - 0.081 = -0.278$  and  $-0.197 + 0.081 = -0.116$ .

This conclusion is appropriate because it accurately represents the range of values within which the true difference in proportions is likely to fall with 90% confidence. Final answer: D

\*\*\*\*\*

Answer Option 9:

The correct conclusion is D. 'We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between -0.278 and -0.116.'

The given confidence interval is  $-0.197 \pm 0.081$ . This means that we are 90% confident that the true difference in proportions between satisfied Toyota and Subaru car owners lies between  $-0.197 - 0.081 = -0.278$  and  $-0.197 + 0.081 = -0.116$ .

Option A is incorrect because probabilities can be negative when comparing proportions.

Option B is incorrect because the interval does contain zero, as  $-0.197 - 0.081 = -0.278 < 0 < -0.197 + 0.081 = -0.116$ .

Option C is incorrect because the interval does not indicate a percentage difference, but rather a range of values for the difference in proportions.

Therefore, the correct conclusion is D. Final Answer: D

\*\*\*\*\*

Question 123:

'Which of the following statements about the correlation coefficient is true?', A. 'The correlation coefficient and the slope of the regression line may have opposite signs.', B. 'A correlation of 1 indicates a perfect cause-and-effect relationship between the variables.', C. 'Correlations of +0.87 and -0.87 indicate the same degree of clustering around the regression line.', D. 'Correlation applies equally well to quantitative and categorical data.'

=====

Answer Option 1:

The correct answer is A. 'The correlation coefficient and the slope of the regression line may have opposite signs.'

The correlation coefficient measures the strength and direction of a linear relationship between two variables. It ranges from -1 to 1. A positive correlation coefficient indicates a positive linear relationship, and a negative correlation coefficient indicates a negative linear relationship.

The slope of the regression line represents the change in the dependent variable for a one-unit increase in the independent variable. If the slope is positive, there is a positive relationship between the variables, and if the slope is negative, there is a negative relationship.

Therefore, it is possible for the correlation coefficient and the slope of the regression line to have opposite signs. This occurs when the relationship between the variables is linear but the direction of the relationship is opposite. Final answer: A

\*\*\*\*\*

Answer Option 2:

The correct statement about the correlation coefficient is A. 'The correlation coefficient and the slope of the regression line may have opposite signs.'

The correlation coefficient measures the strength and direction of the linear relationship between two quantitative variables. It ranges between -1 and 1. When the correlation coefficient is positive, it indicates a positive linear relationship, meaning that as one variable increases, the other variable tends to increase as well. When the correlation coefficient is negative, it indicates a negative linear relationship, meaning that as one variable increases, the other variable tends to decrease.

The slope of the regression line, on the other hand, represents the change in the response variable (dependent variable) for a one-unit increase in the predictor variable (independent variable). The slope can be positive or negative, depending on the direction of the relationship.

Therefore, it is possible for the correlation coefficient and the slope of the regression line to have opposite signs. This occurs when there is a negative linear relationship (correlation coefficient is negative) but a positive slope (increase in predictor variable leads to an increase in response variable on average) or vice versa.

Option A. 'The correlation coefficient and the slope of the regression line may have opposite signs.' is the correct statement. Final Answer: A

\*\*\*\*\*

Answer Option 3:

The correct statement about the correlation coefficient is: A. 'The correlation coefficient and the slope of the regression line may have opposite signs.'

The correlation coefficient measures the strength and direction of the linear relationship between two variables. It takes values between -1 and 1, where -1 indicates a perfect negative linear relationship, 1 indicates a perfect positive linear relationship, and 0 indicates no linear relationship.

The slope of the regression line represents the change in the dependent variable for a one-unit increase in the independent variable. The slope can be positive or negative.

Therefore, it is possible for the correlation coefficient and the slope of the regression line to have opposite signs. This occurs when there is a negative relationship between the variables (correlation coefficient close to -1) but a positive slope (increase in the dependent variable with an increase in the independent variable).

The other statements are not true. B is incorrect because correlation does not imply causation. C is incorrect because correlations of +0.87 and -0.87 indicate the same strength of linear relationship but opposite direction. D is incorrect because correlation is primarily used for quantitative data, although it can also be calculated for ordinal categorical data. The best answer is therefore A. Final answer: A

\*\*\*\*\*

Answer Option 4:

The correct statement about the correlation coefficient is: A. 'The correlation coefficient and the slope of the regression line may have opposite signs.'

This statement is true because the correlation coefficient measures the strength and direction of the linear relationship between two variables, while the slope of the regression line represents the change in the dependent variable for a one-unit increase in the independent variable. The sign of the correlation coefficient indicates the direction of the relationship (positive or negative), while the sign of the slope indicates the direction of change in the dependent variable. Therefore, it is possible for the correlation coefficient and the slope of the regression line to have opposite signs.

The other statements are incorrect:

B. 'A correlation of 1 indicates a perfect cause-and-effect relationship between the variables.' - A correlation of 1 indicates a perfect positive linear relationship, but it does not imply a cause-and-effect relationship.

C. 'Correlations of +0.87 and -0.87 indicate the same degree of clustering around the regression line.' - Correlations of +0.87 and -0.87 indicate the same strength of linear relationship, but they have opposite directions.

D. 'Correlation applies equally well to quantitative and categorical data.' - Correlation is mainly used to measure the linear relationship between two quantitative variables. It may not be appropriate or meaningful for categorical data.

Therefore, the correct answer is A.

\*\*\*\*\*

Answer Option 5:

The correct statement about the correlation coefficient is A. 'The correlation coefficient and the slope of the regression line may have opposite signs.'

The correlation coefficient measures the strength and direction of the linear relationship between two variables. It ranges from -1 to 1, where -1 indicates a perfect negative linear relationship, 1 indicates a perfect positive linear relationship, and 0 indicates no linear relationship.

The slope of the regression line represents the change in the dependent variable for a one-unit change in the independent variable. It can be positive or negative depending on the direction of the relationship. Therefore, it is possible for the correlation coefficient and the slope of the regression line have opposite signs. Final Answer: A.

\*\*\*\*\*

Answer Option 6:

The correct statement about the correlation coefficient is A. 'The correlation coefficient and the slope of the regression line may have opposite signs.'

The correlation coefficient measures the strength and direction of the linear relationship between two variables. It takes values between -1 and 1. A positive correlation coefficient indicates a positive linear relationship, while a negative correlation coefficient indicates a negative linear relationship.

However, the slope of the regression line represents the change in the dependent variable for a one-unit increase in the independent variable. It can have a different sign from the correlation coefficient because it also takes into account the scale of the variables.

Statement B is incorrect because a correlation coefficient of 1 indicates a perfect positive linear relationship, not necessarily a cause-and-effect relationship.

Statement C is incorrect because correlations of +0.87 and -0.87 have the same magnitude but opposite direction, indicating different degrees of clustering around the regression line.

Statement D is incorrect because correlation is primarily used to measure the relationship between quantitative variables, not categorical data.

Therefore, the correct statement is A. Final Answer: A

\*\*\*\*\*

Answer Option 7:

The correct answer is A. 'The correlation coefficient and the slope of the regression line may have opposite signs.'

The correlation coefficient measures the strength and direction of the linear relationship between two variables. It ranges between -1 and 1, where positive values indicate a positive linear relationship, negative values indicate a negative linear relationship, and zero indicates no linear relationship.

The slope of the regression line represents the change in the dependent variable for a one-unit change in the independent variable. It can be positive or negative, indicating the direction of the relationship between the variables.

In some cases, the correlation coefficient and the slope of the regression line may have opposite signs. This occurs when the relationship between the variables is nonlinear. In such cases, the correlation coefficient can still be meaningful in describing the strength of the relationship, even if the slope of the regression line has a different sign.

Therefore, statement A is true. The other statements are false. The correlation coefficient does not imply causation (statement B), correlations of +0.87 and -0.87 do not indicate the same degree of clustering around the regression line (statement C), and correlation is primarily used for quantitative data (statement D).

\*\*\*\*\*

Answer Option 8:

The correct statement about the correlation coefficient is C. 'Correlations of +0.87 and -0.87 indicate the same degree of clustering around the regression line.'

The correlation coefficient measures the strength and direction of the linear relationship between two variables. A correlation coefficient of +0.87 indicates a strong positive linear relationship, and a correlation coefficient of -0.87 indicates a strong negative linear relationship. Both values indicate a similar degree of clustering around the regression line, but in opposite directions. Therefore, statement C is true.

Statement A is false. The correlation coefficient and the slope of the regression line always have the same sign.

Statement B is false. A correlation coefficient of 1 indicates a perfect positive linear relationship, but it does not necessarily imply a cause-and-effect relationship between the variables.

Statement D is false. Correlation is primarily used to measure the linear relationship between two quantitative variables, and it may not be meaningful or applicable for categorical data.

Therefore, the correct statement is C. Final answer: C.

\*\*\*\*\*

Answer Option 9:

The correct statement about the correlation coefficient is A. 'The correlation coefficient and the slope of the regression line may have opposite signs.'

The correlation coefficient measures the strength and direction of the linear relationship between two variables. It ranges from -1 to 1, where -1 indicates a perfect negative linear relationship, 1 indicates a perfect positive linear relationship, and 0 indicates no linear relationship.

The slope of the regression line, on the other hand, represents the change in the dependent variable for a one-unit change in the independent variable. It can be positive or negative depending on the direction of the relationship.

Therefore, it is possible for the correlation coefficient and the slope of the regression line to have opposite signs. This occurs when the relationship between the variables is nonlinear or when there are outliers in the data.

Statement B is incorrect because a correlation of 1 does not necessarily imply a cause-and-effect relationship between the variables. Correlation measures association, not causation.

Statement C is incorrect because correlations of +0.87 and -0.87 do not indicate the same degree of clustering around the regression line. A correlation of +0.87 indicates a strong positive linear relationship, while a correlation of -0.87 indicates a strong negative linear relationship.

Statement D is incorrect because correlation is primarily used for quantitative data, where it measures the strength and direction of the linear relationship between two numerical variables. Correlation does not apply to categorical data in the same way.

\*\*\*\*\*



Question 124:

'A school board of a large school district is proposing a new dress code for students. Some students feel that this dress code unfairly targets female students. To see if there is a difference between boys and girls in their opposition to the new dress code, they conduct a poll of 60 randomly selected male and 70 randomly selected female high school students in the district. They find that 66 females oppose the dress code and 50 males oppose the dress code. Which of the following explains why a two-proportion z-test is not appropriate?', A. 'The sample sizes are different.', B. 'The sample sizes are too large.', C. 'The number of successes and the number of failures for the two groups are not all large enough.', D. 'The shapes of the population distributions are not known.'

=====

Answer Option 1:

The correct answer is C. 'The number of successes and the number of failures for the two groups are not all large enough.'

In order to use a two-proportion z-test, both groups being compared should have a sufficiently large number of successes and failures. The general rule of thumb is that both groups should have at least 10 successes and 10 failures. In this case, the number of successes and failures for both the male and female groups are not specified, so we cannot determine if the conditions for the test are met. Therefore, we cannot conclude that a two-proportion z-test is appropriate. Final answer: C.

\*\*\*\*\*

Answer Option 2:

The correct answer is C. 'The number of successes and the number of failures for the two groups are not all large enough.'

In order for a two-proportion z-test to be appropriate, the sample sizes for both groups should be large enough and the number of successes and failures for each group should also be large enough.

Specifically, the number of successes (students who oppose the dress code) and the number of failures (students who do not oppose the dress code) in each group should be at least 10.

In this scenario, the sample sizes for both male and female students are relatively large (60 and 70 respectively), but the number of male students who oppose the dress code is 50, which is not large enough to satisfy the requirement. Therefore, a two-proportion z-test is not appropriate.

\*\*\*\*\*

Answer Option 3:

The correct answer is A. 'The sample sizes are different.' A two-proportion z-test assumes that the sample sizes are the same for both groups. In this case, the sample size for males is 60 and the sample size for females is 70, so the sample sizes are different. Therefore, a two-proportion z-test is not appropriate in this situation. Final answer: A.

\*\*\*\*\*

Answer Option 4:

The correct answer is C. 'The number of successes and the number of failures for the two groups are not all large enough.' In order for a two-proportion z-test to be appropriate, the number of successes and failures in each group should be at least 10. In this case, the number of successes (opposing the dress code) for both males and females is less than 10. Therefore, a two-proportion z-test is not appropriate in this situation. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The correct answer is C. 'The number of successes and the number of failures for the two groups are not all large enough.'

To apply a two-proportion z-test, both groups should have a sufficient number of successes ( $\geq 10$ ) and failures ( $\geq 10$ ) to ensure that the normal approximation is valid. In this case, the number of successes and failures for each group are 66 (female opposition) and 4 (female support) for the females, and 50 (male opposition) and 10 (male support) for the males.

Since the number of successes (opposition) and failures (support) for the males are not all large enough, a two-proportion z-test is not appropriate.

Hence, the correct answer is C. Final answer: C

\*\*\*\*\*

Answer Option 6:

The correct answer is A. 'The sample sizes are different.'

A two-proportion z-test assumes that the two samples being compared have the same sample size. In this case, the sample size for male students (60) is different from the sample size for female students (70). Therefore, a two-proportion z-test is not appropriate in this situation. Final answer: A.

\*\*\*\*\*

Answer Option 7:

The correct answer is C. 'The number of successes and the number of failures for the two groups are not all large enough.'

In order for a two-proportion z-test to be appropriate, both the number of successes and the number of failures for each group should be at least 10. In this case, the number of successes for females is 66, which is greater than 10, but the number of successes for males is only 50, which is less than 10. Therefore, a two-proportion z-test is not appropriate in this situation. Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct answer is C. 'The number of successes and the number of failures for the two groups are not all large enough.'

In order to use a two-proportion z-test, the number of successes and the number of failures for each group should be large enough. Specifically, we need at least 10 successes and 10 failures for each group.

In this case, the number of successes for both males and females is below 10 (50 and 66 respectively), which means the conditions for a two-proportion z-test are not met. Therefore, option C is the correct explanation. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct answer is C. 'The number of successes and the number of failures for the two groups are not all large enough.'

In order to use a two-proportion z-test, we need the number of successes and the number of failures in each group to be large enough. Specifically, we need both  $np$  and  $n(1-p)$  to be greater than or equal to 10 for each group, where  $n$  is the sample size and  $p$  is the proportion of successes in each group.

In this case, the number of successes in the female group is 66, and the number of failures is  $70 - 66 = 4$ . However,  $np = 70 * (66/70) = 66$ , which is not greater than or equal to 10. Similarly, the number of successes in the male group is 50, and the number of failures is  $60 - 50 = 10$ . Again,  $np = 60 * (50/60) = 50$ , which is not greater than or equal to 10.

Therefore, the two-proportion z-test is not appropriate because the number of successes and the number of failures for the two groups are not all large enough. Answer choice C correctly explains this.

\*\*\*\*\*

Question 125:

'A soft drink dispenser can be adjusted to deliver any fixed number of ounces. If the machine is operating with a standard deviation in delivery equal to 0.3 ounce, what should be the mean setting so that a 12-ounce cup will overflow less than 1% of the time? Assume a normal distribution for ounces delivered.', A. '12 - 0.99(0.3) ounces', B. '12 - 2.326(0.3) ounces', C. '12 - 2.576(0.3) ounces', D. '12 + 2.326(0.3) ounces'

=====

Answer Option 1:

To ensure that a 12-ounce cup will overflow less than 1% of the time, we need to find the mean setting such that the probability of delivering more than 12 ounces is less than 1%.

Since the delivery follows a normal distribution, we can use z-scores to find the appropriate mean setting. The z-score formula is given by:

$$Z = \frac{x - \mu}{\sigma}$$

where  $x$  is the desired value (12 ounces),  $\mu$  is the mean setting, and  $\sigma$  is the standard deviation (0.3 ounces).

We want to find the z-score corresponding to a probability of less than 1%. Looking up the z-score in the normal distribution table, we find that the z-score for a probability of 1% is approximately -2.326.

Solving the z-score formula for  $\mu$ , we have:

$$-2.326 = \frac{12 - \mu}{0.3}$$

Simplifying, we find:

$$-2.326(0.3) = 12 - \mu$$

$$-0.6978 = 12 - \mu$$

$$-\mu = -12.6978$$

$$\mu = 12.6978$$

Therefore, the mean setting should be approximately 12 + 2.326(0.3) ounces, which gives us an answer of  $\boxed{12 + 2.326(0.3)}$ .

\*\*\*\*\*

Answer Option 2:

We want to find the mean setting  $\mu$  such that the probability that a 12-ounce cup overflows is less than 1%.

Let  $X$  be the random variable representing the ounces delivered by the machine. We are given that the standard deviation  $\sigma$  of  $X$  is 0.3 ounces.

We know that the probability of a 12-ounce cup overflowing is equivalent to the probability of  $X$  being greater than 12.

Using the properties of a normal distribution, we can convert this probability into a standard normal probability. We have:

$$P(X > 12) = P\left(Z > \frac{12 - \mu}{\sigma}\right)$$

We need to find the value of  $\mu$  such that  $P\left(Z > \frac{12 - \mu}{0.3}\right) < 0.01$ .

Looking up the value in the standard normal distribution table, we find that  $P(Z > 2.326) \approx 0.01$ .

Therefore, we have:

$$\frac{12 - \mu}{0.3} = 2.326$$

Solving for  $\mu$ , we get:

$$12 - \mu = 2.326(0.3)$$

$$12 - \mu = 0.6978$$

$$\mu = 12 - 0.6978$$

Hence, the mean setting should be approximately 12 - 0.6978 ounces, which is equivalent to  $\boxed{12 - 0.99(0.3)}$ . Final Answer: A

\*\*\*\*\*

Answer Option 3:

We want to find the mean setting such that a 12-ounce cup will overflow less than 1% of the time. Let  $X$  represent the number of ounces delivered by the machine. We know that  $X$  follows a normal distribution with mean  $\mu$  (the mean setting) and standard deviation  $\sigma = 0.3$ .

To find the mean setting  $\mu$ , we need to find the value such that  $P(X > 12) < 0.01$ . We can standardize the variable by subtracting the mean and dividing by the standard deviation. This gives us:

$$\frac{X - \mu}{\sigma} = \frac{X - \mu}{0.3} \sim N(0, 1)$$

Since we want to find  $P(X > 12) < 0.01$ , we can rewrite this as  $P\left(\frac{X - \mu}{0.3} > \frac{12 - \mu}{0.3}\right) < 0.01$ .

Using standard normal tables or a calculator, we can find that  $P\left(Z > \frac{12 - \mu}{0.3}\right) < 0.01$  corresponds to  $Z < -2.326$  (approximately).

Therefore, we have:

$$-2.326 < \frac{12 - \mu}{0.3}$$

Solving for  $\mu$ , we get:

$$12 - \mu > -2.326(0.3)$$

$$\mu < 12 + 2.326(0.3)$$

Thus, the mean setting should be  $\boxed{12 + 2.326(0.3)}$  ounces. Final Answer: D

\*\*\*\*\*

Answer Option 4:

To find the mean setting that will ensure that a 12-ounce cup will overflow less than 1% of the time, we need to find the cutoff point below which 99% of the distribution lies.

Since the standard deviation is 0.3 ounces, we need to find the z-score corresponding to the 1% cutoff point for a normal distribution.

Using a standard normal table or calculator, we find that the z-score corresponding to the 1% cutoff point is approximately 2.326.

The mean setting that ensures the 12-ounce cup will overflow less than 1% of the time is therefore  $12 - 2.326(0.3)$  ounces.

This simplifies to  $12 - 0.6978$  ounces, which is approximately equal to 11.3022 ounces.

Therefore, the correct answer is  $\boxed{\text{(B) } 12 - 2.326(0.3) \text{ ounces}}$ .

\*\*\*\*\*

Answer Option 5:

To find the mean setting for the soft drink dispenser, we need to determine the cutoff point at which a 12-ounce cup will overflow less than 1% of the time. This cutoff point corresponds to the z-score at which the area to the right under the normal distribution curve is 0.01.

Using a standard normal distribution table, we find that the z-score corresponding to an area of 0.01 to the right is approximately 2.326.

Given that the standard deviation is 0.3 ounces, we can calculate the mean setting as follows:

Mean = Desired Value - Z-Score \* Standard Deviation

Mean =  $12 - 2.326 * 0.3 = 12 - 0.6978$

Therefore, the mean setting that will ensure a 12-ounce cup overflows less than 1% of the time is approximately  $12 - 0.6978$  ounces.

The correct answer is A.  $12 - 0.99(0.3)$  ounces.

\*\*\*\*\*

Answer Option 6:

We want to find the mean setting such that a 12-ounce cup will overflow less than 1% of the time.

Let X be the random variable representing the ounces delivered by the machine. We want to find the mean setting such that  $P(X > 12) < 0.01$ .

We know that the standard deviation,  $\sigma$ , is 0.3 ounces. This means that the probability  $P(X > 12)$  can be calculated using the standard normal distribution.

Using the standard normal distribution table, we find that the z-score corresponding to a cumulative probability of 0.01 is approximately -2.326.

The formula for calculating the z-score is:

$$z = \frac{(X - \mu)}{\sigma}$$

Rearranging the formula, we have:

$$X = \mu + z\sigma$$

Substituting the given values, we have:

$$12 = \mu + (-2.326)(0.3)$$

$$12 = \mu - 0.6978$$

Solving for  $\mu$ , we find:

$$\mu = 12 + 0.6978$$

$$\mu = 12.6978$$

Therefore, the mean setting should be approximately 12.6978 ounces.

The closest answer choice is B. ' $12 - 2.326(0.3)$  ounces'.

Final Answer: B

\*\*\*\*\*

Answer Option 7:

We want to find the mean setting of the dispenser such that a 12-ounce cup will overflow less than 1% of the time.

Since ounces delivered follows a normal distribution, we can use the z-score formula to calculate the mean setting.

The z-score formula is given by:

$$z = \frac{(x - \mu)}{\sigma}$$

where z is the z-score, x is the value we are interested in (in this case, 12 ounces),  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

We want to find the mean setting such that the probability of overflowing is less than 1%. This is equivalent to finding the z-score such that the probability of z being greater than or equal to that z-score is less than 1%.

Using a standard normal distribution table, we find that the z-score corresponding to a probability of 1% is approximately 2.326.

Substituting the values into the z-score formula, we have:

$$2.326 = \frac{(12 - \mu)}{0.3}$$

Solving for  $\mu$ , we find:

$$12 - \mu = 2.326 \times 0.3$$

$$\mu = 12 - 2.326 \times 0.3$$

Therefore, the mean setting should be approximately  $\boxed{\text{(B) } 12 - 2.326(0.3) \text{ ounces}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

To find the mean setting of the soft drink dispenser, we need to determine the z-score corresponding to the 1% cutoff point. Since the normal distribution is symmetric, we can find the z-score for the lower tail and then subtract it from the mean to get the desired value.

Using a z-table or calculator, we find that the z-score associated with the 1% cutoff is approximately -2.326. The formula for the z-score is given by:

$$z = (x - \mu) / \sigma$$

where x is the value we want to find the z-score for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case, we want to find the mean setting ( $\mu$ ) that corresponds to a z-score of -2.326 when x is 12 ounces (the desired cup size) and  $\sigma$  is 0.3 ounces (the standard deviation). Rearranging the formula, we have:

$$-2.326 = (12 - \mu) / 0.3$$

Solving for  $\mu$ , we get:

$$12 - \mu = -2.326 * 0.3$$

$$\mu = 12 - (-2.326 * 0.3)$$

$$\mu = 12 + 0.6978$$

$$\mu = 12.6978$$

Therefore, the mean setting should be approximately 12.6978 ounces. To get the answer in the given choices, we can round it to the nearest tenth, resulting in 12.7 ounces. Thus, the correct answer is:  $\mu = 12.7$ , which is not among the given choices.

Since none of the given choices are correct, we can conclude that there is an error in the answer choices provided.

\*\*\*\*\*

Answer Option 9:

We want to find the mean setting  $\mu$  such that  $P(X > 12) < 0.01$ , where  $X$  follows a normal distribution with mean  $\mu$  and standard deviation 0.3 ounces.

Using the standard normal distribution (Z-score), we have  $P(X > 12) = P\left(Z > \frac{12 - \mu}{0.3}\right)$ .

To find the Z-score corresponding to a probability of 0.01, we can look up the value in the standard normal distribution table or use a calculator. The Z-score that corresponds to a probability of 0.01 is approximately 2.326.

Setting  $2.326 = \frac{12 - \mu}{0.3}$  and solving for  $\mu$ , we get  $\mu = 12 - 2.326(0.3)$ .

Therefore, the mean setting should be  $\boxed{\text{B. } 12 - 2.326(0.3) \text{ ounces}}$ . Final Answer: B

\*\*\*\*\*

Question 126:

The average number of calories in Yum-Yum Good candy bars is 210, with a standard deviation of 10. If the number of calories per candy bar is normally distributed, what percent of candy bars contain more than 225 calories?, A. '66.80%', B. '47.70%', C. '43.30%', D. '6.68%'

=====

Answer Option 1:

To find the percent of candy bars that contain more than 225 calories, we need to find the area under the normal distribution curve to the right of 225.

First, we need to calculate the z-score for 225 calories. The z-score formula is given by  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Plugging in the values, we get  $z = \frac{225 - 210}{10} = 1.5$ .

Next, we need to find the area to the right of the z-score of 1.5. Using a standard normal distribution table or a calculator, we find that the area to the right of 1.5 is approximately 0.0668, or 6.68%.

Therefore, the percent of candy bars that contain more than 225 calories is approximately  $\boxed{\text{6.68\%}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

To find the percent of candy bars that contain more than 225 calories, we need to find the area to the right of 225 on the normal distribution curve. We can standardize the value of 225 by using the formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we get  $z = \frac{225 - 210}{10} = 1.5$ .

Using a standard normal distribution table or calculator, we can find that the area to the left of 1.5 is approximately 0.9332. Therefore, the area to the right of 1.5 is  $1 - 0.9332 = 0.0668$ .

Converting this to a percentage, we get  $0.0668 \times 100 = 6.68\%$ . Therefore, the answer is  $\boxed{\text{(D) 6.68\%}}$ .

\*\*\*\*\*

Answer Option 3:

To find the percentage of candy bars that contain more than 225 calories, we can use the standard normal distribution. First, we calculate the z-score corresponding to 225 calories:

$$z = \frac{225 - 210}{10} = 1.5$$

Using a z-table or a calculator, we can find that the probability of a z-score greater than 1.5 is approximately 0.0668. This corresponds to 6.68%. Therefore, the percentage of candy bars that contain more than 225 calories is approximately  $\boxed{\text{6.68\%}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

We want to find the probability that a randomly selected candy bar contains more than 225 calories. We can use the standard normal distribution to solve this problem.

First, we standardize the value 225 using the formula:  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we get  $z = \frac{225 - 210}{10} = 1.5$ .

$$z = \frac{225 - 210}{10} = 1.5$$

Next, we find the probability that a standard normal random variable is greater than 1.5. Using a standard normal table or calculator, we find that this probability is approximately 0.0668.

Finally, we convert this probability to a percentage by multiplying by 100. The answer is therefore approximately 6.68%.

Therefore, the answer is  $\boxed{\text{(D) '6.68\%'}}$ .

\*\*\*\*\*

Answer Option 5:

To find the percentage of candy bars that contain more than 225 calories, we need to find the area under the normal distribution curve to the right of 225.

First, we standardize the value of 225 using the formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value we want to standardize,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case,  $x = 225$ ,  $\mu = 210$ , and  $\sigma = 10$ . Substituting these values into the formula, we get  $z = \frac{225 - 210}{10} = 1.5$ .

Next, we find the area to the right of  $z = 1.5$  under the standard normal curve. Using a standard normal distribution table or a calculator, we find that the area to the right of  $z = 1.5$  is approximately 0.0668.

This means that approximately 0.0668, or 6.68%, of candy bars contain more than 225 calories. Therefore, the answer is  $\boxed{\text{(D. '6.68\%'})}$ . Final Answer: D

\*\*\*\*\*

Answer Option 6:

To solve this problem, we need to standardize the value of 225 calories. The standardized value, also known as the z-score, is given by:

$$z = \frac{x - \mu}{\sigma}$$

Where  $x$  is the value we want to standardize,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we have:

$$z = \frac{225 - 210}{10} = \frac{15}{10} = 1.5$$

Now, we need to find the area under the normal distribution curve to the right of this z-score, which represents the percentage of candy bars that contain more than 225 calories. Using a standard normal distribution table or calculator, we find that the area to the left of  $z = 1.5$  is approximately 0.9332.

To find the area to the right of  $z = 1.5$ , we subtract this value from 1:

$$1 - 0.9332 = 0.0668$$

Converting this to a percentage, we get 6.68%. Therefore, the percentage of candy bars that contain more than 225 calories is approximately 6.68%.

The correct answer is  $\boxed{\text{(D. '6.68\%'})}$ .

\*\*\*\*\*

Answer Option 7:

To find the percentage of candy bars that contain more than 225 calories, we need to find the area under the normal distribution curve to the right of 225.

First, we calculate the z-score for 225 calories using the formula:

$$z = \frac{x - \mu}{\sigma} = \frac{225 - 210}{10} = 1.5$$

Next, we look up the area under the standard normal distribution curve to the right of a z-score of 1.5. Consulting a standard normal distribution table or using a calculator, we find that the area to the right of 1.5 is approximately 0.0668.

This means that approximately 6.68% of candy bars contain more than 225 calories.

Therefore, the answer is  $\boxed{\text{D, '6.68\%'}}$ .

\*\*\*\*\*

Answer Option 8:

To find the percentage of candy bars that contain more than 225 calories, we need to find the area to the right of 225 in the standard normal distribution.

First, we calculate the z-score for 225 calories using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we are interested in,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we get:

$$z = \frac{225 - 210}{10} = 1.5$$

Next, we look up the area to the right of  $z = 1.5$  in the standard normal distribution table or using a calculator. The area to the right of 1.5 is approximately 0.0668 or 6.68%.

Therefore, the percentage of candy bars that contain more than 225 calories is approximately 6.68%, which corresponds to answer choice D. Final Answer: D

\*\*\*\*\*

Answer Option 9:

To determine the percent of candy bars that contain more than 225 calories, we need to find the area under the normal distribution curve to the right of 225.

Let's calculate the z-score for 225 calories using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $\mu$  is the mean and  $\sigma$  is the standard deviation.

In this case,  $x = 225$ ,  $\mu = 210$ , and  $\sigma = 10$ .

$$z = \frac{225 - 210}{10} = 1.5$$

Using a standard normal distribution table or calculator, we can find the area to the right of  $z = 1.5$ .

The area to the right of  $z = 1.5$  is approximately 0.0668 or 6.68%.

Therefore, the percent of candy bars that contain more than 225 calories is approximately  $\boxed{6.68\%}$ .

Final Answer: D

\*\*\*\*\*

Question 127:

The mean daily demand for bread at a popular bakery is 2,500 loaves, with a standard deviation of 225 loaves. Every morning the bakery bakes 3,000 loaves. What is the probability that today it will run out of bread? Assume that the mean daily demand for bread at this bakery is normally distributed., A. '0.8333', B. '0.1667', C. '0.9869', D. '0.0132'

=====

Answer Option 1:

To find the probability that the bakery will run out of bread, we need to find the probability that the demand for bread exceeds the amount baked.

Since the mean daily demand is 2,500 loaves with a standard deviation of 225 loaves, we can use the standard normal distribution to find the probability.

First, we calculate the z-score of the amount baked:

$$z = (3000 - 2500) / 225 = 2.2222$$

Next, we look up the area under the standard normal curve to the right of  $z = 2.2222$ . Using a standard normal table or a calculator, we find that the area is approximately 0.0132.

Therefore, the probability that the bakery will run out of bread is approximately 0.0132 or 1.32%.

The answer is  $\boxed{\text{(D) } 0.0132}$ .

\*\*\*\*\*

Answer Option 2:

To find the probability that the bakery will run out of bread, we need to find the probability that the demand exceeds the supply of 3,000 loaves.

Let's calculate the z-score for a demand of 3,000 loaves:

$$z = \frac{3000 - 2500}{225} = \frac{500}{225} = \frac{20}{9} \approx 2.22$$

Using a standard normal distribution table or calculator, we can find that the probability of a z-score greater than 2.22 is approximately 0.0132.

Therefore, the probability that the bakery will run out of bread is approximately  $\boxed{0.0132}$ .

Final Answer: D

\*\*\*\*\*

Answer Option 3:

To solve this problem, we need to find the probability that the demand for bread exceeds 3,000 loaves.

First, we calculate the z-score for a demand of 3,000 loaves:

$$z = \frac{3000 - 2500}{225} = \frac{500}{225} \approx 2.22$$

Next, we look up the area under the standard normal curve to the right of  $z = 2.22$ . Using a standard normal distribution table or a calculator, we find that the area to the right of  $z = 2.22$  is approximately 0.0132.

Therefore, the probability that the bakery will run out of bread today is approximately 0.0132, or  $\boxed{\text{(D) } 0.0132}$ .

\*\*\*\*\*

Answer Option 4:

To find the probability that the bakery will run out of bread, we need to find the probability that the demand for bread exceeds the available supply of 3,000 loaves.

First, we need to calculate the z-score for a demand of 3,000 loaves. The z-score is given by the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we are interested in (3,000 loaves),  $\mu$  is the mean daily demand (2,500 loaves), and  $\sigma$  is the standard deviation (225 loaves).

Plugging in the values, we get:

$$z = \frac{3000 - 2500}{225} = \frac{500}{225} \approx 2.22$$

Next, we need to find the probability corresponding to this z-score using a standard normal distribution table or a calculator. Looking up the z-score of 2.22, we find that the probability is approximately 0.9869.

Therefore, the probability that the bakery will run out of bread is approximately  $\boxed{\text{(D) } 0.0132}$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

To solve this problem, we need to calculate the z-score of the demand for bread. The z-score is calculated as:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the demand for bread,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case, we want to find the z-score for the demand to be greater than 3,000 loaves, which is the amount the bakery bakes every morning. We use the formula:

$$z = \frac{3000 - 2500}{225}$$

Simplifying, we have:

$$z = \frac{500}{225} = 2.2222$$

We can then find the probability of the demand being greater than 3,000 loaves by looking up the z-score in the standard normal distribution table. From the table, we find that the probability of a z-score being greater than 2.2222 is approximately 0.0132.

Therefore, the probability that the bakery will run out of bread today is approximately 0.0132, or  $\boxed{\text{(D) } 0.0132}$ . Final Answer: D

\*\*\*\*\*

Answer Option 6:



To solve this problem, we need to calculate the z-score of the demand for bread on a given day.

The z-score formula is given by:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to find the z-score for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case, we want to find the z-score for a demand of 3,000 loaves, given that the mean is 2,500 loaves and the standard deviation is 225 loaves. Plugging in the values, we get:

$$z = \frac{3000 - 2500}{225} = \frac{500}{225} = 2.22$$

To find the probability that the bakery will run out of bread, we need to find the area under the normal curve to the right of the z-score. Using a standard normal distribution table or a calculator, we find that the probability is approximately 0.0132.

Therefore, the answer is  $\boxed{\text{D. } 0.0132}$ .

\*\*\*\*\*

Answer Option 7:

To solve this problem, we need to find the z-score for the demand of 3,000 loaves. The z-score is given by:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value of interest,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. In this case,  $x = 3,000$ ,  $\mu = 2,500$ , and  $\sigma = 225$ .

$$z = \frac{3000 - 2500}{225} = \frac{500}{225} = \frac{20}{9}$$

Next, we need to find the probability of the demand being less than or equal to 3,000 loaves. We can use the standard normal distribution table or a calculator to find this probability. Using a calculator, we find that the probability is approximately 0.8333.

Therefore, the probability that the bakery will run out of bread today is 1 minus the probability that the demand is less than or equal to 3,000 loaves. So, the probability is approximately  $1 - 0.8333 = 0.1667$ .

Therefore, the answer is  $\boxed{\text{B. } 0.1667}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

To solve this problem, we need to calculate the z-score for the demand that exceeds the bakery's supply. The z-score is calculated as follows:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the demand,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. In this case, the demand that exceeds the supply is 3,000 loaves, the mean is 2,500 loaves, and the standard deviation is 225 loaves.

Plugging these values into the formula, we get:

$$z = \frac{3000 - 2500}{225} = 2$$

Next, we need to find the probability that the z-score is greater than 2. We can use a standard normal distribution table or a calculator to find this probability. The probability of a z-score greater than 2 is approximately 0.0228. However, we need to find the probability of running out of bread, which is the complement of this probability. Therefore, the probability of running out of bread is approximately  $1 - 0.0228 = 0.9772$ .

Therefore, the answer is approximately 0.9772, which is closest to option C, '0.9869'. Final answer: C

\*\*\*\*\*

Answer Option 9:

To solve this problem, we need to calculate the z-score of the demand today. The z-score measures how many standard deviations away from the mean a particular value is. In this case, the mean daily demand is 2,500 loaves with a standard deviation of 225 loaves.

The z-score can be calculated using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we are interested in,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case, we are interested in the demand today, which is 3,000 loaves. Plugging in the values, we get:

$$z = \frac{3000 - 2500}{225} = \frac{500}{225} \approx 2.22$$

Next, we need to find the probability associated with this z-score. We can look up this value in a standard normal distribution table or use a calculator. The probability associated with a z-score of 2.22 is approximately 0.9869.

Therefore, the probability that the bakery will run out of bread today is approximately  $1 - 0.9869 = 0.0131$  or 1.31%.

The answer is option D: 0.0132.

\*\*\*\*\*

Question 128:

Two dice are rolled simultaneously. If both dice show 6, then the player wins \$20; otherwise the player loses the game. It costs \$2.00 to play the game. What is the expected gain or loss per game?', A. 'The player will gain about \$0.55.', B. 'The player will gain about \$1.44.', C. 'The player will lose about \$0.55.', D. 'The player will lose about \$1.44.'

=====

Answer Option 1:

There are 36 possible outcomes when rolling two dice, and only one of those outcomes results in both dice showing 6. Therefore, the probability of winning is  $\frac{1}{36}$ . The player wins \$20 and loses \$2, so the expected gain or loss per game is

$$\left(\frac{1}{36}\right)(20) + \left(\frac{35}{36}\right)(-2) = -\frac{70}{36} \approx -1.94.$$

This means the player is expected to lose about \$1.94 per game, so the answer is approximately  $\boxed{\text{(D)}}$  \text{The player will lose about \$1.44.}

\*\*\*\*\*

Answer Option 2:

The probability of rolling two dice and getting two sixes is  $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$ . In this case, the player wins \$20.00.

The probability of rolling any other outcome is  $1 - \frac{1}{36} = \frac{35}{36}$ . In this case, the player loses \$2.00.

So, the expected gain or loss per game is

$$\left(\frac{1}{36}\right)(20.00) + \left(\frac{35}{36}\right)(-2.00) = \frac{20.00 - 70.00}{36} = -\frac{50.00}{36} \approx -1.39.$$

Therefore, the player will lose about \$1.39 per game, which is approximately equal to  $\boxed{\text{(D)}}$  \text{The player will lose about \$1.44.}

\*\*\*\*\*

Answer Option 3:

The probability of rolling two sixes is  $\left(\frac{1}{6}\right)^2 = \frac{1}{36}$ . In this case, the player wins \$20.

The probability of not rolling two sixes is  $1 - \frac{1}{36} = \frac{35}{36}$ . In this case, the player loses \$2.

Therefore, the expected gain or loss per game is  $\left(\frac{1}{36}\right)(20) + \left(\frac{35}{36}\right)(-2) = \frac{20}{36} - \frac{70}{36} = -\frac{50}{36} \approx -1.39$ . This means that the player is expected to lose about \$1.39 per game.

Since the question asks for the expected gain or loss in terms of dollars, we can round \$1.39 to \$1.44. Therefore, the correct answer is  $\boxed{\text{(D)}}$  \text{The player will lose about \$1.44.}

\*\*\*\*\*

Answer Option 4:

The probability of rolling two dice and getting two 6s is  $\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}$ . So the player wins \$20 with a probability of  $\frac{1}{36}$  and loses \$2 with a probability of  $\frac{35}{36}$ .

The expected gain or loss per game is  $\left(\frac{1}{36}\right)(20) + \left(\frac{35}{36}\right)(-2) = \frac{20}{36} - \frac{70}{36} = -\frac{50}{36}$ . Simplifying, we find that the expected gain or loss per game is approximately \$-1.39.

Therefore, the player will lose about \$1.39 per game, and the answer is  $\boxed{\text{(D)}}$  \text{The player will lose about \$1.44.} Final Answer: D

\*\*\*\*\*

Answer Option 5:

The probability of rolling two sixes is  $\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}$ . The probability of losing the game is  $1 - \frac{1}{36} = \frac{35}{36}$ .

If the player wins, they gain  $20 - 2 = 18$ . If the player loses, they lose \$2. Therefore, the expected gain or loss per game is  $\left(\frac{1}{36}\right)(18) + \left(\frac{35}{36}\right)(-2) = -\frac{1}{2}$ . That means the player is expected to lose \$0.50 per game. Since the player loses \$2.00 to play the game, the player is expected to lose  $2.00 - 0.50 = 1.50$  per game.

Therefore, the expected gain or loss per game is  $\boxed{\text{(D)}}$  \text{The player will lose about \$1.44.} Final Answer: D

\*\*\*\*\*

Answer Option 6:

The probability of rolling two dice and getting two sixes is  $\left(\frac{1}{6}\right)^2 = \frac{1}{36}$ . The probability of not rolling two sixes is  $1 - \frac{1}{36} = \frac{35}{36}$ .

If the player wins, they receive \$20.00. If the player loses, they lose \$2.00.

The expected gain or loss per game can be calculated as:

$$\left(\frac{1}{36}\right)(20) + \left(\frac{35}{36}\right)(-2) = \frac{20}{36} - \frac{70}{36} = -\frac{50}{36} = -\frac{25}{18} \approx -1.39.$$

Therefore, the player is expected to lose about \$1.39 per game, which is closest to  $\boxed{\text{(D)}}$  \text{The player will lose about \$1.44.} Final Answer: D

\*\*\*\*\*

Answer Option 7:

The probability of rolling two sixes on two fair six-sided dice is  $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$ . The player wins \$20 if this event occurs, and loses \$2 if it does not occur. Therefore, the expected gain or loss per game is

$$\left(\frac{1}{36}\right)(20) + \left(\frac{35}{36}\right)(-2) = \frac{20}{36} - \frac{70}{36} = -\frac{50}{36}.$$

Simplifying, we find that the expected gain or loss per game is approximately \$-1.39. Since the player loses money on average, the answer is  $\boxed{\text{(D)}}$  \text{The player will lose about \$1.44.} Final

answer: D

\*\*\*\*\*

Answer Option 8:

To find the expected gain or loss per game, we need to calculate the expected value.

There are 36 possible outcomes when rolling two dice, and only 1 outcome results in a win (both dice showing 6). So the probability of winning is  $\frac{1}{36}$ , and the probability of losing is  $\frac{35}{36}$ .

If the player wins, they gain \$20, and if they lose, they lose \$2.

The expected value can be calculated as follows:

$$E = \left(\frac{1}{36}\right) \cdot 20 + \left(\frac{35}{36}\right) \cdot (-2) = \frac{20}{36} - \frac{70}{36} = -\frac{50}{36} \approx -1.39$$

Therefore, the expected gain or loss per game is approximately -\$1.39.

Since the player loses about \$1.39 per game, the answer is  $\boxed{\text{The player will lose about } \$1.44}$ .  $\text{Final Answer: D}$

\*\*\*\*\*

Answer Option 9:

There are 36 possible outcomes when two dice are rolled. Out of these, only one outcome results in a win of \$20, and the remaining 35 outcomes result in a loss of \$2 each.

The probability of winning is  $\frac{1}{36}$  and the probability of losing is  $\frac{35}{36}$ .

The expected gain or loss per game can be calculated as follows:

Expected gain or loss = (Probability of winning) \* (Gain from winning) + (Probability of losing) \* (Loss from losing)

$$\text{Expected gain or loss} = \left(\frac{1}{36}\right)(20) + \left(\frac{35}{36}\right)(-2) = \frac{20}{36} - \frac{70}{36} = -\frac{50}{36} = -\frac{25}{18}$$

Therefore, the player is expected to lose  $\frac{25}{18}$  dollars per game.

To convert this into a dollar amount, we multiply by \$2.00 to get  $-\frac{25}{18} \times 2.00 = -\frac{50}{18} \approx -2.7778$ .

Since the player is losing money, the expected loss per game is  $-\$2.7778$ .

Rounded to the nearest cent, the player will lose about  $-\$2.78$  per game.

Therefore, the correct answer is  $\boxed{\text{D. The player will lose about } \$1.44.}$

\*\*\*\*\*

Question 129:

'Suppose you toss a fair die three times and it comes up an even number each time. Which of the following is a true statement?', A. 'By the law of large numbers, the next toss is more likely to be an odd number than another even number.', B. 'Based on the properties of conditional probability the next toss is more likely to be an even number given that three in a row have been even.', C. 'Dice actually do have memories, and thus the number that comes up on the next toss will be influenced by the previous tosses.', D. 'The probability that the next toss will again be even is 0.5.'

=====

Answer Option 1:

The outcome of each individual die toss is independent of the previous tosses since we are assuming a fair die. Therefore, the previous tosses being even does not affect the probability of the next toss being even. So, option B is incorrect.

Similarly, option C is incorrect because dice do not have memories. The outcome of each toss is not influenced by the previous tosses.

Option D is also incorrect because the probability of rolling an even number on a fair die is  $\frac{1}{2}$ , but the fact that the previous tosses were even does not change this probability.

By process of elimination, the only remaining option is A. By the law of large numbers, the likelihood of getting an odd number on the next toss is higher because the previous tosses have all been even.

Therefore, the correct answer is A. Final answer: A

\*\*\*\*\*

Answer Option 2:

The outcomes of each toss of a fair die are independent events, meaning that the result of one toss does not affect the outcome of any other toss. Therefore, the previous tosses of the die being even do not influence the outcome of the next toss.

Since the die is fair, the probability of getting an even number on any single toss is  $\frac{1}{2}$ . This probability remains the same for each toss, regardless of the outcomes of previous tosses.

Therefore, the correct statement is  $\boxed{\text{D. 'The probability that the next toss will again be even is 0.5.'}}$

\*\*\*\*\*

Answer Option 3:

The outcome of each individual die toss is independent of the previous tosses. Tossing an even number three times in a row does not change the probability of getting an even or odd number on the next toss. Therefore, option A, B, and C are not true.

The probability of getting an even number on a single die toss is  $\frac{3}{6} = \frac{1}{2}$ . Since each toss is independent, the probability of getting an even number on the next toss is still  $\frac{1}{2}$ .

Therefore, the true statement is option D: 'The probability that the next toss will again be even is 0.5.' Final answer: D

\*\*\*\*\*

Answer Option 4:

The correct answer is B. Based on the properties of conditional probability, the next toss is more likely to be an even number given that three in a row have been even. Each toss of the die is independent of the previous tosses, so the probability of getting an even number on the next toss is still 1/2, regardless of the outcome of the previous three tosses. The law of large numbers does not apply in this case because it deals with the long-term behavior of independent random events, not one specific outcome based on previous outcomes. Dice do not have memories, so the outcome of the next toss is not influenced by the previous tosses. Therefore, C is incorrect. The probability that the next toss will again be even remains 1/2, so D is also incorrect. Final Answer: B

\*\*\*\*\*

Answer Option 5:

The outcome of each individual toss of a fair die is independent of the previous tosses. This means that the previous tosses have no influence on the outcome of the next toss. Therefore, statement C is incorrect.

Since each toss is independent, the probability of getting an even number on the next toss remains 1/2, regardless of the outcomes of the previous tosses. Hence, statement D is incorrect.

The law of large numbers states that as the number of trials (in this case, tosses) increases, the observed probability of an event (in this case, getting an even number) will approach the theoretical probability.

However, it does not imply that the next toss is more likely to be an odd number. Therefore, statement A is incorrect.

Statement B correctly states that based on the properties of conditional probability, the probability of getting an even number on the next toss, given that the previous three tosses were even, remains 1/2.

Each toss is independent, so the previous outcomes do not affect the probability of the next toss. Therefore, the correct statement is B.

Final Answer: B

\*\*\*\*\*

Answer Option 6:

The previous outcomes of tossing a fair die have no influence on the outcome of future tosses. Each toss of the die is an independent event, and the probability of obtaining an even number on each toss is always  $\frac{1}{2}$ . Therefore, the correct statement is  $\boxed{\text{D. 'The probability that the next toss will again be even is 0.5.'}}$

\*\*\*\*\*

Answer Option 7:

The correct answer is B. Based on the properties of conditional probability, the next toss is more likely to be an even number given that three in a row have been even. Each toss of a fair die is independent of the others, so the outcome of the previous tosses does not affect the probability of the next toss. Therefore, the probability that the next toss will be even is still  $\frac{1}{2}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

The outcome of each die toss is independent of the previous tosses because the die is fair. Therefore, the previous tosses (all resulting in even numbers) do not affect the outcome of the next toss. So, the probability of getting an even number on the next toss is still  $\frac{1}{2}$ . Thus, the true statement is  $\boxed{\text{The probability that the next toss will again be even is 0.5.}}$  (Choice D). Final Answer: D

\*\*\*\*\*

Answer Option 9:

The correct answer is B. 'Based on the properties of conditional probability the next toss is more likely to be an even number given that three in a row have been even.'

The probability of getting an even number on a fair die is  $1/2$ . The probability of getting three even numbers in a row is  $(1/2)^3 = 1/8$ .

Given that three even numbers have already been obtained, the sample space is reduced to the remaining 5 outcomes (1, 3, 5) on the die. Out of these 5 outcomes, only 2 are even numbers. Therefore, the probability of getting an even number on the next toss, given that three in a row have been even, is  $2/5$ .

Therefore, it is more likely that the next toss will be an even number, which is why statement B is true.

\*\*\*\*\*

Question 130:

Taxicabs in a metropolitan area are driven an average of 75,000 miles per year with a standard deviation of 12,000 miles. What is the probability that a randomly selected cab has been driven less than 100,000 miles if it is known that it has been driven over 80,000 miles? Assume a normal distribution of miles per year among cabs.', A. '0.06', B. '0.34', C. '0.66', D. '0.94'

=====

Answer Option 1:

To solve this problem, we can use the concept of standardized scores (z-scores).

First, we need to find the z-score for 100,000 miles using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to find the z-score for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For 100,000 miles:

$$z = \frac{100,000 - 75,000}{12,000} = 2.08$$

Next, we need to find the z-score for 80,000 miles:

$$z = \frac{80,000 - 75,000}{12,000} = 0.42$$

Now, we want to find the probability that a randomly selected cab has been driven less than 100,000 miles, given that it has been driven over 80,000 miles. This is equivalent to finding the area under the normal distribution curve between the z-scores of 0.42 and 2.08.

Using a standard normal distribution table or a calculator, we find that the probability is approximately 0.660.

Therefore, the answer is  $\boxed{C}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

We can standardize the values using the z-score formula:  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case, the mean is  $\mu = 75,000$  miles and the standard deviation is  $\sigma = 12,000$  miles. We want to find the probability that a randomly selected cab has been driven less than 100,000 miles given that it has been driven over 80,000 miles.

To find this probability, we need to find the area under the normal distribution curve to the left of  $x = 100,000$  miles, given that  $x > 80,000$  miles.

First, we standardize the values:

For  $x = 100,000$  miles:

$$z = \frac{100,000 - 75,000}{12,000} = 2.08$$

For  $x = 80,000$  miles:

$$z = \frac{80,000 - 75,000}{12,000} = 0.42$$

Next, we find the probability  $P(z < 2.08)$  and  $P(z > 0.42)$  using a standard normal distribution table or calculator.

$$P(z < 2.08) \approx 0.9821$$

$$P(z > 0.42) \approx 0.6631$$

To find the probability that a randomly selected cab has been driven less than 100,000 miles given that it has been driven over 80,000 miles, we subtract the probability that it has been driven between 80,000 and 100,000 miles from 1.

$$P(\text{less than } 100,000 \text{ miles} \mid \text{over } 80,000 \text{ miles}) = 1 - P(80,000 \leq x \leq 100,000 \text{ miles})$$

$$= 1 - (P(z \leq 2.08) - P(z \leq 0.42))$$

$$= 1 - (0.9821 - 0.6631)$$

$$= 0.681$$

Therefore, the probability that a randomly selected cab has been driven less than 100,000 miles given that it has been driven over 80,000 miles is approximately \$0.681\$. This is closest to  $\boxed{C}$  (0.66). Final Answer: C

\*\*\*\*\*

Answer Option 3:

We are given that a cab has been driven over 80,000 miles. We can calculate the z-score for this value using the formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the given values, we have  $z = \frac{80,000 - 75,000}{12,000} = 0.4167$ .

Next, we want to find the probability that a randomly selected cab has been driven less than 100,000 miles. We can calculate the z-score for this value as well:  $z = \frac{100,000 - 75,000}{12,000} = 2.0833$ .

To find the probability between these two z-scores, we can look up the values in the standard normal table or use a calculator. The area between these two z-scores represents the probability that a randomly selected cab has been driven between 80,000 and 100,000 miles.

Using the standard normal table or a calculator, we find that the area to the left of  $z = 0.4167$  is approximately 0.6628, and the area to the left of  $z = 2.0833$  is approximately 0.9819. Therefore, the probability that a randomly selected cab has been driven less than 100,000 miles, given that it has been driven over 80,000 miles, is  $0.9819 - 0.6628 = 0.3191$ .

The answer is not one of the given choices.

\*\*\*\*\*

Answer Option 4:

Let  $X$  be the number of miles driven by a randomly selected cab. We are given that  $X$  follows a normal distribution with mean  $\mu = 75,000$  and standard deviation  $\sigma = 12,000$ .

We want to find the probability that  $X < 100,000$  given that  $X > 80,000$ . In other words, we are looking for  $P(80,000 < X < 100,000)$ .

To find this probability, we can standardize the distribution by converting  $X$  to a standard normal random variable  $Z$  using the formula  $Z = \frac{X - \mu}{\sigma}$ .

For  $X = 80,000$ , we have  $Z = \frac{80,000 - 75,000}{12,000} = \frac{5,000}{12,000} = \frac{5}{12}$ .

For  $X = 100,000$ , we have  $Z = \frac{100,000 - 75,000}{12,000} = \frac{25,000}{12,000} = \frac{25}{12}$ .

Using a standard normal distribution table or a calculator, we can find the area under the standard normal curve between  $Z = \frac{5}{12}$  and  $Z = \frac{25}{12}$ . This probability represents the probability that  $X$  falls between 80,000 and 100,000.

The probability is approximately \$0.3409. Therefore, the answer is  $\boxed{\text{B}}$ . Final answer: B

\*\*\*\*\*

Answer Option 5:

To solve this problem, we need to use the concept of conditional probability. We are given that the cab has been driven over 80,000 miles. From the information given, we know that the mean is 75,000 miles, and the standard deviation is 12,000 miles. To find the probability that the cab has been driven less than 100,000 miles, given that it has been driven over 80,000 miles, we can use the standard normal distribution.

First, we need to find the z-score for 80,000 miles. The z-score is calculated as  $(x - \text{mean}) / \text{standard deviation}$ . In this case, it is  $(80,000 - 75,000) / 12,000 = 0.4167$ .

Next, we need to find the z-score for 100,000 miles. Again, the z-score is calculated as  $(x - \text{mean}) / \text{standard deviation}$ . In this case, it is  $(100,000 - 75,000) / 12,000 = 2.0833$ .

Using a standard normal distribution table, we can find the probability of a z-score less than 0.4167, which is approximately 0.6631. Similarly, we can find the probability of a z-score less than 2.0833, which is approximately 0.9821.

To find the conditional probability, we subtract the probability of the z-score less than 0.4167 from the probability of the z-score less than 2.0833. This gives us  $0.9821 - 0.6631 = 0.3190$ .

Therefore, the probability that a randomly selected cab has been driven less than 100,000 miles, given that it has been driven over 80,000 miles, is 0.3190.

The answer is not among the given choices, so there may be a mistake in the problem statement or answer choices.

\*\*\*\*\*

Answer Option 6:

To find the probability that a randomly selected cab has been driven less than 100,000 miles given that it has been driven over 80,000 miles, we can use the concept of conditional probability.

Let  $X$  be the random variable representing the number of miles driven by a randomly selected cab. We are given that  $X$  follows a normal distribution with a mean of 75,000 miles and a standard deviation of 12,000 miles.

We want to find  $P(X < 100,000 \mid X > 80,000)$ , which is equal to the conditional probability  $P(X < 100,000 \text{ and } X > 80,000)$  divided by  $P(X > 80,000)$ .

To find  $P(X < 100,000 \text{ and } X > 80,000)$ , we subtract the probability of the complement event from 1:

$$P(X < 100,000 \text{ and } X > 80,000) = 1 - P(X \leq 80,000 \text{ or } X \geq 100,000).$$

Since  $X$  follows a continuous distribution, we can use the complement rule  $(1 - P(A))$  to find  $P(X \leq 80,000 \text{ or } X \geq 100,000)$ :

$$P(X \leq 80,000 \text{ or } X \geq 100,000) = 1 - P(80,000 < X < 100,000).$$

To find  $P(80,000 < X < 100,000)$ , we standardize the values using the z-score formula:

$$P(80,000 < X < 100,000) = P\left(\frac{80,000 - 75,000}{12,000} < \frac{X - 75,000}{12,000} < \frac{100,000 - 75,000}{12,000}\right)$$

$$= P(0.4167 < Z < 2.0833).$$

Using a standard normal distribution table or calculator, we can find that  $P(0.4167 < Z < 2.0833)$  is approximately 0.3366.

$$\text{Therefore, } P(X < 100,000 \mid X > 80,000) = 1 - P(X \leq 80,000 \text{ or } X \geq 100,000) = 1 - 0.3366 = 0.6634.$$

The answer is not provided in the answer choices, but the closest option is C. '0.66'.

\*\*\*\*\*

Answer Option 7:

To solve this problem, we need to standardize the given values using the z-score formula. The z-score is calculated as follows:

$$Z = \frac{x - \mu}{\sigma}$$

where  $x$  is the observed value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case, we are given that the average is 75,000 miles and the standard deviation is 12,000 miles. We want to find the probability that a randomly selected cab has been driven less than 100,000 miles, given that it has been driven over 80,000 miles.

First, we calculate the z-score for the values of 80,000 miles and 100,000 miles:

$$Z_1 = \frac{80,000 - 75,000}{12,000} = 0.4167$$

$$Z_2 = \frac{100,000 - 75,000}{12,000} = 2.0833$$

Next, we use a standard normal distribution table or calculator to find the probabilities associated with these z-scores.

The probability that a randomly selected cab has been driven less than 100,000 miles is the area to the left of  $z = 2.0833$ , which corresponds to a probability of approximately 0.9816.

The probability that a randomly selected cab has been driven less than 80,000 miles is the area to the left of  $z = 0.4167$ , which corresponds to a probability of approximately 0.6628.

Finally, we calculate the conditional probability by subtracting the probability of being driven less than 80,000 miles from the probability of being driven less than 100,000 miles:

$$P(\text{less than 100,000 miles} \mid \text{over 80,000 miles}) = P(\text{less than 100,000 miles}) - P(\text{less than 80,000 miles})$$

$$= 0.9816 - 0.6628$$

$$= 0.3188$$

Therefore, the probability that a randomly selected cab has been driven less than 100,000 miles, given that it has been driven over 80,000 miles, is approximately 0.3188.

The answer is not one of the given options.

\*\*\*\*\*

Answer Option 8:

We can solve this problem by using the concept of standard deviation and z-scores. The z-score is calculated by subtracting the mean from the given value and then dividing by the standard deviation. In this case, the z-score is  $\frac{100000 - 75000}{12000} = 2.08$ .

Using a normal distribution table or calculator, we can find that the probability of a z-score less than 2.08 is approximately 0.9826. However, we need to find the probability that the cab has been driven less than 100,000 miles given that it has been driven over 80,000 miles.

Since we know that the cab has been driven over 80,000 miles, we need to find the probability that it has been driven less than 100,000 miles, but more than 80,000 miles. This can be calculated by finding the difference between the probabilities of the two z-scores:  $\approx 0.9826 - 0.8849 = 0.0977$ .

Therefore, the probability that a randomly selected cab has been driven less than 100,000 miles given that it has been driven over 80,000 miles is approximately \$0.0977. Final answer: None of the given choices.

\*\*\*\*\*

Answer Option 9:

To solve this problem, we will use the concept of conditional probability. We want to find the probability that the cab has been driven less than 100,000 miles given that it has been driven over 80,000 miles.

Let X be the random variable representing the number of miles driven by a randomly selected cab. We know that X follows a normal distribution with mean 75,000 miles and standard deviation 12,000 miles.

First, we can standardize the values of 80,000 miles and 100,000 miles using the formula  $z = (X - \mu) / \sigma$ , where  $\mu$  is the mean and  $\sigma$  is the standard deviation.

For 80,000 miles:

$$z = (80,000 - 75,000) / 12,000 = 0.4167$$

For 100,000 miles:

$$z = (100,000 - 75,000) / 12,000 = 2.0833$$

Now, we can find the probability of X being less than 100,000 miles given that it has been driven over 80,000 miles. This is equivalent to finding the probability that X is less than 100,000 miles within the range of  $z > 0.4167$ .

Using a standard normal distribution table or calculator, we can find that the probability of  $z < 0.4167$  is approximately 0.66.

Therefore, the probability that a randomly selected cab has been driven less than 100,000 miles given that it has been driven over 80,000 miles is approximately 0.66.

The correct answer is C. 0.66.

\*\*\*\*\*



Question 131:

'The registrar's office at a university has noticed that a large number of students fail to report a change of address. The registrar decides to take a random sample of 150 students from the current directory of students and determine the number of students with the correct addresses on record. He then uses this information to construct a 95 percent confidence interval. Which of the following statements must be true?', A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.', B. 'Ninety-five percent of students at this university have the correct address on record.', C. 'The true proportion of students at this university with the correct addresses on record is within the confidence interval 95 percent of the time.', D. 'The sample proportion of students at this university with the correct address on record is within the confidence interval.'

=====

Answer Option 1:

The correct answer is A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.'

A confidence interval is an interval estimate of a population parameter. In this case, the parameter of interest is the proportion of students at the university with the correct address on record. The confidence interval is constructed based on the sample proportion, and it provides a range of plausible values for the true proportion.

The confidence level, in this case 95 percent, refers to the long-run success rate of the method used to construct the interval. It means that if we were to repeat the sampling process and construct the confidence interval many times, approximately 95 percent of the intervals would contain the true proportion.

Therefore, it is accurate to say that the true proportion of students with the correct address on record is within the confidence interval constructed. This statement correctly reflects the interpretation of a confidence interval.

=====

Answer Option 2:

The correct answer is A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.'

A confidence interval is an interval estimate of a population parameter. In this case, the parameter of interest is the proportion of students at the university with the correct address on record. The confidence interval is constructed based on the sample data, and it provides an estimate of the range of values within which the true population proportion is likely to fall.

Therefore, statement A is correct because the true proportion of students with the correct address on record is indeed within the confidence interval constructed. The other statements (B, C, and D) are not necessarily true because they make generalizations about the entire student population based on the sample or the confidence interval, which may not always hold true.

=====

Answer Option 3:

The correct answer is A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.'

A confidence interval provides an estimate of the range in which we can reasonably expect the true population parameter (in this case, the proportion of students with correct addresses) to fall. The confidence interval is constructed based on the sample data and a chosen level of confidence (in this case, 95%). It does not guarantee that the true proportion is within the interval 95% of the time or that 95% of students have the correct address on record. The statement in option B is an incorrect interpretation of a confidence interval. Option C is incorrect because it confuses the interpretation of a confidence interval with the concept of repeated sampling. Option D is incorrect because the confidence interval is constructed for the population parameter, not the sample proportion.

=====

Answer Option 4:

The correct statement is A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.'

A confidence interval is a range of values that provides an estimate of an unknown population parameter. In this case, the unknown population parameter is the proportion of students with the correct address on record. The confidence interval constructed using the random sample provides a range of values within which the true proportion is likely to fall.

Statement B is not necessarily true because the confidence interval only provides an estimate for the true proportion, it does not guarantee that 95 percent of students have the correct address on record.

Statement C is not true because the confidence interval is a fixed range and does not change with repeated sampling. It provides a range of values that is likely to contain the true proportion, but it does not tell us anything about the probability of the true proportion falling within that range.

Statement D is also not true because the confidence interval is based on the sample proportion, not the other way around. The sample proportion is used to construct the confidence interval, but it is not guaranteed to be within the interval.

Therefore, the correct statement is A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.' Final Answer: A

=====

Answer Option 5:

The correct statement is A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.'

A confidence interval provides an estimate of the range within which the true population parameter (in this case, the proportion of students with correct addresses) is likely to fall. The 95 percent confidence interval means that if we were to take many random samples and construct confidence intervals for each sample, approximately 95 percent of those intervals would contain the true population proportion.

Therefore, we can say with 95 percent confidence that the true proportion of students with correct addresses is within the confidence interval constructed.

Statement B, 'Ninety-five percent of students at this university have the correct address on record,' is incorrect because the confidence interval is about the true proportion, not the individual students.

Statement C, 'The true proportion of students at this university with the correct addresses on record is within the confidence interval 95 percent of the time,' is incorrect because the confidence interval provides information about the likelihood that the interval contains the true proportion, not about the proportion of time the interval will contain the true proportion.

Statement D, 'The sample proportion of students at this university with the correct address on record is within the confidence interval,' is incorrect because the confidence interval is not about the sample proportion itself, but rather the true proportion in the population.

=====

Answer Option 6:

The correct answer is C. 'The true proportion of students at this university with the correct addresses on record is within the confidence interval 95 percent of the time.'

A confidence interval is a range of values that is likely to contain the true population parameter. In this case, the confidence interval is constructed based on the sample proportion of students with the correct addresses on record. The confidence interval provides an estimate of the true proportion of students with correct addresses, but it is not guaranteed to capture the true proportion with 100% certainty.

Statement A is incorrect because the true proportion may or may not be within the confidence interval.

Statement B is incorrect because we cannot determine the proportion of students with correct addresses based solely on the information given.

Statement D is incorrect because the confidence interval is constructed for the population parameter, not for the sample proportion.

\*\*\*\*\*

Answer Option 7:

The correct statement is option A: 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.'

A confidence interval is an estimate of an unknown population parameter (in this case, the proportion of students with the correct address on record) based on a sample from that population. The confidence interval gives a range of values within which the true population parameter is likely to fall.

Option B is incorrect because the confidence interval is a statement about the proportion of students in the sample, not the entire population.

Option C is incorrect because the confidence interval is a statement about the likelihood that the true population parameter falls within the interval, not about the interval itself.

Option D is incorrect because the confidence interval is a range of values for the population parameter, not the sample proportion itself.

\*\*\*\*\*

Answer Option 8:

The correct statement is A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.'

A confidence interval is a range estimate of a population parameter. In this case, the population parameter of interest is the proportion of students at the university with the correct address on record. The confidence interval is constructed based on the sample proportion (the number of students with the correct address on record in the sample divided by the total sample size) and the margin of error.

The confidence interval represents a range of values within which we can be confident, at a certain level (in this case, 95% confidence), that the true population proportion falls. Therefore, statement A is correct as it accurately reflects the interpretation and purpose of constructing a confidence interval.

\*\*\*\*\*

Answer Option 9:

The correct option is A. 'The true proportion of students at this university with the correct address on record is within the confidence interval constructed.'

A confidence interval is a range of values that is likely to contain the true population parameter with a certain level of confidence. In this case, the registrar is constructing a confidence interval for the proportion of students with the correct address on record.

Option B cannot be determined because the confidence interval is not a statement about the entire population, but only about the proportion of students in the sample.

Option C is not true because the confidence interval is constructed based on a single random sample, not multiple samples.

Option D is not true because the confidence interval is constructed for the population parameter (proportion of students) and not for the sample proportion. The sample proportion is a single value that is used to estimate the population proportion. The confidence interval provides a range of values within which the true population proportion is likely to fall.

\*\*\*\*\*

Question 132:

'A botanist is interested in testing  $H_0: \mu = 3.5$  cm versus  $H_a: \mu > 3.5$ , where  $\mu$  = the mean petal length of one variety of flowers. A random sample of 50 petals gives significant results at a 5 percent level of significance. Which of the following statements about the confidence interval to estimate the mean petal length is true?', A. 'The specified mean length of 3.5 cm is within a 90 percent confidence interval.', B. 'The specified mean length of 3.5 cm is not within a 90 percent confidence interval.', C. 'The specified mean length of 3.5 cm is below the lower limit of a 90 percent confidence interval.', D. 'The specified mean length of 3.5 cm is below the lower limit of a 95 percent confidence interval.'

=====

Answer Option 1:

Since the significance level is 5 percent, it means that the p-value is less than 0.05. This implies that the null hypothesis is rejected at a 5 percent level of significance. Therefore, the confidence interval will not include the null value of 3.5 cm. Thus, the specified mean length of 3.5 cm is not within a 90 percent confidence interval. Hence, the correct statement is  $\boxed{\text{B. "The specified mean length of 3.5 cm is not within a 90 percent confidence interval."}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 2:

The null hypothesis states that the mean petal length is equal to 3.5 cm. The alternative hypothesis states that the mean petal length is greater than 3.5 cm.

Since the test gives significant results at a 5% level of significance, we can reject the null hypothesis in favor of the alternative hypothesis. This means that there is evidence to suggest that the mean petal length is greater than 3.5 cm.

A confidence interval is used to estimate the true population mean with a certain level of confidence. In this case, the question asks about a 90% confidence interval.

If the specified mean length of 3.5 cm is within the 90% confidence interval, it would suggest that there is no evidence to suggest that the mean petal length is greater than 3.5 cm, which contradicts the significant results obtained in the hypothesis test.

Therefore, the correct statement is: "The specified mean length of 3.5 cm is not within a 90 percent confidence interval." Answer choice B is correct.

\*\*\*\*\*

Answer Option 3:

Since the test gives significant results at a 5 percent level of significance, we can reject the null hypothesis  $H_0: \mu = 3.5$  cm in favor of the alternative hypothesis  $H_a: \mu > 3.5$  cm. This means that the sample mean petal length is greater than 3.5 cm.

A confidence interval is used to estimate the population mean with a certain level of confidence. In this case, since we are interested in estimating the mean petal length, the confidence interval will provide a range of values within which we can be confident that the true mean falls.

The statement "The specified mean length of 3.5 cm is not within a 90 percent confidence interval" is the correct statement. This is because the rejection of the null hypothesis implies that the sample mean is greater than 3.5 cm, and therefore, the specified mean length of 3.5 cm is not within the confidence interval.

Therefore, the correct statement is B. 'The specified mean length of 3.5 cm is not within a 90 percent confidence interval.' Final answer: B

\*\*\*\*\*

Answer Option 4:

Since the null hypothesis is  $\mu = 3.5$  cm and the alternative hypothesis is  $\mu > 3.5$  cm, a significant result at a 5 percent level of significance suggests that the sample mean is significantly higher than 3.5 cm.

A confidence interval is constructed to estimate the population mean. If the specified mean length of 3.5 cm is within the confidence interval, it suggests that the null hypothesis cannot be rejected and the population mean could potentially be 3.5 cm. On the other hand, if the specified mean length is not within the confidence interval, it suggests that the null hypothesis is not plausible.

Since the significant result implies that the sample mean is significantly higher than 3.5 cm, it follows that the specified mean length of 3.5 cm is not within a 90 percent confidence interval. Therefore, the correct statement is: 'The specified mean length of 3.5 cm is not within a 90 percent confidence interval.' This corresponds to answer choice B. Final Answer: B

\*\*\*\*\*

Answer Option 5:

Since the sample gives significant results at a 5 percent level of significance, we reject the null hypothesis  $H_0: \mu = 3.5$  in favor of the alternative hypothesis  $H_a: \mu > 3.5$ . This means that the sample mean is significantly greater than 3.5.

A confidence interval is meant to estimate the population mean with a certain level of confidence. In this case, the confidence interval would be one-sided, with the lower limit being the sample mean and the upper limit being the upper bound of the interval. Since the sample mean is significantly greater than 3.5, it is not within the confidence interval. Therefore, the specified mean length of 3.5 cm is not within a 90 percent confidence interval.

Answer: B

\*\*\*\*\*

Answer Option 6:

The hypothesis test is testing whether the mean petal length is greater than 3.5 cm. If the null hypothesis is rejected at a 5 percent level of significance, it means that the sample provides strong evidence that the mean petal length is greater than 3.5 cm.

A confidence interval provides a range of values within which the true population mean is likely to fall. Since the hypothesis test favors the alternative hypothesis that the mean petal length is greater than 3.5 cm, the confidence interval will also reflect this.

Therefore, the specified mean length of 3.5 cm is not within a 90 percent confidence interval. The correct statement is: B. 'The specified mean length of 3.5 cm is not within a 90 percent confidence interval.'

\*\*\*\*\*

Answer Option 7:

Since the test gave significant results at a 5 percent level of significance, we can reject the null hypothesis in favor of the alternative hypothesis. This means that we have evidence to support the claim that the true mean petal length is greater than 3.5 cm.

For the confidence interval, since we are testing  $H_a: \mu > 3.5$ , we are interested in estimating the upper limit of the mean petal length. Since the test gave significant results, the upper limit of the confidence interval should be greater than 3.5 cm.

Therefore, the statement 'The specified mean length of 3.5 cm is not within a 90 percent confidence interval' is true.

The correct answer is B.

\*\*\*\*\*

Answer Option 8:

The statement "The specified mean length of 3.5 cm is within a 90 percent confidence interval" is not necessarily true. A confidence interval estimates the range of values within which the true population mean is likely to fall. The fact that the null hypothesis was rejected at a 5 percent level of significance does not provide information about the confidence interval. Additionally, the statement does not consider the direction of the alternative hypothesis.

Since the alternative hypothesis is one-tailed ( $\mu > 3.5$ ), the confidence interval will also be one-tailed. Thus, the lower limit of the confidence interval will be the value below which we can be confident that the true mean lies with a certain level of confidence. Therefore, the specified mean length of 3.5 cm is below the lower limit of the 90 percent confidence interval. Thus, the correct statement is: 'The specified mean length of 3.5 cm is below the lower limit of a 90 percent confidence interval.' Therefore, the answer is  $\boxed{\text{C}}$ .

\*\*\*\*\*

Answer Option 9:

Since the sample gives significant results at a 5 percent level of significance, we can conclude that the null hypothesis  $H_0: \mu = 3.5$  cm is rejected in favor of the alternative hypothesis  $H_a: \mu > 3.5$  cm. This means that the data provides evidence that the mean petal length is greater than 3.5 cm.

A 90 percent confidence interval is narrower than a 95 percent confidence interval. Therefore, since we have rejected the null hypothesis at a 5 percent level of significance, it means that the specified mean length of 3.5 cm is not within a 90 percent confidence interval. Hence, the correct option is  $\boxed{\text{B}}$ . 'The specified mean length of 3.5 cm is not within a 90 percent confidence interval.'

\*\*\*\*\*

Question 133:

'A dentist has noticed that about two children in every seven whom he sees professionally develop cavities before they turn 10 years old. Last week he examined the teeth of five unrelated children younger than 10. Let  $X$  be the number of children who develop cavities before turning 10. Which of the following gives the probability that at least one will develop a cavity before turning 10?'. A. ' $P(X = 2, 3, 4, 5, 6, 7)$ '. B. ' $P(X = 2 \text{ out of } 7)$ '. C. ' $P(X = 1)$ '. D. ' $1 - P(X = 0)$ '

=====

Answer Option 1:

The probability that a child does not develop a cavity before turning 10 is  $1 - \frac{2}{7} = \frac{5}{7}$ . Since the dentist examined five children, the probability that none of them develop a cavity before turning 10 is  $\left(\frac{5}{7}\right)^5$ . Therefore, the probability that at least one child will develop a cavity before turning 10 is  $1 - \left(\frac{5}{7}\right)^5$ , which is the same as  $\boxed{1 - P(X = 0)}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

The probability that a child does not develop a cavity before turning 10 is  $1 - \frac{2}{7} = \frac{5}{7}$ . Since the examinations of the five children are unrelated, the probability that none of the children develop a cavity before turning 10 is  $\left(\frac{5}{7}\right)^5$ . Therefore, the probability that at least one child will develop a cavity before turning 10 is  $1 - \left(\frac{5}{7}\right)^5$ . Hence, the answer is  $\boxed{1 - \text{text}(P)(X = 0)}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

The dentist notices that the probability that a child develops a cavity before turning 10 is  $\frac{2}{7}$ . Therefore, the probability that a child does not develop a cavity before turning 10 is  $1 - \frac{2}{7} = \frac{5}{7}$ .

Let's calculate the probability that none of the five children develop cavities before turning 10. Since the children are unrelated, we can calculate this probability by multiplying the probability that each child does not develop a cavity:

$$\left(\frac{5}{7}\right)^5$$

Therefore, the probability that at least one child will develop a cavity before turning 10 is  $1 - \left(\frac{5}{7}\right)^5$ , which corresponds to option D.

Hence, the correct answer is ' $1 - P(X = 0)$ '. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The probability that a child does not develop a cavity before turning 10 is  $1 - \frac{2}{7} = \frac{5}{7}$ . The probability that all five children do not develop cavities is  $\left(\frac{5}{7}\right)^5$ . Therefore, the probability that at least one child will develop a cavity is  $1 - \left(\frac{5}{7}\right)^5$ . The correct answer is  $\boxed{\text{text}(D) \cdot 1 - P(X = 0)}$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

To find the probability that at least one child will develop a cavity before turning 10, we can find the probability that none of the children will develop a cavity and subtract it from 1.

The probability that a child will not develop a cavity before turning 10 is  $\frac{5}{7}$  (since two children out of seven develop cavities). Therefore, the probability that all five children will not develop a cavity is  $\left(\frac{5}{7}\right)^5$ .

Thus, the probability that at least one child will develop a cavity before turning 10 is  $1 - \left(\frac{5}{7}\right)^5$ .

Therefore, the answer is  $\boxed{1 - P(X = 0)}$ . Final Answer: D

\*\*\*\*\*

Answer Option 6:

To find the probability that at least one child will develop a cavity before turning 10, we can find the probability that none of them develop cavities and subtract it from 1. The probability that a child does not develop a cavity is  $1 - \frac{2}{7} = \frac{5}{7}$ . Since the children are unrelated, the probability that none of the five children develop cavities is  $\left(\frac{5}{7}\right)^5$ . Therefore, the probability that at least one child will develop a cavity is  $1 - \left(\frac{5}{7}\right)^5 = \boxed{\text{text}(D) \cdot 1 - P(X = 0)}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

The probability that a child develops a cavity before turning 10 is  $\frac{2}{7}$ . Let's consider the probability that none of the five children develop a cavity before turning 10. For each child, the probability of not developing a cavity is  $\frac{5}{7}$ . Since the children are unrelated, we multiply the probabilities together to get  $\left(\frac{5}{7}\right)^5$ . Therefore, the probability that at least one child will develop a cavity before turning 10 is  $1 - \left(\frac{5}{7}\right)^5$ , which corresponds to option  $\boxed{\text{text}(D) \cdot 1 - P(X = 0)}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

The probability that a child does not develop a cavity before turning 10 is  $1 - \frac{2}{7} = \frac{5}{7}$ . Since the events are independent (we are assuming that one child's oral health does not affect another's), the probability that none of the five children develop a cavity before turning 10 is  $\left(\frac{5}{7}\right)^5$ .

Therefore, the probability that at least one child develops a cavity before turning 10 is  $1 - \left(\frac{5}{7}\right)^5$ . This matches answer choice  $\boxed{\text{text}(D) \cdot 1 - P(X = 0)}$ . Final answer: D

\*\*\*\*\*

Answer Option 9:

The probability that a child develops a cavity before turning 10 is  $\frac{2}{7}$ . To find the probability that at least one child develops a cavity out of the five children examined, we can find the probability that none of the children develop a cavity and subtract it from 1.

The probability that a child does not develop a cavity is  $1 - 2/7 = 5/7$ . Since the examinations are independent, the probability that all five children do not develop a cavity is  $(5/7)^5$ . Therefore, the probability that at least one child develops a cavity is  $1 - (5/7)^5$ .

Thus, the correct answer is  $\boxed{1 - P(X = 0)}$ . Final answer: D

\*\*\*\*\*

Question 134:

'Which of the following is NOT true of the  $\chi^2$  probability distribution function?', A. 'The area under the  $\chi^2$  curve is 1.', B. ' $\chi^2$  is defined only for nonnegative values of the variable.', C. 'For small degrees of freedom, the curve displays strong right-skewness.', D. 'For the same  $\alpha$ , as the number of degrees of freedom increases, the critical value for the rejection region decreases.'

=====

Answer Option 1:

The correct answer is D. 'For the same  $\alpha$ , as the number of degrees of freedom increases, the critical value for the rejection region decreases.'

In fact, for the  $\chi^2$  probability distribution function, as the number of degrees of freedom increases, the critical value for the rejection region increases. This means that it becomes more difficult to reject the null hypothesis as the degrees of freedom increase. Thus, option D is the statement that is NOT true about the  $\chi^2$  probability distribution function. Final answer: D

\*\*\*\*\*

Answer Option 2:

The correct answer is C. 'For small degrees of freedom, the curve displays strong right-skewness.'

The chi-squared probability distribution function is a right-skewed distribution. However, as the degrees of freedom increase, the chi-squared distribution approaches a normal distribution. Therefore, for large degrees of freedom, the curve becomes more symmetric and less skewed. This means that for small degrees of freedom, the curve displays strong right-skewness, but as the degrees of freedom increase, the curve becomes less skewed.

All the other statements are true. The area under the chi-squared curve is indeed 1, the chi-squared variable is defined only for nonnegative values, and for the same alpha level, as the number of degrees of freedom increases, the critical value for the rejection region decreases.

Therefore, the correct answer is C. 'For small degrees of freedom, the curve displays strong right-skewness.' Final answer: C

\*\*\*\*\*

Answer Option 3:

The correct answer is C. 'For small degrees of freedom, the curve displays strong right-skewness.'

The chi-square distribution is a right-skewed distribution, meaning it has a longer right tail. However, as the number of degrees of freedom increases, the shape of the chi-square distribution approaches a symmetrical shape.

Therefore, the statement in option C is NOT true, making it the correct answer. Final Answer: C

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'For the same  $\alpha$ , as the number of degrees of freedom increases, the critical value for the rejection region decreases.' This statement is NOT true.

For the chi-square probability distribution function, the area under the chi-square curve is indeed 1, so statement A is true. The chi-square variable is defined only for nonnegative values, so statement B is true as well. For small degrees of freedom, the chi-square distribution does display strong right-skewness, so statement C is also true.

However, for the same significance level  $\alpha$ , as the number of degrees of freedom increases, the critical value for the rejection region increases, not decreases. This is because as the number of degrees of freedom increases, the chi-square distribution approaches a normal distribution, and the critical value for the rejection region becomes larger to account for the increased variability. Therefore, statement D is NOT true.

So the answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The statement that is NOT true of the  $\chi^2$  probability distribution function is:

D. 'For the same  $\alpha$ , as the number of degrees of freedom increases, the critical value for the rejection region decreases.'

In fact, for the same  $\alpha$  (significance level), as the number of degrees of freedom increases, the critical value for the rejection region increases. This is because as the number of degrees of freedom increases, the  $\chi^2$  distribution approaches a normal distribution, and the critical values are further out in the tails of the distribution. Therefore, the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The correct answer is D. 'For the same  $\alpha$ , as the number of degrees of freedom increases, the critical value for the rejection region decreases.'

In the chi-square probability distribution function, the critical value for the rejection region increases as the number of degrees of freedom increases. This means that as the number of degrees of freedom increases, it becomes more difficult to reject the null hypothesis.

Option A is true because the area under the chi-square curve is always equal to 1.

Option B is true because the chi-square variable is defined only for nonnegative values.

Option C is true because for small degrees of freedom, the chi-square curve displays strong right-skewness.

Therefore, the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The  $\chi^2$  probability distribution function is a continuous probability distribution function that takes nonnegative values. Therefore, statement B is true.

Statement A is also true. The area under the  $\chi^2$  curve represents the total probability, and for any probability distribution function, the total probability should be equal to 1.

Statement C is true as well. For small degrees of freedom, the  $\chi^2$  distribution is highly right-skewed, meaning that it has a longer tail on the right side of the distribution.

However, statement D is NOT true. As the number of degrees of freedom increases, the critical value for the rejection region increases, not decreases. This is because as the degrees of freedom increase, the  $\chi^2$  distribution approaches a normal distribution. In a normal distribution, the critical values for the rejection region become larger as  $\alpha$  stays the same.

Therefore, the answer is  $\boxed{\text{D. 'For the same } \alpha, \text{ as the number of degrees of freedom increases, the critical value for the rejection region decreases.}'}$

\*\*\*\*\*

Answer Option 8:

The  $\chi^2$  probability distribution function (PDF) is a continuous probability distribution that is commonly used in statistics. It is defined for nonnegative values of the variable, so option B is true, not false.

Option A is true. The area under the  $\chi^2$  curve is indeed 1, as is the case for any probability distribution function.

Option C is true. For small degrees of freedom, the  $\chi^2$  distribution does display strong right-skewness. As the degrees of freedom increase, the distribution becomes more symmetrical.

Option D is true. As the number of degrees of freedom increases, the critical value for the rejection region decreases. This is because with a larger sample size, the distribution becomes more normal, leading to a smaller critical value necessary to reject the null hypothesis.

Therefore, the statement that is NOT true of the  $\chi^2$  probability distribution function is option B. Final answer: B

\*\*\*\*\*

Answer Option 9:

The chi-squared probability distribution function (PDF) is a continuous probability distribution defined only for nonnegative values of the variable. This means that option B, ' $\chi^2$  is defined only for nonnegative values of the variable', is true.

Option A, 'The area under the  $\chi^2$  curve is 1', is also true. The area under the chi-squared curve represents the total probability of the distribution, and since it is a probability distribution, the total probability must sum up to 1.

Option C, 'For small degrees of freedom, the curve displays strong right-skewness', is also true. The chi-squared distribution is skewed to the right, and as the degrees of freedom increase, the distribution becomes more symmetric.

Option D, 'For the same  $\alpha$ , as the number of degrees of freedom increases, the critical value for the rejection region decreases', is also true. As the degrees of freedom increase, the chi-squared distribution becomes less spread out, which means that the critical value for a given alpha level decreases.

Therefore, the answer is  $\boxed{\text{C. 'For small degrees of freedom, the curve displays strong right-skewness.'}}$  Final Answer: C

\*\*\*\*\*



Question 135:

The relation between the selling price of a car (in \$1,000) and its age (in years) is estimated from a random sample of cars of a specific model. The relation is given by the following formula: Selling price =  $15.9 - 0.983(\text{age})$  Which of the following can we conclude from this equation? A. 'For every year the car gets older, the selling price goes down by approximately 9.83 percent.', B. 'A new car costs on the average \$9,830.', C. 'For every year the car gets older, the selling price drops by approximately \$1,590.', D. 'For every year the car gets older, the selling price drops by approximately \$983.'

=====

Answer Option 1:

From the given formula, we can conclude that for every year the car gets older, the selling price drops by approximately \$983. Therefore, the correct answer is  $\boxed{\text{D}}$ . 'For every year the car gets older, the selling price drops by approximately \$983.'

\*\*\*\*\*

Answer Option 2:

The equation given is Selling price =  $15.9 - 0.983(\text{age})$ . From this equation, we can conclude that for every year the car gets older, the selling price drops by approximately \$983. This matches option D. Therefore, the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 3:

From the given equation, we can conclude that for every year the car gets older, the selling price drops by \$0.983 ( $0.983 \times 1000 = \$983$ ). Therefore, the correct statement is option D: 'For every year the car gets older, the selling price drops by approximately \$983.' Final answer: D

\*\*\*\*\*

Answer Option 4:

From the equation Selling price =  $15.9 - 0.983(\text{age})$ , we can conclude that for every year the car gets older, the selling price drops by approximately \$983. This corresponds to choice D. Final Answer: D

\*\*\*\*\*

Answer Option 5:

From the equation Selling price =  $15.9 - 0.983(\text{age})$ , we can conclude that for every year the car gets older, the selling price drops by approximately \$983. This corresponds to option D. Final answer: D

\*\*\*\*\*

Answer Option 6:

From the given equation, we can conclude that for every year the car gets older, the selling price drops by approximately  $\$0.983 \times 1,000 = \$983$ . Therefore, the correct statement is option D: 'For every year the car gets older, the selling price drops by approximately \$983.' Final Answer: D

\*\*\*\*\*

Answer Option 7:

From the given equation, we can conclude that for every year the car gets older, the selling price decreases by approximately \$983. This corresponds to answer choice  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

The equation given is Selling price =  $15.9 - 0.983(\text{age})$ .

From this equation, we can conclude that for every year the car gets older, the selling price drops by approximately \$983. This corresponds to choice D.

It is important to note that this equation does not provide information about the percentage decrease in the selling price or the average cost of a new car. Therefore, choices A, B, and C are incorrect.

Therefore, the answer is D.

\*\*\*\*\*

Answer Option 9:

From the given equation, we see that for every year the car gets older, the selling price decreases by 0.983, not 9.83 percent. Therefore, we can conclude that option A is incorrect.

The coefficient of the age term in the equation is -0.983, which means that for every year the car gets older, the selling price decreases by approximately \$983. Therefore, option D is correct.

Option B cannot be concluded from the equation because the average cost of a new car is not given. Option C cannot be concluded because the selling price drops by \$983, not \$1,590. Therefore, the only correct conclusion is option D. Final answer: D

\*\*\*\*\*

Question 136:

'Consider a hypothesis test with  $H_0: \mu = 70$  and  $H_a: \mu < 70$ . Which of the following choices of significance level and sample size results in the greatest power of the test when  $\mu = 65$ ?', A. ' $\alpha = 0.05$ ,  $n = 15$ ', B. ' $\alpha = 0.01$ ,  $n = 15$ ', C. ' $\alpha = 0.05$ ,  $n = 30$ ', D. ' $\alpha = 0.01$ ,  $n = 30$ '

=====

Answer Option 1:

To determine the power of the test, we need to consider the probability of rejecting the null hypothesis when the alternative hypothesis is true. In this case, the alternative hypothesis is that the population mean is less than 70.

The power of the test depends on the significance level, the sample size, and the effect size (the difference between the true population mean and the hypothesized mean under the alternative hypothesis).

Since the true population mean is 65, a smaller significance level and a larger sample size would result in a greater power. Therefore, we should choose the option with the smallest significance level and the largest sample size, which is option D:  $\alpha = 0.01$ ,  $n = 30$ .

Note that the significance level affects the probability of a Type I error (rejecting the null hypothesis when it is true), while the sample size affects the precision of the estimate and the variability of the test statistic. Final answer: D

\*\*\*\*\*

Answer Option 2:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the alternative hypothesis is  $\mu < 70$ , and we want to find the combination of significance level ( $\alpha$ ) and sample size ( $n$ ) that gives the greatest power when  $\mu = 65$ .

To maximize power, we should choose a smaller significance level and a larger sample size. This is because a smaller significance level reduces the chance of making a Type I error (rejecting the null hypothesis when it is true), and a larger sample size increases the ability to detect a difference between the sample mean and the hypothesized mean.

Therefore, the combination that results in the greatest power is ' $\alpha = 0.01$ ,  $n = 30$ '. This is because a significance level of 0.01 is smaller than 0.05, and a sample size of 30 is larger than 15. Thus, the answer is  $\boxed{\text{D. '}\alpha = 0.01, n = 30\text{'}}.$

\*\*\*\*\*

Answer Option 3:

To determine the greatest power of a hypothesis test, we need to consider both the significance level ( $\alpha$ ) and the sample size ( $n$ ).

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis ( $H_0$ ) when the alternative hypothesis ( $H_a$ ) is true. In this case,  $H_a: \mu < 70$  means that the true population mean is less than 70.

A lower significance level ( $\alpha$ ) increases the likelihood of making a Type II error (failing to reject  $H_0$  when it is false). Therefore, a higher significance level is preferred to increase the power of the test.

Additionally, a larger sample size ( $n$ ) increases the power of the test because it provides more information and reduces the variability of the sample mean.

Comparing the given choices, we can see that option D, ' $\alpha = 0.01$ ,  $n = 30$ ', has both a higher significance level and a larger sample size compared to the other options. Therefore, option D will result in the greatest power of the test.

Final Answer: D

\*\*\*\*\*

Answer Option 4:

To determine which choice of significance level and sample size results in the greatest power of the test, we need to consider the critical region and the probability of rejecting the null hypothesis when the alternative hypothesis is true.

Since  $H_a: \mu < 70$ , we are conducting a left-tailed test. For a lower-tailed test, the critical region is in the left tail of the sampling distribution.

A lower significance level  $\alpha$  corresponds to a smaller critical region and a higher probability of making a Type II error (failing to reject the null hypothesis when it is false). In this case, a Type II error would mean not detecting that the true mean is less than 70 when it is actually 65.

A larger sample size  $n$  leads to a smaller standard error and a more precise estimate of the population mean. This results in a higher power, as the test statistic will more likely fall in the critical region when the alternative hypothesis is true.

To maximize power, we should choose the smallest significance level  $\alpha$  and the largest sample size  $n$ .

Therefore, the choice that results in the greatest power is  $\alpha = 0.01$  and  $n = 30$ , which corresponds to option D. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the alternative hypothesis  $H_a: \mu < 70$  means that we are testing for a left-tailed test.

To maximize the power of the test, we want to minimize the type II error, which is the probability of failing to reject the null hypothesis when the alternative hypothesis is true. This means we want to minimize the probability of observing a sample mean greater than or equal to 70 when the true population mean is 65.

To achieve this, we want a smaller significance level  $\alpha$  and a larger sample size  $n$ . Intuitively, a smaller significance level allows us to have stronger evidence against the null hypothesis, and a larger sample size reduces the variability of the sample mean.

Among the given choices, option C with  $\alpha = 0.05$  and  $n = 30$  has the largest sample size, which helps reduce variability and increase the power. Therefore, the answer is C. Final answer: C

\*\*\*\*\*

Answer Option 6:

To determine which choice results in the greatest power of the test, we need to consider the significance level and the sample size.

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the alternative hypothesis is that the population mean is less than 70, and we want to maximize the power of the test when the true population mean is 65.

The power of the test depends on the difference between the true population mean and the hypothesized mean, as well as the variability of the data (standard deviation). However, in this case, we are not given the standard deviation or any other information about the variability of the data.

Therefore, we can only compare the significance level and sample size. Generally, a lower significance level ( $\alpha$ ) and a larger sample size ( $n$ ) result in a greater power of the test. This means that we should choose the option with the lower significance level and larger sample size.

Among the given choices, option D has the lowest significance level ( $\alpha = 0.01$ ) and the larger sample size ( $n = 30$ ). Therefore, the option ' $\alpha = 0.01$ ,  $n = 30$ ' results in the greatest power of the test. The answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 7:

To determine which choice of significance level and sample size results in the greatest power of the test, we need to consider the relationship between significance level, sample size, and power.

Power is the probability of correctly rejecting the null hypothesis when it is false. In this case, the null hypothesis is that the population mean ( $\mu$ ) is equal to 70, and the alternative hypothesis is that the population mean is less than 70.

The power of the test depends on the difference between the true population mean and the hypothesized mean, as well as the significance level and the sample size.

In this case, the true population mean is 65, which is less than the hypothesized mean of 70.

To maximize the power of the test, we want to choose the smallest significance level and the largest sample size. This is because a smaller significance level reduces the probability of a Type I error (incorrectly rejecting the null hypothesis), and a larger sample size increases the precision of the estimate and reduces the variability.

Among the given choices, the option with the smallest significance level and the largest sample size is ' $\alpha = 0.01$ ,  $n = 30$ '. Therefore, this choice will result in the greatest power of the test.

The answer is \boxed{D}.

\*\*\*\*\*

Answer Option 8:

To determine the test with the greatest power, we need to consider the significance level, sample size, and the alternative hypothesis.

A higher significance level means that we are more likely to reject the null hypothesis when it is true. Therefore, a higher significance level generally results in a higher power.

A larger sample size allows us to obtain more information about the population, making it easier to detect a difference from the null hypothesis. Hence, a larger sample size generally results in a higher power.

Since the alternative hypothesis is  $\mu < 70$ , a smaller population mean of 65 falls in favor of rejecting the null hypothesis. Therefore, a smaller population mean generally results in a higher power.

Comparing the choices:

For choices A and C, the significance level is the same at  $\alpha = 0.05$ . However, choice C has a larger sample size of  $n = 30$ , which generally leads to a higher power. Therefore, choice C is a better option.

For choices B and D, the sample size is the same at  $n = 15$ . However, choice D has a lower significance level of  $\alpha = 0.01$ , which generally leads to a higher power. Therefore, choice D is a better option.

Comparing choices C and D, both have a sample size of  $n = 30$ , but choice D has a lower significance level of  $\alpha = 0.01$ . Therefore, choice D results in the greatest power of the test.

Hence, the answer is choice D. Final answer: D

\*\*\*\*\*

Answer Option 9:

To determine the power of the test, we need to consider the probability of rejecting the null hypothesis when the alternative hypothesis is true. In this case, the alternative hypothesis is that the true mean is less than 70 ( $\mu < 70$ ).

The power of the test depends on the significance level ( $\alpha$ ), the sample size ( $n$ ), the true mean ( $\mu$ ), and the standard deviation ( $\sigma$ ) of the population.

Since the true mean is given as  $\mu = 65$ , we can assume that the population standard deviation is known and does not affect the power of the test.

Now let's compare the given choices:

For options A and B, the sample size is the same ( $n = 15$ ), but the significance level is different ( $\alpha = 0.05$  for A and  $\alpha = 0.01$  for B). The smaller the significance level, the higher the power of the test.

Therefore, option B ( $\alpha = 0.01$ ,  $n = 15$ ) has a higher power than option A.

For options C and D, the significance level is the same ( $\alpha = 0.05$ ), but the sample size is different ( $n = 30$  for C and  $n = 15$  for D). A larger sample size generally results in a higher power of the test. Therefore, option C ( $\alpha = 0.05$ ,  $n = 30$ ) has a higher power than option D.

Comparing the results, we find that option C ( $\alpha = 0.05$ ,  $n = 30$ ) has the greatest power of the test.

Therefore, the answer is option C. Final answer: C.

\*\*\*\*\*

Question 137:

'A farmer who raises hens for egg production wants his eggs to have a mean mass of 56 grams. He is considering the purchase of a different type of hen, so he took a random sample of 18 eggs laid by this type of hen. The distribution of the masses is symmetric and mound-shaped with a mean of 54.1 grams and no outliers. The farmer conducted a t-test to see if there is evidence that the eggs from these hens have a mean mass that is different from 56 g and got a test statistic of  $t = -1.973$ . If he uses a 5% significance level, which is the correct conclusion and reason?', A. 'Because  $t$  is more extreme than  $\pm 1.96$ , he should reject the null hypothesis. He has convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.', B. 'Because  $t$  is less extreme than the critical value of  $t$  for 17 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.', C. 'Because  $t$  is less extreme than the critical value of  $t$  for 18 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.', D. 'Because  $t$  is more extreme than the critical value of  $t$  for 18 degrees of freedom, he should reject the null hypothesis. He has convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.'

=====

Answer Option 1:

The critical value of  $t$  for a two-tailed test at the 5% significance level with 18 degrees of freedom is approximately  $\pm 2.101$ . Since  $-1.973$  falls between  $-2.101$  and  $2.101$ , we fail to reject the null hypothesis. Therefore, the correct conclusion is: 'Because  $t$  is less extreme than the critical value of  $t$  for 18 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.' Hence, the answer is  $\boxed{\text{C. 'Because } t \text{ is less extreme than the critical value of } t \text{ for 18 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5\% significance level that the mean mass of eggs from these hens is different from 56 grams.'}}$

\*\*\*\*\*

Answer Option 2:

To determine the correct conclusion, we compare the test statistic  $t$  to the critical value of  $t$  for the given significance level and degrees of freedom.

In this case, the test statistic is  $t = -1.973$  and the critical value of  $t$  for a 5% significance level with 18 degrees of freedom is approximately  $\pm 2.101$  (obtained from a t-table or calculator).

Since  $|-1.973| < 2.101$ , the test statistic does not exceed the critical value. Therefore, we fail to reject the null hypothesis.

The correct conclusion is: 'Because  $t$  is less extreme than the critical value of  $t$  for 18 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.'

The answer is  $\boxed{\text{C.}}$

\*\*\*\*\*

Answer Option 3:

The correct conclusion is C. 'Because  $t$  is less extreme than the critical value of  $t$  for 18 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.'

The farmer conducted a t-test to compare the sample mean (54.1 grams) to the hypothesized population mean (56 grams). The test statistic is  $t = -1.973$ . To determine whether to reject the null hypothesis (that the mean mass is equal to 56 grams), the farmer needs to compare the test statistic to the critical value of  $t$ .

At a 5% significance level and with  $(n-1)$  degrees of freedom ( $n$  = sample size), the critical value of  $t$  for an upper-tailed test is approximately 1.740 (calculated using statistical tables or software). Since the test statistic ( $-1.973$ ) is less extreme (more negative) than the critical value, the farmer does not have enough evidence to reject the null hypothesis. Therefore, he does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams. Final Answer: C

\*\*\*\*\*

Answer Option 4:

The correct answer is C. 'Because  $t$  is less extreme than the critical value of  $t$  for 18 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.'

To determine the correct conclusion, we need to compare the test statistic  $t$  with the critical value of  $t$  for the given significance level and degrees of freedom.

In this case, the significance level is 5% and the sample size is 18, which gives us  $18-1=17$  degrees of freedom.

The critical value of  $t$  for a two-tailed test at the 5% significance level with 17 degrees of freedom is approximately  $\pm 2.110$ .

Since the test statistic  $t$  ( $-1.973$ ) is less extreme than the critical value of  $t$  ( $-2.110$ ), we fail to reject the null hypothesis. This means that we do not have enough evidence to conclude that the mean mass of eggs from these hens is different from 56 grams at the 5% significance level. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The correct conclusion is B. 'Because  $t$  is less extreme than the critical value of  $t$  for 17 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.'

To make a conclusion in a t-test, we compare the test statistic with the critical value. The critical value for a two-tailed test with a 5% significance level and 17 degrees of freedom is approximately  $\pm 2.11$ . Since the test statistic  $t = -1.973$  is less extreme than the critical value of  $t$ , we fail to reject the null hypothesis. This means that we do not have convincing evidence to conclude that the mean mass of eggs from these hens is different from 56 grams at the 5% significance level.

\*\*\*\*\*

Answer Option 6:

The correct conclusion is C. 'Because  $t$  is less extreme than the critical value of  $t$  for 18 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.'

When conducting a t-test, we compare the calculated t statistic to the critical value of t for the given degrees of freedom (in this case, 18). If the calculated t statistic is less extreme (closer to 0) than the critical value, we fail to reject the null hypothesis. This means that we do not have enough evidence to conclude that the mean mass of eggs from these hens is different from 56 grams at the 5% significance level.

Final Answer: C

\*\*\*\*\*

Answer Option 7:

To determine the correct conclusion, we need to compare the test statistic  $t$  to the critical value of  $t$  for the given significance level and degrees of freedom.

The null hypothesis states that the mean mass of eggs from these hens is 56 grams. The alternative hypothesis is that the mean mass is different from 56 grams.

Since the test statistic  $t = -1.973$  is negative, we need to compare its absolute value to the critical value of  $t$  with 17 degrees of freedom.

For a two-tailed test at a 5% significance level with 17 degrees of freedom, the critical value of  $t$  is approximately  $\pm 2.110$ .

Since  $|-1.973| = 1.973 < 2.110$ , we do not reject the null hypothesis. This means that we do not have convincing evidence at the 5% significance level to suggest that the mean mass of eggs from these hens is different from 56 grams.

Therefore, the correct conclusion is:  $\boxed{\text{B. 'Because } t \text{ is less extreme than the critical value of } t \text{ for 17 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5\% significance level that the mean mass of eggs from these hens is different from 56 grams.'}}$

\*\*\*\*\*

Answer Option 8:

The null hypothesis in this case is that the mean mass of eggs from these hens is equal to 56 grams. The alternative hypothesis is that the mean mass is different from 56 grams.

To determine the correct conclusion, we compare the test statistic ( $t$ ) with the critical value of  $t$  for the given significance level and degrees of freedom.

Since the significance level is 5%, we need to find the critical value of  $t$  for a two-tailed test with 17 degrees of freedom. The critical value for a 5% significance level and 17 degrees of freedom is approximately  $\pm 2.110$ .

The test statistic,  $t = -1.973$ , is less extreme (in absolute value) than the critical value of  $t$ . Therefore, we fail to reject the null hypothesis. This means that we do not have convincing evidence at the 5% significance level to conclude that the mean mass of eggs from these hens is different from 56 grams.

Therefore, the correct conclusion is: 'Because  $t$  is less extreme than the critical value of  $t$  for 17 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5% significance level that the mean mass of eggs from these hens is different from 56 grams.' This corresponds to option C. Final answer: C.

\*\*\*\*\*

Answer Option 9:

The correct conclusion is  $\boxed{\text{C. 'Because } t \text{ is less extreme than the critical value of } t \text{ for 18 degrees of freedom, he should not reject the null hypothesis. He does not have convincing evidence at the 5\% significance level that the mean mass of eggs from these hens is different from 56 grams.'}}$

When conducting a t-test, we compare the test statistic ( $t$ -value) to the critical value. In this case, the test statistic is  $t = -1.973$ . To determine the critical value, we need to consider the significance level and the degrees of freedom. Since the sample size is 18, the degrees of freedom is  $18 - 1 = 17$ .

At a 5% significance level, the critical value for a two-tailed test with 17 degrees of freedom is approximately  $\pm 2.110$ . Since the test statistic ( $-1.973$ ) is less extreme (in absolute value) than the critical value, we fail to reject the null hypothesis. This means that we do not have convincing evidence at the 5% significance level to conclude that the mean mass of eggs from these hens is different from 56 grams.

\*\*\*\*\*

Question 138:

'A school committee member is lobbying for an increase in the gasoline tax to support the county school system. The local newspaper conducted a survey of county residents to assess their support for such an increase. What is the population of interest here?', A. 'All school-aged children', B. 'All county residents', C. 'All county residents with school-aged children', D. 'All county residents with children in the county school system'

=====

Answer Option 1:

The population of interest in this scenario is best described as 'All county residents'. This is because the survey is conducted to assess the support for an increase in the gasoline tax among all residents of the county, not just limited to specific groups such as school-aged children or parents of children in the county school system. Therefore, the correct answer is option B: 'All county residents'. Final Answer: B

\*\*\*\*\*

Answer Option 2:

The population of interest here would be all county residents, as the survey is conducted to assess the support of county residents for an increase in the gasoline tax to support the county school system. This includes all residents regardless of age or whether they have children in the school system. Therefore, the correct answer is  $\boxed{\text{B. 'All county residents'}}.$

\*\*\*\*\*

Answer Option 3:

The population of interest in this scenario is the group of people who are being surveyed, which is the target population from which the survey seeks to draw conclusions. In this case, the local newspaper conducted a survey of county residents to assess their support for an increase in the gasoline tax. Therefore, the population of interest is likely to be option B: 'All county residents'. Final Answer: B

\*\*\*\*\*

Answer Option 4:

The population of interest in this scenario is typically defined as the group that the survey results are meant to represent. In this case, the survey is conducted to assess the support for an increase in the gasoline tax to support the county school system. Therefore, the population of interest would most likely be option B, 'All county residents'. This is because the survey is meant to capture the opinions and support of all residents in the county, not just a specific subset such as school-aged children or parents with children in the county school system. Final Answer: B

\*\*\*\*\*

Answer Option 5:

The population of interest in this scenario is most likely 'All county residents' (Option B). The school committee member is lobbying for an increase in the gasoline tax to support the county school system, which would likely impact all county residents, regardless of whether they have school-aged children or children in the county school system. Therefore, the population of interest would include all county residents. Final Answer: B

\*\*\*\*\*

Answer Option 6:

The population of interest in this scenario is 'All county residents'. This is because the survey is conducted to assess the support of county residents for an increase in the gasoline tax to support the county school system. It is not limited to a specific group such as school-aged children or parents with children in the school system. Therefore, the answer is option B: 'All county residents'. Final answer: B.

\*\*\*\*\*

Answer Option 7:

The population of interest is the group of individuals for whom the survey is intended to represent. In this case, the survey is conducted to assess the support for an increase in the gasoline tax to support the county school system. Therefore, the population of interest would be the group of individuals who are directly affected by the county school system. This would include not only the school-aged children and their parents/guardians but also all county residents who may have an interest in the county school system, such as teachers, administrators, and other community members who care about the quality of education in the county.

Therefore, the correct answer is option B. 'All county residents'. Final Answer: B

\*\*\*\*\*

Answer Option 8:

The population of interest in this case is the group of people who are being surveyed to assess their support for an increase in the gasoline tax. Since the goal is to assess the support of county residents, the population of interest is therefore all county residents. Thus, the correct answer is  $\boxed{\text{B. 'All county residents'}}.$

\*\*\*\*\*

Answer Option 9:

The population of interest in this scenario should be D. 'All county residents with children in the county school system'. The committee member is lobbying for an increase in the gas tax to support the county school system, so the focus is on the residents who have children in the county school system. The survey aims to assess their support for the tax increase, making this specific group the population of interest. Final answer: D.

\*\*\*\*\*

Question 139:

"To survey the opinions of the students at your high school, a researcher plans to select every twenty-fifth student entering the school in the morning. Assuming there are no absences, will this result in a simple random sample of students attending your school?", A. "Yes, because every student has the same chance of being selected.", B. "Yes, but only if there is a single entrance to the school.", C. "Yes, because the 24 out of every 25 students who are not selected will form a control group.", D. "No, because not every sample of the intended size has an equal chance of being selected."

=====

Answer Option 1:

The correct answer is D. 'No, because not every sample of the intended size has an equal chance of being selected.'

In a simple random sample, every individual in the population has an equal chance of being selected. However, in this scenario, only every twenty-fifth student entering the school is selected. This means that students who enter the school earlier in the morning have a higher chance of being selected compared to students who enter later.

Therefore, this method of selection does not result in a simple random sample.

\*\*\*\*\*

Answer Option 2:

The correct answer is D. 'No, because not every sample of the intended size has an equal chance of being selected.'

In a simple random sample, every individual in the population should have an equal chance of being selected. However, in this scenario, only every twenty-fifth student entering the school in the morning is selected. This means that not all students have an equal chance of being selected, as only those who enter the school at specific times will be included in the sample. Therefore, this method does not result in a simple random sample. Final answer: D

\*\*\*\*\*

Answer Option 3:

The correct answer is D. 'No, because not every sample of the intended size has an equal chance of being selected.'

In a simple random sample, every possible sample of the intended size has an equal chance of being selected. However, in this case, only every twenty-fifth student entering the school in the morning is selected. This means that students who enter the school earlier have a higher chance of being selected, while students who enter later have a lower chance, or no chance at all, of being selected. Therefore, the selected sample is not a simple random sample.

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'No, because not every sample of the intended size has an equal chance of being selected.'

In a simple random sample, each individual in the population should have an equal chance of being selected. However, in this scenario, only every twenty-fifth student entering the school in the morning is selected. This means that students who enter the school earlier have a higher chance of being selected compared to students who enter later. Therefore, not every sample of the intended size has an equal chance of being selected.

\*\*\*\*\*

Answer Option 5:

The correct answer is D. 'No, because not every sample of the intended size has an equal chance of being selected.'

In a simple random sample, every possible sample of the intended size has an equal chance of being selected. However, in this case, only every twenty-fifth student entering the school in the morning is selected. This means that students who enter the school earlier have a higher chance of being selected compared to students who enter later in the morning. Therefore, not every sample of the intended size has an equal chance of being selected, indicating that this is not a simple random sample.

\*\*\*\*\*

Answer Option 6:

D. 'No, because not every sample of the intended size has an equal chance of being selected.'

In a simple random sample, every individual in the population should have an equal chance of being selected. However, in this case, only every twenty-fifth student entering the school in the morning is selected. The students who enter at other times or through other entrances are not included in the sample. Therefore, the selection method described does not result in a simple random sample. Final answer: D

\*\*\*\*\*

Answer Option 7:

The correct answer is D. 'No, because not every sample of the intended size has an equal chance of being selected.'

In a simple random sample, every individual in the population should have an equal chance of being selected. In this case, only every twenty-fifth student entering the school in the morning is selected. This means that students entering earlier in the morning have a higher chance of being selected compared to those entering later in the morning. Therefore, not every sample of the intended size has an equal chance of being selected, making it not a simple random sample. Final answer: D

\*\*\*\*\*

Answer Option 8:

The answer is D. No, because not every sample of the intended size has an equal chance of being selected. In this case, only students entering the school in the morning are selected, while students who enter the school at other times are not included in the sample. Additionally, the selection process does not ensure that every student has an equal chance of being selected, as only every twenty-fifth student is chosen. Therefore, this sampling method does not result in a simple random sample. Final answer: D

\*\*\*\*\*

Answer Option 9:

The correct answer is D. 'No, because not every sample of the intended size has an equal chance of being selected.' In a simple random sample, every individual in the population should have an equal chance of being selected. However, in this scenario, only every twenty-fifth student entering the school in the morning is selected. Therefore, students who enter the school at different times, such as those who arrive later in the morning, have no chance of being selected. This introduces bias into the sample and makes it not representative of the entire student population. Final Answer: D

\*\*\*\*\*



Question 140:

'Which of the following is a true statement?', A. 'While properly designed experiments can strongly suggest cause-and-effect relationships, a complete census is the only way of establishing such a relationship.', B. 'If properly designed, observational studies can establish cause-and-effect relationships just as strongly as properly designed experiments.', C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.', D. 'A useful approach to overcome bias in observational studies is to increase the sample size.'

=====

Answer Option 1:

The correct statement is D. 'A useful approach to overcome bias in observational studies is to increase the sample size.' Increasing the sample size in observational studies can help reduce the impact of bias and provide more reliable and accurate results. Final Answer: D

\*\*\*\*\*

Answer Option 2:

The correct answer is C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.' Controlled experiments are often conducted after initial observational studies have suggested a potential cause-and-effect relationship. This allows researchers to further investigate and establish a more robust causality through controlled conditions.

Final answer: C

\*\*\*\*\*

Answer Option 3:

The correct answer is C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.' Observational studies can provide evidence of a potential association between variables, but they cannot establish causation. Controlled experiments are then conducted to further investigate and establish cause-and-effect relationships. Final Answer: C

\*\*\*\*\*

Answer Option 4:

The correct statement is: C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.'

Observational studies can provide valuable insights and generate hypotheses about potential cause-and-effect relationships. However, they cannot establish causal relationships with certainty due to potential confounding factors and biases. Controlled experiments are typically conducted later to further investigate and establish cause-and-effect relationships suggested by observational studies. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The correct answer is C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.' This statement is true because observational studies can suggest potential cause-and-effect relationships, but they cannot establish causation definitively. Controlled experiments are designed to manipulate variables and establish cause-and-effect relationships more rigorously. Therefore, controlled experiments are often conducted after observational studies to further investigate and confirm the initial findings. Final Answer: C

\*\*\*\*\*

Answer Option 6:

Among the given statements, the most accurate and true statement is option D. 'A useful approach to overcome bias in observational studies is to increase the sample size.'

Increasing the sample size in observational studies can help reduce bias and increase the precision of the estimates. A larger sample size provides more representative data and reduces the impact of random variation. While increasing the sample size cannot completely eliminate bias, it is a practical and effective way to improve the validity and reliability of observational studies. Therefore, option D is the correct answer. Final answer: D

\*\*\*\*\*

Answer Option 7:

The correct option is C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.'

Option A is incorrect because a complete census is not the only way to establish cause-and-effect relationships. Properly designed experiments can also provide strong evidence for cause-and-effect relationships.

Option B is incorrect because while properly designed observational studies can provide valuable insights and suggest cause-and-effect relationships, they cannot establish them as strongly as properly designed experiments due to potential confounding factors and lack of control over variables.

Option D is incorrect because increasing the sample size alone does not directly overcome bias in observational studies. While a larger sample size can help improve the precision of estimates, it does not address underlying biases caused by confounding variables or other sources of error.

Therefore, the correct option is C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.' Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct answer is D. 'A useful approach to overcome bias in observational studies is to increase the sample size.'

Increasing the sample size in observational studies can help to reduce the impact of bias and increase the precision of estimates. While it cannot completely eliminate bias, a larger sample size can help to provide more reliable results and reduce the influence of random variation. This is a commonly used approach in observational studies to improve the validity and generalizability of the findings.

Option A is incorrect because a complete census is not the only way to establish cause-and-effect relationships. Controlled experiments can also be used to establish such relationships.

Option B is incorrect because observational studies, even if properly designed, cannot establish cause-and-effect relationships as strongly as properly designed experiments. Experiments have the advantage of random assignment and controlled manipulation, which allows for stronger causal inferences.

Option C is incorrect because controlled experiments are typically conducted before observational studies in order to establish cause-and-effect relationships. Observational studies are often used when conducting experiments is not feasible or ethical.

\*\*\*\*\*

Answer Option 9:

The correct statement is C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.'

Option A is incorrect because while properly designed experiments can provide strong evidence for cause-and-effect relationships, a complete census is not the only way to establish such relationships. Other rigorous study designs, such as randomized controlled trials, can also establish cause-and-effect relationships.

Option B is incorrect because observational studies, by their nature, cannot establish cause-and-effect relationships as strongly as properly designed experiments. Observational studies can provide valuable insights and suggest possible relationships, but they cannot provide the same level of control and eliminate confounding factors as experiments.

Option D is incorrect because increasing the sample size may help reduce random error and improve the precision of estimates in observational studies, but it does not directly address or overcome bias. To overcome bias in observational studies, researchers need to carefully control for potential confounding factors and use appropriate statistical methods.

Therefore, the correct statement is C. 'Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.' Final Answer: C.

\*\*\*\*\*

Question 141:

'An automobile service station performs oil changes and tire replacements, as well as other services. Sixty percent of its customers request an oil change, 30 percent request tire replacements, and 10 percent request both. A customer requests an oil change. What is the probability this customer does not request tire replacements?', A. '0.42', B. '0.5', C. '0.7', D. '0.833'

=====

Answer Option 1:

Let's assume that there are 100 customers in total.

Since 60% of the customers request an oil change, there are 60 customers who request an oil change.

Since 30% of the customers request tire replacements, there are 30 customers who request tire replacements.

Since 10% request both, there are 10 customers who request both an oil change and tire replacements.

Therefore, the number of customers who request only an oil change is  $60 - 10 = 50$ .

The probability that a customer who requests an oil change does not request tire replacements is given by the number of customers who request only an oil change divided by the total number of customers who request an oil change. In this case, it is  $50/60 = 5/6$ .

Therefore, the answer is  $\boxed{\text{D. } 0.833}$ .

\*\*\*\*\*

Answer Option 2:

Let's assume there are \$100\$ customers in total. \$60\$ of them request an oil change, \$30\$ request tire replacements, and \$10\$ request both.

If a customer requests an oil change, this means they belong to the group of \$60\$ customers who request an oil change. Out of these \$60\$ customers, \$10\$ also request tire replacements. Thus, out of the \$60\$ customers who request an oil change, \$60 - 10 = 50\$ do not request tire replacements.

Therefore, the probability that a customer who requested an oil change does not request tire replacements is  $\frac{50}{60} = \boxed{0.833}$ . Final answer: D

\*\*\*\*\*

Answer Option 3:

Let's assume that the total number of customers is 100. 60% of the customers request an oil change, which means that 60 customers request an oil change. 10% of the customers request both, which means that 10 customers request both an oil change and a tire replacement.

Therefore, the number of customers who only request an oil change is  $60 - 10 = 50$ .

The percentage of customers who only request an oil change is  $50/100 = 50\%$ .

Therefore, the probability that a customer who requests an oil change does not request a tire replacement is  $50/100 = 0.5$ .

Hence, the answer is  $\boxed{\text{B. } 0.5}$ . Final Answer: B

\*\*\*\*\*

Answer Option 4:

Let's assume that there are 100 customers in total. 60% of them request an oil change, which is 60 customers. 30% of them request tire replacements, which is 30 customers. 10% of them request both, which is 10 customers.

If a customer requests an oil change, there are 60 customers in total who requested an oil change, out of which 10 customers requested both. Therefore, there are 50 customers who requested only an oil change.

The probability that a customer who requested an oil change does not request tire replacements is 50 customers who requested only an oil change divided by 60 customers who requested an oil change. This is equal to  $5/6$ , which is equal to  $\boxed{0.833}$ .

Final Answer: D

\*\*\*\*\*

Answer Option 5:

Let  $O$  be the event that a customer requests an oil change and  $T$  be the event that a customer requests a tire replacement. We want to find the probability of  $O$  given that  $\neg T$  (not  $T$ ) occurs.

We can use conditional probability:

$\begin{aligned} P(O \mid \neg T) &= \frac{P(O \cap \neg T)}{P(\neg T)} \\ &= \frac{P(O) - P(O \cap T)}{1 - P(T)} \\ &= \frac{0.6 - 0.1}{1 - 0.3} \\ &= \frac{0.5}{0.7} \\ &= \boxed{0.7142857142857143}. \end{aligned}$

$\end{aligned}$

Final Answer: Approximately 0.714.

\*\*\*\*\*

Answer Option 6:

Let's assume that there are 100 customers in total. Given that 60% of customers request oil changes, there are 60 customers who request an oil change. Similarly, given that 30% of customers request tire replacements, there are 30 customers who request tire replacements. Since 10% of customers request both an oil change and tire replacements, there are 10 customers who request both.

If a customer requests an oil change, they can either request only an oil change or request both an oil change and tire replacements. Therefore, out of the 60 customers who request an oil change, 10 of them also request tire replacements. This means that there are 50 customers who request only an oil change.

The probability that a customer who requests an oil change does not request tire replacements is therefore  $50/60 = 5/6$ .

Thus, the answer is  $\boxed{\text{(D) } 0.833}$ .

\*\*\*\*\*

Answer Option 7:

Let's assume the total number of customers is 100 (for ease of calculation).

Since 60% of the customers request an oil change, there are 60 customers who request an oil change.

Similarly, 30% of the customers request tire replacements, so there are 30 customers who request tire replacements.

Out of these 30 customers who request tire replacements, 10% also request an oil change. So, there are 10 customers who request both oil change and tire replacements.

Therefore, out of the 60 customers who request an oil change, the number of customers who do not request tire replacements is  $60 - 10 = 50$ .

The probability that a customer who requests an oil change does not request tire replacements is given by:

$$\frac{\text{Number of customers who request only oil change}}{\text{Total number of customers who request oil change}} = \frac{50}{60} = \frac{5}{6} = \boxed{0.833}$$

Final Answer: D

\*\*\*\*\*

Answer Option 8:

Let  $O$  be the event that a customer requests an oil change and  $T$  be the event that a customer requests tire replacements. We are given that  $P(O) = 0.60$ ,  $P(T) = 0.30$ , and  $P(O \cap T) = 0.10$ . We want to find  $P(\neg T | O)$ , which is the conditional probability that a customer does not request tire replacements given that the customer requests an oil change.

Recall that  $P(A | B) = \frac{P(A \cap B)}{P(B)}$ . Using this formula, we have

$$P(\neg T | O) = \frac{P(\neg T \cap O)}{P(O)}$$

Since  $P(\neg T \cap O) = P(O) - P(O \cap T)$ , we can substitute the given values to find

$$P(\neg T | O) = \frac{0.60 - 0.10}{0.60} = \frac{1}{3}$$

Therefore, the probability that a customer does not request tire replacements given that the customer requests an oil change is  $\frac{1}{3}$ . This corresponds to  $\boxed{\text{(C) } 0.7}$ . Final Answer: C

\*\*\*\*\*

Answer Option 9:

Let  $O$  be the event that a customer requests an oil change, and let  $\neg T$  be the event that a customer requests tire replacements. We are given  $P(O) = 0.6$ ,  $P(T) = 0.3$ , and  $P(O \cap T) = 0.1$ . We want to find  $P(O | \neg T)$ , the probability that a customer requests an oil change given that they do not request tire replacements.

By the definition of conditional probability, we have

$$P(O | \neg T) = \frac{P(O \cap \neg T)}{P(\neg T)}$$

Note that  $\neg T$  represents the event that a customer does not request tire replacements. We can express  $\neg T$  as  $T^c$ , where  $T^c$  represents the complement of  $T$ . Then we have

$$P(O | \neg T) = \frac{P(O \cap T^c)}{P(T^c)}$$

Since  $T$  and  $T^c$  are mutually exclusive, we have  $P(O \cap T^c) = P(O) - P(O \cap T) = 0.6 - 0.1 = 0.5$ . Also,  $P(T^c) = 1 - P(T) = 1 - 0.3 = 0.7$ . Substituting these values into the expression, we get

$$P(O | \neg T) = \frac{0.5}{0.7} = \frac{5}{7} \approx \boxed{0.714}$$

Final Answer: C

\*\*\*\*\*

Question 142:

'Which of the following statements is false?', A. 'Like the normal distribution, the t-distributions are symmetric.', B. 'The t-distributions are lower at the mean and higher at the tails, and so are more spread out than the normal distribution.', C. 'The greater the df, the closer the t-distributions are to the normal distribution.', D. 'The smaller the df, the better the 68-95-99.7 Rule works for t-models.'

=====

Answer Option 1:

Statement B is false. The t-distributions are actually higher at the mean and lower at the tails compared to the normal distribution. This is because the t-distributions have fatter tails, which means they have more extreme values compared to the normal distribution. Therefore, statement B is false. Final Answer: B

\*\*\*\*\*

Answer Option 2:

The statement that is false is statement D. The smaller the degrees of freedom (df), the worse the 68-95-99.7 Rule works for t-models. The 68-95-99.7 Rule is based on the assumption of a normal distribution, but as the degrees of freedom decrease, the t-distributions become more spread out and have heavier tails compared to the normal distribution. Therefore, the 68-95-99.7 Rule, which is derived from the normal distribution, is less accurate for t-distributions with smaller degrees of freedom. Final answer: D

\*\*\*\*\*

Answer Option 3:

The correct answer is D. 'The smaller the df, the better the 68-95-99.7 Rule works for t-models.' This statement is false. The 68-95-99.7 Rule, also known as the Empirical Rule, is a rule of thumb that applies to a normal distribution, not t-distributions. For t-distributions, as the degrees of freedom (df) decrease, the distribution becomes more spread out and has heavier tails compared to the normal distribution. Therefore, the 68-95-99.7 Rule does not apply well to t-distributions with small degrees of freedom. Final Answer: D

\*\*\*\*\*

Answer Option 4:

Statement B is false.

The t-distributions are actually higher at the mean and lower at the tails compared to the normal distribution. This is because the t-distributions have heavier tails, meaning that extreme values are more likely to occur compared to the normal distribution. Therefore, the t-distributions are more spread out than the normal distribution.

Therefore, the answer is  $\boxed{\text{B}}$ . 'The t-distributions are lower at the mean and higher at the tails, and so are more spread out than the normal distribution.')

\*\*\*\*\*

Answer Option 5:

The statement "The smaller the df, the better the 68-95-99.7 Rule works for t-models" is false. The 68-95-99.7 Rule, also known as the Empirical Rule, applies to the normal distribution, not the t-distributions. The t-distributions have heavier tails compared to the normal distribution, especially when the degrees of freedom (df) are small. This means that the probability of extreme values occurring in the tails of the t-distributions is higher compared to the normal distribution. Therefore, the 68-95-99.7 Rule does not hold as well for t-models with small df. The correct answer is  $\boxed{\text{D}}$ .

\*\*\*\*\*

Answer Option 6:

The false statement is statement C. 'The greater the df, the closer the t-distributions are to the normal distribution.' This statement is incorrect because as the degrees of freedom (df) increase, the t-distributions actually become closer to the standard normal distribution. This is known as the central limit theorem. Therefore, the correct answer is  $\boxed{\text{C}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 7:

The correct answer is D. 'The smaller the df, the better the 68-95-99.7 Rule works for t-models.' This statement is false. The 68-95-99.7 Rule (also known as the Empirical Rule) is a rule that applies to normal distributions, not t-distributions. The 68-95-99.7 Rule states that for a normal distribution, approximately 68% of the data falls within one standard deviation of the mean, approximately 95% falls within two standard deviations, and approximately 99.7% falls within three standard deviations. This rule does not hold exactly for t-distributions, especially when the degrees of freedom (df) are small. As the degrees of freedom decrease, the t-distribution becomes more spread out and has heavier tails, making the 68-95-99.7 Rule less accurate for t-distributions. Final answer: D.

\*\*\*\*\*

Answer Option 8:

The correct answer is D. 'The smaller the df, the better the 68-95-99.7 Rule works for t-models.' This statement is false. The 68-95-99.7 Rule is based on the normal distribution and does not apply directly to t-distributions. The 68-95-99.7 Rule states that approximately 68%, 95%, and 99.7% of observations in a normal distribution lie within one, two, and three standard deviations of the mean, respectively. While the t-distributions approach the normal distribution as the degrees of freedom (df) increase, they have heavier tails and more spread out distributions compared to the normal distribution. Thus, the 68-95-99.7 Rule does not work as well for t-distributions, especially for smaller df. Final answer: D.

\*\*\*\*\*

Answer Option 9:

The statement that is false is D. 'The smaller the df, the better the 68-95-99.7 Rule works for t-models.'

The 68-95-99.7 rule, also known as the empirical rule, states that for a normal distribution, approximately 68% of the data falls within one standard deviation of the mean, approximately 95% falls within two standard deviations, and approximately 99.7% falls within three standard deviations.

However, for t-distributions, which have fatter tails compared to the normal distribution, the percentage of data within each standard deviation is larger than what the 68-95-99.7 rule would predict. Therefore, the statement that the 68-95-99.7 rule works better for t-models with smaller degrees of freedom (df) is false.

The correct answer is D. Final Answer: D

\*\*\*\*\*



Question 143:

'A kidney dialysis center periodically checks a sample of its equipment and performs a major recalibration if readings are sufficiently off target. Similarly, a fabric factory periodically checks the sizes of towels coming off an assembly line and halts production if measurements are sufficiently off target. In both situations, we have the null hypothesis that the equipment is performing satisfactorily. For each situation, which is the more serious concern, a Type I or Type II error?', A. 'Dialysis center: Type I error, towel manufacturer: Type I error', B. 'Dialysis center: Type I error, towel manufacturer: Type II error', C. 'Dialysis center: Type II error, towel manufacturer: Type I error', D. 'Dialysis center: Type II error, towel manufacturer: Type II error'

=====

Answer Option 1:

A Type I error occurs when we reject the null hypothesis when it is actually true. In the case of the kidney dialysis center, a Type I error would mean that the center performs a major recalibration even though the equipment is actually performing satisfactorily.

A Type II error occurs when we fail to reject the null hypothesis when it is actually false. In the case of the fabric factory, a Type II error would mean that the factory does not halt production even though the towel measurements are actually off target.

Therefore, the correct answer is  $\boxed{\text{A. 'Dialysis center: Type I error, towel manufacturer: Type I error'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 2:

A Type I error occurs when the null hypothesis is rejected when it is actually true. In the case of the kidney dialysis center, a Type I error would mean that the center performs a major recalibration even when the equipment is performing satisfactorily. This would be a more serious concern because unnecessary recalibrations could be costly and time-consuming, leading to potential disruptions in patient care.

On the other hand, a Type II error occurs when the null hypothesis is not rejected when it is actually false. In the case of the fabric factory, a Type II error would mean that production is not halted even when the towel sizes are sufficiently off target. This would be a less serious concern because towels with incorrect sizes could still be produced and potentially cause dissatisfaction among customers.

Therefore, the correct answer is  $\boxed{\text{A. Dialysis center: Type I error, towel manufacturer: Type I error}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 3:

In hypothesis testing, a Type I error occurs when the null hypothesis is rejected when it is actually true. On the other hand, a Type II error occurs when the null hypothesis is accepted when it is actually false. In the context of the given scenarios, the more serious concern for the dialysis center would be a Type II error. This means that the center fails to detect and take action when the equipment is not performing satisfactorily. This can have serious consequences for the patients relying on the dialysis equipment.

For the towel manufacturer, the more serious concern would be a Type I error. This means that the production is halted when the measurements are actually satisfactory. While this may result in a temporary loss in production, it is a necessary precaution to ensure the quality of the towels.

Therefore, the correct answer is  $\boxed{\text{C. 'Dialysis center: Type II error, towel manufacturer: Type I error'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 4:

A Type I error occurs when we reject the null hypothesis when it is actually true. In the case of the dialysis center, a Type I error would mean that the center thinks the equipment is not performing satisfactorily and performs a major recalibration, even though the equipment is actually working fine. This is a serious concern because it could lead to unnecessary expenses and disruptions in the dialysis process for patients.

On the other hand, a Type II error occurs when we fail to reject the null hypothesis when it is actually false. In the case of the towel manufacturer, a Type II error would mean that the manufacturer does not halt production even though the towel sizes are actually off target. This is also a serious concern because it could result in the production of defective towels that do not meet quality standards and could negatively impact the reputation of the manufacturer.

Therefore, the more serious concern for the dialysis center is a Type I error, and the more serious concern for the towel manufacturer is a Type II error. The correct answer is  $\boxed{\text{A. 'Dialysis center: Type I error, towel manufacturer: Type I error'}}$ .

\*\*\*\*\*

Answer Option 5:

In hypothesis testing, a Type I error occurs when we reject the null hypothesis when it is actually true. A Type II error occurs when we fail to reject the null hypothesis when it is actually false.

In the case of the kidney dialysis center, a Type I error would mean that the center recalibrates the equipment even though it is actually performing satisfactorily. This would be a waste of resources and unnecessary interruption of the treatment process. Therefore, a Type I error is the more serious concern for the dialysis center.

In the case of the fabric factory, a Type I error would mean that the production is halted even though the towel sizes are actually within the acceptable range. This would result in unnecessary disruptions to the production process and potential loss of revenue. Therefore, a Type I error is the more serious concern for the towel manufacturer.

Therefore, the correct answer is A. 'Dialysis center: Type I error, towel manufacturer: Type I error'.

\*\*\*\*\*

Answer Option 6:

A Type I error refers to rejecting the null hypothesis when it is actually true. In the context of the kidney dialysis center, this would mean concluding that the equipment is not performing satisfactorily when it actually is. This would lead to unnecessary recalibration. Therefore, a Type I error is the more serious concern for the dialysis center.

On the other hand, a Type II error refers to failing to reject the null hypothesis when it is actually false. In the context of the towel manufacturer, this would mean failing to halt production when the towel sizes are off target. This could result in producing defective towels. Therefore, a Type II error is the more serious concern for the towel manufacturer.

Therefore, the correct answer is  $\boxed{\text{A. 'Dialysis center: Type I error, towel manufacturer: Type I error'}}$ . Final answer: A

\*\*\*\*\*

Answer Option 7:

A Type I error occurs when the null hypothesis is rejected, even though it is true. In the context of the dialysis center, a Type I error would mean that the center performs a major recalibration even though the equipment is actually performing satisfactorily. This would be a serious concern because unnecessary recalibration could be expensive and time-consuming.

A Type II error occurs when the null hypothesis is not rejected, even though it is false. In the context of the towel manufacturer, a Type II error would mean that the production is not halted even though the towel sizes are off target. This would be a serious concern because it means that defective towels could be produced and distributed to customers.

Therefore, the more serious concern for the dialysis center is a Type I error, while the more serious concern for the towel manufacturer is a Type II error. The corresponding answer choice is  $\boxed{\text{A. 'Dialysis center: Type I error, towel manufacturer: Type I error'}}.$

\*\*\*\*\*

Answer Option 8:

A Type I error occurs when the null hypothesis is rejected, but it is actually true. In the context of the dialysis center, this would mean that the center performs a major recalibration even though the equipment is actually performing satisfactorily. A Type II error occurs when the null hypothesis is not rejected, but it is actually false. In the context of the towel manufacturer, this would mean that the production is not halted even though the towel sizes are actually off target.

Therefore, for the dialysis center, the more serious concern is a Type II error because it means that the equipment may not be performing satisfactorily but the center fails to detect it. For the towel manufacturer, the more serious concern is a Type I error because it means that towels with incorrect sizes are being produced even though the production should have been halted.

Hence, the correct answer is  $\boxed{\text{C. Dialysis center: Type II error, towel manufacturer: Type I error}}.$  Final answer: C.

\*\*\*\*\*

Answer Option 9:

In hypothesis testing, a Type I error occurs when we reject the null hypothesis when it is actually true. This means that we conclude there is a problem or issue when there really isn't.

On the other hand, a Type II error occurs when we fail to reject the null hypothesis when it is actually false. This means that we fail to identify a problem or issue when there really is one.

In the context of the dialysis center, the null hypothesis is that the equipment is performing satisfactorily. A Type I error would mean that the center concludes there is a problem with the equipment and performs a major recalibration when, in reality, the equipment is actually performing fine. This would be a more serious concern because it would result in unnecessary recalibrations and potentially disrupt the operations of the center.

In the context of the fabric factory, the null hypothesis is that the towels are being produced with the correct sizes. A Type I error would mean that the factory concludes there is a problem with the towel sizes and halts production when, in reality, the towel sizes are actually within the acceptable range. This would also be a serious concern as it would result in unnecessary halting of production and potential financial losses.

Therefore, the correct answer is  $\boxed{\text{A. 'Dialysis center: Type I error, towel manufacturer: Type I error'}}.$  Final Answer: A

\*\*\*\*\*



Question 144:

'A company with offices in five different countries is interested in estimating the proportion of its employees in favor of banning smoking on the office premises. It is known that the views of people from different countries on issues like smoking tend to vary due to the influence of different local social structures. Which of the following is an appropriate sampling technique to use in this situation, and why?', A. 'Stratified sample, because the population is divided into five strata—namely, five offices in five different countries', B. 'Cluster sample, because the population is divided into five clusters—namely, five offices in five different countries', C. 'Simple random sample, because this is the simplest sampling method', D. 'Simple random sample, because this is the only method that gives unbiased results'

=====

Answer Option 1:

The appropriate sampling technique to use in this situation is A. Stratified sample, because the population is divided into five strata—namely, five offices in five different countries.

In a stratified sample, the population is divided into distinct groups or strata. In this case, the strata would be the five offices in the five different countries. By sampling from each stratum, we can ensure that we have representation from each country and capture the variation in views on banning smoking in each country.

This method allows for more accurate estimates of the proportion of employees in favor of banning smoking within each country, as well as the overall population.

Option B, a cluster sample, would involve randomly selecting entire offices or groups of employees within offices, which may not necessarily provide representative samples from each country.

Option C, a simple random sample, may not provide enough representation from each country and may not capture the variation in views across different countries.

Option D, stating that simple random sample is the only method that gives unbiased results, is incorrect as stratified sampling can also provide unbiased results if done properly.

Therefore, the most appropriate sampling technique for this situation is A. Stratified sample.

\*\*\*\*\*

Answer Option 2:

The appropriate sampling technique in this situation is A. Stratified sample, because the population is divided into five strata - the five offices in five different countries.

In a stratified sample, the population is divided into subgroups or strata, and then a random sample is taken from each stratum. This ensures that each stratum is represented proportionally in the sample, which is important in this case because the company wants to estimate the proportion of employees in favor of banning smoking in each country separately.

By using a stratified sample, the company can ensure that the views of employees from each country are taken into account and representatively sampled. This allows for more accurate estimates of the proportion of employees in favor of banning smoking in each country, taking into consideration the variation in views across different cultural and social contexts.

Therefore, option A is the most appropriate sampling technique in this situation.

\*\*\*\*\*

Answer Option 3:

The appropriate sampling technique to use in this situation is A. Stratified sample because the population is divided into five strata, which are the five offices in five different countries.

Using a stratified sample allows us to ensure that each country is represented proportionally in the sample. This is important because the views on banning smoking may vary between countries due to different local social structures. By dividing the population into strata based on the offices in different countries, we can take a sample from each stratum and obtain a representative sample that reflects the diversity of views on smoking across the different countries.

A cluster sample (option B) would involve randomly selecting entire offices as clusters, which may not necessarily represent the diversity of views within each office. A simple random sample (option C) or a simple random sample (option D) may not guarantee representation from each country in proportion to their population. Therefore, option A is the most appropriate sampling technique in this situation. Final answer: A

\*\*\*\*\*

Answer Option 4:

The appropriate sampling technique in this situation would be a 'Stratified sample, because the population is divided into five strata—namely, five offices in five different countries'.

Stratified sampling involves dividing the population into distinct subgroups, or strata, based on certain characteristics. In this case, the countries where the company has offices serve as the natural strata. By selecting a proportionate number of employees from each country, we can ensure that the sample is representative of the entire population and takes into account the potential variability in views on smoking across different countries.

This method allows for a more accurate estimation of the proportion of employees in favor of banning smoking on the office premises within each country, as well as for the overall company. Final answer: A

\*\*\*\*\*

Answer Option 5:

The appropriate sampling technique to use in this situation is A. Stratified sample, because the population is divided into five strata—namely, five offices in five different countries.

A stratified sample is a sampling technique that divides the population into subgroups or strata based on certain characteristics or attributes. In this case, the population of employees is divided into five strata based on the five different countries where the offices are located. This ensures that each country is represented in the sample and allows for a more accurate estimation of the proportion of employees in favor of banning smoking in each country.

Using a cluster sample (option B) would involve randomly selecting entire offices or clusters, which may not represent the employees from each country proportionally. This could lead to biased results if certain offices have significantly different proportions of employees in favor of banning smoking.

Simple random sampling (option C) and simple random sample (option D) are not ideal in this situation because they do not take into account the variability in views on smoking among employees from different countries. A simple random sample would randomly select employees without considering the strata or countries they belong to, which could result in an unrepresentative sample.

Therefore, the best option is A. Stratified sample, because it allows for a more accurate estimation of the proportion of employees in favor of banning smoking in each country. Final answer: A.

\*\*\*\*\*

Answer Option 6:

The appropriate sampling technique in this situation would be a 'Stratified sample, because the population is divided into five strata—namely, five offices in five different countries' (option A).

In this case, the company is interested in estimating the proportion of employees in favor of banning smoking on the office premises. Since the views of people from different countries tend to vary, it is important to ensure that the sample is representative of each country's population.

By using a stratified sample, the company can divide the population into five strata based on the five different countries. From each stratum, a random sample of employees can be selected. This ensures that the sample includes employees from each country in proportion to their representation in the population, allowing for more accurate estimates of the proportion of employees in favor of banning smoking in each country.

Therefore, option A is the most appropriate sampling technique in this situation. Final Answer: A

\*\*\*\*\*

Answer Option 7:

The appropriate sampling technique to use in this situation is A. Stratified sample, because the population is divided into five strata—namely, the five offices in five different countries.

In a stratified sample, the population is divided into subgroups or strata, and then a random sample is taken from each stratum. This method ensures that each stratum is represented in the sample proportionally to its size in the population.

In this case, since the company has offices in five different countries, each country represents a different stratum. By using a stratified sample, we can ensure that the proportion of employees from each country is represented in the sample accurately, taking into account the variation in views on smoking due to different social structures in each country.

The other options, B. Cluster sample, C. Simple random sample, and D. Simple random sample because this is the only method that gives unbiased results, are not the most appropriate sampling techniques in this situation as they do not take into account the variation due to different countries (strata) in the population. Final Answer: A

\*\*\*\*\*

Answer Option 8:

The most appropriate sampling technique in this situation would be a stratified sample, as stated in option A. The reason for this is that the population (employees of the company) is divided into five strata, which correspond to the five different offices in five different countries.

By selecting a stratified sample, the company can ensure representation from each office, accounting for the potential variation in views on banning smoking due to different local social structures. This sampling method allows for a more accurate estimation of the proportion of employees in favor of banning smoking, as it takes into account the diversity of opinions among employees from different countries.

Option B, a cluster sample, would not be as appropriate in this case because it divides the population into clusters based on geographical proximity rather than considering the different countries. Option C, a simple random sample, may not provide sufficient representation from each country, leading to potential bias in the estimation. Option D is incorrect because other sampling methods can also provide unbiased results, such as stratified sampling.

Therefore, the most appropriate sampling technique in this situation would be a stratified sample, option A. Final Answer: A

\*\*\*\*\*

Answer Option 9:

The appropriate sampling technique to use in this situation is the cluster sample because the population is divided into five clusters, which are the five offices in five different countries.

In a cluster sample, the population is first divided into clusters or groups, and then a random sample of clusters is selected. In this case, the clusters are the five offices in different countries.

The advantage of using a cluster sample in this situation is that it is more practical and cost-effective. Instead of trying to reach all employees in all offices in all countries, the company can randomly select a few offices and survey the employees in those offices. This reduces the time and resources needed for data collection.

On the other hand, the other options are not appropriate for this situation. A stratified sample would be used if the population was divided into different groups based on specific characteristics. In this case, the offices in different countries are not based on any specific characteristic related to smoking preferences. A simple random sample would be inefficient and would require surveying all employees in all offices, which may not be feasible.

Therefore, the most appropriate sampling technique in this situation is the cluster sample, making option B, 'Cluster sample, because the population is divided into five clusters—namely, five offices in five different countries', the correct answer. Final Answer: B

\*\*\*\*\*

Question 145:

'Semir rolls a six-sided die every morning to determine what he will have for breakfast. If he rolls a 1 or 2, he takes time to cook himself a big breakfast. If he rolls a 3 or larger he grabs a quick lighter breakfast. When he cooks himself a big breakfast, there is a 15% chance he will be late for school. If he has a lighter breakfast, there is a 6% chance he will be late for school. What is the probability Semir will be on time for school any given day?', A. '0.09', B. '0.21', C. '0.8', D. '0.91'

=====

Answer Option 1:

The probability that Semir will be on time for school is the complement of the probability that he will be late for school.

The probability that he will cook a big breakfast and be late for school is  $\frac{1}{6} \cdot (0.15) = \frac{1}{40}$ .

The probability that he will have a lighter breakfast and be late for school is  $\frac{5}{6} \cdot (0.06) = \frac{5}{100} = \frac{1}{20}$ .

Therefore, the probability that Semir will be late for school is  $\frac{1}{40} + \frac{1}{20} = \frac{3}{40}$ .

Hence, the probability that Semir will be on time for school is  $1 - \frac{3}{40} = \boxed{\frac{37}{40}}$ .

Final answer: D

\*\*\*\*\*

Answer Option 2:

The probability that Semir will cook himself a big breakfast is  $\frac{2}{6} = \frac{1}{3}$ , and the probability that he will have a lighter breakfast is  $\frac{4}{6} = \frac{2}{3}$ .

The probability that he will be late for school given that he cooked himself a big breakfast is \$0.15\$, and the probability that he will be late for school given that he had a lighter breakfast is \$0.06\$.

Therefore, the probability that Semir will be late for school on any given day is  $\left(\frac{1}{3}\right)(0.15) + \left(\frac{2}{3}\right)(0.06) = 0.09 + 0.12 = 0.21$ .

Thus, the probability that Semir will be on time for school any given day is  $1 - 0.21 = \boxed{0.79}$ . Final Answer: C

\*\*\*\*\*

Answer Option 3:

The probability that Semir cooks himself a big breakfast is  $\frac{2}{6} = \frac{1}{3}$  and the probability that he grabs a lighter breakfast is  $\frac{4}{6} = \frac{2}{3}$ . The probability that he will be late for school given that he cooks himself a big breakfast is \$0.15\$ and the probability that he will be late for school given that he grabs a lighter breakfast is \$0.06\$. Therefore, the probability that he will be on time for school any given day is  $(1 - 0.15) \cdot \frac{1}{3} + (1 - 0.06) \cdot \frac{2}{3} = 0.85 \cdot \frac{1}{3} + 0.94 \cdot \frac{2}{3} = \frac{17}{20} = \boxed{0.85}$ . Final Answer: A

\*\*\*\*\*

Answer Option 4:

The probability of rolling a 1 or 2 is  $\frac{2}{6} = \frac{1}{3}$ , and the probability of rolling a 3 or larger is  $\frac{4}{6} = \frac{2}{3}$ . If Semir rolls a 1 or 2, there is a 15% chance he will be late, so the probability of being on time if he cooks a big breakfast is  $1 - 0.15 = 0.85$ . If he rolls a 3 or larger, there is a 6% chance he will be late, so the probability of being on time if he has a lighter breakfast is  $1 - 0.06 = 0.94$ . Therefore, the overall probability of being on time is

$\frac{1}{3} \cdot 0.85 + \frac{2}{3} \cdot 0.94 = \frac{85}{100} \cdot \frac{1}{3} + \frac{94}{100} \cdot \frac{2}{3} = \frac{459}{600} = \frac{153}{200} = \boxed{0.765}$ . Final answer: None of the given answer choices are correct.

\*\*\*\*\*

Answer Option 5:

The probability of rolling a 1 or 2 is  $\frac{2}{6} = \frac{1}{3}$ , and the probability of being late for school given that he cooks himself a big breakfast is \$0.15\$. So the probability of being late for school when he rolls a 1 or 2 is  $\frac{1}{3} \cdot 0.15 = \frac{1}{20}$ .

The probability of rolling a 3 or larger is  $\frac{4}{6} = \frac{2}{3}$ , and the probability of being late for school given that he grabs a quick lighter breakfast is \$0.06\$. So the probability of being late for school when he rolls a 3 or larger is  $\frac{2}{3} \cdot 0.06 = \frac{1}{15}$ .

Therefore, the probability of being on time for school any given day is  $1 - \left(\frac{1}{20} + \frac{1}{15}\right) = 1 - \frac{7}{60} = \boxed{\frac{53}{60}}$ . Final Answer: None of the above.

\*\*\*\*\*

Answer Option 6:

The probability of Semir rolling a 1 or 2 is  $\frac{2}{6} = \frac{1}{3}$ , and the probability of him being late when he cooks himself a big breakfast is \$0.15\$. Therefore, the probability of him rolling a 1 or 2 and being late is  $\frac{1}{3} \cdot 0.15 = \frac{1}{20}$ .

The probability of Semir rolling a 3, 4, 5, or 6 is  $\frac{4}{6} = \frac{2}{3}$ , and the probability of him being late when he grabs a quick lighter breakfast is \$0.06\$. Therefore, the probability of him rolling a 3, 4, 5, or 6 and being late is  $\frac{2}{3} \cdot 0.06 = \frac{1}{15}$ .

The probability of Semir being on time for school is equal to \$1\$ minus the probability of him being late. Therefore, the probability of Semir being on time for school any given day is  $1 - \left(\frac{1}{20} + \frac{1}{15}\right) = \frac{17}{20}$ .

Thus, the answer is  $\boxed{\text{(D) } 0.91}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

The probability of Semir rolling a 1 or 2 is  $\frac{2}{6} = \frac{1}{3}$ , and the probability of him rolling a 3 or larger is  $\frac{4}{6} = \frac{2}{3}$ .

If Semir rolls a 1 or 2 and cooks himself a big breakfast, there is a 15% chance he will be late for school. Therefore, the probability of him being late given that he cooks a big breakfast is  $\frac{1}{3} \cdot \frac{15}{100} = \frac{1}{20}$ .

If Semir rolls a 3 or larger and has a lighter breakfast, there is a 6% chance he will be late for school. Therefore, the probability of him being late given that he has a lighter breakfast is  $\frac{2}{3} \cdot \frac{6}{100} = \frac{1}{50}$ .

To find the probability of Semir being on time for school, we need to find the complement of him being late. Therefore, the probability of him being on time is  $1 - \left(\frac{1}{20} + \frac{1}{50}\right) = 1 - \frac{7}{100} = \frac{93}{100}$ .

Thus, the probability of Semir being on time for school any given day is  $\boxed{\frac{93}{100}}$ . Final answer: D

\*\*\*\*\*

Answer Option 8:

To find the probability that Semir will be on time for school, we need to consider the probabilities of him rolling a 1 or 2 and being late, and the probabilities of him rolling a 3 or larger and being on time.

The probability of him rolling a 1 or 2 is  $\frac{1}{3}$ , and the probability of being late when he cooks a big breakfast is \$0.15. Therefore, the probability of him rolling a 1 or 2 and being late is  $\frac{1}{3} \times 0.15 = 0.05$ .

The probability of him rolling a 3 or larger is  $\frac{2}{3}$ , and the probability of being on time when he has a lighter breakfast is  $1 - 0.06 = 0.94$ . Therefore, the probability of him rolling a 3 or larger and being on time is  $\frac{2}{3} \times 0.94 = 0.6267$ .

Finally, we add the probabilities of the two cases together to find the overall probability of him being on time:  $0.05 + 0.6267 = 0.6767$ . Therefore, the probability Semir will be on time for school any given day is  $\boxed{\text{text{(B) } 0.21}}$ .

\*\*\*\*\*

Answer Option 9:

The probability that Semir rolls a 1 or 2 is  $\frac{2}{6} = \frac{1}{3}$ , and the probability that he rolls a 3 or larger is  $\frac{4}{6} = \frac{2}{3}$ .

If Semir rolls a 1 or 2, there is a 15% chance he will be late for school. Therefore, the probability of him being on time if he rolls a 1 or 2 is  $1 - 0.15 = 0.85$ .

If Semir rolls a 3 or larger, there is a 6% chance he will be late for school. Therefore, the probability of him being on time if he rolls a 3 or larger is  $1 - 0.06 = 0.94$ .

To find the overall probability of Semir being on time, we need to multiply the probabilities of the two events (rolling a 1 or 2 and rolling a 3 or larger) by their respective probabilities of him being on time. Thus, the overall probability is:

$$\left(\frac{1}{3}\right)(0.85) + \left(\frac{2}{3}\right)(0.94) = 0.28 + 0.63 = \boxed{0.91}$$

Therefore, the answer is D.

\*\*\*\*\*

Question 146:

'A teacher's union and a school district are negotiating salaries for the coming year. The teachers want more money, and the district, claiming, as always, budget constraints, wants to pay as little as possible. The district, like most, has a large number of moderately paid teachers and a few highly paid administrators. The salaries of all teachers and administrators are included in trying to figure out, on average, how much the professional staff currently earn. Which of the following would the teachers' union be most likely to quote during negotiations?', A. 'The mean of all the salaries.', B. 'The mode of all the salaries.', C. 'The standard deviation of all the salaries.', D. 'The median of all the salaries.'

=====

Answer Option 1:

The teachers' union would most likely quote the median of all the salaries during negotiations. The median represents the middle value in a data set when arranged in ascending or descending order. It is not affected by extreme values or outliers, which makes it a good measure of central tendency to represent the typical salary of the teachers. The district may have a few highly paid administrators, but the median would not be significantly affected by these outliers, giving a more accurate representation of the salaries of the majority of teachers. Therefore, the answer is  $\boxed{\text{(D) 'The median of all the salaries.'}}$ .

=====

Answer Option 2:

The teachers' union would be most likely to quote the median of all the salaries during negotiations. The median represents the middle value of a set of salaries when they are arranged in increasing or decreasing order. By using the median, the union can highlight the value that divides the salaries into two equal groups, suggesting that the majority of teachers' salaries fall around this middle value. This is advantageous for the union because it emphasizes the income of the majority of teachers, rather than the influence of any outliers or highly paid administrators. Therefore, the answer is  $\boxed{\text{(D) 'The median of all the salaries.'}}$ .

=====

Answer Option 3:

The most likely quote for the teachers' union during negotiations would be option D, "The median of all the salaries."

The median is the middle value of a set of numbers when they are arranged in order. It represents the salary that separates the higher-paid half from the lower-paid half of the staff. By quoting the median, the teachers' union is emphasizing the salary that is most representative of the majority of teachers, rather than the mean (option A) which can be influenced by a few highly paid administrators. The mode (option B) is the value that appears most frequently and may not be a good measure of the typical salary. The standard deviation (option C) measures the spread of the salaries and may not be relevant to the union's argument for higher wages. Therefore, the teachers' union would most likely quote the median of all the salaries. Final Answer: D

=====

Answer Option 4:

The teachers' union would be most likely to quote the median of all the salaries during negotiations. The median represents the middle value of a set of data when arranged in order. By quoting the median, the union can focus on the salary that separates the higher-paid administrators from the moderately paid teachers, which is more relevant to the majority of union members. The mean (option A) can be easily skewed by a few extremely high or low salaries, making it an unreliable measure for negotiations. The mode (option B) represents the most frequently occurring salary, which may not accurately represent the overall salary distribution. The standard deviation (option C) measures the dispersion of salaries, but it may not be a primary concern for the union during negotiations. Thus, the most relevant measure for the union to quote would be the median (option D). Final Answer: D

=====

Answer Option 5:

The teachers' union would most likely quote the median of all the salaries.

The mean can be easily influenced by a few highly paid administrators, so it may not accurately represent the salaries of the teachers. The mode is the value that appears most frequently, but it may not be a good representation of the overall distribution of salaries. The standard deviation measures the spread of the salaries, but it may not be the most relevant measure for negotiations.

On the other hand, the median represents the middle value of the salaries when they are arranged in increasing order. It is not affected by extreme values or outliers, so it provides a more accurate representation of the typical salary for the teachers. Therefore, the teachers' union would be most likely to quote the median of all the salaries.

Therefore, the answer is  $\boxed{\text{(D) 'The median of all the salaries.'}}$ .

=====

Answer Option 6:

The teachers' union would most likely quote the median of all the salaries during negotiations. The median is the middle value when the salaries are arranged in ascending or descending order. It represents the salary that divides the distribution into two equal halves. By using the median, the union can highlight that half of the staff earns below this value, implying that many teachers are not being adequately compensated. Additionally, the median is not influenced by extreme values, such as the highly paid administrators, which would be the case with the mean. Therefore, the union would choose option D, 'The median of all the salaries,' as the most favorable statistic to support their argument. Final Answer: D

=====

Answer Option 7:

The most appropriate measure for the teachers' union to quote during negotiations would be the median of all the salaries, which is represented by option D.

The median represents the middle value of a data set when the values are arranged in ascending or descending order. By quoting the median, the teachers' union would be focusing on the salary that divides the staff into two equal groups: half of the staff earning more and half earning less.

This measure is relevant to the teachers' union because it captures the salary that is most representative of the majority of teachers. By highlighting the median, the union emphasizes the salary that is not influenced by extreme values or outliers, such as the highly paid administrators. It reflects the salary that is most typical for the majority of teachers, providing a fair representation of their earnings.

Hence, the teachers' union would be most likely to quote the median of all the salaries. Final Answer: D

\*\*\*\*\*

Answer Option 8:

During negotiations, the teachers' union would most likely quote the median of all the salaries, represented by option D. The median represents the middle value when the salaries are arranged in increasing order. By quoting the median, the union is focusing on the middle value, which is less likely to be influenced by extreme outliers such as highly paid administrators. This allows the union to argue for a salary increase that benefits the majority of teachers. Option A, the mean of all the salaries, can be easily skewed by the presence of a few highly paid administrators and may not accurately represent the typical teacher's salary. Option B, the mode of all the salaries, represents the most frequently occurring salary and may not be a good representation of the overall salary distribution. Option C, the standard deviation of all the salaries, measures the spread of the salaries but may not directly relate to the negotiation goals of the teachers' union. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The teachers' union would most likely quote the median of all the salaries during negotiations.

The mean, or average, of all the salaries can be easily influenced by the highly paid administrators and may not accurately represent the salary of the majority of teachers.

The mode refers to the most frequent salary, but it may not provide a good estimate of the typical salary if there are multiple modes or if the mode is not representative of the majority of teachers.

The standard deviation measures the dispersion or variability of the salaries, but it may not be a priority for the teachers' union during negotiations.

The median, however, represents the middle value of the salaries when they are arranged in ascending order. It is not affected by extreme values such as the salaries of highly paid administrators, and it provides a better estimate of the typical salary for the majority of teachers. Therefore, the teachers' union would most likely quote the median of all the salaries.

So the answer is  $\boxed{\text{D. The median of all the salaries.}}$

\*\*\*\*\*

Question 147:

The Hardcore Construction Company has two offices, one in Atlanta and one in New Orleans. Fifteen engineers work in the Atlanta office, and 14 engineers work in the New Orleans office. The business manager decided to use a 2-sample t-test to compare the mean salaries of engineers in the two offices. Because there were only 15 engineers in one office and 14 engineers in the other, he used the salaries of all the engineers in the computation. Is the 2-sample t-test an appropriate inferential technique in this situation?, A. 'Yes, because he is comparing the means of two small groups.', B. 'Yes. Both Atlanta and New Orleans are large cities, so the salaries are comparable.', C. 'Yes. Because Atlanta and New Orleans are about 500 miles apart, the two groups of engineers can be assumed to be independent.', D. 'No, because the entire population information was used from both offices. Because no samples were taken, a t-test should not be used.'

=====

Answer Option 1:

The correct answer is C. 'Yes. Because Atlanta and New Orleans are about 500 miles apart, the two groups of engineers can be assumed to be independent.'

The 2-sample t-test is an appropriate inferential technique in this situation because the engineers in the Atlanta and New Orleans offices can be assumed to be independent. The distance between the two offices suggests that the engineers from each office are not directly influenced by each other and can be treated as separate groups for comparison purposes. The sample sizes of 15 and 14 are also sufficient for the t-test to be valid. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The correct answer is D. 'No, because the entire population information was used from both offices. Because no samples were taken, a t-test should not be used.'

In a 2-sample t-test, two separate samples are taken from two different populations to compare their means. In this scenario, the business manager used the salaries of all the engineers in both offices, which means that the entire population information was used. Since no separate samples were taken, the assumptions necessary for a t-test are not met. Therefore, a 2-sample t-test is not an appropriate inferential technique in this situation. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The correct answer is D. 'No, because the entire population information was used from both offices. Because no samples were taken, a t-test should not be used.'

In a 2-sample t-test, we compare the means of two independent samples. The key word here is "independent." In this scenario, the business manager used the salaries of all the engineers in both offices, meaning that no sampling was done. This violates the assumption of independence required for a t-test. Additionally, the fact that one group has 15 engineers and the other has 14 does not make the groups "small." The appropriate inferential technique in this situation would be a different type of statistical analysis, such as a one-sample t-test or a confidence interval for the difference in means.

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'No, because the entire population information was used from both offices. Because no samples were taken, a t-test should not be used.'

In a two-sample t-test, the goal is to compare the means of two independent groups by taking random samples from each group. However, in this situation, the business manager used the salaries of all the engineers in both offices, which means that no sampling was done. The entire population information was used, which is not appropriate for a t-test. Additionally, the size of the groups (15 engineers in one office and 14 engineers in the other) does not determine whether a t-test is appropriate or not. The appropriateness of a t-test depends on the method of sampling and the assumptions of independence and normality of the data.

\*\*\*\*\*

Answer Option 5:

The correct answer is D. 'No, because the entire population information was used from both offices. Because no samples were taken, a t-test should not be used.'

In a 2-sample t-test, the goal is to compare the means of two populations based on samples taken from each population. However, in this scenario, the business manager used the salaries of all the engineers in both offices, effectively using the entire population information. Since no samples were taken and the entire population information was used, a t-test is not appropriate. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The correct answer is D. 'No, because the entire population information was used from both offices. Because no samples were taken, a t-test should not be used.'

In a 2-sample t-test, the assumption is that we are taking independent random samples from two populations. However, in this situation, the business manager used the salaries of all the engineers in the computation, which means that there were no samples taken. The entire population information was used, and therefore, a t-test should not be used. Final answer: D.

\*\*\*\*\*

Answer Option 7:

The correct answer is D. 'No, because the entire population information was used from both offices. Because no samples were taken, a t-test should not be used.'

In a 2-sample t-test, we compare the means of two independent samples to determine if there is a significant difference between them. However, in this scenario, the business manager used the salaries of all the engineers in the computation. This means that the data used for the test is the entire population, rather than a sample. Since no sampling was done, a t-test is not appropriate. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The appropriate answer is D. 'No, because the entire population information was used from both offices. Because no samples were taken, a t-test should not be used.' The 2-sample t-test is used to compare the means of two independent samples. In this situation, the business manager used the salaries of all the engineers in both offices, which means the entire population information was used. Since no sampling was done, the use of a t-test is not appropriate. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The correct answer is A. 'Yes, because he is comparing the means of two small groups.'

The 2-sample t-test is appropriate when comparing the means of two groups, regardless of the sample size. In this case, the manager is comparing the mean salaries of engineers in the two offices, which is a valid application of the 2-sample t-test. The fact that the sample sizes are small does not make the test inappropriate. It is important to note that the t-test is used to make inferences about the population based on sample data, and it does not require the entire population information to be used. Final answer: A

\*\*\*\*\*



Question 148:

'For the regression line, which of the following statements about residuals is true?', A. 'Influential scores have large residuals.', B. 'If the linear model is good, the number of positive residuals will be the same as the number of negative residuals.', C. 'The mean of the residuals is always zero.', D. 'If the correlation is 0, there will be a distinct pattern in the residual plot.'

=====

Answer Option 1:

The correct statement about residuals is C. 'The mean of the residuals is always zero.'

In a regression model, the residuals represent the differences between the observed values and the predicted values. The mean of the residuals is always equal to zero because the regression line is fitted in such a way that it minimizes the sum of the squared residuals. This means that the positive residuals cancel out the negative residuals, resulting in a mean of zero.

Statement A is incorrect because influential scores, or outliers, can have a large impact on the regression line but not necessarily have large residuals.

Statement B is incorrect because the number of positive residuals and negative residuals can be different even in a good linear model. It depends on the distribution of the data points and the relationship between the predictor and response variables.

Statement D is incorrect because a correlation of 0 does not imply a distinct pattern in the residual plot. The residual plot can have a distinct pattern even when the correlation is not 0, depending on the nature of the relationship between the variables.

Therefore, the correct statement is C. 'The mean of the residuals is always zero.' Final Answer: C

\*\*\*\*\*

Answer Option 2:

The correct statement about residuals is C. 'The mean of the residuals is always zero.'

The residuals represent the difference between the observed values and the predicted values from the regression line. The mean of the residuals is always zero because the regression line is fitted to minimize the sum of the squared residuals. This means that the positive and negative residuals cancel each other out, resulting in a mean of zero.

Statement A is incorrect because influential scores, or outliers, can have large residuals but not all large residuals are due to influential scores.

Statement B is also incorrect because the number of positive residuals does not have to be the same as the number of negative residuals in a good linear model. It depends on the distribution of the data and the accuracy of the model.

Statement D is incorrect because if the correlation is 0, there will be no linear relationship between the variables and the residual plot will show random scatter rather than a distinct pattern.

Therefore, the correct statement is C. 'The mean of the residuals is always zero.' Final Answer: C

\*\*\*\*\*

Answer Option 3:

The correct statement about residuals is C. 'The mean of the residuals is always zero.'

The residuals are the differences between the observed values and the predicted values from the regression line. If the regression line is a good fit for the data, the errors should average out to zero. This means that the mean of the residuals will always be zero.

Option A is incorrect because influential scores refer to data points that have a large effect on the regression line, not the residuals. Option B is not necessarily true since the number of positive residuals does not have to be the same as the number of negative residuals. Option D is also incorrect because a correlation of 0 does not necessarily imply a distinct pattern in the residual plot.

Therefore, the correct answer is C. 'The mean of the residuals is always zero.' Final Answer: C

\*\*\*\*\*

Answer Option 4:

The correct statement about residuals is: C. 'The mean of the residuals is always zero.'

In a regression analysis, the residuals represent the difference between the observed values and the predicted values from the regression line. The mean of the residuals is always zero because the regression line is fitted in a way that minimizes the sum of the squared residuals. This means that the positive residuals and negative residuals will balance out, resulting in a mean of zero.

Option A is incorrect because influential scores can have either large or small residuals. An influential score is an outlier that has a large effect on the regression line.

Option B is incorrect because the number of positive residuals does not necessarily have to be the same as the number of negative residuals. It depends on the distribution of the data and the relationship between the predictor and response variables.

Option D is incorrect because the presence or absence of a distinct pattern in the residual plot is related to the presence or absence of a linear relationship between the predictor and response variables, not just the correlation. It is possible to have a correlation of 0 and still have a distinct pattern in the residual plot if there is a non-linear relationship between the variables.

Therefore, the correct statement is C. 'The mean of the residuals is always zero.' Final answer: C.

\*\*\*\*\*

Answer Option 5:

The correct statement about residuals is C. 'The mean of the residuals is always zero.'

The residuals are the differences between the observed values and the predicted values from the regression line. By definition, the regression line is fitted in a way that minimizes the sum of the squared residuals, which means that the mean of the residuals will always be zero. This is because the regression line is centered around the mean of the observed values.

Option A is incorrect because influential scores refer to outliers that significantly affect the regression line, not necessarily the size of the residuals.

Option B is incorrect because the number of positive residuals does not have to be the same as the number of negative residuals even if the linear model is good. It depends on the distribution of the data points and the relationship between the independent and dependent variables.

Option D is incorrect because a correlation of 0 does not necessarily indicate a distinct pattern in the residual plot. The residual plot can have a distinct pattern even with non-zero correlation, and it can be random even with zero correlation.

\*\*\*\*\*

Answer Option 6:

The correct statement about residuals is C. 'The mean of the residuals is always zero.'

Residuals are defined as the differences between the observed values and the predicted values from the regression line. These differences can be positive or negative, depending on whether the observed value is above or below the predicted value. However, if we take the average of all the residuals, the positive and negative differences will cancel out, resulting in a mean of zero. Therefore, the mean of the residuals is always zero.

Statement A is incorrect because influential scores refer to outliers that have a large impact on the regression line, not necessarily large residuals.

Statement B is not always true. In a good linear model, the number of positive residuals may not always be the same as the number of negative residuals. It is more important to consider the overall pattern of the residuals, such as whether they are randomly scattered around zero.

Statement D is incorrect. A correlation of 0 means there is no linear relationship between the variables, so the residual plot would not show a distinct pattern. A distinct pattern in the residual plot suggests the presence of a linear relationship.

\*\*\*\*\*

Answer Option 7:

The correct statement about residuals is C. 'The mean of the residuals is always zero.'

The residual for each data point is the vertical distance between the actual data point and the predicted value on the regression line. Some residuals will be positive, and some will be negative, but their sum will always be zero. This is because the regression line is chosen to minimize the sum of the squared residuals, so the positive residuals will cancel out the negative residuals when averaged.

Option A is incorrect because influential scores can have either large or small residuals, depending on their position relative to the regression line.

Option B is not necessarily true. In a good linear model, the number of positive residuals may not be the same as the number of negative residuals. It is possible for a linear model to have more positive residuals than negative or vice versa.

Option D is also not true. If the correlation is 0, it means there is no linear relationship between the variables. In this case, the residual plot will not show a distinct pattern.

Therefore, the correct statement is C. 'The mean of the residuals is always zero.' Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct statement about residuals is C. 'The mean of the residuals is always zero.'

The residuals represent the vertical distances between the observed data points and the predicted values on the regression line. By definition, the regression line is the line that minimizes the sum of the squared residuals. Therefore, the mean of the residuals is zero, as the line is chosen to minimize the overall vertical distance.

Statement A is incorrect because influential scores, which are data points that have a strong effect on the regression line, may or may not have large residuals.

Statement B is incorrect because even if the linear model is good, the number of positive residuals may not be the same as the number of negative residuals. It is possible for the residuals to be skewed towards one side if there is a systematic bias in the model.

Statement D is incorrect because the presence or absence of a distinct pattern in the residual plot is not determined solely by the correlation coefficient. The residual plot can have a distinct pattern even if the correlation is not zero, and it can also show no pattern even if the correlation is zero.

Therefore, the correct statement is C. 'The mean of the residuals is always zero.' Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct statement about residuals is C. 'The mean of the residuals is always zero.'

The residuals are the differences between the observed values and the predicted values of the response variable based on the regression line. If the regression line is a good fit for the data, the residuals should have a mean of zero, indicating that, on average, the predicted values are equal to the observed values.

Option A is incorrect because influential scores, which are data points that have a strong influence on the slope of the regression line, may or may not have large residuals. It depends on the specific data and the relationship between the predictor and response variables.

Option B is incorrect because the number of positive residuals does not have to be the same as the number of negative residuals even if the linear model is good. It is possible for the residuals to be skewed, meaning that there may be more positive or negative residuals depending on the nature of the data.

Option D is incorrect because even if the correlation between the predictor and response variables is 0, there may not necessarily be a distinct pattern in the residual plot. The presence of a pattern in the residual plot indicates that the linear model may not be appropriate for the data, but a lack of correlation does not guarantee the presence of a pattern in the residuals.

Therefore, the correct statement is C. 'The mean of the residuals is always zero.' Final answer: C

\*\*\*\*\*

Question 149:

'A reading specialist in a large public school system believes that the more time students spend reading, the better they will do in school. She plans a middle school experiment in which an SRS of 30 eighth graders will be assigned four extra hours of reading per week, an SRS of 30 seventh graders will be assigned two extra hours of reading per week, and an SRS of 30 sixth graders with no extra assigned reading will be a control group. After one school year, the mean GPAs from each group will be compared. Is this a good experimental design?', A. 'Yes.', B. 'No, because while this design may point out an association between reading and GPA, it cannot establish a cause-and-effect relationship.', C. 'No, because without blinding, there is a strong chance of a placebo effect.', D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

=====

Answer Option 1:

The answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

In this experimental design, grade level is not controlled for. Grade level is a lurking variable that is likely to be confounded with the variables under consideration, such as the impact of extra reading time on GPA. It is possible that the observed differences in GPA among the three groups are due to factors other than the extra reading time. For example, eighth graders may have higher GPAs on average compared to sixth graders regardless of the extra reading time. Therefore, it is not possible to determine whether the observed differences in GPA are solely due to the extra reading time or to the grade level.

Final answer: D.

\*\*\*\*\*

Answer Option 2:

The correct answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

Grade level is a lurking variable because it is related to both the amount of reading assigned and the GPA. The design of the experiment groups students based on grade level, which means that any observed differences in GPA between the groups may be due to differences in grade level rather than the amount of assigned reading. Therefore, grade level is confounded with the variables under consideration, making it difficult to determine the true effect of the extra assigned reading on GPA. To address this issue, a randomized controlled trial with random assignment of students to different reading groups would be a more appropriate design. Final answer: D.

\*\*\*\*\*

Answer Option 3:

The answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

In this experimental design, grade level is not controlled for. Grade level can potentially have a significant impact on GPA, and it is possible that any observed differences in GPA between the groups could be influenced by the differences in grade level rather than the extra assigned reading. Therefore, the design does not allow for a clear determination of the relationship between reading and GPA. To establish a cause-and-effect relationship, other potential confounding variables should be controlled for, such as prior academic performance or socioeconomic status. Final answer: D

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

In this experiment, the reading specialist plans to assign different amounts of extra reading time to different grade levels. However, grade level itself could be a lurking variable that affects both the amount of extra reading time and the GPA. For example, if eighth graders typically have higher GPAs than sixth graders, the observed difference in GPAs between the groups may be due to grade level rather than the extra reading time. Therefore, this design does not allow us to establish a cause-and-effect relationship between reading time and GPA. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The correct answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

In this experimental design, the grade level of the students is not being controlled or accounted for. Grade level can be a lurking variable that affects both the amount of time spent reading and the students' GPAs. For example, older students (eighth graders) may have more advanced reading skills and higher GPAs compared to younger students (sixth graders), regardless of the amount of assigned reading. Therefore, the results may be confounded by grade level, making it difficult to establish a cause-and-effect relationship between reading and GPA. To improve the experimental design, the researcher should consider implementing randomized block design or stratified sampling to account for grade level as a confounding variable. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The correct answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

The experimental design described in the scenario does not account for the potential influence of grade level on the relationship between reading time and GPA. Grade level is a lurking variable that could confound the results and lead to incorrect conclusions. For example, if the eighth-grade group ends up with higher GPAs on average, it could be attributed to the extra reading time, but it may simply be because they are in a higher grade.

To address this issue, a better experimental design would involve randomly assigning students from all three grade levels to different reading time conditions. This would help eliminate the potential confounding effect of grade level and provide a more accurate assessment of the impact of reading time on GPA.

\*\*\*\*\*

Answer Option 7:

The correct answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

In this experimental design, the grade level of the students is not controlled or accounted for. Grade level is a lurking variable that may have a significant impact on students' GPAs. The experiment compares students from different grade levels, which introduces confounding variables that can affect the results. For example, eighth graders may have more advanced academic skills compared to sixth graders, which could lead to higher GPAs regardless of the extra reading time.

To establish a cause-and-effect relationship between reading time and GPA, the experiment should randomly assign students from the same grade level into different groups, with one group receiving extra reading time and the other group not. This would help control for the lurking variable of grade level and provide more meaningful results.

\*\*\*\*\*

Answer Option 8:

The correct answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

In this experimental design, the variable being tested is the amount of extra reading assigned to students. However, the grade level of the students is not controlled for. Grade level could be a lurking variable that is correlated with both the amount of extra reading assigned and GPA. For example, if eighth graders tend to have higher GPAs than seventh graders or sixth graders, the results of the experiment may be influenced by the grade level rather than the extra reading assigned.

To eliminate the confounding effect of grade level, a better experimental design would be to randomly assign students from all three grade levels into different treatment groups. This would ensure that any differences in GPA observed between the groups can be attributed to the amount of extra reading assigned, rather than grade level. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The answer is D. 'No, because grade level is a lurking variable which may well be confounded with the variables under consideration.'

In this experiment, the independent variable is the extra hours of reading assigned per week, and the dependent variable is the students' GPAs. However, the grade level of the students is a lurking variable that can confound the results. There may be inherent differences in academic performance between eighth graders, seventh graders, and sixth graders that are unrelated to the amount of extra reading assigned. Therefore, any observed differences in GPAs between the groups cannot be solely attributed to the extra reading.

To improve the experimental design, it would be better to randomize the assignment of extra reading hours across all grade levels to ensure that any differences in GPAs can be attributed to the amount of reading rather than grade level. Additionally, blinding the students and teachers to the specific condition (extra reading or control) can help mitigate any potential placebo effect.

\*\*\*\*\*

Question 150:

'What is a sampling distribution?', A. 'A distribution of all the statistics that can be found in a given sample', B. 'A histogram, or other such visual representation, showing the distribution of a sample', C. 'A normal distribution of some statistic', D. 'A distribution of all the values taken by a statistic from all possible samples of a given size'

=====

Answer Option 1:

A sampling distribution refers to the distribution of a statistic, such as the mean or standard deviation, that is calculated from different samples of the same size taken from a population. This distribution shows how the statistic varies across different samples. Therefore, the correct answer is D. 'A distribution of all the values taken by a statistic from all possible samples of a given size'. Final Answer: D

\*\*\*\*\*

Answer Option 2:

A sampling distribution refers to the distribution of a statistic calculated from multiple samples of the same size, taken from the same population. It represents all the possible values that the statistic can take when calculated from different samples.

Therefore, the correct answer is D. 'A distribution of all the values taken by a statistic from all possible samples of a given size'.

\*\*\*\*\*

Answer Option 3:

A sampling distribution is a theoretical distribution that represents all the possible values of a statistic that can be obtained from repeated sampling of a population. In other words, it is a distribution of all the values taken by a statistic from all possible samples of a given size. Therefore, the correct answer is D: 'A distribution of all the values taken by a statistic from all possible samples of a given size'. Final

answer: D

\*\*\*\*\*

Answer Option 4:

A sampling distribution is the distribution of all the values taken by a statistic from all possible samples of a given size. Therefore, the correct answer is  $\boxed{\text{D. 'A distribution of all the values taken by a statistic from all possible samples of a given size'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 5:

A sampling distribution is a theoretical distribution that represents all the possible values or outcomes of a statistic, calculated from all possible samples of a given size, taken from a population. Therefore, the correct answer is D. 'A distribution of all the values taken by a statistic from all possible samples of a given size'. Final answer: D

\*\*\*\*\*

Answer Option 6:

A sampling distribution refers to option D, 'A distribution of all the values taken by a statistic from all possible samples of a given size'. In other words, it represents the distribution of statistics that can be calculated from different samples of the same population. This distribution provides information about the variability and properties of the statistic being measured. Final answer: D.

\*\*\*\*\*

Answer Option 7:

A sampling distribution refers to the distribution of a statistic calculated from multiple samples taken from the same population. In other words, it is the distribution of all the values taken by a statistic from all possible samples of a given size. Therefore, the correct answer is  $\boxed{\text{D. 'A distribution of all the values taken by a statistic from all possible samples of a given size'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

A sampling distribution refers to the distribution of a statistic calculated from multiple samples of the same size drawn from the same population. This means that the sampling distribution represents all the possible values that the statistic can take when calculated from different samples. Therefore, the correct answer is D, 'A distribution of all the values taken by a statistic from all possible samples of a given size'. Final answer: D

\*\*\*\*\*

Answer Option 9:

A sampling distribution refers to option D, which is "A distribution of all the values taken by a statistic from all possible samples of a given size." In other words, a sampling distribution represents the distribution of a statistic (such as the mean or proportion) that is calculated from different samples of the same size taken from a population. It provides information about how the statistic varies across different samples and helps to draw inferences about the population based on sample data. Final Answer: D

\*\*\*\*\*

Question 151:

A farmer wants to know whether a new fertilizer has increased the mean weight of his apples. With the old fertilizer, the mean weight was 4.0 ounces per apple. The farmer decides to test  $H_0: \mu = 4.0$  ounces versus  $H_a: \mu > 4.0$  ounces, at a 5 percent level of significance, where  $\mu$  = the mean weight of apples using the new fertilizer. The weights of apples are approximately normally distributed. The farmer takes a random sample of 16 apples and computes a mean of 4.3 ounces and a standard deviation of 0.6 ounces. Which of the following gives the p-value for this test?, A. 'P(Z > 2)', B. 'P(Z < 2)', C. 'P(t > 2) with 15 degrees of freedom', D. 'P(t < 2) with 15 degrees of freedom'

=====

Answer Option 1:

To test the hypothesis, we calculate the test statistic, which is the z-score in this case. The formula for the z-score is given by:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

where  $\bar{x}$  is the sample mean,  $\mu$  is the hypothesized population mean,  $\sigma$  is the population standard deviation, and  $n$  is the sample size.

Plugging in the values from the problem, we have:

$$\bar{x} = 4.3, \mu = 4.0, \sigma = 0.6, n = 16$$

$$Z = \frac{4.3 - 4.0}{0.6 / \sqrt{16}} = \frac{0.3}{0.15} = 2$$

Since we are testing the hypothesis that the mean weight is greater than 4.0 ounces, our alternative hypothesis is one-sided and we are looking for the probability that the z-score is greater than 2. The p-value for this test is the probability of observing a z-score greater than 2.

Therefore, the correct answer is  $P(Z > 2)$ .

\*\*\*\*\*

Answer Option 2:

To test the hypothesis, we can use a one-sample t-test since the standard deviation of the population is not known. We compute the test statistic:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{4.3 - 4.0}{0.6 / \sqrt{16}} = \frac{0.3}{0.15} = 2$$

Since the alternative hypothesis is that the mean weight is greater than 4.0 ounces, we are interested in the right tail area. The p-value is the probability of observing a t-value greater than or equal to 2. With 15 degrees of freedom, the t-distribution is symmetric, so we have

$$P(t > 2) = P(t > 2) \text{ with 15 degrees of freedom}$$

Final answer: C

\*\*\*\*\*

Answer Option 3:

To test the hypothesis, we can use a one-sample t-test since we have the sample mean, sample standard deviation, and the sample size. The test statistic is given by

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

where  $\bar{x}$  is the sample mean,  $\mu$  is the hypothesized mean,  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case, we have  $\bar{x} = 4.3$ ,  $\mu = 4.0$ ,  $s = 0.6$ , and  $n = 16$ . Substituting these values into the formula, we get

$$t = \frac{4.3 - 4.0}{0.6 / \sqrt{16}} = \frac{0.3}{0.15} = 2$$

To find the p-value, we need to calculate the probability of obtaining a t-value greater than 2 (since we are testing  $H_a: \mu > 4.0$ ). Since the sample size is 16, we have 15 degrees of freedom.

Therefore, the p-value for this test is given by  $P(t > 2)$  with 15 degrees of freedom. Thus, the correct answer is  $P(t > 2)$  with 15 degrees of freedom.

\*\*\*\*\*

Answer Option 4:

To test the null hypothesis  $H_0: \mu = 4.0$  versus the alternative hypothesis  $H_a: \mu > 4.0$ , we can use a one-sample t-test since the population standard deviation is unknown and we have a sample size of 16, which is large enough to apply the Central Limit Theorem.

The test statistic is calculated as

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

where  $\bar{X}$  is the sample mean,  $\mu_0$  is the hypothesized mean under the null hypothesis,  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case,  $\bar{X} = 4.3$ ,  $\mu_0 = 4.0$ ,  $s = 0.6$ , and  $n = 16$ . Therefore,

$$t = \frac{4.3 - 4.0}{0.6 / \sqrt{16}} = \frac{0.3}{0.15} = 2$$

To find the p-value, we compare the test statistic to the t-distribution with  $n-1 = 15$  degrees of freedom. Since the alternative hypothesis is one-sided ( $\mu > 4.0$ ), we are interested in the upper tail probability.

Therefore, the p-value for this test is given by  $P(t > 2)$  with 15 degrees of freedom. Final Answer: C

\*\*\*\*\*

Answer Option 5:

To test the hypothesis, we can use a one-sample t-test since the population standard deviation is not known. The test statistic is given by

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

where  $\bar{x}$  is the sample mean,  $\mu$  is the hypothesized mean,  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case, we have  $\bar{x} = 4.3$ ,  $\mu = 4.0$ ,  $s = 0.6$ , and  $n = 16$ . Plugging in these values, we get

$$t = \frac{4.3 - 4.0}{0.6 / \sqrt{16}} = \frac{0.3}{0.15} = 2$$

Since we are testing whether the mean weight is greater than 4.0 ounces, we are interested in the right-tailed p-value. The p-value is the probability of observing a test statistic as extreme as the one calculated (or more extreme) under the null hypothesis.

Therefore, the p-value for this test is the probability that a t-distribution with 15 degrees of freedom is greater than 2. The correct choice is  $\boxed{\text{P}(t > 2) \text{ with 15 degrees of freedom}}$ . Final Answer: C

Answer Option 6:

To test the null hypothesis  $H_0: \mu = 4.0$  versus the alternative hypothesis  $H_a: \mu > 4.0$ , we will use a one-sample t-test because the population standard deviation is unknown and the sample size is small ( $n = 16$ ).

The test statistic is given by

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where  $\bar{x}$  is the sample mean,  $\mu_0$  is the hypothesized mean under the null hypothesis,  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case, we have  $\bar{x} = 4.3$ ,  $\mu_0 = 4.0$ ,  $s = 0.6$ , and  $n = 16$ . Plugging these values into the formula, we get

$$t = \frac{4.3 - 4.0}{0.6/\sqrt{16}} = \frac{0.3}{0.6/4} = 2$$

Since we are testing  $\mu > 4.0$ , we are interested in the right-tail of the t-distribution. The p-value is the probability of observing a t-value greater than or equal to 2. The p-value can be found using a t-table or a calculator, and it corresponds to the area under the t-distribution curve to the right of 2.

Therefore, the correct answer is  $\boxed{\text{P}(t > 2) \text{ with 15 degrees of freedom}}$ . Final answer: C

Answer Option 7:

To determine the p-value for this test, we first calculate the test statistic using the formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where  $\bar{x}$  is the sample mean,  $\mu$  is the hypothesized population mean,  $s$  is the sample standard deviation, and  $n$  is the sample size.

Plugging in the given values, we have:

$$t = \frac{4.3 - 4.0}{0.6/\sqrt{16}} = \frac{0.3}{0.15} = 2$$

Since we are testing the hypothesis that the mean weight is greater than 4.0 ounces, we are only concerned with the right tail of the t-distribution.

To find the p-value, we compare the test statistic to the t-distribution with  $n-1$  degrees of freedom. In this case,  $n = 16$ , so we use the t-distribution with 15 degrees of freedom.

The p-value is defined as the probability of observing a more extreme value than the test statistic, given the null hypothesis is true.

Since we are testing against the alternative hypothesis that the mean is greater than 4.0 ounces, we are interested in the probability of observing a t-value greater than 2. The p-value for this test is denoted as  $P(t > 2)$  with 15 degrees of freedom.

Therefore, the correct answer is  $\boxed{\text{C. } P(t > 2) \text{ with 15 degrees of freedom}}$ .

Answer Option 8:

To test the null hypothesis  $H_0: \mu = 4.0$  versus the alternative hypothesis  $H_a: \mu > 4.0$ , we can use a one-sample t-test since the population standard deviation is unknown.

The test statistic for the one-sample t-test is given by:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where  $\bar{x}$  is the sample mean,  $\mu_0$  is the hypothesized population mean,  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case,  $\bar{x} = 4.3$ ,  $\mu_0 = 4.0$ ,  $s = 0.6$ , and  $n = 16$ . Substituting these values into the formula, we get:

$$t = \frac{4.3 - 4.0}{0.6/\sqrt{16}} = \frac{0.3}{0.15} = 2$$

The p-value for this one-tailed test can be calculated as the probability of observing a t-value greater than 2. Since we are testing the alternative hypothesis that the mean weight of the apples is greater than 4.0 ounces, the p-value is given by  $P(t > 2)$  with 15 degrees of freedom.

Therefore, the correct answer is  $\boxed{\text{C. } P(t > 2) \text{ with 15 degrees of freedom}}$ . Final Answer: C

Answer Option 9:

To test the null hypothesis  $H_0: \mu = 4.0$  versus the alternative hypothesis  $H_a: \mu > 4.0$ , we can use a one-sample t-test since the population standard deviation is unknown.

The test statistic is given by  $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ , where  $\bar{x}$  is the sample mean,  $\mu$  is the hypothesized mean,  $s$  is the sample standard deviation, and  $n$  is the sample size.

In this case, we have  $\bar{x} = 4.3$ ,  $\mu = 4.0$ ,  $s = 0.6$ , and  $n = 16$ . Plugging these values in, we get:

$$t = \frac{4.3 - 4.0}{0.6/\sqrt{16}} = \frac{0.3}{0.15} = 2$$

The p-value is the probability of observing a test statistic as extreme or more extreme than the observed test statistic, assuming the null hypothesis is true. Since we are testing  $H_a: \mu > 4.0$ , we are interested in the upper tail of the t-distribution.

Since the sample size is small (16), we use the t-distribution instead of the standard normal distribution. The degrees of freedom is  $n - 1 = 16 - 1 = 15$ .

Therefore, the p-value for this test is given by  $P(t > 2)$  with 15 degrees of freedom. Hence, the answer is  $\boxed{\text{C. } P(t > 2) \text{ with 15 degrees of freedom}}$ . Final Answer: C

Question 152:

'In a study on the effect of music on worker productivity, employees were told that a different genre of background music would be played each day and the corresponding production outputs noted. Every change in music resulted in an increase in production. This is an example of', A. 'the effect of a treatment unit.', B. 'the placebo effect.', C. 'the control group effect.', D. 'sampling error.'

=====

Answer Option 1:

The scenario described is an example of the placebo effect. The employees' belief that a different genre of background music would result in an increase in production led to an actual increase in production, even though the music itself may not have had any direct impact on productivity. Therefore, the correct answer is B. 'the placebo effect.' Final Answer: B

\*\*\*\*\*

Answer Option 2:

The given scenario describes a situation where the change in music resulted in an increase in production. This suggests that the observed effect is not due to the actual treatment (the genre of background music) but rather due to the participants' belief or expectation that the change in music would lead to an increase in productivity. This phenomenon is known as the placebo effect.

Therefore, the correct answer is B. 'the placebo effect.' Final answer: B

\*\*\*\*\*

Answer Option 3:

The given scenario describes a situation where a change in the independent variable (in this case, the genre of background music) leads to a change in the dependent variable (in this case, worker productivity). This suggests that the observed increase in production is likely a result of the actual effect of the music on productivity, rather than any other factors. Therefore, the correct answer is A. 'the effect of a treatment unit.' Final Answer: A

\*\*\*\*\*

Answer Option 4:

The scenario described in the question is an example of the placebo effect. The employees were told that a different genre of background music would be played each day, and believing that the change in music would affect their productivity, their production outputs increased. The actual effect of the music on productivity may be negligible or non-existent, but the employees' belief in the music's effect led to an increase in production. Therefore, the correct answer is B. 'the placebo effect.' Final Answer: B

\*\*\*\*\*

Answer Option 5:

The scenario described is an example of the placebo effect. The employees were told that different genres of background music would be played each day, implying that the music itself had an effect on their productivity. However, the increase in production was not actually due to the music, but rather due to the employees' belief that the music would have a positive impact. The placebo effect occurs when a person's belief or expectation influences their response to a treatment or intervention, even if the treatment itself has no real effect. Therefore, the correct answer is  $\boxed{\text{B. 'the placebo effect.'}}$ .

\*\*\*\*\*

Answer Option 6:

The given scenario describes a situation where the change in music is associated with an increase in production. However, there is no mention of any actual treatment or intervention being applied. Therefore, the correct answer is not A. "the effect of a treatment unit."

The scenario also does not involve the use of a placebo, where a placebo treatment is given to participants and they experience improvements due to the belief that they are receiving an effective treatment.

Therefore, the correct answer is not B. "the placebo effect."

Similarly, the scenario does not involve a control group, which is a group that does not receive any treatment or intervention. Therefore, the correct answer is not C. "the control group effect."

Sampling error refers to the error or discrepancy that may occur when the sample chosen for a study does not accurately represent the population. This term is not applicable to the given scenario. Therefore, the correct answer is not D. "sampling error."

Based on the information provided, none of the given options accurately describe the scenario. It is possible that the correct answer is not included in the options provided.

\*\*\*\*\*

Answer Option 7:

This scenario describes the placebo effect. The employees believed that the change in music would improve their productivity, and as a result, their production outputs increased. However, in reality, the change in music had no direct effect on their productivity. The improvement was purely psychological, driven by their expectation or belief that the music would make them more productive. Therefore, the correct answer is  $\boxed{\text{B. the placebo effect}}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

The given scenario describes a situation where the change in music resulted in an increase in production. This suggests that the increase in productivity is directly attributed to the change in the independent variable, which is the music genre. Therefore, this is an example of "the effect of a treatment unit." Hence, the correct answer is A. "the effect of a treatment unit."

\*\*\*\*\*

Answer Option 9:

The scenario described is an example of the placebo effect. The employees were told that different genres of background music were being played, and this belief alone resulted in an increase in production.

The actual music itself may not have had any direct effect on productivity, but rather the employees' perception of the music influenced their performance. Therefore, the correct answer is  $\boxed{\text{B.}}$  the placebo effect.

\*\*\*\*\*



Question 153:

'In a clinical trial, 30 sickle cell anemia patients are randomly assigned to two groups. One group receives the currently marketed medicine, and the other group receives an experimental medicine. Each week, patients report to the clinic where blood tests are conducted. The lab technician is unaware of the kind of medicine the patient is taking. This design can be described as', A. 'a completely randomized design, with the currently marketed medicine and the experimental medicine as two treatments', B. 'a matched-pairs design, with the currently marketed medicine and the experimental medicine forming a pair', C. 'a randomized block design, with the currently marketed medicine and the experimental medicine as two blocks', D. 'a randomized block design, with the currently marketed medicine and the experimental medicine as two treatments'

=====

Answer Option 1:

The design described in the given scenario is a randomized block design. In a randomized block design, the subjects are divided into blocks based on certain characteristics that may influence the response variable. In this case, the patients are divided into two groups based on the type of medicine they receive. This grouping into blocks (groups) allows for better control and comparison between the two treatments. Each group (block) receives a different treatment (currently marketed medicine or experimental medicine), and within each group, the patients are randomly assigned to their respective treatments. Therefore, the correct answer is  $\boxed{\text{C. a randomized block design, with the currently marketed medicine and the experimental medicine as two blocks}}$ . Final answer: C

\*\*\*\*\*

Answer Option 2:

The design described in the scenario is a randomized block design. In a randomized block design, the participants are divided into blocks based on certain characteristics that may affect the outcome. In this case, the two groups (currently marketed medicine and experimental medicine) can be considered as the blocks. The participants within each block are then randomly assigned to the treatments.

The design is not a matched-pairs design because a matched-pairs design involves pairing participants who are similar in certain characteristics and then assigning each pair to different treatments.

Therefore, the correct answer is  $\boxed{\text{C. a randomized block design, with the currently marketed medicine and the experimental medicine as two blocks}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 3:

The design described in the problem is a randomized block design. In a randomized block design, subjects are divided into blocks based on certain characteristics that may affect the response variable. In this case, the blocks are the sickle cell anemia patients, and they are divided into two groups. The currently marketed medicine and the experimental medicine are the treatments being compared. The random assignment of patients to the two groups ensures that any potential confounding variables are distributed evenly between the groups. Therefore, the correct answer is  $\boxed{\text{(C) a randomized block design, with the currently marketed medicine and the experimental medicine as two blocks}}$ .

\*\*\*\*\*

Answer Option 4:

The design described in the problem can be characterized as a randomized block design, with the currently marketed medicine and the experimental medicine as two treatments.

In a randomized block design, subjects are divided into blocks based on certain characteristics that may affect the outcome. In this case, the patients are divided into two groups based on their condition of having sickle cell anemia. Within each block, the treatments (currently marketed medicine and experimental medicine) are randomly assigned to the patients. This helps to account for any potential differences or confounding factors between the two groups.

Therefore, the correct answer is C. 'a randomized block design, with the currently marketed medicine and the experimental medicine as two blocks'. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The design described in the scenario is a randomized block design. In a randomized block design, the subjects are divided into blocks based on specific characteristics, and then within each block, the treatments are randomly assigned. In this case, the blocks are the sickle cell anemia patients, and the two treatments are the currently marketed medicine and the experimental medicine. Therefore, the answer is  $\boxed{\text{C. 'a randomized block design, with the currently marketed medicine and the experimental medicine as two blocks'}}$ . Final answer: C

\*\*\*\*\*

Answer Option 6:

The design described in the scenario is a randomized block design, with the currently marketed medicine and the experimental medicine as two treatments. In a randomized block design, the patients are divided into blocks based on certain characteristics (in this case, sickle cell anemia patients) and then randomly assigned to different treatments within each block. This design helps to control for any potential confounding variables or differences among the patients in each block. Therefore, the correct answer is  $\boxed{\text{D. a randomized block design, with the currently marketed medicine and the experimental medicine as two treatments}}$ . Final Answer: D.

\*\*\*\*\*

Answer Option 7:

The given design can be described as a randomized block design, with the currently marketed medicine and the experimental medicine as two treatments. In a randomized block design, subjects are divided into blocks (in this case, the blocks are the two groups of patients receiving different medicines) and then randomly assigned to treatments within each block. This design helps to control for any potential confounding variables or differences between the groups by ensuring that each treatment is represented within each block. Therefore, the correct answer is  $\boxed{\text{D. a randomized block design, with the currently marketed medicine and the experimental medicine as two treatments}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

The design described in the scenario is a randomized block design, with the currently marketed medicine and the experimental medicine as two treatments. In a randomized block design, subjects are divided into blocks based on certain characteristics, and then each treatment is randomly assigned within each block. In this case, the two groups (currently marketed medicine and experimental medicine) are the blocks, and within each group, the patients are randomly assigned to receive one of the two treatments. Therefore, the correct answer is  $\boxed{\text{(C) a randomized block design, with the currently}}$

marketed medicine and the experimental medicine as two blocks))\$. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The design described in the problem can be classified as a completely randomized design, with the currently marketed medicine and the experimental medicine as two treatments. This is because the patients are randomly assigned to the two groups, and there is no specific pairing or blocking involved in the assignment process. Therefore, the correct answer is  $\boxed{\text{A. a completely randomized design, with the currently marketed medicine and the experimental medicine as two treatments}})$ \$.  
\*\*\*\*\*

Question 154:

'Suppose that the weights of trucks traveling on the interstate highway system are normally distributed. If 70% of the trucks weigh more than 12,000 pounds and 80% weigh more than 10,000 pounds, what are the mean and standard deviation for the weights of trucks traveling on the interstate system?', A. ' $\mu = 14,900$ ;  $\sigma = 6100$ ', B. ' $\mu = 15,100$ ;  $\sigma = 6200$ ', C. ' $\mu = 15,300$ ;  $\sigma = 6300$ ', D. ' $\mu = 15,500$ ;  $\sigma = 6400$ '

=====

Answer Option 1:

Let  $X$  be a random variable representing the weight of the trucks. We are given that 70% of the trucks weigh more than 12,000 pounds. This can be expressed as  $P(X > 12,000) = 0.70$ . Using the standard normal distribution, we can find that the z-score corresponding to a probability of 0.70 is approximately 0.524. Therefore, we have:

$$\frac{12,000 - \mu}{\sigma} = 0.524$$

Similarly, we are given that 80% of the trucks weigh more than 10,000 pounds, which can be expressed as  $P(X > 10,000) = 0.80$ . Using the standard normal distribution, the z-score corresponding to a probability of 0.80 is approximately 0.842. Therefore, we have:

$$\frac{10,000 - \mu}{\sigma} = 0.842$$

Solving the system of equations formed by these two equations, we can find the values of  $\mu$  and  $\sigma$ . After solving the system, we find that  $\mu = 15,100$  and  $\sigma = 6,200$ . Therefore, the answer is  $\boxed{\text{B}}$   $\mu = 15,100$ ;  $\sigma = 6,200$ . Final answer: B

\*\*\*\*\*

Answer Option 2:

Let  $X$  be the weight of a truck traveling on the interstate system. We are given that 70% of trucks weigh more than 12,000 pounds, so  $P(X > 12,000) = 0.70$ . Similarly, 80% of trucks weigh more than 10,000 pounds, so  $P(X > 10,000) = 0.80$ .

From the properties of the normal distribution, we know that  $P(X > 12,000) = 1 - P(X \leq 12,000)$  and  $P(X > 10,000) = 1 - P(X \leq 10,000)$ . Using these equations, we can find the values of  $P(X \leq 12,000)$  and  $P(X \leq 10,000)$ .

$$P(X \leq 12,000) = 1 - P(X > 12,000) = 1 - 0.70 = 0.30$$

$$P(X \leq 10,000) = 1 - P(X > 10,000) = 1 - 0.80 = 0.20$$

Next, we need to find the z-scores corresponding to these probabilities. Using a standard normal distribution table or calculator, we can find that the z-score for  $P(X \leq 12,000)$  is approximately -0.524 and the z-score for  $P(X \leq 10,000)$  is approximately -0.842.

Now, we can use the z-score formula to find the mean and standard deviation:

$$z = \frac{x - \mu}{\sigma}$$

For  $P(X \leq 12,000)$ ,

$$-0.524 = \frac{12,000 - \mu}{\sigma}$$

For  $P(X \leq 10,000)$ ,

$$-0.842 = \frac{10,000 - \mu}{\sigma}$$

Solving this system of equations, we find  $\mu \approx 15,100$  and  $\sigma \approx 6,200$ .

Therefore, the mean and standard deviation for the weights of trucks traveling on the interstate system are approximately  $\mu = 15,100$  and  $\sigma = 6,200$ .

The correct answer is  $\boxed{\text{B}}$   $\mu = 15,100$ ;  $\sigma = 6,200$ .

\*\*\*\*\*

Answer Option 3:

Let  $X$  be a random variable representing the weight of trucks traveling on the interstate system. We are given that 70% of the trucks weigh more than 12,000 pounds, which means that  $P(X > 12,000) = 0.70$ . Similarly, we are given that 80% of the trucks weigh more than 10,000 pounds, which means that  $P(X > 10,000) = 0.80$ .

Using the properties of the normal distribution, we know that  $P(X > 12,000) = 1 - P(X \leq 12,000)$ , and  $P(X > 10,000) = 1 - P(X \leq 10,000)$ . Plugging in the given values, we have  $1 - P(X \leq 12,000) = 0.70$  and  $1 - P(X \leq 10,000) = 0.80$ .

Subtracting  $1 - P(X \leq 12,000) = 0.70$  from  $1 - P(X \leq 10,000) = 0.80$ , we get  $P(10,000 < X \leq 12,000) = 0.10$ . This means that 10% of the trucks weigh between 10,000 and 12,000 pounds.

To find the mean and standard deviation, we can use the Z-score formula. Let  $Z$  be the standard normal variable. We have

$$P(10,000 < X \leq 12,000) = P\left(\frac{10,000 - \mu}{\sigma} < Z \leq \frac{12,000 - \mu}{\sigma}\right) = 0.10$$

Since we know that 10% of the standard normal distribution lies between  $Z = \frac{10,000 - \mu}{\sigma}$  and  $Z = \frac{12,000 - \mu}{\sigma}$ , we can use a standard normal table to find that the Z-scores are approximately -1.28 and 1.28.

Therefore, we have

$$\frac{10,000 - \mu}{\sigma} = -1.28 \quad \text{and} \quad \frac{12,000 - \mu}{\sigma} = 1.28$$

Solving these equations, we find that  $\mu \approx 15,100$  and  $\sigma \approx 6,200$ .

Thus, the mean and standard deviation for the weights of trucks traveling on the interstate system are approximately  $\mu = 15,100$  and  $\sigma = 6,200$ . The correct answer choice is  $\boxed{\text{B}}$ .

Final Answer: B

\*\*\*\*\*

Answer Option 4:

To solve this problem, we can use the properties of the normal distribution.

Let  $\mu$  be the mean and  $\sigma$  be the standard deviation of the weights of trucks. We know that 70% of the trucks weigh more than 12,000 pounds, which means that the area to the right of 12,000 under the normal distribution curve is 0.70. Similarly, 80% of the trucks weigh more than 10,000 pounds, so the area to the right of 10,000 under the normal distribution curve is 0.80.

From the given information, we can calculate the z-scores corresponding to these areas using a standard normal distribution table or a calculator. The z-score corresponding to an area of 0.70 is approximately 0.524 and the z-score corresponding to an area of 0.80 is approximately 0.841.

Using the z-score formula, we have:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value of the observation,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For  $x = 12,000$ , we have:

$$0.524 = \frac{12,000 - \mu}{\sigma}$$

and for  $x = 10,000$ , we have:

$$0.841 = \frac{10,000 - \mu}{\sigma}$$

We can solve these two equations to find  $\mu$  and  $\sigma$ . Dividing the first equation by the second equation, we get:

$$\frac{0.524}{0.841} = \frac{12,000 - \mu}{10,000 - \mu}$$

Simplifying, we obtain:

$$0.524(10,000 - \mu) = 0.841(12,000 - \mu)$$

Expanding and rearranging, we have:

$$5240 - 0.524\mu = 10092 - 0.841\mu$$

Solving for  $\mu$ , we find:

$$0.317\mu = 513.908$$

$$\mu \approx 15,300$$

Substituting this value of  $\mu$  into either of the original equations, we can solve for  $\sigma$ . Using the equation with  $x = 12,000$ , we have:

$$0.524 = \frac{12,000 - 15,300}{\sigma}$$

Simplifying, we obtain:

$$0.524\sigma = -3,300$$

$$\sigma \approx 6,300$$

Therefore, the mean and standard deviation for the weights of trucks traveling on the interstate system are approximately  $\mu = 15,300$  and  $\sigma = 6,300$ , so the correct answer is  $\boxed{\text{(C)}}$

$\mu = 15,300$ ;  $\sigma = 6,300$ ).

\*\*\*\*\*

Answer Option 5:

Let  $X$  be a random variable representing the weights of the trucks. We know that  $P(X > 12,000) = 0.70$  and  $P(X > 10,000) = 0.80$ .

Using the standard normal distribution, we can find the corresponding z-scores for these probabilities.

For  $P(X > 12,000) = 0.70$ , the z-score is given by  $Z = \frac{X - \mu}{\sigma}$ , where  $\mu$  is the mean and  $\sigma$  is the standard deviation. Solving for the z-score, we have  $Z = \frac{12,000 - \mu}{\sigma}$ .

We can use the z-score table or calculator to find that the z-score corresponding to a probability of 0.70 is approximately 0.5244.

Similarly, for  $P(X > 10,000) = 0.80$ , the z-score is given by  $Z = \frac{10,000 - \mu}{\sigma}$ . Using the z-score table or calculator, we find that the z-score corresponding to a probability of 0.80 is approximately 0.8416.

Now we have two equations:  $0.5244 = \frac{(12000 - \mu)}{\sigma}$  and  $0.8416 = \frac{(10000 - \mu)}{\sigma}$ .

We can solve these equations to find the values of  $\mu$  and  $\sigma$ . Subtracting the second equation from the first, we get:

$$0.5244 - 0.8416 = \frac{(12000 - \mu)}{\sigma} - \frac{(10000 - \mu)}{\sigma}$$

$$-0.3172 = \frac{(2000 - \mu)}{\sigma}$$

Simplifying, we find:

$$2000 - \mu = -0.3172\sigma$$

$$\mu = 2000 + 0.3172\sigma$$

Substituting this value of  $\mu$  into the second equation, we have:

$$0.8416 = \frac{(10000 - (2000 + 0.3172\sigma))}{\sigma}$$

Simplifying and solving for  $\sigma$ , we find:

$$0.8416 = \frac{(8000 - 0.3172\sigma)}{\sigma}$$

$$0.8416\sigma = 8000 - 0.3172\sigma$$

$$1.1588\sigma = 8000$$

$$\sigma = 6904.3$$

Substituting this value of  $\sigma$  back into the equation for  $\mu$ , we find:

$$\mu = 2000 + 0.3172(6904.3)$$

$$\mu = 2200 + 2191.9$$

$$\mu = 4391.9$$

Therefore, the mean and standard deviation for the weights of trucks traveling on the interstate system are approximately  $\mu = 4391.9$  and  $\sigma = 6904.3$ .

The closest option is  $\mu = 4391.9$ ;  $\sigma = 6904.3$ , which corresponds to answer choice D.

Final Answer: D

\*\*\*\*\*

Answer Option 6:

Let  $X$  be a random variable representing the weights of the trucks. We are given that  $P(X > 12,000) = 0.70$  and  $P(X > 10,000) = 0.80$ .

From the first condition, we have:

$$P(X > 12,000) = P\left(\frac{X - \mu}{\sigma} > \frac{12,000 - \mu}{\sigma}\right) = 0.70$$

where  $\mu$  is the mean and  $\sigma$  is the standard deviation.

Using the standard normal distribution table, we find that  $\frac{12,000 - \mu}{\sigma}$  is approximately equal to  $0.524$ .

From the second condition, we have:

$$P(X > 10,000) = P\left(\frac{X - \mu}{\sigma} > \frac{10,000 - \mu}{\sigma}\right) = 0.80$$

Using the standard normal distribution table, we find that  $\frac{10,000 - \mu}{\sigma}$  is approximately equal to  $0.841$ .

Now, we can solve for  $\mu$  and  $\sigma$  by setting up a system of equations:

$$\begin{cases}$$

$$\frac{12,000 - \mu}{\sigma} = 0.524$$

$$\frac{10,000 - \mu}{\sigma} = 0.841$$

$$\end{cases}$$

Solving this system, we find that  $\mu \approx 15,100$  and  $\sigma \approx 6,200$ . Therefore, the correct answer is  $\boxed{\text{(B) } \mu = 15,100; \sigma = 6,200}$ . Final Answer: B

\*\*\*\*\*

Answer Option 7:

Let  $X$  be the weight of a truck traveling on the interstate system. We are given that 70% of the trucks weigh more than 12,000 pounds, so  $P(X > 12,000) = 0.7$ . Similarly, we are given that 80% of the trucks weigh more than 10,000 pounds, so  $P(X > 10,000) = 0.8$ .

Using the standard normal distribution, we can find the corresponding z-scores for these probabilities. From the z-table, we find that  $P(Z > 0.524) = 0.7$  and  $P(Z > 0.842) = 0.8$ .

Now, we can use the z-score formula to find the corresponding values of  $X$ :

$$0.524 = \frac{12,000 - \mu}{\sigma} \quad \text{and} \quad 0.842 = \frac{10,000 - \mu}{\sigma}$$

Solving these two equations simultaneously, we can find the values of  $\mu$  and  $\sigma$ . Multiplying the first equation by  $0.842$  and the second equation by  $0.524$ , and then subtracting the two equations, we get:

$$0.842(12,000 - \mu) - 0.524(10,000 - \mu) = 0.318\mu - 4,400 = 0$$

Solving for  $\mu$ , we find  $\mu = 13,836.48$ . Substituting this value back into one of the original equations, we can solve for  $\sigma$ :

$$\sigma = \frac{12,000 - \mu}{0.524} = \frac{12,000 - 13,836.48}{0.524} \approx 6,101.92$$

Therefore, the mean and standard deviation for the weights of trucks traveling on the interstate system are approximately  $\mu = 13,836.48$  and  $\sigma = 6,101.92$ , respectively. The closest answer choice is  $\boxed{\text{(A) } \mu = 14,900; \sigma = 6,100}$ . Final Answer: A

\*\*\*\*\*

Answer Option 8:

Let  $X$  be the weight of a truck. We know that  $P(X > 12,000) = 0.70$  and  $P(X > 10,000) = 0.80$ .

From the first condition, we have  $P\left(Z > \frac{12,000 - \mu}{\sigma}\right) = 0.70$ , where  $Z$  is a standard normal random variable. Using a standard normal table or calculator, we find that  $Z \approx 0.5244$ . Solving for  $\frac{12,000 - \mu}{\sigma}$ , we get  $\frac{12,000 - \mu}{\sigma} \approx 0.5244$ .

Similarly, from the second condition, we have  $P\left(Z > \frac{10,000 - \mu}{\sigma}\right) = 0.80$ , which implies  $\frac{10,000 - \mu}{\sigma} \approx 0.8416$ .

Dividing the two equations, we have

$$\frac{\frac{12,000 - \mu}{\sigma}}{\frac{10,000 - \mu}{\sigma}} \approx \frac{0.5244}{0.8416}$$

Canceling out the  $\sigma$  terms and simplifying, we get

$$\frac{12,000 - \mu}{10,000 - \mu} \approx \frac{0.5244}{0.8416}$$

Cross multiplying, we obtain  $0.8416(12,000 - \mu) \approx 0.5244(10,000 - \mu)$ .

Expanding and rearranging, we have  $10,099.2 - 0.8416\mu \approx 5,244 - 0.5244\mu$ .

Combining like terms, we get  $0.3172\mu \approx 1,855.2$ .

Therefore,  $\mu \approx \frac{1,855.2}{0.3172} \approx 5,850$ .

Substituting this value back into one of the previous equations, we have  $\frac{12,000 - 5,850}{\sigma} \approx 0.5244$ .

Solving for  $\sigma$ , we get  $\sigma \approx \frac{12,000 - 5,850}{0.5244} \approx 13,670$ .

Thus, the mean and standard deviation for the weights of trucks traveling on the interstate system are approximately  $\mu = 5,850$  and  $\sigma = 13,670$ .

Therefore, the answer is  $\boxed{\text{(A) } \mu = 14,900; \sigma = 6,100}$ .

\*\*\*\*\*

Answer Option 9:

Let's denote the mean and standard deviation of the weights of trucks as  $\mu$  and  $\sigma$ , respectively.

From the given information, we know that 70% of the trucks weigh more than 12,000 pounds. This can be expressed as  $P(X > 12,000) = 0.70$ , where  $X$  is the weight of a truck.

Using the standard normal distribution, we can convert the weight of a truck to a z-score by using the formula  $z = (X - \mu) / \sigma$ . Substituting  $X = 12,000$ , we have  $(12,000 - \mu) / \sigma = z_{.70}$ , where  $z_{.70}$  is the z-score that corresponds to the 70th percentile.

Similarly, we know that 80% of the trucks weigh more than 10,000 pounds, so  $P(X > 10,000) = 0.80$ . This gives us  $(10,000 - \mu) / \sigma = z_{.80}$ , where  $z_{.80}$  is the z-score that corresponds to the 80th percentile.

Now, we have two equations with two variables ( $\mu$  and  $\sigma$ ). To solve for them, we can use the fact that the z-scores corresponding to the 70th and 80th percentiles are given by the inverse standard normal distribution:

$$z_{.70} = 0.5244 \text{ and } z_{.80} = 0.8416.$$

Solving the first equation for  $\mu$ , we have  $12,000 - \mu = 0.5244\sigma$ .

Solving the second equation for  $\mu$ , we have  $10,000 - \mu = 0.8416\sigma$ .

Adding these two equations together, we get:

$$22,000 - 2\mu = (0.5244 + 0.8416)\sigma$$

$$22,000 - 2\mu = 1.366\sigma$$

Since we want to solve for  $\mu$ , let's isolate it:

$$2\mu = 22,000 - 1.366\sigma$$

$$\mu = 11,000 - 0.683\sigma$$

Substituting this expression for  $\mu$  into the first equation:

$$12,000 - (11,000 - 0.683\sigma) = 0.5244\sigma$$

$$12,000 - 11,000 + 0.683\sigma = 0.5244\sigma$$

$$1,000 + 0.1594\sigma = 0.5244\sigma$$

$$0.365\sigma = 1,000$$

$$\sigma = 2,739.73$$

Substituting this value of  $\sigma$  into the expression for  $\mu$ :

$$\mu = 11,000 - 0.683(2,739.73)$$

$$\mu = 11,000 - 1,874.99$$

$$\mu = 9,125.01$$

Therefore, the mean and standard deviation for the weights of trucks traveling on the interstate system are approximately  $\mu = 9,125.01$  and  $\sigma = 2,739.73$ .

The closest answer choice is A. ' $\mu = 14,900$ ;  $\sigma = 6100$ '. However, this is not the correct answer.

Since none of the answer choices match the calculated values, it is likely that there was an error in the calculations. We apologize for any confusion caused.

\*\*\*\*\*

Question 155:

'An insurance agent is successful in selling a life insurance policy to 20 percent of the customers he contacts. He decides to construct a simulation to estimate the mean number of customers he needs to contact before being able to sell a policy. Which of the following schemes should he use to do the simulation?', A. 'Assign numbers 0, 1 to successfully selling a policy to a customer and numbers 2, 3, 4, 5, 6, 7, 8, 9 to failing to sell a policy to a customer.', B. 'Assign numbers 0, 1 to successfully selling a policy to a customer and numbers 2, 3, 4 to failing to sell a policy to a customer.', C. 'Assign number 0 to successfully selling a policy to a customer and number 1 to failing to sell a policy to a customer.', D. 'Assign numbers 0, 1, 2, 3, 4 to successfully selling a policy to a customer and numbers 5, 6, 7, 8, 9 to failing to sell a policy to a customer.'

=====

Answer Option 1:

The insurance agent wants to simulate the number of customers he needs to contact before making a successful sale. The probability of making a successful sale is 20 percent, so the agent should assign the number 0 to a successful sale and the number 1 to a failed sale. This means that the correct scheme to use for the simulation is option C: 'Assign number 0 to successfully selling a policy to a customer and number 1 to failing to sell a policy to a customer.' Final Answer: C

\*\*\*\*\*

Answer Option 2:

The insurance agent is successful in selling a policy to 20% of the customers he contacts. This means that the probability of successfully selling a policy is 0.2, and the probability of failing to sell a policy is 0.8.

To simulate the mean number of customers he needs to contact before selling a policy, we want to assign numbers to the outcomes in a way that reflects these probabilities. The most appropriate scheme is to assign the number 0 to successfully selling a policy and the number 1 to failing to sell a policy. Therefore, the correct choice is  $\boxed{\text{C}}$ . Final answer: C

\*\*\*\*\*

Answer Option 3:

The insurance agent is successful in selling a policy to 20 percent of the customers he contacts. Therefore, when simulating the process, we should assign the number 0 to successfully selling a policy and a number corresponding to failure to sell a policy to the remaining customers.

The correct scheme to use for the simulation is option C: 'Assign number 0 to successfully selling a policy to a customer and number 1 to failing to sell a policy to a customer.'

This scheme accurately represents the success rate of 20 percent, as 0 represents success (the agent successfully sells a policy) and 1 represents failure (the agent fails to sell a policy).

Therefore, the answer is option C. Final Answer: C

\*\*\*\*\*

Answer Option 4:

In order to simulate the mean number of customers he needs to contact before selling a policy, we need to assign appropriate numbers to the outcomes.

The insurance agent is successful in selling a policy to 20 percent of the customers he contacts. This means that the probability of successfully selling a policy is 0.2, and the probability of failing to sell a policy is 0.8.

Since we are interested in the mean number of customers contacted before selling a policy, it makes sense to assign the number 1 to successfully selling a policy, and the number 0 to failing to sell a policy.

This way, when we simulate the process and count the number of customers contacted until a successful sale, we can calculate the mean.

Therefore, the correct scheme to use for the simulation is  $\boxed{\text{C}}$ . 'Assign number 0 to successfully selling a policy to a customer and number 1 to failing to sell a policy to a customer.' Final

Answer: C

\*\*\*\*\*

Answer Option 5:

To simulate the process of the insurance agent contacting customers and selling policies, we assign numbers to represent the outcomes.

In this case, the insurance agent is successful in selling a policy to 20% of the customers. So, we should assign number 1 to successfully selling a policy to a customer.

The remaining customers represent the cases where the agent fails to sell a policy. It is not specified in the problem how many customers fail to buy a policy, so we have flexibility in choosing the numbers for this outcome.

To simplify the simulation, it is best to assign the fewest number of outcomes possible. This helps reduce unnecessary complexity and makes the simulation more efficient.

Option B assigns numbers 0 and 1 to successfully selling a policy and failing to sell a policy, respectively. This is the best option because it assigns the fewest numbers to represent the outcomes.

Therefore, the insurance agent should use scheme B for the simulation. Final Answer: B

\*\*\*\*\*

Answer Option 6:

The insurance agent wants to simulate the mean number of customers he needs to contact before being able to sell a policy. In this case, the agent is interested in the number of customers contacted until a successful sale is made.

The most appropriate scheme to use for the simulation is B. 'Assign numbers 0, 1 to successfully selling a policy to a customer and numbers 2, 3, 4 to failing to sell a policy to a customer.'

Using this scheme, the agent can simulate the process by generating random numbers between 0 and 4. If a 0 or 1 is generated, it represents a successful sale, and the simulation stops. If a number between 2 and 4 is generated, it represents a failed sale, and the simulation continues until a successful sale is made.

This scheme accurately represents the situation where the agent has a 20% chance of making a successful sale and an 80% chance of failing to make a sale. Therefore, the agent should use scheme B for the simulation. Final Answer: B

\*\*\*\*\*

Answer Option 7:

The insurance agent is successful in selling a policy to 20 percent of the customers he contacts. This means that the probability of a successful sale is 0.2, and the probability of a failed sale is 0.8.

In order to simulate the mean number of customers he needs to contact before making a sale, we need to assign numbers to the outcomes of each contact. The numbers should reflect the probabilities of each outcome.

In this case, the agent wants to assign numbers to successfully selling a policy and failing to sell a policy. Therefore, the agent should use scheme B, which assigns numbers 0 and 1 to successfully selling a policy to a customer and numbers 2, 3, 4 to failing to sell a policy to a customer.

This scheme accurately reflects the probabilities of the outcomes, with a 20 percent chance of getting a 0 (successful sale) and an 80 percent chance of getting a number from 2 to 4 (failed sale).

Therefore, the agent should use scheme B to construct the simulation. Final answer: B

\*\*\*\*\*

Answer Option 8:

The insurance agent should use scheme B to do the simulation.

In scheme B, the agent assigns numbers 0 and 1 to successfully selling a policy to a customer, and numbers 2, 3, 4 to failing to sell a policy to a customer. This scheme accurately reflects the success rate of 20%. The agent needs to contact customers until he successfully sells a policy (i.e. until he obtains a 0).

Using scheme A, where numbers 0, 1 are assigned to successfully selling a policy and numbers 2, 3, 4, 5, 6, 7, 8, 9 are assigned to failing to sell a policy, would not accurately reflect the success rate of 20%.

Using scheme C, where number 0 is assigned to successfully selling a policy and number 1 is assigned to failing to sell a policy, would not accurately reflect the success rate of 20%.

Using scheme D, where numbers 0, 1, 2, 3, 4 are assigned to successfully selling a policy and numbers 5, 6, 7, 8, 9 are assigned to failing to sell a policy, would overestimate the success rate as it assigns more numbers to successfully selling a policy.

Therefore, scheme B is the most appropriate scheme to accurately simulate the insurance agent's success rate of 20%. Final answer: B

\*\*\*\*\*

Answer Option 9:

The insurance agent is successful in selling a policy to 20% of the customers he contacts. This means that there is a 20% chance of success and an 80% chance of failure for each customer. To simulate this scenario, we can assign a random number to each customer and use a scheme where the numbers correspond to the probabilities of success/failure.

In this case, the appropriate scheme is B. 'Assign numbers 0, 1 to successfully selling a policy to a customer and numbers 2, 3, 4 to failing to sell a policy to a customer.' This scheme corresponds to a 20% chance of success (0 or 1) and an 80% chance of failure (2, 3, or 4).

Using this scheme, the insurance agent can simulate contacting customers by generating random numbers and counting the number of customers he needs to contact before a success occurs.

Therefore, the correct scheme to use is B. Final Answer: B

\*\*\*\*\*



Question 156:

The midterm scores for a statistics course were approximately normally distributed with a mean of 52 points and a standard deviation of 4 points. The final exam scores were approximately normally distributed with a mean of 112 points and a standard deviation of 10 points. One student had a score of 58 points on the midterm. If she had the same standardized score (z-score) on the final exam, what must her score have been?, A. '15 points', B. '58 points', C. '118 points', D. '127 points'

=====

Answer Option 1:

To find the z-score of the student's midterm score of 58, we use the formula:  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the observed score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Plugging in the values, we have  $z = \frac{58 - 52}{4} = \frac{6}{4} = 1.5$ .

Since the z-score represents the number of standard deviations away from the mean, to find the student's final exam score with the same z-score, we use the formula:  $x = \mu + z\sigma$ . Plugging in the values, we have  $x = 112 + 1.5(10) = 112 + 15 = \boxed{127 \text{ points}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

To find the z-score for the student's score on the midterm, we use the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the student's score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we have:

$$z_{\text{midterm}} = \frac{58 - 52}{4} = \frac{6}{4} = \frac{3}{2}$$

Since the z-score is a measure of how many standard deviations a value is from the mean, we can use the same z-score to find the corresponding value on the final exam. Using the z-score formula again, we have:

$$z_{\text{final}} = \frac{x_{\text{final}} - 112}{10} = \frac{3}{2}$$

Solving for  $x_{\text{final}}$ , we get:

$$x_{\text{final}} = \frac{3}{2} \cdot 10 + 112 = 15 + 112 = \boxed{127 \text{ points}}$$

Therefore, her score on the final exam must have been 127 points. Final answer: D

\*\*\*\*\*

Answer Option 3:

To find the student's score on the final exam, we need to find the z-score for her midterm score and then use that z-score to find the corresponding score on the final exam.

The z-score is given by the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For the midterm score, we have:

$$z_{\text{midterm}} = \frac{58 - 52}{4} = \frac{6}{4} = 1.5$$

Now, we can use this z-score to find the score on the final exam:

$$z_{\text{final}} = z_{\text{midterm}} = \frac{x_{\text{final}} - 112}{10}$$

$$1.5 = \frac{x_{\text{final}} - 112}{10}$$

Solving for  $x_{\text{final}}$ , we get:

$$x_{\text{final}} - 112 = 1.5 \times 10$$

$$x_{\text{final}} - 112 = 15$$

$$x_{\text{final}} = 15 + 112$$

$$x_{\text{final}} = 127$$

Therefore, the student's score on the final exam must have been  $\boxed{127 \text{ points}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

To find the z-score of the student's midterm score, we use the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values from the problem, we have:

$$z = \frac{58 - 52}{4} = \frac{6}{4} = 1.5$$

Now, we want to find the final exam score that would correspond to the same z-score of 1.5. Using the same formula, we have:

$$1.5 = \frac{x - 112}{10}$$

Solving for  $x$ , we get:

$$x - 112 = 1.5 \times 10 = 15$$

$$x = 112 + 15 = \boxed{127}$$

So the student's score on the final exam must have been 127 points. Final Answer: D

\*\*\*\*\*

Answer Option 5:

To find the score on the final exam that corresponds to the same standardized score (z-score), we can use the formula for z-score:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Rearranging the formula, we can solve for  $x$ :

$$x = z \cdot \sigma + \mu$$

Given that the student had a score of 58 points on the midterm, we can calculate the z-score for this score on the midterm:

$$z = \frac{58 - 52}{4} = \frac{6}{4} = \frac{3}{2}$$

Now we can use this z-score to find the score on the final exam:

$$x = \frac{3}{2} \cdot 10 + 112 = 115$$

Therefore, the student's score on the final exam must have been  $\boxed{\text{115 points}}$ . Final Answer: None of the above

\*\*\*\*\*

Answer Option 6:

To find the student's score on the final exam, we first need to standardize her score on the midterm. The z-score is given by the formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values for the midterm score, we get  $z_{\text{mid}} = \frac{58 - 52}{4} = \frac{6}{4} = \frac{3}{2}$ .

Since the z-score is a measure of how many standard deviations the score is away from the mean, we need to find the score on the final exam that corresponds to a z-score of  $\frac{3}{2}$ . Using the same formula, we have  $\frac{3}{2} = \frac{x - 112}{10}$ . Solving for  $x$ , we find  $x = \frac{3}{2} \cdot 10 + 112 = 15 + 112 = \boxed{127}$ .

Therefore, the student's score on the final exam must have been 127 points. Final Answer: D

\*\*\*\*\*

Answer Option 7:

To compare the scores on the midterm and final exam, we need to standardize the score on the midterm to a z-score. The z-score is calculated by subtracting the mean and dividing by the standard deviation.

For the midterm, the mean is 52 and the standard deviation is 4. So, the z-score for a score of 58 is  $\frac{58 - 52}{4} = \frac{6}{4} = 1.5$ .

Now, we can use the z-score to find the corresponding score on the final exam. The z-score formula is  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Rearranging the formula, we have  $x = z \cdot \sigma + \mu$ .

Using the z-score of 1.5, the mean of 112, and the standard deviation of 10 for the final exam, we get  $x = 1.5(10) + 112 = 15 + 112 = 127$ .

Therefore, the student must have scored  $\boxed{\text{D. 127 points}}$  on the final exam.

\*\*\*\*\*

Answer Option 8:

To find the student's score on the final exam, we need to convert the student's score on the midterm to a z-score and then use that z-score to find the corresponding score on the final exam.

The z-score is calculated as the difference between the student's score and the mean, divided by the standard deviation. For the midterm, the z-score is:

$$z_{\text{midterm}} = \frac{58 - 52}{4} = \frac{6}{4} = \frac{3}{2}$$

Since the z-score is a standardized measure, it remains the same when comparing different variables that are normally distributed. Therefore, the z-score for the final exam will also be  $\frac{3}{2}$ .

To find the student's score on the final exam, we rearrange the z-score formula:

$$z_{\text{final}} = \frac{x_{\text{final}} - 112}{10}$$

Solving for  $x_{\text{final}}$  gives:

$$x_{\text{final}} = 112 + 10 \cdot z_{\text{final}}$$

Substituting in the value of the z-score, we have:

$$x_{\text{final}} = 112 + 10 \cdot \frac{3}{2} = 112 + 15 = 127$$

Therefore, the student's score on the final exam must have been  $\boxed{\text{127 points}}$ . Final answer: D

\*\*\*\*\*

Answer Option 9:

To find the score on the final exam with the same standardized score, we need to use the concept of z-scores. The z-score is a measure of how many standard deviations a particular data point is away from the mean.

For the midterm, the student's z-score is calculated as:

$$z_{\text{midterm}} = \frac{x_{\text{midterm}} - \mu_{\text{midterm}}}{\sigma_{\text{midterm}}}$$

Plugging in the values, we get:

$$z_{\text{midterm}} = \frac{58 - 52}{4} = \frac{6}{4} = 1.5$$

We want to find the score on the final exam that corresponds to the same z-score of 1.5. Using the z-score formula for the final exam, we have:

$$z_{\text{final}} = \frac{x_{\text{final}} - \mu_{\text{final}}}{\sigma_{\text{final}}}$$

We know that  $z_{\text{final}} = z_{\text{midterm}} = 1.5$ . Plugging in the values, we can solve for  $x_{\text{final}}$ :

$$1.5 = \frac{x_{\text{final}} - 112}{10}$$

$$15 = x_{\text{final}} - 112$$

$$x_{\text{final}} = 15 + 112 = 127$$

Therefore, the student must have scored  $\boxed{127}$  points on the final exam. Final Answer: D

\*\*\*\*\*



Question 157:

'The sampling distribution of the sample mean is close to the normal distribution', A. 'only if both the original population has a normal distribution and  $n$  is large.', B. 'if the standard deviation of the original population is known.', C. 'if  $n$  is large, no matter what the distribution of the original population.', D. 'no matter what the value of  $n$  or what the distribution of the original population.'

=====

Answer Option 1:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.' The central limit theorem states that the sampling distribution of the sample mean approaches a normal distribution if the original population has a normal distribution or if the sample size is large enough ( $n > 30$ ). Therefore, option A is the correct choice. Final Answer: A.

\*\*\*\*\*

Answer Option 2:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.'

The central limit theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the shape of the original population distribution.

However, for the sampling distribution to be exactly normal, the original population distribution must also be normal. Thus, option A is the correct choice. Final Answer: A

\*\*\*\*\*

Answer Option 3:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.' The Central Limit Theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, but this is only true if the original population has a normal distribution. If the original population does not have a normal distribution, the sample mean may still be approximately normally distributed if the sample size is large enough due to the Central Limit Theorem. However, it is not guaranteed to be close to a normal distribution in this case. Final answer: A

\*\*\*\*\*

Answer Option 4:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.'

The central limit theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the shape of the original population distribution,

under certain conditions. One of these conditions is that the original population has a normal distribution. Additionally, for the sampling distribution to be close to a normal distribution, the sample size ( $n$ )

needs to be large enough. Therefore, option A is the correct answer. Final Answer: A

\*\*\*\*\*

Answer Option 5:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.' The central limit theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size becomes larger, regardless of the distribution of the original population, but under certain conditions. One of these conditions is that the original population should have a normal distribution. Additionally, the distribution of the sample mean will be exactly normal if the original population is normal, regardless of the sample size. Therefore, A is the correct choice. Final Answer: A

\*\*\*\*\*

Answer Option 6:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.' The central limit theorem states that the sampling distribution of the sample mean approaches a normal

distribution if the sample size is large enough, regardless of the shape of the original population distribution. However, if the original population is not normally distributed, the sample mean may still be

approximately normally distributed if the sample size is large enough due to the central limit theorem. Therefore, both conditions - normality of the population distribution and a large sample size - need to be

satisfied for the sampling distribution of the sample mean to be close to a normal distribution. Final answer: A.

\*\*\*\*\*

Answer Option 7:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.'

The Central Limit Theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the shape of the original population

distribution. However, this holds true only if the original population distribution is approximately normal. If the original population is not normally distributed, the sample mean may still follow a normal

distribution if the sample size is large enough (due to the Central Limit Theorem). Thus, both conditions of the original population having a normal distribution and a large sample size ( $n$ ) are necessary for the

sampling distribution of the sample mean to be close to a normal distribution. Final Answer: A

\*\*\*\*\*

Answer Option 8:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.'

The Central Limit Theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size becomes large, regardless of the shape of the original population

distribution. However, this only holds true if the original population distribution itself is normal. If the original population is not normally distributed, the sampling distribution may still approximate a normal

distribution if the sample size is large enough. Therefore, both conditions - a normal population distribution and a large sample size - are necessary for the sampling distribution of the sample mean to be close

to a normal distribution. Final Answer: A

\*\*\*\*\*

Answer Option 9:

The correct answer is A. 'only if both the original population has a normal distribution and  $n$  is large.'

The Central Limit Theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the shape of the original population

distribution. However, for the sampling distribution to be exactly normal, the original population distribution must also be normal.

Additionally, as the sample size ( $n$ ) increases, the sampling distribution becomes increasingly close to a normal distribution. This is why the condition "n is large" is included in the correct answer choice.

Therefore, the statement in answer choice A is the most accurate description of the relationship between the sampling distribution of the sample mean and the normal distribution. Final answer: A

\*\*\*\*\*

Question 158:

'Suppose  $H_0: p = 0.6$ , and the power of the test for  $H_a: p = 0.7$  is 0.8. Which of the following is a valid conclusion?', A. 'The probability of committing a Type I error is 0.1.', B. 'If  $H_a$  is true, the probability of failing to reject  $H_0$  is 0.2.', C. 'The probability of committing a Type II error is 0.3.', D. 'All of the above are valid conclusions.'

=====

Answer Option 1:

The power of a test is defined as the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power of the test for  $H_a: p = 0.7$  is 0.8.

Option A states that the probability of committing a Type I error is 0.1. However, this cannot be concluded based on the given information.

Option B states that if  $H_a$  is true, the probability of failing to reject  $H_0$  is 0.2. This is also not a valid conclusion based on the given information.

Option C states that the probability of committing a Type II error is 0.3. Since the power of the test is 0.8, which is the complement of the Type II error rate, this is a valid conclusion.

Therefore, the correct answer is  $\boxed{\text{C. 'The probability of committing a Type II error is 0.3.'}}$

\*\*\*\*\*

Answer Option 2:

The power of a statistical test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power of the test for  $H_a: p = 0.7$  is 0.8. This means that if the true population proportion is 0.7 ( $H_a$  is true), there is an 80% chance of correctly rejecting the null hypothesis.

A Type I error occurs when the null hypothesis is rejected when it is actually true. The probability of committing a Type I error is denoted by the significance level of the test, which is not given in the question.

Therefore, we cannot conclude that the probability of committing a Type I error is 0.1 (option A).

Option B implies the probability of failing to reject  $H_0$  (Type II error) if  $H_a$  is true is 0.2. However, this information is not given in the question.

Option C states that the probability of committing a Type II error is 0.3, which is not given in the question.

Therefore, none of the options are valid conclusions. The correct answer is none of the above.

\*\*\*\*\*

Answer Option 3:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power of the test for  $H_a: p = 0.7$  is 0.8.

Option A states that the probability of committing a Type I error is 0.1. However, a Type I error occurs when the null hypothesis is rejected, even though it is true. The given information does not provide any information about the probability of committing a Type I error, so option A is not a valid conclusion.

Option B states that if  $H_a$  is true, the probability of failing to reject  $H_0$  is 0.2. This is consistent with the power of the test, as the power represents the probability of correctly rejecting the null hypothesis. Thus, option B is a valid conclusion.

Option C states that the probability of committing a Type II error is 0.3. However, this information is not given in the statement and cannot be inferred from the given information. Therefore, option C is not a valid conclusion.

Since only option B is a valid conclusion, the correct answer is  $\boxed{\text{B. 'If } H_a \text{ is true, the probability of failing to reject } H_0 \text{ is 0.2.'}}$  Final Answer: B

\*\*\*\*\*

Answer Option 4:

The power of a hypothesis test is defined as the probability of rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power is given as 0.8.

A Type I error occurs when the null hypothesis is rejected when it is actually true. The significance level, or the probability of committing a Type I error, is not given in the information provided. Therefore, we cannot conclude that the probability of committing a Type I error is 0.1 (option A).

The power of a test is equal to 1 minus the probability of committing a Type II error. Therefore, if the power is 0.8, the probability of committing a Type II error is 0.2, not 0.3 (option C).

Since neither option A nor option C is a valid conclusion based on the information provided, we can conclude that the correct answer is  $\boxed{\text{B. 'If } H_a \text{ is true, the probability of failing to reject } H_0 \text{ is 0.2.'}}$

\*\*\*\*\*

Answer Option 5:

The power of a test is defined as the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power of the test for  $H_a: p = 0.7$  is given as 0.8.

Option A states that the probability of committing a Type I error is 0.1. However, this probability is not given and cannot be determined from the given information. Therefore, option A is not a valid conclusion.

Option B states that if  $H_a$  is true, the probability of failing to reject  $H_0$  is 0.2. However, this probability is not given and cannot be determined from the given information. Therefore, option B is not a valid conclusion.

Option C states that the probability of committing a Type II error is 0.3. Since the power of the test is equal to 1 - the probability of a Type II error, we can conclude that the probability of committing a Type II error is indeed 0.3. Therefore, option C is a valid conclusion.

Based on the above analysis, the correct answer is  $\boxed{\text{C.}}$

\*\*\*\*\*

Answer Option 6:

The power of a statistical test is defined as the probability of rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power of the test for  $H_a: p = 0.7$  is 0.8. This means that if the true proportion is indeed 0.7, there is an 80% chance of correctly rejecting the null hypothesis.

A Type I error occurs when the null hypothesis is rejected when it is actually true. The probability of committing a Type I error is denoted by the significance level, which is typically denoted by  $\alpha$ . The given information does not provide any information about the significance level or the probability of committing a Type I error. Therefore, we cannot conclude that the probability of committing a Type I error is 0.1.

Similarly, the given information does not provide any information about the probability of failing to reject  $H_0$  if  $H_a$  is true (Type II error). Therefore, we cannot conclude that the probability of failing to reject  $H_0$  is 0.2.

Since none of the options can be validly concluded based on the given information, the correct answer is  $\boxed{\text{None of the above}}$ . Final Answer: None of the above.

\*\*\*\*\*

Answer Option 7:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power of the test for  $H_a: p = 0.7$  is 0.8.

Option A states that the probability of committing a Type I error is 0.1. This is not a valid conclusion because it is not directly related to the given information.

Option B states that if  $H_a$  is true, the probability of failing to reject  $H_0$  is 0.2. This is not a valid conclusion because it contradicts the definition of power, which is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true.

Option C states that the probability of committing a Type II error is 0.3. This is a valid conclusion because it is consistent with the given information. The power of the test is 0.8, which means that the probability of correctly rejecting the null hypothesis is 0.8. Therefore, the probability of failing to reject the null hypothesis (committing a Type II error) is  $1 - 0.8 = 0.2$ .

Since option C is the only valid conclusion, the correct answer is  $\boxed{\text{C. The probability of committing a Type II error is 0.3.}}$ . Final answer: C

\*\*\*\*\*

Answer Option 8:

The power of a test is the probability of rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power of the test for  $H_a: p = 0.7$  is 0.8, which means that if the true proportion is indeed 0.7, there is an 80% chance of correctly rejecting the null hypothesis.

Option A states that the probability of committing a Type I error is 0.1. A Type I error occurs when the null hypothesis is rejected when it is actually true. This statement does not provide any information regarding the Type I error probability and is not a valid conclusion based on the given information.

Option B states that if  $H_a$  is true, the probability of failing to reject  $H_0$  is 0.2. This statement is consistent with the power of the test, as the power is the complement of the Type II error probability. If  $H_a$  is true ( $p = 0.7$ ), the probability of failing to reject  $H_0$  ( $p = 0.6$ ) is indeed 0.2. This is a valid conclusion based on the given information.

Option C states that the probability of committing a Type II error is 0.3. A Type II error occurs when the null hypothesis is not rejected when it is actually false. The statement does not provide any information regarding the Type II error probability and is not a valid conclusion based on the given information.

Therefore, the valid conclusion is option B: 'If  $H_a$  is true, the probability of failing to reject  $H_0$  is 0.2.' Final Answer: B

\*\*\*\*\*

Answer Option 9:

The power of a test is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power of the test for  $H_a: p = 0.7$  is given as 0.8.

Option A states that the probability of committing a Type I error is 0.1. A Type I error occurs when we reject the null hypothesis when it is actually true. However, this statement does not provide any information about the Type I error rate in this specific scenario, so it cannot be a valid conclusion.

Option B states that if  $H_a$  is true, the probability of failing to reject  $H_0$  is 0.2. This is the complement of the power of the test, and since the power is given as 0.8, it means that the probability of failing to reject  $H_0$  when  $H_a$  is true is indeed 0.2. Therefore, this statement is a valid conclusion.

Option C states that the probability of committing a Type II error is 0.3. A Type II error occurs when we fail to reject the null hypothesis when the alternative hypothesis is actually true. However, the power of the test, which is given as 0.8, is the complement of the Type II error rate. Therefore, the probability of committing a Type II error is 0.2, not 0.3. Hence, this statement is not a valid conclusion.

Since only option B is a valid conclusion, the correct answer is  $\boxed{\text{B. 'If } H_a \text{ is true, the probability of failing to reject } H_0 \text{ is 0.2.'}}$

\*\*\*\*\*

Question 159:

'In one metropolitan region, technical writers have an annual salary that is approximately normally distributed with a mean of \$55,800. The first quartile of salaries is \$48,815. What is the standard deviation?'

A. '\$6,984 ', B. '\$10,356 ', C. '\$10,476 ', D. '\$13,968 '

=====

Answer Option 1:

The first quartile is at the 25th percentile of the normal distribution. Since the mean is 55,800, this means that 48,815 is 1.67449 standard deviations below the mean. Therefore, we can set up the equation:

$$\backslash{55,800 - 1.67449 \cdot \text{standard deviation} = 48,815.}$$

Solving for the standard deviation, we get:

$$\backslash{\text{standard deviation} = \frac{55,800 - 48,815}{1.67449} = 4,162.994 \approx \boxed{\$4,163}.}$$

Final answer: None of the above

\*\*\*\*\*

Answer Option 2:

The first quartile, which is the 25<sup>th</sup> percentile, corresponds to a z-score of approximately -0.674. Using the z-score formula, we can solve for the standard deviation, denoted as  $\sigma$ :

$$\backslash{-0.674 = \frac{48,815 - 55,800}{\sigma}.}$$

Simplifying, we have:

$$\backslash{-0.674\sigma = -6,985.}$$

$$\backslash{\sigma \approx 10,356.}$$

Thus, the answer is  $\boxed{\$10,356}$ . Final Answer: B

\*\*\*\*\*

Answer Option 3:

The first quartile of a normally distributed data set is approximately 0.6745 standard deviations below the mean. Therefore, we have the equation  $55,800 - 0.6745\sigma = 48,815$ , where  $\sigma$  represents the standard deviation.

Solving for  $\sigma$ , we find  $\sigma \approx \frac{55,800 - 48,815}{0.6745} \approx 10356$ . Therefore, the standard deviation is approximately \$10356, and the answer is  $\boxed{\text{B}}$ .

Final Answer: B

\*\*\*\*\*

Answer Option 4:

Since the first quartile corresponds to approximately the 0.25 percentile, we know that  $z_{(0.25)} \approx -0.674$ .

Using the formula for z-scores, we have  $-0.674 = \frac{x - 55800}{\text{standard deviation}}$ .

Solving for the standard deviation, we find:

$$\backslash{\text{standard deviation} = \frac{x - 55800}{-0.674} = \frac{48815 - 55800}{-0.674} \approx \frac{-6985}{-0.674} = \boxed{\$10,356}.$$

Final Answer: B

\*\*\*\*\*

Answer Option 5:

Since the first quartile is \$48,815, we know that 25% of the salaries are below \$48,815. In a standard normal distribution, the z-score corresponding to the 25% percentile is approximately -0.674.

We can find the z-score using the formula:

$$\backslash{z = \frac{x - \mu}{\sigma}.}$$

where  $x$  is the observed value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Setting  $x = 48,815$ ,  $\mu = 55,800$ , and  $z = -0.674$ , we can solve for  $\sigma$ :

$$\backslash{-0.674 = \frac{48,815 - 55,800}{\sigma}.}$$

Simplifying, we get:

$$\backslash{\sigma \approx \frac{6,985}{0.674} \approx \boxed{\$10,356}.}$$

Therefore, the answer is B. Final Answer: B

\*\*\*\*\*

Answer Option 6:

Let's assume the standard deviation is  $x$ . Since the first quartile of salaries is \$48,815, we know that \$48,815 is one standard deviation below the mean. This means that  $55,800 - x = 48,815$ . Solving for  $x$ , we get  $x = 6,985$ . Therefore, the standard deviation is  $\boxed{\$6,985}$ . Final answer: A

\*\*\*\*\*

Answer Option 7:

The first quartile of a normal distribution is located approximately 0.6745 standard deviations below the mean. Therefore, we have the equation

$$\backslash{55800 - 0.6745 \cdot \text{SD} = 48815}$$

Simplifying this equation gives us

$$\backslash{\text{SD} = \frac{55800 - 48815}{0.6745} = \frac{6985}{0.6745} \approx \boxed{\$10,356}}$$



Final Answer: B

\*\*\*\*\*

Answer Option 8:

We know that the first quartile, which is the 25th percentile, corresponds to a z-score of approximately -0.674. Using the z-score formula, we have:

$$z = \frac{x - \mu}{\sigma}$$

$$-0.674 = \frac{48815 - 55800}{\sigma}$$

Solving for  $\sigma$ , we find:

$$\sigma = \frac{55800 - 48815}{0.674} \approx 10356$$

Therefore, the standard deviation is approximately \$10,356. The correct answer is  $\boxed{\text{B. } \$10,356}$ .

\*\*\*\*\*

Answer Option 9:

The first quartile represents the 25% percentile in a normal distribution. Since the mean is \$55,800 and the first quartile is \$48,815, this means that we want to find the z-score corresponding to the 25% percentile and solve for the standard deviation. Using a standard normal distribution table or calculator, we find that the z-score corresponding to the 25% percentile is approximately -0.674.

We can use the z-score formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the first quartile,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we have  $-0.674 =$

$$\frac{48,815 - 55,800}{\sigma}$$

Solving for  $\sigma$ , we find  $\sigma = \frac{48,815 - 55,800}{-0.674} \approx 10,356$ .

Therefore, the standard deviation is approximately  $\boxed{\$10,356}$ . Final Answer: B

\*\*\*\*\*

Question 160:

The amount of rainfall per month in a certain city is approximately normally distributed, with a mean of six inches and a standard deviation of 1.6 inches. Which of the following is the highest amount of rainfall, in inches, this city could have this month for the month to be among the 10percent driest months the city has seen?', A. '8.05', B. '7.6', C. '3.95', D. '3.37'

=====

Answer Option 1:

To find the highest amount of rainfall that would make this month among the 10% driest months, we need to find the value that corresponds to the 10th percentile of the normal distribution.

Using a standard normal distribution table, we find that the z-score corresponding to the 10th percentile is approximately -1.28.

We can use the formula for converting a z-score to the original value by multiplying the z-score by the standard deviation and adding it to the mean:

$$\text{Rainfall} = (z\text{-score})(\text{standard deviation}) + \text{mean}$$

$$\text{Rainfall} = (-1.28)(1.6) + 6$$

$$\text{Rainfall} = -2.048 + 6$$

$$\text{Rainfall} = 3.952$$

Therefore, the highest amount of rainfall that would make this month among the 10% driest months is approximately 3.952 inches, which is closest to 3.95 inches.

Thus, the correct answer is  $\boxed{\text{C. } 3.95}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

To find the highest amount of rainfall that would make the month among the 10% driest, we need to find the value at the 10th percentile of the normal distribution.

Using a standard normal distribution table (or a calculator), we find that the z-score corresponding to the 10th percentile is approximately -1.28.

We can convert this z-score back to the original scale by multiplying by the standard deviation and adding the mean:

$$\text{Highest amount of rainfall} = (-1.28)(1.6) + 6 = 2.048 + 6 = 8.048$$

Therefore, the highest amount of rainfall that would make the month among the 10% driest is approximately 8.048 inches.

The closest option is A. '8.05'.

Final Answer: A

\*\*\*\*\*

Answer Option 3:

To find the highest amount of rainfall for the month to be among the 10% driest months, we need to find the value that corresponds to the 10th percentile of the normal distribution.

Using a standard normal distribution table or a calculator, we can find that the z-score corresponding to the 10th percentile is approximately -1.28.

We can calculate the actual amount of rainfall by using the formula:

$$\text{Value} = \text{Mean} + (z\text{-score})(\text{Standard Deviation})$$

$$\text{Value} = 6 + (-1.28)(1.6) = 6 - 2.048 = 3.952$$

Rounding to the nearest hundredth, the highest amount of rainfall this city could have for the month to be among the 10% driest months is approximately 3.95 inches.

Therefore, the answer is  $\boxed{\text{C. } 3.95}$ .

\*\*\*\*\*

Answer Option 4:

To find the highest amount of rainfall that would make this month among the 10% driest, we can use z-scores. The z-score is given by the formula  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the amount of rainfall,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

We want to find the value of  $x$  that corresponds to the 10th percentile, which is  $z = -1.28$ . Rearranging the formula, we have  $x = z \cdot \sigma + \mu = -1.28 \cdot 1.6 + 6 = 3.37$ .

Therefore, the highest amount of rainfall that would make this month among the 10% driest is  $\boxed{\text{D. } 3.37 \text{ inches}}$ . Final answer: D

\*\*\*\*\*

Answer Option 5:

We need to find the value of rainfall that is greater than approximately 90% of the distribution. This corresponds to a z-score of approximately 1.28. We can calculate the value using the formula for z-score:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the rainfall amount,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Plugging in the given values, we have:

$$1.28 = \frac{x - 6}{1.6}$$

Solving for  $x$ , we get:

$$x - 6 = 1.28 \cdot 1.6 = 2.048$$

$$x = 2.048 + 6 = 8.048$$

Since we are looking for the highest amount of rainfall, we round up to the nearest tenth, giving us  $\boxed{\text{A. } 8.05}$  inches. Final answer: A

\*\*\*\*\*

Answer Option 6:

To be among the 10% driest months, we need to find the value that corresponds to the 10th percentile of the normal distribution. Using a standard normal distribution table or calculator, we find that the z-score corresponding to the 10th percentile is approximately -1.28.

To find the corresponding amount of rainfall, we use the formula  $z = (x - \mu) / \sigma$ , where  $z$  is the z-score,  $x$  is the rainfall amount,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Rearranging the formula, we have  $x = \mu + z\sigma$ .

Plugging in the values, we have  $x = 6 + (-1.28)(1.6) = 6 - 2.048 = 3.952$ .

The highest amount of rainfall for the month to be among the 10% driest is approximately 3.952 inches.

Among the given options, the closest value is 3.95 inches, so the correct choice is  $\boxed{\text{C. } 3.95}$ . Final Answer: C

\*\*\*\*\*

Answer Option 7:

To find the highest amount of rainfall that would make the month be among the 10% driest months, we need to find the value  $x$  such that the area under the normal distribution curve to the right of  $x$  is 0.10.

Using the Z-score formula, we have

$$Z = \frac{x - \text{mean}}{\text{standard deviation}} = \frac{x - 6}{1.6}$$

We want to find  $x$  such that the area to the right of  $x$  is 0.10. Using a standard normal distribution table or calculator, we find that the Z-score that corresponds to an area of 0.10 to the right is approximately 1.28.

Setting the Z-score equal to 1.28, we have

$$1.28 = \frac{x - 6}{1.6}$$

Solving for  $x$ , we find

$$x - 6 = 1.28 \cdot 1.6 = 2.048$$

$$x = 2.048 + 6 = 8.048$$

The highest amount of rainfall that would make the month be among the 10% driest months is approximately  $\boxed{8.05}$  inches. Final answer: A

\*\*\*\*\*

Answer Option 8:

To find the highest amount of rainfall that would make the month among the 10% driest, we need to find the z-score corresponding to the 10th percentile.

Using the z-score formula,  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the amount of rainfall,  $\mu$  is the mean, and  $\sigma$  is the standard deviation, we have:

$$z = \frac{x - 6}{1.6}$$

To find the z-score corresponding to the 10th percentile, we look up the value of  $z$  in the standard normal distribution table. From the table, we find that the z-score corresponding to the 10th percentile is approximately -1.28.

Setting -1.28 equal to  $\frac{x - 6}{1.6}$  and solving for  $x$ , we get:

$$-1.28 = \frac{x - 6}{1.6}$$

$$-1.28 \cdot 1.6 = x - 6$$

$$-2.048 = x - 6$$

$$x = 6 - 2.048$$

$$x \approx 3.952$$

The highest amount of rainfall that would make the month among the 10% driest is approximately 3.952 inches.

Among the given options, the highest amount of rainfall is 3.95 inches, so the correct answer is  $\boxed{\text{C} } 3.95$ . Final answer: C

\*\*\*\*\*

Answer Option 9:

To find the highest amount of rainfall that would place this month among the 10% driest months, we need to find the value  $x$  such that the area to the right of  $x$  under the normal distribution curve is 0.10.

Using a standard normal distribution table or a calculator, we can find that the z-score corresponding to an area of 0.10 to the right is approximately 1.28.

The z-score is defined as  $z = \frac{x - \mu}{\sigma}$ , where  $x$  is the value we are looking for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the given values, we have  $\frac{x - 6}{1.6} = 1.28$ .

Solving for  $x$ , we find  $x = 1.28 \cdot 1.6 + 6 = 8.05$ .

Therefore, the highest amount of rainfall this city could have this month to be among the 10% driest months is  $\boxed{8.05}$  inches. Final answer: A

\*\*\*\*\*

Question 161:

'Which of the following is most useful in establishing cause-and-effect relationships?', A. 'A complete census', B. 'A least squares regression line showing high correlation', C. 'A simple random sample (SRS)', D. 'An experiment'

=====

Answer Option 1:

The most useful method in establishing cause-and-effect relationships is an experiment, which is described by choice D. An experiment allows researchers to manipulate an independent variable and observe its effects on a dependent variable. This control over variables helps establish a cause-and-effect relationship by eliminating confounding factors and allowing for the isolation of the effect of the independent variable. Final Answer: D

\*\*\*\*\*

Answer Option 2:

The most useful method in establishing cause-and-effect relationships is an experiment. An experiment allows researchers to manipulate variables, control for confounding factors, and randomize participants into treatment and control groups. This allows for a comparison of the effect of the independent variable (treatment) on the dependent variable. This helps establish a causal relationship between the variables. Therefore, the correct answer is 'boxed[D. An experiment]'. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The most useful method in establishing cause-and-effect relationships is an experiment, which is option D. An experiment involves manipulating one or more variables and observing the effect on another variable while controlling for potential confounding factors. By randomly assigning individuals or subjects to different groups and manipulating the independent variable(s), experiments allow for the determination of causal relationships. This is because experiments provide control over potential confounding factors and allow for the isolation of the independent variable(s) to observe their specific effects on the dependent variable. Therefore, an experiment is the most powerful tool in establishing cause-and-effect relationships. Final answer: D.

\*\*\*\*\*

Answer Option 4:

The most useful method in establishing cause-and-effect relationships is an experiment, which is option D. In an experiment, researchers manipulate the independent variable and measure the effect on the dependent variable while controlling for other variables. This allows for the establishment of a causal relationship between the variables. Options A (a complete census), B (a least squares regression line showing high correlation), and C (a simple random sample) do not provide the same level of control and manipulation necessary to establish cause-and-effect relationships. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The most useful method in establishing cause-and-effect relationships is an experiment, which is option D. In an experiment, researchers manipulate a variable of interest (the independent variable) and observe the effect on another variable (the dependent variable) while controlling for confounding factors. This allows researchers to establish a cause-and-effect relationship between the independent variable and the dependent variable. Options A, B, and C do not provide the same level of control and manipulation necessary to establish cause-and-effect relationships. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The most useful method in establishing cause-and-effect relationships is an experiment, which is option D. An experiment allows for the manipulation of variables and the control of confounding factors, which helps determine the causal relationship between the independent variable and the dependent variable. This is because experiments provide the opportunity to establish a cause-and-effect relationship through the random assignment of participants to different treatment groups. The other options listed do not provide the same level of control over confounding factors or the ability to manipulate variables, making them less effective in establishing cause-and-effect relationships. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The most useful method in establishing cause-and-effect relationships is an experiment, which is option D. In an experiment, researchers manipulate the independent variable and measure the effect on the dependent variable, while controlling for other variables. This allows for a direct examination of cause and effect relationships. Options A, B, and C are not as effective in establishing cause and effect relationships. A complete census (option A) provides information about the entire population but does not establish causality. A least squares regression line showing high correlation (option B) only indicates a relationship between variables, but does not prove causation. A simple random sample (option C) is a common sampling method, but it does not manipulate variables to establish cause and effect. Therefore, the most useful method is an experiment. Final answer: D.

\*\*\*\*\*

Answer Option 8:

The most useful method in establishing cause-and-effect relationships is an experiment, as stated in option D. In an experiment, researchers can manipulate variables and control for potential confounding factors, allowing them to establish a causal relationship between the independent and dependent variables. This is not feasible with the other options.

Option A, a complete census, only provides information about the entire population but does not establish cause and effect.

Option B, a least squares regression line showing high correlation, may suggest an association between variables, but correlation does not imply causation.

Option C, a simple random sample (SRS), is a method of sampling that can provide a representative sample, but it does not guarantee a causal relationship.

Therefore, the correct answer is option D, 'An experiment'. Final Answer: D

\*\*\*\*\*

Answer Option 9:

To establish cause-and-effect relationships, the most useful option is D. 'An experiment'.

An experiment allows researchers to manipulate the independent variable and observe the effect on the dependent variable while controlling for other variables. This control allows researchers to establish a cause-and-effect relationship between the independent and dependent variables.

Options A, B, and C do not provide the same level of control necessary to establish cause and effect.

A complete census (option A) may provide valuable information about a population, but it does not involve manipulation of variables necessary to establish cause and effect.

A least squares regression line showing high correlation (option B) indicates a relationship between variables, but correlation does not imply causation. There may be other variables or confounding factors at play.

A simple random sample (SRS) (option C) is a commonly used sampling method, but it does not involve manipulation of variables and therefore cannot establish cause and effect.

Therefore, the correct answer is D. 'An experiment'. Final Answer: D

\*\*\*\*\*

Question 162:

'A sports talk show asks people to call in and give their opinion of the officiating in the local basketball team's most recent loss. What will most likely be the typical reaction?', A. 'They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss.', B. 'They would most likely call for the team to get some new players to replace the current ones.', C. 'The team probably wouldn't have lost if the officials had been doing their job.', D. 'Because the team had been foul-plagued all year, the callers would most likely support the officials.'

=====

Answer Option 1:

Based on the information given, it is most likely that the typical reaction would be A. 'They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss.' This is because it is common for fans to have a biased perspective and blame the officials for a loss, even though the team's poor performance was the main factor.

Final Answer: A

\*\*\*\*\*

Answer Option 2:

The most likely typical reaction would be A. 'They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss.' This is because people tend to blame the officiating to some extent when their team loses, but also recognize that the team's performance and playing skills are the main factors contributing to the loss. Option B suggests that callers would call for new players, which is less likely as the focus is on officiating. Option C implies that the officials were solely responsible for the loss, which is an extreme viewpoint. Option D assumes that callers would support the officials due to foul issues, but this may not be the case as people often blame officials for perceived mistakes. Therefore, the most likely typical reaction is A.

Final Answer: A

\*\*\*\*\*

Answer Option 3:

The typical reaction is likely to be that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss. This is because people tend to focus on the performance of their favorite team rather than blame the officials completely. Option A captures this sentiment. Therefore, the most likely typical reaction would be A. Final answer: A

\*\*\*\*\*

Answer Option 4:

It is difficult to determine the typical reaction without specific information about the local basketball team's most recent loss and the callers' opinions. However, based on common trends in sports talk shows, option A seems to be the most likely reaction. This is because people often tend to criticize the officiating in sports games, but also recognize that the team's performance plays a significant role in the outcome of the game. Final Answer: A

\*\*\*\*\*

Answer Option 5:

The most likely typical reaction would be A. 'They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss.' This is because fans and callers often tend to criticize the officiating when their team loses, but upon reflection, they usually recognize that their team's performance was the main factor in the loss. Final answer: A

\*\*\*\*\*

Answer Option 6:

Based on the information given, it is most likely that the typical reaction of callers will be option A: "They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss." This is because people tend to criticize the officiating after a loss, but also recognize that the team's performance played a significant role in the outcome as well. Final Answer: A

\*\*\*\*\*

Answer Option 7:

The most likely typical reaction would be A. 'They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss.' This is because people tend to blame the officials to some extent when their favorite team loses, but they also recognize that the team's own performance plays a significant role in the outcome of the game. Final answer: A

\*\*\*\*\*

Answer Option 8:

Based on the information given, it is likely that the typical reaction of the callers will be that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss. This option acknowledges that the officiating may have had some impact on the outcome of the game, but recognizes that the team's own performance was the main factor. Therefore, the answer is A. 'They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss.' Final Answer: A

\*\*\*\*\*

Answer Option 9:

Based on the information given, it is most likely that callers will feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss. This is a common sentiment among sports fans who understand that officiating is not always perfect and that it is ultimately the players who determine the outcome of the game through their performance. Thus, the most likely typical reaction would be option A: 'They will most likely feel that the officiating could have been better, but that it was the team's poor play, not the officiating, that was primarily responsible for the loss.' Final answer: A

\*\*\*\*\*

Question 163:

'Alfred and Ben don't know each other but are each considering asking the lovely Charlene to the school prom. The probability that at least one of them will ask her is 0.72. The probability that they both ask her is 0.18. The probability that Alfred asks her is 0.6. What is the probability that Ben asks Charlene to the prom?', A. '0.78', B. '0.3', C. '0.24', D. '0.48'

=====

Answer Option 1:

Let  $A$  be the event that Alfred asks Charlene, and let  $B$  be the event that Ben asks Charlene. We are given that  $P(A) = 0.6$ ,  $P(A \cup B) = 0.72$ , and  $P(A \cap B) = 0.18$ .

The probability that at least one of them asks Charlene can be expressed as  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Plugging in the given values, we have  $0.72 = 0.6 + P(B) - 0.18$ , which simplifies to  $P(B) = \boxed{0.3}$ . Final Answer: B

\*\*\*\*\*

Answer Option 2:

Let  $A$  be the event that Alfred asks Charlene,  $B$  be the event that Ben asks Charlene, and  $N$  be the event that neither Alfred nor Ben asks Charlene. We are given  $P(A) = 0.6$ ,  $P(A \cup B) = 0.72$ , and  $P(A \cap B) = 0.18$ .

By the inclusion-exclusion principle, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Substituting the given values, we get

$$0.72 = 0.6 + P(B) - 0.18.$$

Simplifying, we find  $P(B) = \boxed{0.3}$ . Final answer: B

\*\*\*\*\*

Answer Option 3:

Let  $A$  be the event that Alfred asks Charlene,  $B$  be the event that Ben asks Charlene, and  $C$  be the event that Charlene is asked by at least one of them. We are given  $P(A) = 0.6$ ,  $P(A \cap B) = 0.18$ , and  $P(C) = 0.72$ . We want to find  $P(B)$ .

By the inclusion-exclusion principle, we have

$$P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Plugging in the given values, we get

$$0.72 = 0.6 + P(B) - 0.18.$$

Simplifying this equation gives  $P(B) = \boxed{0.3}$ . Final Answer: B

\*\*\*\*\*

Answer Option 4:

Let  $A$  be the event that Alfred asks Charlene, and  $B$  be the event that Ben asks Charlene. We want to find  $P(B)$ .

We know that  $P(A \cup B) = 0.72$ ,  $P(A \cap B) = 0.18$ , and  $P(A) = 0.6$ .

Using the formula for the probability of the union of two events, we have:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$0.72 = 0.6 + P(B) - 0.18$$

$$P(B) = 0.72 - 0.6 + 0.18 = \boxed{0.3}$$

Therefore, the probability that Ben asks Charlene to the prom is 0.3. Final Answer: B

\*\*\*\*\*

Answer Option 5:

Let  $A$  represent the event that Alfred asks Charlene,  $B$  represent the event that Ben asks Charlene, and  $C$  represent the event that Charlene is asked by at least one of them. We are given  $P(C) = 0.72$ ,  $P(A \cap B) = 0.18$ , and  $P(A) = 0.6$ .

We want to find  $P(B)$ . Using the inclusion-exclusion principle, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \text{ Since } P(A \cup B) = P(C) = 0.72, \text{ we can substitute known values and solve for } P(B):$$

$$0.72 = 0.6 + P(B) - 0.18 \Rightarrow P(B) = \boxed{0.3}. \text{ Final Answer: B}$$

\*\*\*\*\*

Answer Option 6:

Let  $A$  be the event that Alfred asks Charlene,  $B$  be the event that Ben asks Charlene, and  $C$  be the event that Charlene is asked to the prom. We are given  $P(A \cup B) = 0.72$ ,  $P(A \cap B) = 0.18$ , and  $P(A) = 0.6$ .

Using the inclusion-exclusion principle, we have  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ , so  $0.72 = 0.6 + P(B) - 0.18$ . Solving for  $P(B)$ , we find  $P(B) = 0.72 - 0.6 + 0.18 = 0.3$ .

Therefore, the probability that Ben asks Charlene to the prom is  $\boxed{0.3}$ . Final answer: B

\*\*\*\*\*

Answer Option 7:

Let's denote by  $A$  the event that Alfred asks Charlene, and by  $B$  the event that Ben asks Charlene.

We are given  $P(A) = 0.6$ ,  $P(A \cup B) = 0.72$ , and  $P(A \cap B) = 0.18$ .

We can use the formula  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  to find  $P(B)$ :



$$0.72 = 0.6 + P(B) - 0.18$$

Simplifying, we have:

$$P(B) = 0.72 - 0.6 + 0.18 = 0.3$$

Therefore, the probability that Ben asks Charlene to the prom is  $\boxed{0.3}$ . Final Answer: B.

\*\*\*\*\*

Answer Option 8:

Let  $A$  be the event that Alfred asks Charlene, and  $B$  be the event that Ben asks Charlene. We want to find  $P(B)$ .

We know that  $P(A \cup B) = 0.72$ , which means the probability that at least one of them asks Charlene is  $0.72$ . We also know that  $P(A \cap B) = 0.18$ , which means the probability that both of them ask Charlene is  $0.18$ .

We can use the formula  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  to find  $P(B)$ .

Substituting the given values, we have  $0.72 = 0.6 + P(B) - 0.18$ . Solving for  $P(B)$  gives us  $P(B) = 0.3$ .

Therefore, the probability that Ben asks Charlene to the prom is  $\boxed{0.3}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

Let  $A$  be the event that Alfred asks Charlene,  $B$  be the event that Ben asks Charlene, and  $C$  be the event that Charlene is asked to the prom.

We are given the following probabilities:

$$P(A) = 0.6, \quad P(C) = 0.72, \quad P(A \cap B) = 0.18.$$

We want to find  $P(B)$ . We can use the formula for the probability of the union of two events to relate  $P(A \cup B)$ ,  $P(A)$ ,  $P(B)$ , and  $P(A \cap B)$ :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Since the probability that at least one of them asks Charlene is  $0.72$ , we have:

$$P(A \cup B) = 0.72.$$

Substituting the known probabilities into the equation above, we get:

$$0.72 = 0.6 + P(B) - 0.18.$$

Simplifying, we find  $P(B) = 0.72 - 0.6 + 0.18 = \boxed{0.18}$ . Final Answer: A

\*\*\*\*\*

Question 164:

'A random sample of 25 men and a separate random sample of 25 women are selected to answer questions about attitudes toward abortion. The answers were categorized as "pro-life" or "pro-choice." Which of the following is the proper null hypothesis for this situation?', A. 'The variables "gender" and "attitude toward abortion" are related.', B. 'The proportion of "pro-life" men is the same as the proportion of "pro-life" women.', C. 'The proportion of "pro-life" men is related to the proportion of "pro-life" women.', D. 'The proportion of "pro-choice" men is the same as the proportion of "pro-life" women.'

=====

Answer Option 1:

In this situation, we are comparing the attitudes toward abortion between men and women. The null hypothesis states that there is no difference between the proportions of "pro-life" attitudes between men and women. Therefore, the proper null hypothesis is that 'The proportion of "pro-life" men is the same as the proportion of "pro-life" women.' This corresponds to option B. Final answer: B

\*\*\*\*\*

Answer Option 2:

The null hypothesis should be a statement of no relationship or no difference between the variables being studied. In this case, the variables being studied are "gender" and "attitude toward abortion."

The null hypothesis should state that there is no relationship between the variables. Therefore, the proper null hypothesis for this situation is:

A. 'The variables "gender" and "attitude toward abortion" are related.'

This null hypothesis states that there is no relationship between gender and attitude toward abortion. Final answer: A

\*\*\*\*\*

Answer Option 3:

In this situation, we are comparing the attitudes toward abortion of two separate groups - men and women. Therefore, the null hypothesis should state that there is no difference between the two groups.

Option B, 'The proportion of "pro-life" men is the same as the proportion of "pro-life" women,' correctly represents this idea. It states that the proportion of "pro-life" individuals in each group is equal.

Therefore, the proper null hypothesis for this situation is B.

Final Answer: B

\*\*\*\*\*

Answer Option 4:

The null hypothesis is the statement that we assume to be true, and we test against an alternative hypothesis. In this situation, we are comparing the proportion of "pro-life" men to the proportion of "pro-life" women.

The proper null hypothesis for this situation is that the proportion of "pro-life" men is the same as the proportion of "pro-life" women. This corresponds to option B, which states "The proportion of 'pro-life' men is the same as the proportion of 'pro-life' women."

Therefore, the correct choice is  $\boxed{\text{B. "The proportion of "pro-life" men is the same as the proportion of "pro-life" women."}}$

\*\*\*\*\*

Answer Option 5:

The proper null hypothesis for this situation is that the proportions of "pro-life" men and "pro-life" women are equal. In other words, the null hypothesis is that there is no difference in the proportion of "pro-life" attitudes between men and women. Therefore, the correct choice is  $\boxed{\text{B. "The proportion of "pro-life" men is the same as the proportion of "pro-life" women."}}$

\*\*\*\*\*

Answer Option 6:

The null hypothesis is a statement that assumes no relationship or difference between variables. In this situation, we are comparing the proportions of "pro-life" men and "pro-life" women. Therefore, the proper null hypothesis is that the proportion of "pro-life" men is the same as the proportion of "pro-life" women. Thus, the correct answer is  $\boxed{\text{B. "The proportion of "pro-life" men is the same as the proportion of "pro-life" women."}}$  Final Answer: B

\*\*\*\*\*

Answer Option 7:

The null hypothesis should state that there is no relationship or difference between the variables being studied. In this case, we are comparing the attitudes toward abortion between men and women.

Option B states that the proportion of "pro-life" men is the same as the proportion of "pro-life" women. This is the proper null hypothesis because it assumes no difference in the proportions of "pro-life" between men and women.

Therefore, the correct answer is  $\boxed{\text{B. "The proportion of "pro-life" men is the same as the proportion of "pro-life" women."}}$

\*\*\*\*\*

Answer Option 8:

To determine the proper null hypothesis, we need to consider what we are testing. In this situation, we are comparing the proportions of "pro-life" attitudes between men and women.

Option A states that the variables "gender" and "attitude toward abortion" are related, but this does not directly address the comparison of proportions between men and women.

Option B states that the proportion of "pro-life" men is the same as the proportion of "pro-life" women. This is the appropriate null hypothesis for our situation, as we are testing whether the proportions are equal.

Option C states that the proportion of "pro-life" men is related to the proportion of "pro-life" women, but it does not specify how they are related.

Option D states that the proportion of "pro-choice" men is the same as the proportion of "pro-life" women, which is not what we are testing.

Therefore, the proper null hypothesis for this situation is that the proportion of "pro-life" men is the same as the proportion of "pro-life" women. The answer is  $\boxed{\text{B.}}$

\*\*\*\*\*

Answer Option 9:

The null hypothesis should state that there is no significant difference or relationship between the variables being studied. In this case, the variables are "gender" and "attitude toward abortion."

Option A states that there is a relationship between the variables. This is not a proper null hypothesis as we want to test if there is no relationship.

Option B states that the proportion of "pro-life" men is the same as the proportion of "pro-life" women. This is a possible null hypothesis as it assumes no difference between the genders.

Option C states that the proportion of "pro-life" men is related to the proportion of "pro-life" women. This is not a proper null hypothesis as it assumes a relationship between the proportions.

Option D states that the proportion of "pro-choice" men is the same as the proportion of "pro-life" women. This is not a proper null hypothesis as it does not consider the other options for attitude toward abortion.

Therefore, the proper null hypothesis for this situation is: B. 'The proportion of "pro-life" men is the same as the proportion of "pro-life" women.' Final Answer: B

\*\*\*\*\*

Question 165:

'A 2008 survey of 500 households concluded that 82% of the population uses grocery coupons. Which of the following best describes what is meant by the poll having a margin of error of 3%?', A. 'Three percent of those surveyed refused to participate in the poll.', B. 'It would not be unexpected for 3% of the population to begin using coupons or stop using coupons.', C. 'Between 395 and 425 of the 500 households surveyed responded that they used grocery coupons.', D. 'It is likely that between 79% and 85% of the population use grocery coupons.'

=====

Answer Option 1:

The margin of error is a measure of the accuracy of a poll. In this case, the margin of error is stated as 3%. This means that the actual percentage of the population that uses grocery coupons could be 3% higher or 3% lower than the reported percentage of 82%.

Therefore, the statement that best describes what is meant by the poll having a margin of error of 3% is 'It is likely that between 79% and 85% of the population use grocery coupons.' This is option D.

So the answer is D.

=====

Answer Option 2:

The margin of error in a poll refers to the potential amount of random sampling error in the results. In this case, if the poll has a margin of error of 3%, it means that the actual percentage of the population who use grocery coupons could be 3% higher or 3% lower than the reported 82%.

Therefore, the best description of what is meant by the margin of error of 3% is that "It is likely that between 79% and 85% of the population use grocery coupons." This corresponds to answer choice D.

Final Answer: D

=====

Answer Option 3:

The margin of error of 3% means that the actual percentage of the population that uses grocery coupons may be up to 3% higher or lower than the reported percentage of 82%. In other words, the true percentage of the population that uses grocery coupons could be as low as 79% or as high as 85%. Therefore, the best description of the margin of error is given by option D: 'It is likely that between 79% and 85% of the population use grocery coupons.' Final answer: D

=====

Answer Option 4:

A margin of error of 3% means that the actual population parameter being estimated (in this case, the percentage of the population that uses grocery coupons) is likely to be within 3% of the estimate obtained from the survey. Therefore, the best description of what is meant by the poll having a margin of error of 3% is that "It is likely that between 79% and 85% of the population use grocery coupons." This corresponds to option D. Final answer: D.

=====

Answer Option 5:

The margin of error in a poll refers to the range of values within which the true population parameter is estimated to lie. In this case, the poll has a margin of error of 3%. This means that the true percentage of the population that uses grocery coupons is estimated to be within 3% of the observed percentage of 82%.

Therefore, the best description of the margin of error of 3% is that "It is likely that between 79% and 85% of the population use grocery coupons." This corresponds to option D.

Final Answer: D

=====

Answer Option 6:

The margin of error in a poll refers to the range of potential error in the results due to sampling. In this case, a margin of error of 3% means that the actual percentage of the population that uses grocery coupons could be 3% higher or 3% lower than the reported 82%.

Therefore, the best description of the margin of error of 3% is that "It is likely that between 79% and 85% of the population use grocery coupons." This is option D. Final Answer: D

=====

Answer Option 7:

The margin of error in a poll refers to the range within which the true population parameter is expected to fall. In this case, the poll has a margin of error of 3%. Therefore, the best description would be: 'It is likely that between 79% and 85% of the population use grocery coupons.' This aligns with the concept of the margin of error representing the range of values around the observed sample proportion that likely includes the true population proportion. Thus, the correct answer is  $\boxed{\text{D}}$ . Final Answer: D

=====

Answer Option 8:

The margin of error refers to the amount of uncertainty in the survey results due to random sampling. In this case, a margin of error of 3% means that the actual percentage of the population that uses grocery coupons could be 3% higher or 3% lower than the reported 82%.

Therefore, the correct answer is D. 'It is likely that between 79% and 85% of the population use grocery coupons.' This option accurately describes the range within which the true percentage of the population using grocery coupons is likely to fall. Final answer: D

=====

Answer Option 9:

A margin of error of 3% means that the survey results may deviate from the true population value by up to 3%. Since the survey concluded that 82% of the population uses grocery coupons, the margin of error of 3% indicates that the true percentage of the population using grocery coupons could be as low as  $82\% - 3\% = 79\%$  or as high as  $82\% + 3\% = 85\%$ . Therefore, the correct answer is  $\boxed{\text{D}}$ . 'It is likely that between 79% and 85% of the population use grocery coupons.')

\*\*\*\*\*

Question 166:

'A hypothesis test is set up so that  $P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}) = 0.05$  and  $P(\text{failing to reject } H_0 \text{ when } H_0 \text{ is false}) = 0.26$ . What is the power of the test?', A. '0.26', B. '0.05', C. '0.95', D. '0.74'

=====

Answer Option 1:

The power of a hypothesis test is the probability of rejecting the null hypothesis when the null hypothesis is false. In this case, the power is given as  $P(\text{failing to reject } H_0 \text{ when } H_0 \text{ is false}) = 0.26$ . Therefore, the power of the test is  $\boxed{\text{A}} \text{ } 0.26$ \$. Final Answer: A

\*\*\*\*\*

Answer Option 2:

The power of a hypothesis test is the probability of rejecting the null hypothesis when the alternative hypothesis is true. In this case, the power is given by  $P(\text{rejecting } H_0 \text{ when } H_0 \text{ is false})$ . We are given that this probability is 0.26. Therefore, the power of the test is 0.26, which corresponds to answer choice A. Final Answer: A.

\*\*\*\*\*

Answer Option 3:

The power of a hypothesis test is defined as the probability of rejecting the null hypothesis, given that the null hypothesis is false. In this case,  $P(\text{failing to reject } H_0 \text{ when } H_0 \text{ is false}) = 0.26$ , so the power of the test is  $1 - 0.26 = 0.74$ . Thus, the answer is  $\boxed{\text{D}} \text{ } 0.74$ \$.

\*\*\*\*\*

Answer Option 4:

The power of a hypothesis test is defined as the probability of correctly rejecting the null hypothesis ( $H_0$ ) when the alternative hypothesis ( $H_1$ ) is true. In this case, the power is equal to  $1 - P(\text{failing to reject } H_0 \text{ when } H_1 \text{ is true})$ , which is  $1 - 0.26 = 0.74$ \$. Therefore, the power of the test is  $\boxed{\text{D}} \text{ } 0.74$ \$. Final answer: D

\*\*\*\*\*

Answer Option 5:

The power of a hypothesis test is the probability of rejecting the null hypothesis ( $H_0$ ) when the alternative hypothesis ( $H_1$ ) is true. In this case, we are given that the probability of failing to reject  $H_0$  when  $H_0$  is false is 0.26\$. Therefore, the probability of rejecting  $H_0$  when  $H_0$  is false (i.e., the power of the test) is  $1 - 0.26 = 0.74$ \$.

Thus, the power of the test is  $\boxed{\text{D}} \text{ } 0.74$ \$. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The power of a hypothesis test is defined as the probability of rejecting the null hypothesis ( $H_0$ ) when the alternative hypothesis ( $H_1$ ) is true. In this case, rejecting  $H_0$  when  $H_0$  is true is equivalent to making a Type I error, and the probability of this happening is given as 0.05. Failing to reject  $H_0$  when  $H_0$  is false is equivalent to making a Type II error, and the probability of this happening is given as 0.26.

Since the power is the complement of the probability of making a Type II error, we have:

$$\text{Power} = 1 - P(\text{Type II Error}) = 1 - 0.26 = 0.74$$

Therefore, the power of the test is 0.74, which corresponds to answer choice D.

\*\*\*\*\*

Answer Option 7:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis ( $H_0$ ) when the null hypothesis is false. In this case, the probability of failing to reject  $H_0$  when  $H_0$  is false is given as 0.26. So, the probability of correctly rejecting  $H_0$  when  $H_0$  is false (i.e., the power of the test) is  $1 - 0.26 = 0.74$ . Therefore, the correct answer is  $\boxed{\text{D}} \text{ } 0.74$ \$. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The power of a hypothesis test is the probability of rejecting the null hypothesis ( $H_0$ ) when the alternative hypothesis ( $H_1$ ) is true. In this case,  $P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}) = 0.05$ , so the significance level (Type I error) is 0.05. This means that there is a 5% chance of rejecting  $H_0$  when it is actually true.

Now, we are given  $P(\text{failing to reject } H_0 \text{ when } H_0 \text{ is false}) = 0.26$ , which is the complement of the power. Therefore, the power is  $1 - 0.26 = 0.74$ .

Thus, the power of the test is  $\boxed{\text{D}} \text{ } 0.74$ \$. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis ( $H_0$ ) when the alternative hypothesis ( $H_1$ ) is true. In this case, we are given that the probability of rejecting  $H_0$  when  $H_0$  is true is 0.05, which is the significance level (denoted as  $\alpha$ ). The complement of  $\alpha$  is the probability of not rejecting  $H_0$  when  $H_0$  is true, also known as the Type I error rate.

We are also given that the probability of failing to reject  $H_0$  when  $H_0$  is false is 0.26. This is the probability of making a Type II error when  $H_1$  is true. In other words, it is the probability of not correctly rejecting  $H_0$  when  $H_1$  is actually true.

Since the power of a test is equal to 1 minus the probability of a Type II error, the power of the test is  $1 - 0.26 = \boxed{\text{D}} \text{ } 0.74$ \$. Final Answer: D

\*\*\*\*\*

Question 167:

'For a class project, Charlotte recorded the heights of all 28 students in her class and calculated several statistics. She then realized she made an error recording the height of the tallest person in the class. She correctly had him listed as the tallest, but needed to add two inches to his recorded height to correct it. Which of these measures of spread must remain unchanged?', A. 'Mean absolute deviation', B. 'Standard deviation', C. 'Variance', D. 'Interquartile range'

=====

Answer Option 1:

Adding a constant value to each data point does not change the mean, so the 'Mean absolute deviation' remains unchanged. However, since the standard deviation and variance are calculated using the deviations from the mean, adding a constant value to each data point will change the deviation and therefore change both the standard deviation and variance. The interquartile range, on the other hand, only depends on the order statistics and is unaffected by adding a constant to each data point. Therefore, the answer is  $\boxed{\text{(D) Interquartile range}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

Adding two inches to the height of the tallest person in the class will only affect the maximum value and will not affect any other values or the overall distribution of heights. Therefore, the measures of spread that are unaffected by this change are the ones that do not depend on the maximum value. The mean absolute deviation, standard deviation, and variance all take into account the differences between each data point and the mean of the data set, but they do not directly depend on the maximum value. However, the interquartile range, which is the difference between the third quartile (Q3) and the first quartile (Q1), does depend on the actual values of the data set and will be affected by the change in the maximum value. Therefore, the measure of spread that must remain unchanged is the 'Mean absolute deviation' (Option A). Final Answer: A

\*\*\*\*\*

Answer Option 3:

Adding a constant value to all the data values will not change the interquartile range, as the relative position of the data values remains the same. Therefore, the correct answer is  $\boxed{\text{(D) Interquartile range}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

The mean absolute deviation, standard deviation, and variance are all measures of spread that are affected by changes in individual data points. Adding two inches to the height of the tallest person will increase the spread of the data, and thus change these measures.

On the other hand, the interquartile range (IQR) is a measure of spread that is based on the range between the first quartile (25th percentile) and the third quartile (75th percentile) of the data. Adding two inches to the tallest person's height will not affect the quartiles or the IQR, as it only depends on the relative positioning of the middle 50% of the data.

Therefore, the measure of spread that must remain unchanged is the 'Interquartile range' or option D.

\*\*\*\*\*

Answer Option 5:

Adding a constant value to each data point will not change the following measures of spread: mean absolute deviation, standard deviation, and variance. These measures are all based on the differences between the data points and the mean, and adding a constant value to each data point will only shift all these differences by the same amount, resulting in the same spread.

However, the interquartile range (IQR) is based on the values of the first quartile (Q1) and the third quartile (Q3), which are determined by the actual values in the dataset. Adding a constant value to the tallest person's height will affect both Q1 and Q3, and therefore, the IQR will be changed.

Therefore, the measure of spread that must remain unchanged is the mean absolute deviation, so the answer is  $\boxed{\text{(A)}}$ . Final answer: A

\*\*\*\*\*

Answer Option 6:

Adding a constant value to each data point does not change any of the measures of spread listed. Therefore, the answer is  $\boxed{\text{(D. 'Interquartile range')}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

Adding a constant value to each data point will not change the mean absolute deviation, standard deviation, variance, or interquartile range. Therefore, the correct answer is  $\boxed{\text{(A)}}$  'Mean absolute deviation'. Final Answer: A

\*\*\*\*\*

Answer Option 8:

Adding a constant value to each data point will not change the interquartile range. This is because the interquartile range depends only on the order of the data, not the actual values. Since the tallest person in the class remains the tallest after the correction, this means that their position relative to the other data points remains unchanged. Therefore, the interquartile range must remain unchanged. The correct answer is  $\boxed{\text{(D. 'Interquartile range')}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

Adding a constant value to each data point will not change the following measures of spread: 'Mean absolute deviation' and 'Standard deviation'.

Mean absolute deviation (MAD) is the average of the absolute differences between each data point and the mean. Adding a constant value to each data point will add the same constant value to the mean, resulting in the same absolute differences and therefore the same MAD.

Standard deviation is the square root of the variance, and variance is the average of the squared differences between each data point and the mean. Adding a constant value to each data point will add the same constant value to the mean, resulting in the same squared differences and therefore the same variance and standard deviation.

However, the 'Interquartile range' will change because it is based on the quartiles of the data, which depend on the actual values of the data points.

Therefore, the correct answer is  $\boxed{\text{(A) Mean absolute deviation}}$ .

\*\*\*\*\*



Question 168:

'The main purpose of blocking in an experiment is to:', A. 'reduce bias.', B. 'reduce confounding.', C. 'reduce variation within treatments.', D. 'reduce variation between treatments.'

=====

Answer Option 1:

The main purpose of blocking in an experiment is to reduce confounding. Confounding occurs when there are variables other than the treatment variable that affect the response variable. By blocking, we ensure that these confounding variables are evenly distributed among the treatment groups, reducing their impact on the results. Therefore, the correct answer is B. 'reduce confounding.' Final answer: B

\*\*\*\*\*

Answer Option 2:

The main purpose of blocking in an experiment is to reduce confounding. Confounding occurs when there are additional variables that are correlated with both the treatment and the outcome, making it difficult to determine the true effect of the treatment. Blocking helps to control for these additional variables by creating homogeneous groups within which the treatment effect can be more accurately assessed. Therefore, the correct answer is B. 'reduce confounding.'

\*\*\*\*\*

Answer Option 3:

The main purpose of blocking in an experiment is to reduce variation within treatments. This is achieved by grouping similar experimental units together into blocks and then randomly assigning treatments within each block. By doing so, any sources of variability that are specific to a particular block are controlled for, leading to a reduction in variation within treatments. Therefore, the correct answer is C. 'reduce variation within treatments.'

\*\*\*\*\*

Answer Option 4:

The main purpose of blocking in an experiment is to reduce confounding. By blocking, we group similar individuals or subjects together to create blocks, ensuring that any potential confounding variables are evenly distributed within each block. This helps to reduce the influence of confounding factors on the treatment effect and allows for a more accurate assessment of the treatment's true effect. Therefore, the correct answer is B. 'reduce confounding.' Final Answer: B

\*\*\*\*\*

Answer Option 5:

The main purpose of blocking in an experiment is to reduce variation between treatments. By blocking, we group similar experimental units together based on certain characteristics or variables that may affect the outcome of the experiment. This helps to control or reduce the effects of these variables on the response variable, making the treatment effects more apparent and reducing the variation between treatments. Therefore, the correct answer is  $\boxed{\text{D. 'reduce variation between treatments.'}}$

\*\*\*\*\*

Answer Option 6:

The main purpose of blocking in an experiment is to reduce confounding. Confounding occurs when variables other than the treatment being studied affect the outcome of the experiment. By blocking, we group similar subjects together and ensure that each treatment group is represented equally within each block. This helps to eliminate the potential influence of confounding variables, allowing us to draw more valid conclusions about the effect of the treatment. Therefore, the correct answer is B. 'reduce confounding.' Final answer: B

\*\*\*\*\*

Answer Option 7:

The main purpose of blocking in an experiment is to reduce variation within treatments. By blocking, we group similar experimental units together and assign different treatments within each block. This helps account for potential sources of variation that may exist within the blocks, such as differences in initial characteristics or conditions. Blocking helps to ensure that any observed differences between treatments are not due to factors within the blocks, but rather due to the actual treatments themselves. Therefore, the correct answer is  $\boxed{\text{C. 'reduce variation within treatments.'}}$

\*\*\*\*\*

Answer Option 8:

The main purpose of blocking in an experiment is to reduce bias. By blocking, we ensure that each treatment group is representative of the different factors or variables that could potentially affect the outcome of the experiment. This helps to eliminate or reduce any potential confounding variables and ensures that the treatment groups are comparable. Therefore, the correct answer is A. 'reduce bias.'

Final answer: A

\*\*\*\*\*

Answer Option 9:

The main purpose of blocking in an experiment is to reduce confounding. Confounding occurs when an extraneous variable is associated with both the treatment and the outcome, making it difficult to determine the true effect of the treatment. Blocking helps to account for these potential confounding variables by creating homogeneous groups within each treatment group, thus reducing the variability due to these variables. Therefore, the correct answer is B. 'reduce confounding.' Final Answer: B

\*\*\*\*\*

Question 169:

'A manufacturer of motor oil is interested in testing the effects of a newly developed additive on the lifespan of an engine. Twenty-five different engine types are selected at random and each one is tested using oil with the additive and oil without the additive. What type of analysis will yield the most useful information?'. A. 'Matched pairs comparison of population proportions', B. 'Matched pairs comparison of population means', C. 'Independent samples comparison of population proportions', D. 'Independent samples comparison of population means'

=====

Answer Option 1:

The most useful information would be obtained through a 'Matched pairs comparison of population means' analysis. This involves testing the same engine type with and without the additive, allowing for a direct comparison of the lifespan of each engine under different conditions. By using matched pairs, any variability in the results due to engine type is controlled for, and the effect of the additive can be isolated. Thus, the correct choice is B. 'Matched pairs comparison of population means'. Final Answer: B

\*\*\*\*\*

Answer Option 2:

To determine the effects of the newly developed additive on the lifespan of an engine, it would be most useful to conduct a 'Matched pairs comparison of population means' analysis. In this type of analysis, each engine type serves as its own control, as it is tested both with and without the additive. By comparing the lifespans of the engines using oil with the additive to the lifespans of the same engines using oil without the additive, we can directly measure the impact of the additive on each engine type.

Therefore, the correct choice is B. 'Matched pairs comparison of population means'.

\*\*\*\*\*

Answer Option 3:

The most useful analysis in this scenario would be a "Matched pairs comparison of population means" (Option B).

In this analysis, each engine type is tested using both the oil with the additive and the oil without the additive. By comparing the performance of each engine type under both conditions, we can directly observe the effect of the additive on the lifespan of the engine.

This approach eliminates potential confounding factors that may vary between different engine types, such as engine design or operating conditions. It allows for a direct comparison within each engine type, providing more precise and informative results.

Therefore, the correct answer is Option B, "Matched pairs comparison of population means". Final Answer: B

\*\*\*\*\*

Answer Option 4:

The most useful information for this scenario would be obtained through a "Matched pairs comparison of population means" analysis (option B).

In this analysis, each engine type is tested using both the oil with the additive and the oil without the additive. By comparing the lifespans of the engines using the two types of oil, we can directly assess the effect of the additive on each individual engine type. This method allows us to control for the variability between different engine types and focus on the specific impact of the additive.

Option A, "Matched pairs comparison of population proportions", would be more suitable if we were comparing categorical variables or proportions (e.g., the proportion of engines that failed using each type of oil).

Options C and D, "Independent samples comparison of population proportions" and "Independent samples comparison of population means", respectively, involve comparing independent groups, which is not applicable in this scenario since we are interested in analyzing the effect of the additive within the same group of engines.

\*\*\*\*\*

Answer Option 5:

The most useful analysis in this scenario would be a 'Matched pairs comparison of population means' (B). In this type of analysis, each engine type is tested using both the oil with the additive and the oil without the additive. By comparing the lifespan of each engine type with and without the additive, we can directly measure the effect of the additive on the engine lifespan. This analysis eliminates the potential influence of other factors, such as variations between engine types, and allows for a more accurate assessment of the additive's impact.

The other options, 'Matched pairs comparison of population proportions' (A), 'Independent samples comparison of population proportions' (C), and 'Independent samples comparison of population means' (D), are not as appropriate in this context. These options either involve comparing proportions (which may not be applicable in this case) or comparing independent samples (which would not account for the variability among engine types). Therefore, the most useful analysis in this situation is a 'Matched pairs comparison of population means' (B). Final Answer: B

\*\*\*\*\*

Answer Option 6:

The most useful analysis in this situation would be a 'Matched pairs comparison of population means', as we are interested in comparing the lifespan of an engine when using oil with the additive versus oil without the additive. By using a matched pairs design, where each engine type is tested with both types of oil, we can directly compare the results within each engine type and assess the impact of the additive on the engine lifespan. This analysis allows us to control for any potential confounding factors that may vary between different engine types. Therefore, the correct choice is B. 'Matched pairs comparison of population means'. Final answer: B

\*\*\*\*\*

Answer Option 7:

The most useful analysis in this scenario would be a "Matched pairs comparison of population means" (B). In this study design, each engine type is tested using both oil with the additive and oil without the additive. The engine types serve as the matched pairs, and the lifespan of the engine is measured as the outcome variable. By comparing the lifespans of the engines using the two types of oil, we can directly assess the effects of the additive on the lifespan. This design eliminates the potential confounding factors from the variation between different engine types, as each engine type serves as its own control. Therefore, the matched pairs comparison of population means is the most appropriate analysis in this case. Final answer: B

\*\*\*\*\*

Answer Option 8:

The most useful analysis in this scenario would be a "Matched pairs comparison of population means" (B).

In this study, the same engine type is tested using oil with the additive and oil without the additive. The purpose is to compare the lifespan of the engine using the two different types of oil. By using matched pairs (i.e., testing the same engine type twice), any variability in engine lifespan that is not due to the additive can be controlled for. This allows for a more accurate comparison between the two types of oil.

A matched pairs comparison of population proportions (A) would be more appropriate if the study was looking at a categorical outcome (e.g., pass/fail) rather than the lifespan of the engine.

Independent samples comparison of population proportions (C) would be more appropriate if the study was comparing different engine types using the oil with and without the additive.

Independent samples comparison of population means (D) would also be more appropriate if the study was comparing different engine types using the oil with and without the additive, but without matching the engine types.

\*\*\*\*\*

Answer Option 9:

The most useful analysis in this case would be a matched pairs comparison of population means, represented by option B. In this study, each engine type is tested using both oil with the additive and oil without the additive. By comparing the lifespans of the engines under the two conditions, we can directly assess the effect of the additive on each engine type. This analysis takes into account the variability among the engines and allows for a more accurate evaluation of the additive's impact. Option A (matched pairs comparison of population proportions) would be suitable if we were comparing categorical data, but here we are interested in the lifespan, which is a continuous variable. Independent samples comparison of population proportions (option C) and independent samples comparison of population means (option D) would not be as effective in this case, as we are specifically interested in the effect of the additive on each engine type, rather than comparing different types of engines with and without the additive. Therefore, the most appropriate analysis is a matched pairs comparison of population means (option B). Final Answer: B

\*\*\*\*\*

Question 170:

"To determine the average cost of running for a congressional seat, a simple random sample of 50 politicians is chosen and the politicians' records examined. The cost figures show a mean of \$125,000 with a standard deviation of \$32,000. Which of the following is the best interpretation of a 90% confidence interval estimate for the average cost of running for office?", A. "90% of politicians running for a congressional seat spend between \$117,500 and \$132,500.", B. "90% of politicians running for a congressional seat spend a mean dollar amount that is between \$117,500 and \$132,500.", C. "We are 90% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500.", D. "We are 90% confident that politicians running for a congressional seat spend a mean dollar amount between \$117,500 and \$132,500."

=====

Answer Option 1:

The correct interpretation of a 90% confidence interval estimate is: "We are 90% confident that the true average cost of running for a congressional seat falls between \$117,500 and \$132,500." Therefore, the answer is  $\boxed{\text{C. 'We are 90\% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500.'}}$

\*\*\*\*\*

Answer Option 2:

The correct interpretation of a confidence interval is option C: "We are 90% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500."

A confidence interval does not provide information about the proportion of politicians or the mean dollar amount spent by politicians. Instead, it provides a range of values within which we can be confident that the true population parameter (in this case, the average cost of running for office) falls. The confidence level (90% in this case) indicates the level of confidence we have in the interval containing the true parameter. Final Answer: C

\*\*\*\*\*

Answer Option 3:

The best interpretation of a 90% confidence interval estimate is option C - "We are 90% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500."

A confidence interval is a range of values that is likely to contain the true population parameter with a certain level of confidence. In this case, the confidence level is 90%. The interval, \$117,500 to \$132,500, represents the range of likely values for the average cost of running for office, based on the sample data.

It is important to note that the confidence interval provides a range of values, not a specific value. Therefore, it cannot be stated with certainty that "90% of politicians running for a congressional seat spend between \$117,500 and \$132,500." (option A) or that "90% of politicians running for a congressional seat spend a mean dollar amount that is between \$117,500 and \$132,500." (option B). The correct interpretation is that we are 90% confident that the true average cost of running for office falls within the given range. Final answer: C

\*\*\*\*\*

Answer Option 4:

The best interpretation of a confidence interval is C. "We are 90% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500." This statement captures the idea that the true average cost of running for a congressional seat falls within the specified range with 90% confidence. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The best interpretation of a confidence interval is option D, "We are 90% confident that politicians running for a congressional seat spend a mean dollar amount between \$117,500 and \$132,500." This interpretation captures the idea that the confidence interval is an interval estimate for the true population mean, and the level of confidence (90% in this case) represents the percentage of confidence intervals that would contain the true population mean if the sampling process were repeated many times. Final answer: D

\*\*\*\*\*

Answer Option 6:

The best interpretation of a 90% confidence interval estimate is C. "We are 90% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500."

A confidence interval is a range of values that is likely to contain the true population parameter with a certain level of confidence. In this case, the confidence interval estimate for the average cost of running for office is between \$117,500 and \$132,500. This means that if we were to repeat the process of sampling and calculating the confidence interval many times, approximately 90% of the intervals would contain the true average cost of running for office. Final Answer: C

\*\*\*\*\*

Answer Option 7:

The correct interpretation of a confidence interval is option C: "We are 90% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500." This interpretation captures the idea that the true population mean falls within the given interval with 90% confidence. Option A does not mention confidence, option B incorrectly refers to the individual politicians' spending, and option D incorrectly refers to the mean spending as a range. Final answer: C

\*\*\*\*\*

Answer Option 8:

The best interpretation of a 90% confidence interval estimate for the average cost of running for office is option C: "We are 90% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500."

A confidence interval provides a range of values within which the true population parameter (in this case, the average cost of running for office) is likely to fall. The confidence level (in this case, 90%) represents the probability that the interval contains the true population parameter.

Therefore, option C accurately represents the interpretation of the confidence interval estimate. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct interpretation of a 90% confidence interval estimate is given by option C: 'We are 90% confident that politicians running for a congressional seat spend between \$117,500 and \$132,500.'

A confidence interval is a range of values within which the true population parameter is estimated to fall. In this case, we are estimating the average cost of running for office. The 90% confidence interval indicates that if we were to repeat the sampling process many times, about 90% of the resulting intervals would contain the true average cost.

Therefore, we can say with 90% confidence that the average cost of running for a congressional seat falls between \$117,500 and \$132,500. Final answer: C

\*\*\*\*\*

Question 171:

'A filling machine puts an average of four ounces of coffee in jars, with a standard deviation of 0.25 ounces. Forty jars filled by this machine are selected at random. What is the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces?', A. '0.0057', B. '0.0225', C. '0.025', D. '0.05'

=====

Answer Option 1:

The mean amount per jar filled in the sampled jars is a sample mean. The sample mean follows a normal distribution with mean equal to the population mean (4 ounces) and standard deviation equal to the population standard deviation divided by the square root of the sample size (0.25 ounces divided by the square root of 40).

Thus, the standard deviation of the sample mean is  $0.25/\sqrt{40} \approx 0.0396$  ounces.

To find the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces, we need to find the probability that a normally distributed random variable with mean 4 ounces and standard deviation 0.0396 ounces is less than 3.9 ounces.

Using standardization, we find that

$$P(X < 3.9) = P\left(Z < \frac{3.9 - 4}{0.0396}\right) = P(Z < -2.5253)$$

where  $Z$  is a standard normal random variable.

Using a standard normal table or a calculator, we find that  $P(Z < -2.5253) \approx 0.0057$ .

Therefore, the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces is approximately  $\boxed{0.0057}$ . Final answer: A

\*\*\*\*\*

Answer Option 2:

The mean amount per jar filled by the machine is 4 ounces. The standard deviation of the mean, also known as the standard error, is the standard deviation of the individual jars divided by the square root of the sample size. In this case, the standard error is  $\frac{0.25}{\sqrt{40}}$ .

To find the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces, we can use the z-score. The z-score is calculated by subtracting the mean from the desired value and dividing by the standard error:  $z = \frac{3.9 - 4}{\frac{0.25}{\sqrt{40}}}$ .

Using a z-table or calculator, we find that the z-score of -2.83 corresponds to a probability of approximately 0.0023. However, since we want the probability that the mean amount is less than 3.9 ounces, we need to find the probability to the left of the z-score. This is 0.5 minus the probability to the right of the z-score:  $0.5 - 0.0023 = 0.4977$ .

Therefore, the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces is approximately  $\boxed{0.4977}$ . Final Answer: None of the above

\*\*\*\*\*

Answer Option 3:

The mean amount per jar filled by the machine is 4 ounces and the standard deviation is 0.25 ounces. Since we are sampling 40 jars, according to the Central Limit Theorem, the distribution of the mean amount per jar filled in the sampled jars will approach a normal distribution with mean 4 ounces and standard deviation  $\frac{0.25}{\sqrt{40}}$  ounces.

We want to find the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces. We can standardize the value 3.9 using the z-score formula:

$$Z = \frac{X - \mu}{\sigma} = \frac{3.9 - 4}{\frac{0.25}{\sqrt{40}}} = -4\sqrt{10}$$

Using a standard normal distribution table or calculator, we find that the probability that a standard normal random variable is less than  $-4\sqrt{10}$  is approximately  $\boxed{0.0057}$ . Final Answer: A

\*\*\*\*\*

Answer Option 4:

The mean amount of coffee per jar filled by the machine is 4 ounces, and the standard deviation is 0.25 ounces. Since we are sampling from a large population (40 jars), we can approximate the distribution of the mean amount per jar as a normal distribution with mean 4 and standard deviation  $\frac{0.25}{\sqrt{40}} \approx 0.0395$  ounces.

To find the probability that the mean amount per jar is less than 3.9 ounces, we need to find the z-score associated with 3.9 ounces and then find the corresponding cumulative probability from the standard normal distribution.

The z-score is calculated as  $z = \frac{3.9 - 4}{0.0395} \approx -2.53$ . Looking up this z-score in the standard normal distribution table, we find that the cumulative probability is approximately 0.0057.

Therefore, the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces is approximately  $\boxed{0.0057}$ . Final Answer: A

\*\*\*\*\*

Answer Option 5:

We can use the Central Limit Theorem to approximate the distribution of the sample mean. The mean of the sample mean will be equal to the population mean, which is 4 ounces. The standard deviation of the sample mean will be equal to the population standard deviation divided by the square root of the sample size. So, the standard deviation of the sample mean is 0.25 ounces divided by the square root of 40, which is approximately 0.0397 ounces.

We want to find the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces. We can standardize this value using the sample mean and the standard deviation of the sample mean:

$$Z = \frac{X - \mu}{\sigma} = \frac{3.9 - 4}{0.0397} = -2.5189$$

Now, we can find the probability using the standard normal distribution. Looking up the value of -2.5189 in a standard normal distribution table, we find that the probability is approximately 0.0057.

Therefore, the correct answer is  $\boxed{\text{A. } 0.0057}$ . Final answer: A

\*\*\*\*\*

Answer Option 6:

The mean amount per jar filled in the sampled jars is a random variable that follows a normal distribution because of the Central Limit Theorem. The mean of this distribution is equal to the mean of the individual jars, which is four ounces. The standard deviation of this distribution, known as the standard error, is equal to the standard deviation of the individual jars divided by the square root of the sample

size. In this case, the standard error is  $\$0.25/\sqrt{40} \approx 0.0397\$$  ounces.

To find the probability that the mean amount per jar filled is less than 3.9 ounces, we need to find the z-score corresponding to this value and then find the area under the normal curve to the left of this z-score.

The z-score is calculated as  $\$(3.9 - 4) / 0.0397 \approx -2.5189\$$ . Looking up this z-score in a standard normal distribution table or using a calculator, we find that the area to the left of this z-score is approximately 0.0057.

Therefore, the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces is approximately  $\boxed{0.0057}$ %, which corresponds to option A. Final answer: A

\*\*\*\*\*

Answer Option 7:

The mean amount per jar filled by the machine is 4 ounces, and the standard deviation is 0.25 ounces. We are interested in the probability that the mean amount per jar in the sampled jars is less than 3.9 ounces.

To solve this, we can use the Central Limit Theorem, which states that the distribution of sample means approaches a normal distribution as the sample size increases. Since the sample size is 40, we can assume that the distribution of sample means is approximately normal.

The standard deviation of the distribution of sample means, also known as the standard error, can be calculated by dividing the standard deviation of the population by the square root of the sample size. In this case, the standard error is  $\$\frac{0.25}{\sqrt{40}} \approx 0.0396\$$  ounces.

To find the probability that the mean amount per jar in the sampled jars is less than 3.9 ounces, we can standardize the value using the z-score formula:

$$Z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to standardize,  $\mu$  is the mean of the distribution, and  $\sigma$  is the standard deviation of the distribution.

Plugging in the values, we have:

$$Z = \frac{3.9 - 4}{0.0396} \approx -2.5253$$

Using a standard normal distribution table or calculator, we can find that the probability of getting a z-score less than -2.5253 is approximately 0.0057.

Thus, the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces is  $\boxed{0.0057}$ %.

Final Answer: A

\*\*\*\*\*

Answer Option 8:

The mean amount per jar filled in the sampled jars is a random variable with a normal distribution. The mean of this distribution is still four ounces, and the standard deviation is  $\frac{0.25}{\sqrt{40}}$  ounces (by the Central Limit Theorem). We want to find the probability that the mean amount is less than 3.9 ounces.

Converting to a standard normal distribution, we have:

$$P\left(Z < \frac{3.9 - 4}{\frac{0.25}{\sqrt{40}}}\right) = P\left(Z < -2\sqrt{40}\right)$$

Using a table or calculator, we find that  $P(Z < -2\sqrt{40}) \approx 0.0057$ %.

Therefore, the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces is approximately  $\boxed{0.0057}$ %. Final answer: A

\*\*\*\*\*

Answer Option 9:

The mean amount per jar filled by the machine follows a normal distribution with mean 4 ounces and standard deviation 0.25 ounces. The distribution of the sample mean is also normal with the same mean of 4 ounces, but with a standard deviation of  $\frac{0.25}{\sqrt{40}}$  ounces, since we are taking the mean of 40 jars.

We want to find the probability that the sample mean is less than 3.9 ounces. We can standardize this value by subtracting the mean and dividing by the standard deviation:

$$Z = \frac{3.9 - 4}{\frac{0.25}{\sqrt{40}}} = -3.5777$$

Using a standard normal distribution table or calculator, we can find that the probability of Z being less than -3.5777 is approximately 0.00018.

Therefore, the probability that the mean amount per jar filled in the sampled jars is less than 3.9 ounces is approximately  $\boxed{0.0057}$ %. Final answer: A

\*\*\*\*\*

Question 172:

'A town has one high school, which buses students from urban, suburban, and rural communities. Which of the following samples is recommended in studying attitudes toward tracking of students in honors, regular, and below-grade classes?', A. 'Convenience sample', B. 'Simple random sample (SRS)', C. 'Stratified sample', D. 'Systematic sample'

=====

Answer Option 1:

The recommended sample for studying attitudes toward tracking of students in different classes would be a 'Stratified sample'. A stratified sample involves dividing the population into different groups or strata based on certain characteristics (in this case, the type of class), and then randomly selecting individuals from each stratum. This ensures that there is representation from each group in the sample, which is important for studying attitudes towards tracking in different classes. Therefore, the correct answer is C. 'Stratified sample'. Final answer: C

\*\*\*\*\*

Answer Option 2:

The recommended sample in studying attitudes toward tracking of students in different classes would be a stratified sample. This is because we want to ensure that we have representation from different communities (urban, suburban, and rural) in our sample. By dividing the population into distinct strata (in this case, the different communities), we can randomly sample from each stratum to ensure that we have a representative sample that reflects the diversity of the population. Therefore, the answer is  $\boxed{\text{C. Stratified sample}}$ . Final answer: C

\*\*\*\*\*

Answer Option 3:

The recommended sample method for studying attitudes toward tracking of students in different classes would be a 'Stratified sample'. This is because a stratified sample allows for the intentional and proportional representation of different groups or strata within the population. In this case, the population would consist of students from urban, suburban, and rural communities. By using a stratified sample, we can ensure that the attitudes of students from each community are adequately represented in the study, leading to more accurate and meaningful results. Therefore, the answer is option C. 'Stratified sample'. Final Answer: C

\*\*\*\*\*

Answer Option 4:

In order to accurately study attitudes towards tracking of students in different classes, it is important to obtain a representative sample that includes students from all different communities. A convenience sample would not be recommended as it is based on convenience and does not ensure a representative sample. A simple random sample (SRS) would also not be the best choice as it may not guarantee an equal representation from each community. A systematic sample would also not be ideal as it may introduce bias if the sampling interval aligns with any patterns present in the population.

The most suitable option would be a stratified sample, where the population is divided into different groups (urban, suburban, rural) and a random sample is selected from each group. This ensures that students from all communities are represented in the study. Therefore, the recommended option is C. 'Stratified sample'. Final Answer: C

\*\*\*\*\*

Answer Option 5:

In order to study attitudes toward tracking of students in different classes, it is important to ensure that the sample is representative of the entire population.

A. A convenience sample is one that is easily accessible and convenient for the researcher to collect. This type of sample may not be representative of the population, as it may not include individuals from all demographics or areas. Therefore, it is not recommended in this case.

B. A simple random sample (SRS) involves randomly selecting individuals from the population, ensuring that each individual has an equal chance of being selected. This type of sample is more likely to be representative of the population, making it a good choice for studying attitudes toward tracking.

C. A stratified sample involves dividing the population into distinct groups (strata) and then randomly selecting individuals from each group. This helps ensure that individuals from different demographics or areas are included in the sample. This could be a good choice for studying attitudes toward tracking if the researcher wants to specifically compare opinions across different communities (urban, suburban, rural).

D. A systematic sample involves selecting individuals from the population at regular intervals. This might not be ideal for studying attitudes toward tracking, as it may introduce biases if there is a pattern or structure in the population.

Given the information provided, the recommended sample for studying attitudes toward tracking of students in different classes would be either a simple random sample (SRS) or a stratified sample (depending on the specific research goals). Therefore, the answer is either B. 'Simple random sample (SRS)' or C. 'Stratified sample'.

\*\*\*\*\*

Answer Option 6:

The recommended sample in studying attitudes toward tracking of students in different classes would be a 'Stratified sample' (option C).

A stratified sample involves dividing the population into different groups or strata based on certain characteristics (such as urban, suburban, and rural communities in this case). Then, a random sample is taken from each stratum to ensure representation from different groups. This allows for a more accurate representation of the entire population and ensures that each group is adequately represented in the sample.

Options A, B, and D are not the most appropriate choices for this scenario:

- A convenience sample (option A) involves selecting individuals who are readily available or easily accessible. This may introduce bias and may not provide a representative sample.

- A simple random sample (SRS) (option B) involves selecting individuals randomly from the entire population. While this can be effective in some cases, it may not capture the differences in attitudes toward tracking among different communities.

- A systematic sample (option D) involves selecting individuals at regular intervals from a population. This may not capture the diversity of attitudes toward tracking in different communities.

Therefore, the most suitable sample for studying attitudes toward tracking of students in different classes would be a stratified sample (option C). Final Answer: C

\*\*\*\*\*



Answer Option 7:

In order to study attitudes toward tracking of students in different classes, it is important to ensure that the sample is representative of the entire population.

Option A, a convenience sample, is not recommended because it involves selecting individuals who are easily accessible or readily available, which may introduce bias and not provide a representative sample.

Option B, a simple random sample (SRS), involves randomly selecting individuals from the entire population, which is a good way to ensure that the sample is representative.

Option C, a stratified sample, involves dividing the population into subgroups (strata) and then randomly selecting individuals from each subgroup. This can be useful if there are important subgroups in the population that need to be represented in the sample. However, if the goal is to study attitudes across different communities (urban, suburban, and rural), a stratified sample may not be the best choice.

Option D, a systematic sample, involves selecting individuals according to a predetermined systematic approach, such as selecting every  $n$ th individual from a list. While this method can be useful in some cases, it may introduce bias if there is a hidden pattern in the list that is related to the attitudes being studied.

Given the goal of studying attitudes across different communities, the recommended sample is a 'Stratified sample' (Option C). This will ensure representation from urban, suburban, and rural communities.

Thus, the answer is C.

\*\*\*\*\*

Answer Option 8:

In studying attitudes toward tracking of students in different classes, it is important to ensure that the sample is representative of the entire population. A convenience sample (option A) is not recommended because it involves selecting individuals who are easily accessible or readily available, which may result in a biased sample.

A simple random sample (SRS) (option B) involves randomly selecting individuals from the population, giving each member an equal chance of being selected. This method helps to minimize bias and ensure representativeness, making it a good choice for studying attitudes toward tracking.

A stratified sample (option C) involves dividing the population into different groups or strata and then randomly selecting individuals from each group. In this case, the population would need to be divided into urban, suburban, and rural communities, and then individuals would be randomly selected from each group. While this method can be effective for ensuring representation from each group, it may not be necessary if the goal is to study attitudes toward tracking across all communities.

A systematic sample (option D) involves selecting individuals at fixed intervals from an ordered list of the population. This method may introduce bias if there is a pattern in the ordering of the list that relates to the variable being studied, such as attitudes toward tracking.

Therefore, the recommended sample for studying attitudes toward tracking of students in different classes would be a simple random sample (SRS), so the correct answer is  $\boxed{\text{B. Simple random sample (SRS)}}$ .

\*\*\*\*\*

Answer Option 9:

The recommended sample in studying attitudes toward tracking of students in different classes would be a 'Stratified sample'. A stratified sample involves dividing the population into homogeneous subgroups (in this case, the different classes) and then randomly selecting participants from each subgroup. This ensures representation from each class and allows for comparisons and analyses between the different groups. Final Answer: C

\*\*\*\*\*

Question 173:

'A shipment of resistors have an average resistance of 200 ohms with a standard deviation of 5 ohms, and the resistances are normally distributed. Suppose a randomly chosen resistor has a resistance under 194 ohms. What is the probability that its resistance is greater than 188 ohms?', A. '0.07', B. '0.12', C. '0.50', D. '0.93'

=====

Answer Option 1:

To solve this problem, we can use the properties of the normal distribution.

The average resistance is 200 ohms and the standard deviation is 5 ohms. We want to find the probability that the resistance is greater than 188 ohms.

First, we need to calculate the z-score of 188 ohms. The z-score is given by:

$$z = \frac{x - \mu}{\sigma}$$

where x is the value we want to find the probability for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Plugging in the values, we get:

$$z = \frac{188 - 200}{5} = -2.4$$

Next, we can use a standard normal distribution table or a calculator to find the probability corresponding to this z-score.

From the standard normal distribution table, we find that the probability of getting a z-score less than -2.4 is approximately 0.0082.

Since we want the probability that the resistance is greater than 188 ohms, we subtract this probability from 1:

$$P(\text{resistance} > 188) = 1 - 0.0082 = 0.9918$$

Therefore, the probability that the resistance is greater than 188 ohms is approximately 0.9918 or 99.18%.

The answer is not given among the choices.

=====

Answer Option 2:

To solve this problem, we need to use the concept of standard deviation.

First, we need to standardize the values using the formula  $Z = \frac{X - \mu}{\sigma}$ , where X is the value we are interested in,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For the value 194 ohms, the standardized value is  $Z = \frac{194 - 200}{5} = -1.2$ .

For the value 188 ohms, the standardized value is  $Z = \frac{188 - 200}{5} = -2.4$ .

Next, we need to find the area under the normal distribution curve to the left of  $Z = -1.2$  and  $Z = -2.4$  respectively. We can use a standard normal distribution table or a calculator to find these values.

The area to the left of  $Z = -1.2$  is approximately 0.1151.

The area to the left of  $Z = -2.4$  is approximately 0.0082.

Now, we subtract these values from 1 to find the area to the right of each value.

The area to the right of  $Z = -1.2$  is approximately  $1 - 0.1151 = 0.8849$ .

The area to the right of  $Z = -2.4$  is approximately  $1 - 0.0082 = 0.9918$ .

Since we are interested in the probability that the resistance is greater than 188 ohms, we need to find the probability corresponding to  $Z = -2.4$ .

Therefore, the probability that the resistance is greater than 188 ohms is approximately  $\boxed{0.9918}$ . Final Answer: D

=====

Answer Option 3:

To solve this problem, we can use the properties of the normal distribution. Given that the average resistance is 200 ohms and the standard deviation is 5 ohms, we can convert the given resistances to z-scores.

The z-score for a resistance of 194 ohms is calculated as:

$$z_1 = \frac{194 - 200}{5} = -1.2$$

The z-score for a resistance of 188 ohms is calculated as:

$$z_2 = \frac{188 - 200}{5} = -2.4$$

We want to find the probability that the resistance is greater than 188 ohms, which is the probability that the z-score is greater than -2.4. By looking up the z-score in the standard normal distribution table (or using a calculator), we find that the probability is approximately 0.9938.

Therefore, the answer is  $\boxed{0.93}$ . Final answer: D

=====

Answer Option 4:

To solve this problem, we can standardize the values using z-scores. The z-score is calculated as  $\frac{x - \mu}{\sigma}$ , where x is the given value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For a randomly chosen resistor, we are given that  $x = 194$  ohms. The mean is  $\mu = 200$  ohms, and the standard deviation is  $\sigma = 5$  ohms. The z-score for a resistance of 194 ohms is  $\frac{194 - 200}{5} = -1.2$ .

To find the probability that the resistance is greater than 188 ohms, we need to find the area under the normal curve to the right of the z-score -1.2. Using a standard normal distribution table or a calculator, we find that the probability is approximately 0.8849.

However, we are asked for the probability that the resistance is greater than 188 ohms, which is the complement of the probability that the resistance is less than or equal to 188 ohms. Using symmetry properties of the normal distribution, we know that the area to the left of a z-score is equal to the area to the right of the negative of that z-score. Therefore, the probability that the resistance is greater than 188 ohms is approximately  $1 - 0.8849 = 0.1151$ .

Rounding to two decimal places, the probability is approximately 0.12, so the correct answer is  $\boxed{\text{(B) } 0.12}$ .

\*\*\*\*\*

Answer Option 5:

To find the probability that a randomly chosen resistor has a resistance greater than 188 ohms, we need to find the z-score for 194 ohms and calculate the area under the standard normal distribution curve to the right of this z-score.

The z-score is calculated as:

$$z = \frac{X - \mu}{\sigma}$$

where X is the value of interest (194 ohms in this case),  $\mu$  is the mean (200 ohms), and  $\sigma$  is the standard deviation (5 ohms).

Plugging in the values, we get:

$$z = \frac{194 - 200}{5} = -1.2$$

Using a standard normal distribution table or calculator, we find that the area to the left of  $z = -1.2$  is approximately 0.1151. Therefore, the area to the right of  $z = -1.2$  (which represents the probability that a resistor has a resistance greater than 188 ohms) is approximately  $1 - 0.1151 = 0.8849$ .

So the probability that a randomly chosen resistor has a resistance greater than 188 ohms is approximately 0.8849, which is closest to 0.93 (D). Final answer: D

\*\*\*\*\*

Answer Option 6:

To solve this problem, we can use the standard normal distribution.

The average resistance of the resistors is 200 ohms, and the standard deviation is 5 ohms. We can convert the resistance values to z-scores using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where x is the resistance value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For the resistor with a resistance of 194 ohms:

$$z = \frac{194 - 200}{5} = -1.2$$

For the resistor with a resistance of 188 ohms:

$$z = \frac{188 - 200}{5} = -2.4$$

We need to find the probability that the resistance is greater than 188 ohms, which is equivalent to finding the probability that the z-score is greater than -2.4.

Using a z-table or a calculator, we find that the probability of a z-score greater than -2.4 is approximately 0.9938.

Therefore, the probability that the resistance is greater than 188 ohms is approximately 0.9938 or about 99.38%.

The answer is not listed among the choices. There seems to be an error in the answer choices.

\*\*\*\*\*

Answer Option 7:

To find the probability that the resistance of a randomly chosen resistor is greater than 188 ohms, we need to find the area under the normal distribution curve to the right of 188 ohms.

First, we need to calculate the z-score corresponding to 188 ohms. The z-score is calculated using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we are interested in (188 ohms),  $\mu$  is the mean (200 ohms), and  $\sigma$  is the standard deviation (5 ohms).

Plugging in the values, we have:

$$z = \frac{188 - 200}{5} = -\frac{12}{5}$$

Next, we need to find the area to the right of the z-score  $-12/5$  under the normal distribution curve. This can be done using a standard normal distribution table or a calculator. The area to the right of  $-12/5$  is equal to 1 minus the area to the left of  $-12/5$ . From the table or calculator, we find that the area to the left of  $-12/5$  is approximately 0.067.

Therefore, the probability that the resistance of a randomly chosen resistor is greater than 188 ohms is approximately  $1 - 0.067 = 0.933$ , which is closest to  $\boxed{\text{(D) } 0.93}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

To solve this problem, we will use the properties of the normal distribution. Given that the average resistance is 200 ohms and the standard deviation is 5 ohms, we can calculate the z-scores for 194 ohms and 188 ohms.

The z-score of a value is calculated by subtracting the mean and dividing by the standard deviation. For 194 ohms:

$$z = \frac{194 - 200}{5} = -1.2$$

For 188 ohms:

$$z = \frac{188 - 200}{5} = -2.4$$

Now, we need to find the probability that a randomly chosen resistor has a resistance greater than 188 ohms. Since the normal distribution is symmetric, we can find this by subtracting the probability of the resistance being less than 188 ohms from 1. We can use a z-table or calculator to find the corresponding probabilities.

From the z-table, we find that the probability of a z-score being less than -1.2 is 0.1151, and the probability of a z-score being less than -2.4 is 0.0082.

Therefore, the probability that a randomly chosen resistor has a resistance greater than 188 ohms is:

$$1 - 0.0082 = 0.9918 \approx \boxed{0.93}$$

Final Answer: D

\*\*\*\*\*

Answer Option 9:

To solve this problem, we need to standardize the values using z-scores. The z-score of a value is defined as the number of standard deviations it is away from the mean. In this case, we want to find the probability that the resistance is greater than 188 ohms.

The z-score of 194 ohms is given by:

$$z = \frac{194 - 200}{5} = -1.2$$

Similarly, the z-score of 188 ohms is given by:

$$z = \frac{188 - 200}{5} = -2.4$$

Since the resistances are normally distributed, we can use the standard normal distribution table to find the probability associated with each z-score.

The probability that the resistance is greater than 194 ohms is given by:

$$P(z > -1.2) = 1 - P(z \leq -1.2)$$

From the standard normal distribution table, we find that  $P(z \leq -1.2) \approx 0.1151$ . Therefore,

$$P(z > -1.2) \approx 1 - 0.1151 \approx 0.8849$$

Similarly, the probability that the resistance is greater than 188 ohms is given by:

$$P(z > -2.4) = 1 - P(z \leq -2.4)$$

From the standard normal distribution table, we find that  $P(z \leq -2.4) \approx 0.0082$ . Therefore,

$$P(z > -2.4) \approx 1 - 0.0082 \approx 0.9918$$

Since the event of the resistance being greater than 188 ohms is independent of the event of the resistance being greater than 194 ohms, we can multiply the probabilities together:

$$P(\text{resistance} > 194 \text{ ohms}) \times P(\text{resistance} > 188 \text{ ohms}) \approx 0.8849 \times 0.9918 \approx 0.8777$$

Therefore, the probability that the resistance is greater than 188 ohms is approximately 0.8777. The answer is not one of the given choices.

\*\*\*\*\*

Question 174:

'A major polling organization wants to predict the outcome of an upcoming national election (in terms of the proportion of voters who will vote for each candidate). They intend to use a 95% confidence interval with margin of error of no more than 2.5%. What is the minimum sample size needed to accomplish this goal?', A. '1536', B. '39', C. '1537', D. '40'

=====

Answer Option 1:

To calculate the minimum sample size needed to achieve a specific margin of error, we can use the formula:

$$n = \left( \frac{Z \cdot \sigma}{E} \right)^2$$

where:

- $n$  is the sample size,
- $Z$  is the z-score corresponding to the desired level of confidence (in this case, 95% confidence corresponds to a z-score of approximately 1.96),
- $\sigma$  is the estimated standard deviation (we can use 0.5 as a conservative estimate since 0.5 is the maximum variance),
- and  $E$  is the desired margin of error (in this case, 2.5% corresponds to 0.025).

Plugging in these values, we get:

$$n = \left( \frac{1.96 \cdot 0.5}{0.025} \right)^2 \approx 1536$$

Therefore, the minimum sample size needed is 1536, so the answer is  $\boxed{\text{A. 1536}}$ . Final answer: A

\*\*\*\*\*

Answer Option 2:

To determine the minimum sample size needed for a 95% confidence interval with a margin of error of 2.5%, we need to use the following formula:

$$n = \left( \frac{z \cdot \sigma}{E} \right)^2$$

where:

- $n$  is the required sample size,
- $z$  is the z-score corresponding to the desired confidence level (in this case, 95% corresponds to a z-score of 1.96),
- $\sigma$  is the estimated standard deviation (we don't have this information in the problem statement, so we'll use 0.5 as a conservative estimate since we don't know the true proportion),
- $E$  is the desired margin of error (2.5% = 0.025).

Plugging in the values, we get:

$$n = \left( \frac{1.96 \cdot 0.5}{0.025} \right)^2 = 1536.64$$

Since we cannot have a fraction of a person, we round up to the nearest whole number to get the minimum sample size of 1537.

Therefore, the answer is  $\boxed{\text{C. 1537}}$ .

\*\*\*\*\*

Answer Option 3:

To determine the minimum sample size needed, we can use the formula for sample size in a proportion:

$$n = \left( \frac{Z \cdot \sigma}{E} \right)^2$$

where  $n$  is the sample size,  $Z$  is the Z-score corresponding to the desired confidence level (in this case, 95% corresponds to a Z-score of approximately 1.96),  $\sigma$  is the estimated standard deviation of the population, and  $E$  is the desired margin of error.

Since we do not have any information about the population, we can use the worst-case scenario where  $\sigma = 0.5$ . This is because the maximum possible standard deviation occurs when we assume that the proportion of voters who will vote for each candidate is 0.5 (which maximizes the variance). Additionally, we can convert the margin of error to a proportion by dividing it by 100.

Substituting these values into the formula, we get:

$$n = \left( \frac{1.96 \cdot 0.5}{0.025} \right)^2$$

Simplifying, we find that the minimum sample size needed is approximately 1536. Therefore, the correct answer is  $\boxed{\text{A. 1536}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 4:

The margin of error for a 95% confidence interval is given by

$$ME = Z \cdot \sqrt{\frac{p(1-p)}{n}}$$

where  $Z$  is the z-score corresponding to the desired confidence level (in this case, 95%),  $p$  is the estimated proportion of voters who will vote for a candidate, and  $n$  is the sample size.

We want the margin of error to be no more than 2.5%, or 0.025. We can find the z-score corresponding to a 95% confidence level using a standard normal distribution table or calculator. The z-score for a 95% confidence level is approximately 1.96.

Solving the margin of error equation for  $n$ , we have

$$n = \left( \frac{Z \cdot ME}{\sqrt{p(1-p)}} \right)^2$$

Since we don't have an estimate for  $p$ , we can use the worst-case scenario where  $p = 0.5$ , which will give us the largest required sample size. Plugging in the values, we get

$$n = \left( \frac{1.96 \cdot 0.025}{\sqrt{0.5(1-0.5)}} \right)^2 = 1536$$

Therefore, the minimum sample size needed is  $\boxed{1536}$ . Final Answer: A

\*\*\*\*\*

Answer Option 5:

The margin of error is given by

$$\text{ME} = Z \cdot \sqrt{\frac{p(1-p)}{n}}$$

where  $Z$  is the z-score corresponding to a 95% confidence level,  $p$  is the proportion of voters, and  $n$  is the sample size.

We are given that the margin of error should be no more than 2.5%, so we can set up the equation:

$$2.5\% = Z \cdot \sqrt{\frac{p(1-p)}{n}}$$

Squaring both sides and solving for  $n$ , we get:

$$n = \frac{Z^2 \cdot p(1-p)}{(0.025)^2}$$

The largest value of  $p(1-p)$  occurs when  $p = \frac{1}{2}$ , so we can use this to maximize the sample size.

Substituting  $p = \frac{1}{2}$  and  $Z = 1.96$  into the equation, we get:

$$n = \frac{(1.96)^2 \cdot \frac{1}{2} \cdot \frac{1}{2}}{(0.025)^2} = 1536.64$$

Since the sample size must be an integer, the minimum sample size needed is  $\boxed{1537}$ . Final Answer: C

\*\*\*\*\*

Answer Option 6:

The margin of error is given as 2.5%, which means the actual proportion of voters in the population could differ from the predicted proportion by at most 2.5%. We can express this as an inequality:

$$2.5\% = 0.025 \leq z \cdot (\sqrt{p \cdot (1-p)} / \sqrt{n})$$

Where  $z$  is the z-score corresponding to a 95% confidence level (approximately 1.96) and  $p$  is the proportion of voters who will vote for a particular candidate.

Solving for  $n$ , we have:

$$n \geq (z^2 \cdot p \cdot (1-p)) / E^2$$

Where  $E$  is the margin of error.

Since we want the minimum sample size, we can assume  $p = 0.5$ , which gives the maximum possible value for the expression.

$$n \geq (1.96^2 \cdot 0.5 \cdot (1-0.5)) / 0.025^2$$

$$n \geq 3.8416 \cdot 0.25 / 0.000625$$

$$n \geq 15.6416 / 0.000625$$

$$n \geq 25027.44$$

Therefore, the minimum sample size needed is greater than or equal to 25028. The closest answer choice to this is 1536, so the answer is  $\boxed{\text{A. 1536}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 7:

The margin of error is given as 2.5%. This means that the organization wants to be 95% confident that the true proportion of voters falls within 2.5% of the estimated proportion.

To calculate the minimum sample size needed, we can use the formula for the margin of error of a proportion:

$$\text{Margin of Error} = z \cdot \sqrt{\frac{p(1-p)}{n}}$$

where  $z$  is the z-score corresponding to the desired confidence level,  $p$  is the estimated proportion, and  $n$  is the sample size.

Since the margin of error is given as 2.5%, we have:

$$0.025 = z \cdot \sqrt{\frac{p(1-p)}{n}}$$

To find the minimum sample size, we want to maximize the margin of error, which occurs when  $p = 0.5$  (assuming no prior knowledge about the distribution of voters). Plugging in these values, we have:

$$0.025 = z \cdot \sqrt{\frac{0.5(1-0.5)}{n}}$$

Solving for  $n$ , we get:

$$n = \left( \frac{z}{0.025} \right)^2 \cdot 0.5(1-0.5)$$

Using a z-score of 1.96 for a 95% confidence level, we find:

$$n = \left( \frac{1.96}{0.025} \right)^2 \cdot 0.5(1-0.5) \approx 1536$$

Therefore, the minimum sample size needed is  $\boxed{1536}$  (A). Final Answer: A

\*\*\*\*\*

Answer Option 8:

The margin of error for a 95% confidence interval is given by

$$\text{ME} = Z \cdot \sqrt{\frac{p(1-p)}{n}}$$

where  $Z$  is the z-score corresponding to the desired confidence level,  $p$  is the estimated proportion of the population, and  $n$  is the sample size. In this case, we want the margin of error to be no more than 2.5%, so  $\text{ME} \leq 0.025$ . Since we want to find the minimum sample size, we want to maximize the margin of error, so we can rewrite the equation as

$$0.025 \geq Z \cdot \sqrt{\frac{p(1-p)}{n}}$$

Since we do not know the value of  $p$ , we can conservatively assume  $p = 0.5$  (which maximizes the margin of error). We also want a 95% confidence interval, so  $Z = 1.96$ . Substituting these values into the equation, we have

$$0.025 \geq 1.96 \cdot \sqrt{\frac{0.5(1-0.5)}{n}}$$

Simplifying, we get

$$0.025 \geq 1.96 \cdot \sqrt{\frac{0.25}{n}}$$

Squaring both sides, we have

$$\sqrt{0.000625} \geq 3.8416 \cdot \sqrt{\frac{0.25}{n}}$$

Rearranging, we get

$$\sqrt{n} \geq \frac{3.8416 \cdot \sqrt{0.25}}{\sqrt{0.000625}} = 1536$$

Thus, the minimum sample size needed is  $\boxed{1536}$ . Final answer: A

\*\*\*\*\*

Answer Option 9:

To find the minimum sample size needed, we can use the formula for the margin of error in a confidence interval for a proportion:

$$E = Z \cdot \sqrt{\frac{p(1-p)}{n}}$$

where  $E$  is the margin of error,  $Z$  is the z-score corresponding to the desired confidence level (in this case, 95% corresponds to a z-score of approximately 1.96),  $p$  is the estimated proportion of voters who will vote for a particular candidate, and  $n$  is the sample size.

We want the margin of error to be no more than 0.025 (2.5% expressed as a decimal), so we can set up the inequality:

$$0.025 \geq 1.96 \cdot \sqrt{\frac{p(1-p)}{n}}$$

Simplifying, we have:

$$\frac{0.025^2}{1.96^2} \geq \frac{p(1-p)}{n}$$

To maximize the value of the right side of the inequality (and thus minimize the sample size), we can assume  $p = 0.5$  since this yields the largest possible value of  $p(1-p)$ . Plugging in the values, we have:

$$\frac{0.025^2}{1.96^2} \geq \frac{0.5(1-0.5)}{n}$$

Solving for  $n$ , we find:

$$\sqrt{n} \geq \frac{0.5(1-0.5)}{\frac{0.025^2}{1.96^2}}$$

Evaluating the right side of the inequality, we get:

$$\sqrt{n} \geq 1536$$

Therefore, the minimum sample size needed is 1536, so the answer is  $\boxed{\text{A } 1536}$ . Final Answer: A

\*\*\*\*\*

Question 175:

'Given that the sample has a standard deviation of zero, which of the following is a true statement?', A. 'The standard deviation of the population is also zero.', B. 'The sample mean and sample median are equal.', C. 'The sample may have outliers.', D. 'The population has a symmetric distribution.'

=====

Answer Option 1:

If the sample has a standard deviation of zero, it means that all the values in the sample are the same. This implies that the standard deviation of the population is also zero, because the population consists of the same values as the sample. Therefore, the correct statement is A. 'The standard deviation of the population is also zero.' Final answer: A

\*\*\*\*\*

Answer Option 2:

If the sample has a standard deviation of zero, it means that all the values in the sample are the same. Therefore, the population from which the sample is drawn must also consist of identical values, leading to the conclusion that the standard deviation of the population is also zero. Thus, the correct statement is  $\boxed{\text{A. 'The standard deviation of the population is also zero.'}}$  Final Answer: A

\*\*\*\*\*

Answer Option 3:

If the sample has a standard deviation of zero, it means that all the values in the sample are the same.

- A. The standard deviation of the population can also be zero if and only if the population consists of identical values, which is possible if all the values in the sample are from the entire population. However, this statement does not have to be true in general, as the population can have a standard deviation greater than zero even if the sample has a standard deviation of zero. So, statement A is not always true.
- B. If all the values in the sample are the same, then the sample mean and sample median will also be the same. So, statement B is true.
- C. Since all the values in the sample are the same, there cannot be any outliers in the sample. So, statement C is not true.
- D. The standard deviation being zero does not necessarily imply that the population has a symmetric distribution. The population can have a symmetric distribution even if the sample has a standard deviation of zero, but it can also have an asymmetric distribution. So, statement D is not always true.

Therefore, the correct statement is B. 'The sample mean and sample median are equal.' Final Answer: B

\*\*\*\*\*

Answer Option 4:

If the sample has a standard deviation of zero, it means that all the data points in the sample are the same. Therefore, the sample mean and sample median are also equal (since they are both equal to the single value in the sample). So the correct statement is  $\boxed{\text{B. 'The sample mean and sample median are equal.'}}$  Final Answer: B

\*\*\*\*\*

Answer Option 5:

If the sample has a standard deviation of zero, it means that all the values in the sample are the same. Therefore, the sample mean and sample median are equal, so the statement 'The sample mean and sample median are equal' is true. Hence, the correct answer is  $\boxed{\text{B.}}$  Final answer: B

\*\*\*\*\*

Answer Option 6:

If the sample has a standard deviation of zero, it means that every data point in the sample is the same value. This implies that there is no variation or spread in the sample. For the options:

- A. The standard deviation of the population is also zero: This is not necessarily true. The population may have other values that were not included in the sample, so the standard deviation of the population may not be zero.
- B. The sample mean and sample median are equal: This is true since all the data points in the sample are the same value, so the mean and median will also be that value.
- C. The sample may have outliers: This is not possible since all the data points are the same value. There is only one value in the sample, so there cannot be any outliers.
- D. The population has a symmetric distribution: This is not necessarily true. The population distribution can be any shape, not necessarily symmetric, as long as all the data points are the same value.

Therefore, the only true statement is: 'The sample mean and sample median are equal.' This corresponds to option B. Final answer: B

\*\*\*\*\*

Answer Option 7:

If the sample has a standard deviation of zero, it means that all the observations in the sample are the same value.

For option A, it is true that the standard deviation of the population is also zero. Since the sample consists of all the observations in the population, and all the observations are the same, the standard deviation of the population will also be zero.

For option B, it is true that the sample mean and sample median are equal, because all the observations in the sample are the same value.

For option C, it is false that the sample may have outliers. Since all the observations are the same value, there are no extreme values that can be considered outliers.

For option D, it is false that the population has a symmetric distribution. The population could consist of all the same value, which would result in a distribution that is not symmetric.

Therefore, the only true statement is option A: 'The standard deviation of the population is also zero.' Final Answer: A

\*\*\*\*\*

Answer Option 8:

If the sample has a standard deviation of zero, it means that all the values in the sample are equal to each other.

A. The standard deviation of the population can also be zero if all the values in the population are equal to each other. However, we do not have enough information to determine if this is true or not based on the given statement.



B. It is not necessarily true that the sample mean and sample median are equal. The mean is affected by the values of all the observations in the sample, while the median is only affected by the middle value(s).

C. The sample having a standard deviation of zero does not necessarily imply the presence or absence of outliers. Outliers are extreme values that are significantly different from the other values, and they can occur even if the standard deviation is zero.

D. The population does not necessarily have a symmetric distribution just because the sample has a standard deviation of zero. A symmetric distribution means that the values are evenly distributed around the mean, and this does not depend on the standard deviation of the sample.

Based on the given statement, the only true statement is  $\boxed{\text{None of the above.}}$

\*\*\*\*\*

Answer Option 9:

If the sample has a standard deviation of zero, it means that all the values in the sample are the same.

A. The standard deviation of the population cannot be determined from the given information. It is possible for the population to have a standard deviation of zero, but it is also possible for it to have a non-zero standard deviation.

B. The sample mean and sample median will be equal if all the values in the sample are the same. However, this is not always the case. If the sample has a non-zero standard deviation, it means that the values in the sample vary, and in this case, the sample mean and sample median will not necessarily be equal.

C. The sample may have outliers even if the standard deviation is zero. Outliers are extreme values that are significantly different from the other values in the sample. In this case, since all the values in the sample are the same, there are no outliers, but this is not always the case when the standard deviation is zero.

D. The population distribution cannot be determined from the given information. A zero standard deviation in the sample does not necessarily imply a symmetric distribution for the population. It is possible for the population to have a symmetric distribution, but it is not a definite conclusion.

Therefore, the only true statement is  $\boxed{\text{None of the above.}}$

\*\*\*\*\*

Question 176:

The number of customers served per day by a large department store is normally distributed, with a mean of 3,250 customers and a standard deviation of 320. Find the range of customers served on the middle 50 percent of days. A. (3,034, 3,466); B. (2,930, 3,570); C. (2,610, 3,890); D. (2,450, 4,050)

=====

Answer Option 1:

The middle 50 percent of a normally distributed data set falls within 1 standard deviation of the mean on either side.

The lower bound of the middle 50 percent is  $\$3,250 - 320 = 2,930\$$ .

The upper bound of the middle 50 percent is  $\$3,250 + 320 = 3,570\$$ .

Therefore, the range of customers served on the middle 50 percent of days is (2,930, 3,570)\$.

Hence, the answer is  $\boxed{\text{(B) (2,930, 3,570)}}\$$ .

\*\*\*\*\*

Answer Option 2:

To find the range of customers served on the middle 50 percent of days, we need to find the values that correspond to the 25th and 75th percentiles of the normal distribution.

The 25th percentile corresponds to a z-score of -0.6745, and the 75th percentile corresponds to a z-score of 0.6745. We can find the actual values by multiplying these z-scores by the standard deviation and adding/subtracting them to/from the mean.

For the 25th percentile:  $\$3250 + (-0.6745) \cdot 320 = 3034\$$

For the 75th percentile:  $\$3250 + (0.6745) \cdot 320 = 3466\$$

Therefore, the range of customers served on the middle 50 percent of days is \$(3034, 3466)\$.

The correct answer is  $\boxed{\text{(A) (3034, 3466)}}\$$ .

\*\*\*\*\*

Answer Option 3:

To find the range of customers served on the middle 50 percent of days, we need to find the interval that contains the middle 50 percent of the normal distribution.

The middle 50 percent of a normal distribution lies within approximately 1.96 standard deviations from the mean.

So, the lower end of the range is given by  $\$3250 - 1.96(320)\$,$  and the upper end of the range is given by  $\$3250 + 1.96(320)\$$ .

Using a calculator to evaluate these expressions, we get the lower end as approximately 3034 and the upper end as approximately 3466.

Therefore, the range of customers served on the middle 50 percent of days is approximately (3034, 3466).

Thus, the answer is  $\boxed{\text{(A) (3,034, 3,466)}}\$$ .

\*\*\*\*\*

Answer Option 4:

To find the range of customers served on the middle 50 percent of days, we need to find the interval that contains the middle 50 percent of the distribution.

The middle 50 percent of a normal distribution lies within approximately 1.96 standard deviations of the mean. Since the standard deviation is 320, this gives us an interval of  $\$1.96 \times 320 = 627.2\$$ .

The mean of 3250 customers serves as the center of this interval. Therefore, the lower bound of the interval is  $\$3250 - 627.2 = 2622.8\$$  and the upper bound is  $\$3250 + 627.2 = 3872.8\$$ .

Rounding these values to the nearest whole number, we get the range of customers served on the middle 50 percent of days to be (2623, 3873).

Therefore, the correct answer is  $\boxed{\text{(C) (2,610, 3,890)}}\$$ .

\*\*\*\*\*

Answer Option 5:

To find the range of customers served on the middle 50 percent of days, we need to find the range between the 25th percentile and the 75th percentile of the normal distribution.

The z-score corresponding to the 25th percentile is -0.674 (approximately) and the z-score corresponding to the 75th percentile is 0.674 (approximately).

We can use the formula for finding the raw score from a z-score:

$$x = \text{mean} + z \times \text{standard deviation}$$

For the lower bound, we have:

$$x_1 = 3,250 + (-0.674) \times 320 = 3,034.8$$

For the upper bound, we have:

$$x_2 = 3,250 + (0.674) \times 320 = 3,465.2$$

Therefore, the range of customers served on the middle 50 percent of days is approximately (3034, 3466), which corresponds to answer choice  $\boxed{\textbf{(A)}}\$$ . Final Answer: A

\*\*\*\*\*

Answer Option 6:

To find the range of customers served on the middle 50 percent of days, we need to find the values that correspond to the 25th and 75th percentiles of the normal distribution.

Since the mean is 3,250 and the standard deviation is 320, we can use the properties of the normal distribution to find these values.

The 25th percentile corresponds to a z-score of -0.6745, and the 75th percentile corresponds to a z-score of 0.6745. We can find the actual values by multiplying these z-scores by the standard deviation and adding/subtracting the result to/from the mean.

For the 25th percentile:

$$\$3,250 + (-0.6745) \cdot 320 = 3,250 - 215.84 = 3,034.16\$$$

For the 75th percentile:

$$\$3,250 + (0.6745) \cdot 320 = 3,250 + 215.84 = 3,465.84$$

So, the range of customers served on the middle 50 percent of days is approximately (3034, 3466). Therefore, the answer is  $\boxed{\text{(A) } (3,034, 3,466)}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 7:

To find the range of customers served on the middle 50 percent of days, we need to find the range between the 25th and 75th percentiles of the normal distribution.

The 25th percentile corresponds to the z-score  $z_{.25}$  such that  $P(Z < z_{.25}) = 0.25$ . From a standard normal distribution table, we find that  $z_{.25} \approx -0.675$ .

Similarly, the 75th percentile corresponds to the z-score  $z_{.75}$  such that  $P(Z < z_{.75}) = 0.75$ . From the table,  $z_{.75} \approx 0.675$ .

To find the corresponding values in the original distribution, we use the formulas:

$$z = \frac{x - \mu}{\sigma}$$

Solving for  $x$ , we have  $x = z \cdot \sigma + \mu$ .

Substituting the values, we get:

$$x_{.25} = -0.675 \cdot 320 + 3250 \approx 3033.5$$

$$x_{.75} = 0.675 \cdot 320 + 3250 \approx 3466.5$$

Therefore, the range of customers served on the middle 50 percent of days is approximately (3034, 3467). The answer is  $\boxed{\text{(A) } (3,034, 3,466)}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 8:

To find the range of customers served on the middle 50 percent of days, we need to find the z-scores corresponding to the lower and upper quartiles.

First, we find the z-score corresponding to the lower quartile. Since the lower quartile corresponds to the 25th percentile, we know that 25% of the days have fewer customers served. Using the z-table or a calculator, we find that the z-score corresponding to the lower quartile is approximately -0.674.

Next, we find the z-score corresponding to the upper quartile. Since the upper quartile corresponds to the 75th percentile, we know that 75% of the days have fewer customers served. Using the z-table or a calculator, we find that the z-score corresponding to the upper quartile is approximately 0.674.

To find the actual number of customers corresponding to these z-scores, we use the formula:

$$x = \mu + z\sigma$$

where  $x$  is the actual value,  $\mu$  is the mean,  $z$  is the z-score, and  $\sigma$  is the standard deviation.

For the lower quartile:

$$x_{\text{(lower)}} = 3250 + (-0.674)(320) \approx 3034$$

For the upper quartile:

$$x_{\text{(upper)}} = 3250 + (0.674)(320) \approx 3466$$

Therefore, the range of customers served on the middle 50 percent of days is approximately (3034, 3466). The correct answer choice is  $\boxed{\text{(A) } (3,034, 3,466)}}$ .

\*\*\*\*\*

Answer Option 9:

To find the range of customers served on the middle 50 percent of days, we need to find the range between the 25th and 75th percentiles.

First, we need to find the z-scores corresponding to the 25th and 75th percentiles. The cumulative distribution function (CDF) for a normal distribution gives us the probability of getting a value less than or equal to a given value. Using the z-score table or a calculator, we find that the z-score corresponding to the 25th percentile is approximately -0.674 and the z-score corresponding to the 75th percentile is approximately 0.674.

Next, we need to find the actual number of customers corresponding to these z-scores. We can do this by using the formula for the z-score:

$$z = \frac{x - \mu}{\sigma}$$

Rearranging the formula, we get:

$$x = z \cdot \sigma + \mu$$

Plugging in the values, we find that the number of customers corresponding to a z-score of -0.674 is approximately  $(-0.674 \cdot 320 + 3250) \approx 3034$ , and the number of customers corresponding to a z-score of 0.674 is approximately  $(0.674 \cdot 320 + 3250) \approx 3466$ .

Therefore, the range of customers served on the middle 50 percent of days is approximately (3034, 3466). The correct answer is  $\boxed{\text{(A) } (3,034, 3,466)}}$ . Final answer: A

\*\*\*\*\*

Question 177:

'Suppose that 54% of the graduates from your high school go on to 4-year colleges, 20% go on to 2-year colleges, 19% find employment, and the remaining 7% search for a job. If a randomly selected student is not going on to a 2-year college, what is the probability she will be going on to a 4-year college?', A. '0.460', B. '0.540', C. '0.630', D. '0.675'

=====

Answer Option 1:

If a randomly selected student is not going on to a 2-year college, then they must either be going on to a 4-year college, finding employment, or searching for a job. These three options make up  $54\% + 19\% + 7\% = 80\%$  of the students.

Since we are only interested in the probability of going on to a 4-year college, we can calculate the probability as  $\frac{54\%}{80\%} = \frac{54}{80} = \frac{27}{40} = \boxed{0.675}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

Let's assume that there are 100 students in total.

Out of the 100 students, 54 go on to 4-year colleges, 20 go on to 2-year colleges, 19 find employment, and 7 search for a job. This totals to  $54 + 20 + 19 + 7 = 100$  students.

If a randomly selected student is not going on to a 2-year college, then the remaining students are those going on to 4-year colleges, finding employment, and searching for a job. This totals to  $54 + 19 + 7 = 80$  students.

Therefore, the probability that the student will be going on to a 4-year college is  $54/80 = 0.675$ .

The correct answer is  $\boxed{\text{(D) } 0.675}$ .

\*\*\*\*\*

Answer Option 3:

Let  $A$  be the event that a randomly selected student goes on to a 4-year college, and let  $B$  be the event that a randomly selected student does not go on to a 2-year college. We are trying to find  $P(A|B)$ , the probability that a student goes on to a 4-year college given that they do not go on to a 2-year college.

We can use Bayes' theorem to find  $P(A|B)$ :

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

We are given that 54% of the graduates go on to 4-year colleges, so  $P(A) = 0.54$ . We are also given that 20% go on to 2-year colleges, so  $P(B) = 1 - 0.20 = 0.80$ .

Now, we need to find  $P(B|A)$ , the probability that a student does not go on to a 2-year college given that they go on to a 4-year college. Since these events are independent, we have  $P(B|A) = P(\text{goes on to a 2-year college} \mid \text{goes on to a 4-year college})$ . We are not given the conditional probability  $P(\text{goes on to a 2-year college} \mid \text{goes on to a 4-year college})$ , but we can find it by subtracting the probability of going on to a 4-year college from 1:  $P(\text{goes on to a 2-year college} \mid \text{goes on to a 4-year college}) = 1 - 0.54 = 0.46$ .

Plugging in the values, we have:

$$P(A|B) = \frac{0.46 \cdot 0.54}{0.80} = 0.3105$$

Thus, the probability that a randomly selected student, who is not going on to a 2-year college, will be going on to a 4-year college is approximately  $\boxed{0.3105}$ . Final Answer: None of the above

\*\*\*\*\*

Answer Option 4:

Let's consider 100 students as a sample size for simplicity. Out of the 100 students, 54 students go on to 4-year colleges, 20 students go on to 2-year colleges, 19 students find employment, and 7 students search for a job. This means that  $100 - 20 = 80$  students are not going on to a 2-year college. Out of these 80 students, 54 students go on to 4-year colleges. Therefore, the probability that a randomly selected student who is not going on to a 2-year college will be going on to a 4-year college is  $54/80 = 0.675$ , so the answer is  $\boxed{\text{(D) } 0.675}$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

Let's consider the complementary event. If a randomly selected student is not going on to a 2-year college, then either they are going on to a 4-year college, finding employment, or searching for a job.

The probability of going on to a 4-year college is 54% and the probability of finding employment is 19%. The probability of searching for a job is 7%.

Therefore, the probability that the student will be going on to a 4-year college is  $\frac{54\%}{54\% + 19\% + 7\%} = \frac{54\%}{80\%} = \frac{54}{80} = \frac{27}{40} = \boxed{0.675}$ .

Final Answer: D

\*\*\*\*\*

Answer Option 6:

Let's assume that there are 100 graduates from your high school. Out of these 100 graduates, 54 go on to 4-year colleges, 20 go on to 2-year colleges, 19 find employment, and 7 search for a job.

Since we know that the student is not going on to a 2-year college, we can subtract the 20 students going on to 2-year colleges from the total number of graduates, which leaves us with  $100 - 20 = 80$  graduates.

Out of these 80 graduates, 54 go on to 4-year colleges. Therefore, the probability that a randomly selected student who is not going on to a 2-year college will be going on to a 4-year college is  $54/80 = 0.675$ .

Thus, the answer is  $\boxed{\text{(D) } 0.675}$ .

\*\*\*\*\*

Answer Option 7:

If a randomly selected student is not going on to a 2-year college, then the probability that she will be going on to a 4-year college is the ratio of the number of students going on to a 4-year college to the total number of students not going on to a 2-year college.

Let's assume there are 100 students in total. Out of these, 54 students go on to 4-year colleges, 20 students go on to 2-year colleges, 19 students find employment, and 7 students are searching for a job.

This means that  $100 - 20 = 80$  students are not going on to a 2-year college.

Out of these 80 students, 54 students go on to 4-year colleges. Therefore, the probability that a randomly selected student who is not going on to a 2-year college will be going on to a 4-year college is  $\frac{54}{80} = 0.675$ .

Thus, the answer is  $\boxed{\text{D}}$ . Final answer: D

\*\*\*\*\*

Answer Option 8:

Let's create a sample space with four events: going to a 4-year college, going to a 2-year college, finding employment, and searching for a job. We're given the following probabilities:  $P(\text{4-year college}) = 0.54$ ,

$P(\text{2-year college}) = 0.20$ ,  $P(\text{employment}) = 0.19$ , and  $P(\text{searching for a job}) = 0.07$ .

We need to find the probability that a randomly selected student, given that they are not going to a 2-year college, is going to a 4-year college. To do this, we'll divide the probability of going to a 4-year college and not going to a 2-year college by the probability of not going to a 2-year college.

$P(\text{not going to a 2-year college}) = 1 - P(\text{2-year college}) = 1 - 0.20 = 0.80$

$P(\text{4-year college and not 2-year college}) = P(\text{4-year college}) - P(\text{4-year college and 2-year college})$

$= 0.54 - 0.20 = 0.34$

$P(\text{4-year college} \mid \text{not going to 2-year college}) = P(\text{4-year college and not 2-year college}) / P(\text{not going to 2-year college})$

$= 0.34 / 0.80$

$= 17/40$

Therefore, the probability that a randomly selected student, given that they are not going to a 2-year college, is going on to a 4-year college is  $17/40$ , which is approximately equal to 0.425 or 42.5%.

The answer choice closest to 0.425 is A. '0.460'.

Final answer: A

\*\*\*\*\*

Answer Option 9:

Let's assume that there are 100 graduates from the high school.

Out of these 100 graduates, 54 go on to 4-year colleges, 20 go on to 2-year colleges, 19 find employment, and 7 search for a job.

If a randomly selected student is not going on to a 2-year college, it means that the student can either go on to a 4-year college, find employment, or search for a job.

Out of the remaining 80 students, 54 go on to 4-year colleges, so the probability that a randomly selected student is not going on to a 2-year college and is going on to a 4-year college is  $\frac{54}{80} =$

$\frac{27}{40} = \boxed{0.675}$ .

Final Answer: D

\*\*\*\*\*

Question 178:

The Department of Health plans to test the lead level in a specific park. Because a high lead level is harmful to children, the park will be closed if the lead level exceeds the allowed limit. The department randomly selects several locations in the park, gets soil samples from those locations, and tests the samples for their lead levels. Which of the following decisions would result from the type I error?, A. 'Closing the park when the lead levels are within the allowed limit', B. 'Keeping the park open when the lead levels are in excess of the allowed limit', C. 'Closing the park when the lead levels are in excess of the allowed limit', D. 'Keeping the park open when the lead levels are within the allowed limit'

=====

Answer Option 1:

A type I error occurs when we reject the null hypothesis (the lead levels are within the allowed limit) when it is actually true. In this case, it means closing the park when the lead levels are within the allowed limit. Therefore, the decision that would result from a type I error is  $\boxed{\text{A. 'Closing the park when the lead levels are within the allowed limit'}}$ . Final answer: A

\*\*\*\*\*

Answer Option 2:

A type I error occurs when we falsely reject a true null hypothesis. In this scenario, the null hypothesis would be that the lead level in the park is within the allowed limit.

So, the decision that would result from a type I error is to close the park when the lead levels are within the allowed limit.

Therefore, the correct answer is  $\boxed{\text{A. 'Closing the park when the lead levels are within the allowed limit'}}$ .

\*\*\*\*\*

Answer Option 3:

A type I error occurs when a null hypothesis is rejected even though it is true. In this case, the null hypothesis would be that the lead levels in the park are within the allowed limit.

Option A, 'Closing the park when the lead levels are within the allowed limit', would result in a type I error because the park is being closed even though the null hypothesis is true.

Therefore, the decision that would result from the type I error is:  $\boxed{\text{A. 'Closing the park when the lead levels are within the allowed limit'}}$ .

\*\*\*\*\*

Answer Option 4:

A type I error occurs when we falsely reject a true null hypothesis. In this case, the null hypothesis would be that the lead levels in the park are within the allowed limit.

Therefore, if we make a type I error, it means that we would be closing the park (taking action) when the lead levels are actually within the allowed limit (true null hypothesis).

So the decision that would result from a type I error is: 'Closing the park when the lead levels are within the allowed limit'. Therefore, the answer is  $\boxed{\text{A}}$ .

\*\*\*\*\*

Answer Option 5:

A Type I error occurs when we mistakenly reject a true null hypothesis. In this case, the null hypothesis is that the lead levels in the park are within the allowed limit.

Option A, 'Closing the park when the lead levels are within the allowed limit', would result in a Type I error because we would be closing the park even though the lead levels are actually within the allowed limit.

Therefore, the decision that would result from a Type I error is  $\boxed{\text{A. 'Closing the park when the lead levels are within the allowed limit'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 6:

A type I error occurs when a null hypothesis is rejected even though it is true. In this case, the null hypothesis would be that the lead levels in the park are within the allowed limit. Option C, 'Closing the park when the lead levels are in excess of the allowed limit,' would result from a type I error because the park would be closed even though the lead levels are actually within the allowed limit. Therefore, the correct answer is  $\boxed{\text{C. 'Closing the park when the lead levels are in excess of the allowed limit'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 7:

A type I error occurs when we mistakenly reject a true null hypothesis. In this case, the null hypothesis would be that the lead levels in the park are within the allowed limit.

Option C, 'Closing the park when the lead levels are in excess of the allowed limit', would result from a type I error because the park is being closed (rejecting the null hypothesis) even though the lead levels are actually within the allowed limit.

Therefore, the answer is  $\boxed{\text{C. 'Closing the park when the lead levels are in excess of the allowed limit'}}$ .

\*\*\*\*\*

Answer Option 8:

A type I error occurs when the null hypothesis is true, but we reject it. In this case, the null hypothesis would be that the lead levels in the park are within the allowed limit. Therefore, a type I error would result from the decision to close the park when the lead levels are actually within the allowed limit. So the correct answer is  $\boxed{\text{A. Closing the park when the lead levels are within the allowed limit'}}$ . Final answer: A

\*\*\*\*\*

Answer Option 9:

A Type I error occurs when a null hypothesis is rejected, even though it is true. In this case, the null hypothesis would be that the lead levels are within the allowed limit.

If the park is closed when the lead levels are within the allowed limit, it means that the null hypothesis is rejected incorrectly, resulting in a Type I error. Therefore, the decision of 'Closing the park when the lead levels are within the allowed limit' would result from a Type I error.

So, the answer is  $\boxed{\text{A. 'Closing the park when the lead levels are within the allowed limit'}}$ .

\*\*\*\*\*



Question 179:

'For a class project, a student wants to see if boys and girls at their large high school differ in the number of contacts they have stored in their phone. The student conducts a survey of 50 randomly sampled boys and 40 randomly selected girls, and asks them to report the number of contacts. Which of the following is true about this situation?', A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.', B. 'Because the sample sizes are different, the student should not use a two-sample t-test.', C. 'Because the sample sizes are both greater than 30, the student should not use a two-sample t-test.', D. 'Because the shape of the population distribution is not known, the student should not use a two-sample t-test.'

=====

Answer Option 1:

The correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'

When comparing the means of two populations, we typically use a two-sample t-test if the population standard deviations are unknown and the conditions are met. In this case, the student collected random samples from both the boys and girls at their high school, so the conditions for inference are satisfied. Since the population standard deviations are not known, a two-sample t-test is appropriate. Final

Answer: A

\*\*\*\*\*

Answer Option 2:

The correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'

In this situation, the student wants to compare the means of two independent groups (boys and girls) to determine if there is a significant difference in the number of contacts stored in their phones. The sample sizes (50 and 40) are large enough to use a two-sample t-test, as a general guideline, a sample size of at least 30 or greater is considered adequate.

Furthermore, the population standard deviations are not known, which is a requirement to utilize a z-test. Therefore, a two-sample t-test is appropriate in this scenario. Final Answer: A

\*\*\*\*\*

Answer Option 3:

The correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'

In this situation, the student wants to compare the means of two independent samples (boys and girls) to see if they differ in the number of contacts they have stored in their phone. Since the population standard deviations are not known, a two-sample t-test is appropriate.

The conditions for using a two-sample t-test are:

1. Independence: The samples must be randomly selected and the boys' and girls' responses must be independent of each other.
2. Nearly Normal: The distribution of the number of contacts for each group should be approximately normal. Since the sample sizes are reasonably large (50 boys and 40 girls), we can assume that the sampling distributions of the means are normal by the Central Limit Theorem.
3. Equal Variances (optional): If the variances of the two groups are equal, we can use the pooled t-test. If the variances are different, we can use the separate variances t-test.

Based on the given information, there is no indication of unequal variances, so we can assume equal variances and proceed with a two-sample t-test.

Therefore, the correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'

\*\*\*\*\*

Answer Option 4:

The correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'

In this situation, the student is comparing the means of two independent samples (boys and girls) to see if there is a difference in the number of contacts stored in their phones. Since the population standard deviations are unknown, a two-sample t-test is appropriate.

The sample sizes being different (option B) does not affect the use of a two-sample t-test. As long as the samples are independent, a two-sample t-test can still be used.

The sample sizes being greater than 30 (option C) is not a requirement for using a two-sample t-test. It is usually a guideline for when the Central Limit Theorem can be applied to approximate the sampling distribution of the difference in sample means. However, even if the sample sizes are smaller than 30, as long as the samples are approximately normally distributed or the sample sizes are large enough for the t-distribution to be a good approximation, a two-sample t-test can still be used.

The shape of the population distribution being unknown (option D) is not a reason to not use a two-sample t-test. The t-test is robust to deviations from normality, especially when the sample sizes are large.

As long as the assumptions of independence and random sampling are met, a two-sample t-test can still be used.

Therefore, the correct statement is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.' Final Answer: A

\*\*\*\*\*

Answer Option 5:

The correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'

In this situation, the student wants to compare the means of two independent populations (boys and girls) based on their respective samples. Since the population standard deviations are not known, a two-sample t-test is appropriate. Conditions for using a two-sample t-test include random sampling, independence between the two samples, normality of the populations (or large sample sizes), and equal population standard deviations (which are not known in this case).

The fact that the sample sizes are different (50 boys and 40 girls) does not necessarily prevent the use of a two-sample t-test. As long as the conditions for a t-test are met, including random sampling and independence between the samples, a two-sample t-test can still be used.

The condition for the sample sizes being both greater than 30 is not a requirement for using a two-sample t-test. It is a guideline for when the normality assumption can be relaxed, but it does not prevent the use of a t-test.

The shape of the population distribution being unknown is not a reason to avoid using a two-sample t-test. The t-test is robust to deviations from normality, especially with larger sample sizes.

Therefore, the correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.' Final Answer: A



\*\*\*\*\*

Answer Option 6:

The correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'

A two-sample t-test is appropriate when comparing the means of two independent samples when the population standard deviations are not known. The conditions for a two-sample t-test include random sampling, independence of the two samples, approximately normal distributions, and similar variability in the populations.

In this case, the student has randomly sampled boys and girls, and the samples are assumed to be independent. The student does not have information about the population standard deviations, so a two-sample t-test is appropriate.

The fact that the sample sizes are different (50 boys and 40 girls) does not affect the choice of using a two-sample t-test. As long as the other conditions are met, a two-sample t-test can still be used.

The sample sizes being greater than 30 is not a condition for using a two-sample t-test. It is a guideline for when the t-distribution can be approximated by a normal distribution. In this case, even if the sample sizes were less than 30, a two-sample t-test could still be used if the other conditions were met.

The shape of the population distribution being unknown is not a condition for using a two-sample t-test. The t-test is robust against violations of normality as long as the sample sizes are large enough.

Therefore, the correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.' Final Answer: A.

\*\*\*\*\*

Answer Option 7:

The appropriate test to use in this situation depends on the conditions being met. In this case, both sample sizes are greater than 30, so the conditions for using a two-sample t-test are satisfied. Therefore, the correct option is  $\boxed{\text{A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'}}$  Final Answer: A

\*\*\*\*\*

Answer Option 8:

The correct statement about this situation is B. 'Because the sample sizes are different, the student should not use a two-sample t-test.'

In order to use a two-sample t-test, the sample sizes should be approximately equal. However, in this case, the sample sizes are different (50 boys and 40 girls). Therefore, a two-sample t-test is not appropriate for testing if there is a difference in the number of contacts between boys and girls. Final Answer: B

\*\*\*\*\*

Answer Option 9:

The correct answer is A. 'Because the population standard deviations are not known and conditions are met, the student should use a two-sample t-test.'

In this situation, the student wants to compare the means of two independent groups (boys and girls) to see if there is a significant difference in the number of contacts stored in their phones. The conditions for applying a two-sample t-test are:

1. The data for each group is independent and randomly sampled.
2. The sampling distribution of each group is approximately normal, or the sample sizes are large enough (greater than 30).
3. The population standard deviations are not known.

The student has randomly sampled boys and girls, satisfying the first condition. Although the sample sizes are different (50 boys and 40 girls), this does not violate the second condition since both sample sizes are greater than 30. Since the population standard deviations are not known, this satisfies the third condition. Therefore, the student should use a two-sample t-test to compare the means of the two groups. Final Answer: A

\*\*\*\*\*

Question 180:

The probability is 0.2 that a value selected at random from a normal distribution with mean 600 and standard deviation 15 will be above what number?, A. '0.84', B. '603.8', C. '612.6', D. '587.4'

=====

Answer Option 1:

To find the value above which the probability is 0.2, we can use the standard normal distribution. We convert the given normal distribution with mean 600 and standard deviation 15 into a standard normal distribution with mean 0 and standard deviation 1 using the formula  $Z = \frac{x - \mu}{\sigma}$ .

Let  $x$  be the value we are looking for. We want to find the value  $x$  such that the probability of a standard normal distribution being above  $x$  is 0.2. Using a standard normal distribution table, we find that the cumulative probability for a z-score of 0.84 is approximately 0.7995. We want to find the value of  $x$  such that the probability is 0.2, so we subtract 0.2 from 0.7995 to get 0.5995.

Now, we solve for  $z$  using the formula  $Z = \frac{x - \mu}{\sigma}$ . We plug in the given values  $\mu = 600$ ,  $\sigma = 15$ , and the calculated probability into the formula:

$$0.5995 = \frac{x - 600}{15}$$

$$9.4925 = x - 600$$

$$x \approx 609.4925$$

Therefore, the value above which the probability is 0.2 is approximately 609.4925. The closest answer choice is  $\boxed{\text{(C) } 612.6}$ . Final Answer: C

\*\*\*\*\*

Answer Option 2:

In a standard normal distribution, the area to the left of  $z=0.84$  is approximately 0.8. Since we have a normal distribution with mean 600 and standard deviation 15, we can convert the value of  $z=0.84$  to the corresponding value in the original distribution using the formula  $x = z \cdot \text{standard deviation} + \text{mean}$ .

Substituting in the values, we get  $x = 0.84 \cdot 15 + 600 = 612.6$ . Therefore, the probability is 0.2 that a value selected at random will be above  $x=612.6$ .

Thus, the answer is  $\boxed{\text{(C. '612.6')}}$ .

\*\*\*\*\*

Answer Option 3:

To find the value above which the probability is 0.2, we need to find the z-score that corresponds to a cumulative probability of 0.8 (since the area under the curve to the left of this value is 0.8). Using the standard normal distribution table or a calculator, we find that the z-score corresponding to a cumulative probability of 0.8 is approximately 0.8416.

The z-score formula is given by:  $z = \frac{(x - \mu)}{\sigma}$ , where  $x$  is the value we are looking for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Rearranging the formula, we have:  $x = z \cdot \sigma + \mu$ .

Plugging in the values, we get:  $x = 0.8416 \cdot 15 + 600 = 612.6$ .

Therefore, the answer is C. '612.6'.

\*\*\*\*\*

Answer Option 4:

To find the value above which 20% of the values lie, we need to find the z-score that corresponds to the 80th percentile. From a standard normal distribution, we know that the z-score corresponding to the 80th percentile is approximately 0.84.

To find the corresponding value in the original distribution, we can use the z-score formula:

$$z = \frac{(x - \text{mean})}{\text{standard deviation}}$$

$$0.84 = \frac{(x - 600)}{15}$$

Solving for  $x$ , we get:

$$x - 600 = 0.84 \cdot 15$$

$$x - 600 = 12.6$$

$$x = 612.6$$

Therefore, the value above which 20% of the values lie is approximately 612.6. The correct answer choice is C. Final answer: C.

\*\*\*\*\*

Answer Option 5:

To find the value above which 20% of the values lie, we can use the Z-score formula.

The Z-score is given by  $Z = \frac{(X - \mu)}{\sigma}$ , where  $X$  is the value we are looking for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

We want to find the Z-score such that the area to the right of it is 0.2. Using a standard normal distribution table or calculator, we find that the Z-score corresponding to an area of 0.2 to the right is approximately 0.84.

So, we can set up the equation  $0.84 = \frac{(X - 600)}{15}$  and solve for  $X$ :

$$0.84 = \frac{(X - 600)}{15}$$

$$12.6 = X - 600$$

$$X = 612.6$$

Thus, the value above which 20% of the values lie is approximately 612.6. Therefore, the correct answer is  $\boxed{\text{(C)} \ 612.6}$ .

\*\*\*\*\*

Answer Option 6:

To find the value above which 20% of the values lie, we need to find the z-score corresponding to the 0.8 percentile. This z-score is given by the inverse standard normal cumulative distribution function (also known as the inverse normal distribution or the quantile function) of 0.8.

Using a calculator or a table, we find that the z-score corresponding to a cumulative probability of 0.8 is approximately 0.842.

The z-score formula is given by:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we are interested in,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Rearranging the formula, we have:

$$x = z \cdot \sigma + \mu$$

Substituting the values, we get:

$$x = 0.842 \cdot 15 + 600 \approx 612.63$$

Therefore, the value above which 20% of the values lie is approximately 612.6. The answer is C. Final Answer: C

\*\*\*\*\*

Answer Option 7:

To find the value above which the probability is 0.2, we can use the Z-score formula. The Z-score is calculated as follows:

$$Z = \frac{X - \mu}{\sigma}$$

where  $X$  is the value we want to find,  $\mu$  is the mean of the distribution, and  $\sigma$  is the standard deviation. We can rearrange the formula to solve for  $X$ :

$$X = Z \cdot \sigma + \mu$$

Given that the probability is 0.2, the corresponding Z-score can be found using a standard normal distribution table. The Z-score that corresponds to a probability of 0.2 is approximately -0.84. Plugging this value into the formula, we get:

$$X = -0.84 \cdot 15 + 600 = 587.4$$

Therefore, the value above which the probability is 0.2 is 587.4. The correct answer is  $\boxed{\text{(D) } 587.4}$ .

\*\*\*\*\*

Answer Option 8:

We can use the standard normal distribution to solve this problem. First, we need to standardize the given value by using the z-score formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to find,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Plugging in the values, we have:

$$z = \frac{x - 600}{15}$$

Since we want to find the value above which the probability is 0.2, we can find the corresponding z-score using the inverse cumulative distribution function (CDF) of the standard normal distribution. The inverse CDF for a probability of 0.2 is approximately -0.8416.

Setting the z-score equal to -0.8416, we can solve for  $x$ :

$$-0.8416 = \frac{x - 600}{15}$$

Solving for  $x$ , we get:

$$x = (-0.8416)(15) + 600 = 587.4$$

Therefore, the answer is  $\boxed{\text{(D) } 587.4}$ . Final answer: D

\*\*\*\*\*

Answer Option 9:

To determine the value above which a random value from the normal distribution will fall with a probability of 0.2, we can use the standard normal distribution table.

First, we need to convert the given values to the standard normal distribution by calculating the z-score. The z-score is given by the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to convert,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case,  $x$  is the value above which we want to find the probability. The mean  $\mu$  is 600, and the standard deviation  $\sigma$  is 15. Thus, the z-score is:

$$z = \frac{x - 600}{15}$$

Next, we need to find the z-score corresponding to a cumulative probability of 0.2. From the standard normal distribution table, we find that the closest z-score to 0.2 is approximately -0.84.

Now, we can solve for  $x$  using the z-score formula:

$$-0.84 = \frac{x - 600}{15}$$

$$-12.6 = x - 600$$

$$x = 587.4$$

Therefore, the value above which a random value from the normal distribution will be with a probability of 0.2 is approximately 587.4.

The correct answer is  $\boxed{\text{(D) } 587.4}$ .

\*\*\*\*\*

Question 181:

'In a group of 10 scores, the largest score is increased by 40 points. What will happen to the mean?', A. 'It will remain the same.', B. 'It will increase by 4 points.', C. 'It will increase by 10 points.', D. 'It will increase by 40 points.'

=====

Answer Option 1:

The mean of a group of scores is calculated by summing all the scores and dividing by the number of scores. If the largest score is increased by 40 points, the total sum of the scores will increase by 40. Since the number of scores remains the same, the mean will increase by  $\frac{40}{10} = 4$  points. Therefore, the answer is  $\boxed{\text{B. It will increase by 4 points.}}$

\*\*\*\*\*

Answer Option 2:

The mean is calculated by summing all the scores and then dividing by the number of scores. When the largest score is increased by 40 points, the sum of all the scores increases by 40. Since the number of scores remains the same, the mean will increase by  $\frac{40}{10} = \boxed{4}$  points. Final Answer: B

\*\*\*\*\*

Answer Option 3:

The mean of a set of numbers is the sum of all the numbers divided by the count of numbers. When the largest score is increased by 40 points, the sum of the scores increases by 40 points. Since the count of scores remains the same, the mean will also increase by  $\frac{40}{10} = \boxed{\text{(C) } 10 \text{ points}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 4:

The mean is calculated by summing all the scores and dividing by the number of scores. If the largest score is increased by 40 points, the sum of all the scores will increase by 40. However, since the number of scores remains the same, the mean will also increase by  $\frac{40}{10} = 4$  points. Therefore, the answer is  $\boxed{\text{B. It will increase by 4 points.}}$

\*\*\*\*\*

Answer Option 5:

The mean of a set of numbers is calculated by summing all the numbers and dividing by the total count. When the largest score is increased by 40 points, the sum of the scores will increase by 40. Since the total count remains the same, the mean will increase by  $\frac{40}{10} = 4$  points. Therefore, the answer is  $\boxed{\text{B. It will increase by 4 points.}}$  Final Answer: B

\*\*\*\*\*

Answer Option 6:

The mean of a set of scores is calculated by summing all the scores and dividing by the total number of scores. When the largest score is increased by 40 points, the sum of all the scores increases by 40. However, the total number of scores remains the same. Therefore, the mean will increase by  $\frac{40}{10} = 4$  points. Therefore, the answer is  $\boxed{\text{B. It will increase by 4 points.}}$

\*\*\*\*\*

Answer Option 7:

The mean of a group of numbers is the sum of the numbers divided by the total number of numbers. Increasing the largest score by 40 points will increase the sum of the scores by 40 points, but the total number of scores remains the same. Therefore, the mean will increase by  $\frac{40}{10} = 4$  points. Thus, the answer is  $\boxed{\text{B. It will increase by 4 points.}}$  Final Answer: B

\*\*\*\*\*

Answer Option 8:

To find the impact on the mean, we need to consider the sum of all the scores. Let's call the original largest score  $L$  and the sum of all the scores  $SS$ . Since we have 10 scores, the mean is  $\frac{SS}{10}$ . When the largest score is increased by 40 points, the new sum of all the scores becomes  $SS+40$ . However, since all the other scores remain the same, the new sum is just  $SS+40-L$ .

The new mean is then  $\frac{SS+40-L}{10}$ .

Now, notice that  $SS+40-L$  is equal to  $SS+(40-L)$ . Since  $40-L$  is a negative value (since  $L$  is the largest score), adding it to the sum will decrease the overall sum. Therefore, the new mean will be smaller than the original mean.

Hence, the correct answer is A. 'It will remain the same.' Final answer: A

\*\*\*\*\*

Answer Option 9:

The mean of a set of numbers is equal to the sum of the numbers divided by the total number of numbers.

Let's consider the initial set of scores. Suppose the scores are  $x_1, x_2, \dots, x_{10}$ , where  $x_{10}$  is the largest score. The mean of these scores is given by

$$\text{Mean} = \frac{x_1 + x_2 + \dots + x_{10}}{10}$$

When the largest score,  $x_{10}$ , is increased by 40 points, the new set of scores becomes  $x_1, x_2, \dots, x_9, x_{10} + 40$ . The mean of these new scores is given by

$$\text{New Mean} = \frac{x_1 + x_2 + \dots + x_9 + (x_{10} + 40)}{10}$$

Notice that subtracting the initial mean from the new mean gives

$$\begin{aligned} \text{New Mean} - \text{Mean} &= \frac{x_1 + x_2 + \dots + x_9 + (x_{10} + 40)}{10} - \frac{x_1 + x_2 + \dots + x_{10}}{10} \\ &= \frac{x_1 + x_2 + \dots + x_9 + (x_{10} + 40) - x_1 - x_2 - \dots - x_{10}}{10} \\ &= \frac{40}{10} \\ &= 4 \end{aligned}$$

Therefore, the new mean is increased by  $\boxed{\text{4 points}}$ \$. Final answer: B  
\*\*\*\*\*

Question 182:

The distribution of salaries of a county school system with 4,752 employees is known to be right skewed, with the superintendent's salary an outlier on the higher side. A random sample of 20 employees was taken and their salaries recorded. A 95 percent t-confidence interval for the mean salary of the county school system employees is (\$15,360, \$32,470). The t-confidence interval is not appropriate in this situation because', A. 'the sample size is too small compared to the number of employees in the system to give a fair representation', B. 'the skewed salary distribution tells us that assumption of normality of the sampled population will not be satisfied', C. 'the population standard deviation of the salaries of employees of this county school system is not known', D. 'the teacher salaries are not negotiable, but the superintendent's salary is negotiable, so the superintendent's salary should be excluded from the population sampled'

=====

Answer Option 1:

The correct answer is C.

The t-confidence interval assumes that the population standard deviation is unknown and estimates it using the sample standard deviation. In this case, the fact that the population standard deviation of the salaries of employees of this county school system is not known does not make the t-confidence interval inappropriate.

Option A is incorrect because the sample size of 20 employees can still provide a fair representation of the population, especially if the sampling is done randomly.

Option B is incorrect because the assumption of normality is not required for the t-confidence interval. The t-confidence interval is robust to violations of normality assumption as long as the sample size is reasonably large.

Option D is unrelated to the appropriateness of the t-confidence interval calculation. The inclusion or exclusion of specific salaries, such as the superintendent's salary, does not affect the validity of the t-confidence interval calculation.

=====

Answer Option 2:

The correct answer is B. The t-confidence interval assumes that the sampled population follows a normal distribution. However, in this case, the given information states that the distribution of salaries is right-skewed, indicating that the assumption of normality is not satisfied. Therefore, the t-confidence interval is not appropriate in this situation.

=====

Answer Option 3:

The correct answer is B. The t-confidence interval assumes that the sampled population follows a normal distribution. However, since the distribution of salaries is known to be right skewed, the assumption of normality is not satisfied. Therefore, the t-confidence interval is not appropriate for this situation. The other options are not valid reasons for the t-confidence interval to be inappropriate in this situation.

=====

Answer Option 4:

The correct answer is B. The t-confidence interval assumes that the sampled population follows a normal distribution. However, in this case, it is stated that the distribution of salaries is right skewed, which means it does not satisfy the assumption of normality. Therefore, the t-confidence interval is not appropriate in this situation. Final answer: B

=====

Answer Option 5:

The correct answer is C. In a t-confidence interval, the population standard deviation is estimated using the sample standard deviation. Therefore, the population standard deviation does not need to be known. The other options are not valid reasons for the t-confidence interval being inappropriate. Option A is not a valid reason because the sample size of 20 is sufficient for a t-confidence interval. Option B is also not a valid reason because the t-confidence interval does not assume normality of the population, but rather assumes that the sampling distribution of the mean is approximately normal due to the Central Limit Theorem. Option D is not a valid reason because the inclusion of the superintendent's salary does not affect the appropriateness of the t-confidence interval. Final answer: C.

=====

Answer Option 6:

The correct answer is B. 'the skewed salary distribution tells us that assumption of normality of the sampled population will not be satisfied'.

A t-confidence interval assumes that the sample mean follows a normal distribution. However, in this case, the given information states that the salary distribution of the county school system is right skewed. Therefore, the assumption of normality is not satisfied, and a t-confidence interval is not appropriate in this situation.

Answer choice A is incorrect because the sample size of 20 employees is sufficient for a t-confidence interval, as long as the other assumptions are met.

Answer choice C is not the reason why the t-confidence interval is not appropriate. Whether or not the population standard deviation is known does not affect the use of a t-confidence interval; it only affects the width of the interval.

Answer choice D is not the reason why the t-confidence interval is not appropriate. The negotiability of salaries does not affect the use of a t-confidence interval.

=====

Answer Option 7:

The correct answer is C. 'the population standard deviation of the salaries of employees of this county school system is not known'.

The t-confidence interval assumes that the population standard deviation is unknown and that the sample size is small. However, the other options mentioned are not valid reasons for the t-confidence interval to be inappropriate in this situation.

Option A is incorrect because the sample size of 20 is sufficient to provide a reasonable estimate of the population mean, especially when the population size is much larger.

Option B is incorrect because the assumption of normality is not necessary for the t-confidence interval. The t-confidence interval is robust to violations of normality as long as the sample size is sufficiently large.

Option D is incorrect because the inclusion or exclusion of the superintendent's salary does not affect the appropriateness of the t-confidence interval. The sample is a random sample of employees, and the inclusion of the superintendent's salary does not violate any assumptions of the t-confidence interval.

Therefore, the correct answer is C. 'the population standard deviation of the salaries of employees of this county school system is not known'. Final answer: C

\*\*\*\*\*

Answer Option 8:

The correct answer is C. 'the population standard deviation of the salaries of employees of this county school system is not known'.

The reason why the t-confidence interval is not appropriate in this situation is because the population standard deviation is not known. In order to use the t-distribution for constructing a confidence interval, we typically need to know the population standard deviation. However, in this case, the information provided does not give us any information about the population standard deviation of the salaries of the employees.

Options A, B, and D are not correct because they do not explain why the t-confidence interval is not appropriate in this situation. Option A is incorrect because the sample size of 20 is actually relatively large and would typically be sufficient for using a t-confidence interval. Option B is incorrect because even though the salary distribution is skewed, we can still use the t-confidence interval as long as other assumptions, such as random sampling, are met. Option D is incorrect because the negotiability of salaries does not affect the appropriateness of using a t-confidence interval.

\*\*\*\*\*

Answer Option 9:

The t-confidence interval assumes that the sampled population follows a normal distribution. Option B states that the skewed salary distribution tells us that the assumption of normality of the sampled population will not be satisfied. Therefore, the t-confidence interval is not appropriate in this situation. The correct answer is  $\boxed{\text{B}}$ . Final answer: B

\*\*\*\*\*

Question 183:

'Assume that the masses of chicken eggs are normally distributed with a mean of 45 g and a standard deviation of 4 g. What mass of egg would be the 25th percentile of the masses of all the eggs?', A. '42.2 g', B. '42.3 g', C. '42.4 g', D. '42.5 g'

=====

Answer Option 1:

To find the 25th percentile of a normally distributed variable, we need to find the corresponding z-score and then convert it back to the original variable using the mean and standard deviation.

The z-score corresponding to the 25th percentile is approximately -0.674. We can look this up in a standard normal distribution table or use a calculator.

The formula to convert a z-score back to the original variable is:

$$x = z * \text{standard deviation} + \text{mean}$$

Plugging in the values, we have:

$$x = -0.674 * 4 + 45 = 42.504$$

Rounding to the nearest tenth, we get  $x = 42.5$  g.

Therefore, the mass of the egg that would be the 25th percentile is approximately 42.5 g. The correct answer is D.

\*\*\*\*\*

Answer Option 2:

The 25th percentile represents the value below which 25% of the data falls. To find this value, we can use the standard normal distribution table.

First, we need to standardize the value using the z-score formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we are looking for,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Rearranging the formula, we have:

$$x = z\sigma + \mu$$

The z-score associated with the 25th percentile can be found by looking up the corresponding value in the standard normal distribution table. From the table, we find that the z-score for the 25th percentile is approximately -0.674.

Substituting the values into the formula, we have:

$$x = -0.674 \cdot 4 + 45 = 42.504$$

Rounding to one decimal place, we get 42.5 g. Therefore, the mass of the egg at the 25th percentile is  $\boxed{\text{(D) 42.5 g}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

To find the 25th percentile of a normal distribution, we need to find the value that separates the lowest 25% of the data from the rest.

Using a standard normal distribution table or a calculator, we can find that the z-score associated with the 25th percentile is approximately -0.674.

The z-score formula is given by:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value we want to find,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Rearranging the formula, we have:

$$x = z\sigma + \mu$$

Plugging in the values, we get:

$$x = -0.674 \cdot 4 + 45 = 42.304$$

Rounding to one decimal place, the mass of the egg that would be the 25th percentile is approximately  $\boxed{\text{42.3 g}}$ . Final answer: B

\*\*\*\*\*

Answer Option 4:

To find the 25th percentile, we need to find the z-score corresponding to the 25th percentile and then convert it back to a mass using the mean and standard deviation of the distribution.

The z-score corresponding to the 25th percentile is approximately -0.6745 (which can be obtained from a standard normal distribution table or using statistical software).

The formula to convert a z-score to a raw score given the mean and standard deviation is: raw score = (z-score \* standard deviation) + mean.

Plugging in the values, we get: raw score = (-0.6745 \* 4) + 45 = 42.302 g.

Rounding to the nearest tenth, the mass of the egg at the 25th percentile is approximately 42.3 g.

Therefore, the answer is  $\boxed{\text{(B. 42.3 g)}}$ .

\*\*\*\*\*

Answer Option 5:

To find the mass that corresponds to the 25th percentile, we need to find the z-score that corresponds to this percentile and then use it to find the corresponding value in the normal distribution.

The z-score is given by the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In this case, we want to find the value  $x$  such that the area to the left of  $x$  under the normal curve is 0.25.



Using a standard normal distribution table, we find that the z-score corresponding to the 25th percentile is approximately -0.675.

Substituting this value into the z-score formula, we get:

$$-0.675 = (x - 45) / 4$$

Solving for x, we have:

$$x - 45 = -2.7$$

$$x = 42.3$$

Therefore, the mass that corresponds to the 25th percentile is approximately 42.3 g.

The correct answer is B. 42.3 g.

\*\*\*\*\*

Answer Option 6:

To find the 25th percentile, we need to find the z-score corresponding to the 25th percentile and then use that z-score to find the corresponding mass of the egg.

The z-score can be found using the formula:  $z = (x - \text{mean}) / \text{standard deviation}$ .

For the 25th percentile, the area to the left of the z-score is 0.25. Using a standard normal distribution table or a calculator, we find that the z-score is approximately -0.6745.

Substituting the values into the z-score formula, we have:

$$-0.6745 = (x - 45) / 4.$$

Solving for x, we get:

$$x - 45 = -0.6745 \cdot 4,$$

$$x - 45 = -2.698,$$

$$x = 42.302.$$

Rounding to one decimal place, the mass of the egg at the 25th percentile is approximately 42.3 g. Therefore, the answer is  $\boxed{\text{B. } 42.3 \text{ g}}$ .

\*\*\*\*\*

Answer Option 7:

To find the 25th percentile of a normally distributed variable, we need to find the z-score associated with the 25th percentile and then convert it back to the original scale. The z-score can be found using the standard normal distribution table or a calculator.

The z-score for the 25th percentile corresponds to an area of 0.25 to the left of it. Looking up this area in the standard normal distribution table, we find that the z-score is approximately -0.6745.

To convert the z-score back to the original scale, we use the formula:

$$z = (x - \mu) / \sigma,$$

where x is the value on the original scale,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Rearranging the formula, we have:

$$x = z \cdot \sigma + \mu.$$

Plugging in the values, we get:

$$x = -0.6745 \cdot 4 + 45 = 42.5.$$

Therefore, the mass of the egg at the 25th percentile is approximately 42.5 g, which corresponds to answer choice D. Final answer: D

\*\*\*\*\*

Answer Option 8:

To find the mass that corresponds to the 25th percentile, we need to find the z-score that corresponds to the 25th percentile and then use that z-score to find the corresponding mass.

The z-score is calculated using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where  $x$  is the observed value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

For the 25th percentile, we look for the z-score that corresponds to a cumulative probability of 0.25. Using a standard normal distribution table, we find that this z-score is approximately -0.674.

We can rearrange the z-score formula to solve for  $x$ :

$$x = \mu + z \cdot \sigma$$

Plugging in the values, we get:

$$x = 45 + (-0.674) \cdot 4$$

$$x \approx 45 - 2.696$$

$$x \approx 42.304$$

Rounding to the nearest tenth, the mass of the egg that corresponds to the 25th percentile is approximately 42.3 g. Therefore, the answer is  $\boxed{\text{B. } 42.3 \text{ g}}$ .

\*\*\*\*\*

Answer Option 9:

To find the 25th percentile, we need to find the value of the egg mass such that 25% of the data lies below it. In a normal distribution, we can use the z-score to find the corresponding value.

The z-score formula is given by:

$$z = \frac{x - \mu}{\sigma}$$

where x is the value we want to find,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

Rearranging the formula, we have:

$$x = z \cdot \sigma + \mu$$

Substituting the given values, we have:

$$x = -0.674 \cdot 4 + 45 = 42.504$$

Rounding to one decimal place, we get 42.5 g.

Therefore, the mass of the egg at the 25th percentile is  $\boxed{\text{42.5 g}}$ . Final Answer: D

\*\*\*\*\*

Error with subject\_matters/high\_school\_statistics\_test/prompts\_9\_184.jsonl: 'user'

Question 185:

'What is the probability of a Type II error when a hypothesis test is being conducted at the 10% significance level ( $\alpha = 0.10$ )?'. A. '0.05', B. '0.10', C. '0.90', D. 'There is insufficient information to answer this question.'

=====

Answer Option 1:

To determine the probability of a Type II error, we need additional information about the power of the test and the effect size. Without this information, we cannot calculate the probability of a Type II error.

Therefore, the answer is D. 'There is insufficient information to answer this question.'

\*\*\*\*\*

Answer Option 2:

The probability of a Type II error depends on the specific hypothesis test being conducted, as well as the sample size, effect size, and other factors. Without additional information about the specific test and its parameters, we cannot determine the probability of a Type II error. Therefore, the answer is D. 'There is insufficient information to answer this question.'

\*\*\*\*\*

Answer Option 3:

The probability of a Type II error cannot be determined solely based on the significance level ( $\alpha$ ) of the hypothesis test. The probability of a Type II error depends on several factors, including the sample size, the effect size, and the specific alternative hypothesis being tested. Therefore, the answer is: D. 'There is insufficient information to answer this question.'

\*\*\*\*\*

Answer Option 4:

The probability of a Type II error depends on the specific hypothesis test being conducted, as well as the true value of the parameter being tested. Without any additional information about the alternative hypothesis, the power of the test, or the specific test statistic being used, it is not possible to determine the probability of a Type II error. Therefore, the correct answer is D. 'There is insufficient information to answer this question.'

\*\*\*\*\*

Answer Option 5:

The probability of a Type II error cannot be determined solely based on the significance level ( $\alpha$ ) of the hypothesis test. The probability of a Type II error depends on factors such as the effect size, sample size, and the specific alternative hypothesis being tested. Therefore, the correct answer is D. 'There is insufficient information to answer this question.' Final Answer: D

\*\*\*\*\*

Answer Option 6:

The probability of a Type II error depends on the specific hypothesis test being conducted, as well as the population parameters and sample size. Therefore, without further information, it is not possible to determine the probability of a Type II error. The correct answer is D: 'There is insufficient information to answer this question.' Final answer: D

\*\*\*\*\*

Answer Option 7:

The probability of a Type II error is the probability of failing to reject the null hypothesis when the alternative hypothesis is true. This probability depends on factors such as the sample size, the effect size, and the specific test being conducted. Without more information about the specific hypothesis test being conducted, we cannot determine the probability of a Type II error. Therefore, the answer is  $\boxed{\text{D. There is insufficient information to answer this question.}}$

\*\*\*\*\*

Answer Option 8:

The probability of a Type II error depends on the power of the test, which in turn depends on the specific alternative hypothesis being tested. Without any information about the alternative hypothesis or the details of the hypothesis test, it is not possible to determine the probability of a Type II error. Therefore, the correct answer is D: 'There is insufficient information to answer this question.'

\*\*\*\*\*

Answer Option 9:

The probability of a Type II error is dependent on the specific situation and cannot be determined solely based on the significance level ( $\alpha$ ) of the hypothesis test. The probability of a Type II error is the probability of failing to reject the null hypothesis when it is actually false. It is influenced by factors such as the sample size, the effect size, and the variability of the data.

Since there is insufficient information given to determine these factors, the correct answer is D. 'There is insufficient information to answer this question.' Final answer: D

\*\*\*\*\*

Question 186:

'In comparing the life expectancies of two models of refrigerators, the average years before complete breakdown of 10 model A refrigerators is compared with that of 15 model B refrigerators. The 90% confidence interval estimate of the difference is (6, 12). Which of the following is the most reasonable conclusion?', A. 'The mean life expectancy of one model is twice that of the other.', B. 'The mean life expectancy of one model is 6 years, while the mean life expectancy of the other is 12 years.', C. 'The probability that the life expectancies are different is 0.90.', D. 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.'

=====

Answer Option 1:

The 90% confidence interval estimate (6, 12) means that we are 90% confident that the true difference in mean life expectancies of the two models falls between 6 and 12 years. Therefore, the most reasonable conclusion is option D: 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.' Final Answer: D.

\*\*\*\*\*

Answer Option 2:

The correct conclusion is D. 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.'

The confidence interval (6, 12) means that we are 90% confident that the true difference in life expectancies of the two models falls within the interval (6, 12). This does not provide any information about the specific values of the mean life expectancies of the two models or the probability of them being different.

\*\*\*\*\*

Answer Option 3:

The correct conclusion is D. 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.' The 90% confidence interval estimate of the difference in life expectancies of the two refrigerator models is (6, 12), which means that based on the sample data, we can be 90% confident that the true difference in life expectancies lies between 6 and 12 years. This does not imply the other options provided. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The correct conclusion is D. 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.'

The given confidence interval (6, 12) means that we are 90% confident that the true difference in life expectancies between the two models falls within this range. It does not imply that one model's life expectancy is necessarily twice that of the other, or that one model has a life expectancy of exactly 6 years and the other has a life expectancy of exactly 12 years. Additionally, we cannot conclude that the probability that the life expectancies are different is 0.90, as the confidence interval is about the range of the difference, not the probability of a difference. The most reasonable conclusion is that we have a 90% confidence that the difference in life expectancies falls between 6 and 12 years. Final answer: D

\*\*\*\*\*

Answer Option 5:

The correct answer is D. 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.'

The given information states that the 90% confidence interval estimate of the difference between the average years before complete breakdown of model A and model B refrigerators is (6, 12). This means that we are 90% confident that the true difference between the mean life expectancies of the two models falls within the interval (6, 12) years.

Answer choices A and B make definitive statements about the mean life expectancies, which cannot be accurately determined based on the given information. Answer choice C is incorrect because the confidence interval estimate does not provide information about the probability that the life expectancies are different.

Therefore, the most reasonable conclusion is that we should be 90% confident that the difference in life expectancies of the two refrigerator models is between 6 and 12 years. Final Answer: D.

\*\*\*\*\*

Answer Option 6:

The most reasonable conclusion is  $\boxed{\text{D. 'We should be 90\% confident that the difference in life expectancies is between 6 and 12 years.'}}$ .

Explanation: The 90% confidence interval (6, 12) means that we are 90% confident that the true difference in the mean life expectancies of the two models falls between 6 and 12 years. This does not provide conclusive evidence that one model's life expectancy is twice that of the other (option A) or that the mean life expectancies of one model is 6 years and the other is 12 years (option B). Additionally, it does not state that the probability that the life expectancies are different is 0.90 (option C). The correct interpretation is that we can be 90% confident that the difference in life expectancies lies between 6 and 12 years.

Final Answer: D

\*\*\*\*\*

Answer Option 7:

The correct conclusion is D. 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.'

The confidence interval (6, 12) means that we are 90% confident that the true difference in the mean life expectancies of the two models falls between 6 and 12 years. This means that, based on the given data, the most reasonable conclusion is that there is a 90% probability that the difference in life expectancies is between 6 and 12 years. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The correct conclusion is D. 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.'

The confidence interval estimate (6, 12) means that we are 90% confident that the true difference in the mean life expectancies of the two models falls between 6 and 12 years. It does not imply that one model has twice the life expectancy of the other (option A), or that the mean life expectancies are exactly 6 and 12 years (option B). It also does not mean that the probability of the life expectancies being different is 0.90 (option C). The correct interpretation is that based on the given data, we can be 90% confident that the true difference in life expectancies is within the provided interval. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The correct conclusion is D. 'We should be 90% confident that the difference in life expectancies is between 6 and 12 years.'

The confidence interval (6, 12) means that based on the sample data, we can be 90% confident that the true difference in life expectancies between the two models falls between 6 and 12 years. This does not mean that the mean life expectancy of one model is exactly 6 years and the other is exactly 12 years (option B). It also does not mean that one model has twice the mean life expectancy of the other (option A).

Option C is incorrect because the confidence interval refers to the difference in life expectancies, not whether they are different or not. The confidence interval indicates the range of plausible values for the difference, not a probability that the life expectancies are different. The correct interpretation is that we are 90% confident that the true difference in life expectancies is between 6 and 12 years. Final answer: D.

\*\*\*\*\*

Question 187:

'A medicine is known to produce side effects in one in five patients taking it. Suppose a doctor prescribes the medicine to four unrelated patients. What is the probability that none of the patients will develop side effects?', A. '0.8', B. '0.4096', C. '0.25', D. '0.2'

=====

Answer Option 1:

The probability that a patient will not develop side effects is  $1 - \frac{1}{5} = \frac{4}{5}$ . Since the patients are unrelated, the probability that all four patients will not develop side effects is

$\left(\frac{4}{5}\right)^4 = 0.4096$ . Final answer: B

\*\*\*\*\*

Answer Option 2:

The probability that a patient does not develop side effects from the medicine is 1 minus the probability that the patient does develop side effects. Since the medicine is known to produce side effects in one in five patients, the probability that a patient does develop side effects is  $\frac{1}{5}$ . Therefore, the probability that a patient does not develop side effects is  $1 - \frac{1}{5} = \frac{4}{5}$ .

Since the four patients are unrelated, the probability that none of them develop side effects is the product of the individual probabilities that each patient does not develop side effects. Therefore, the probability that none of the patients develop side effects is  $\left(\frac{4}{5}\right)^4 = \frac{256}{625}$ .

Thus, the answer is  $\boxed{\text{(B. ) } 0.4096}$ .

\*\*\*\*\*

Answer Option 3:

The probability that a patient will not develop side effects from the medicine is  $1 - \frac{1}{5} = \frac{4}{5}$ . Since the patients are unrelated, the probability that all four patients will not develop side effects is

$\left(\frac{4}{5}\right)^4 = \frac{256}{625}$ . Therefore, the answer is  $\boxed{\frac{256}{625}}$ . Final answer: None of the above.

\*\*\*\*\*

Answer Option 4:

The probability that a patient does not develop side effects from the medicine is  $\frac{4}{5}$ . Since the four patients are unrelated, the probability that none of them develop side effects is

$\left(\frac{4}{5}\right)^4 = \frac{256}{625}$ . Therefore, the answer is  $\boxed{\text{(B) } 0.4096}$ .

\*\*\*\*\*

Answer Option 5:

The probability that any one patient does not experience side effects is  $\frac{4}{5}$ . Since the patients are unrelated, the probability that all four do not experience side effects is

$\left(\frac{4}{5}\right)^4 = \frac{256}{625}$ . Thus, the answer is  $\boxed{\text{(B)}} \backslash 0.4096$ . Final Answer: B

\*\*\*\*\*

Answer Option 6:

The probability that a patient does not develop side effects from taking the medicine is  $1 - \frac{1}{5} = \frac{4}{5}$ . Since the patients are unrelated, the probability that all four patients do not develop side

effects is  $\left(\frac{4}{5}\right)^4 = \frac{256}{625}$ . Therefore, the probability that none of the patients will develop side effects is  $\boxed{\frac{256}{625}}$ . Final Answer: None

\*\*\*\*\*

Answer Option 7:

The probability that each patient does not develop side effects is  $1 - \frac{1}{5} = \frac{4}{5}$ . Since the patients are unrelated, the probability that all four patients do not develop side effects is simply the

product of the probabilities for each patient:  $\left(\frac{4}{5}\right)^4 = \frac{256}{625}$ . Thus, the answer is  $\boxed{\text{(B) } 0.4096}$ . Final answer: B

\*\*\*\*\*

Answer Option 8:

The probability that a patient does not develop side effects from the medicine is  $1 - \frac{1}{5} = \frac{4}{5}$ . Since the patients are unrelated, the probability that none of them will develop side effects is the

product of the probabilities that each individual patient does not develop side effects. Therefore, the probability is  $\left(\frac{4}{5}\right)^4 = \frac{256}{625}$ . Simplifying, we get  $\frac{256}{625} =$

$\boxed{0.4096}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

The probability that a patient does not develop side effects is  $\frac{4}{5}$ . Since the patients are unrelated, the probability that all four patients do not develop side effects is  $\left(\frac{4}{5}\right)^4 =$

$\frac{256}{625}$ . Therefore, the probability that none of the patients will develop side effects is  $\boxed{\frac{256}{625}}$ . Final Answer: None

\*\*\*\*\*

Question 188:

'Which of the following is not an advantage of stratified random sampling over simple random sampling?'. A. 'When done correctly, a stratified random sample is less biased than a simple random sample.'. B. 'When done correctly, a stratified random sampling process has less variability from sample to sample than a simple random sample.'. C. 'When done correctly, a stratified random sample can provide, with a smaller sample size, an estimate that is just as reliable as that of a simple random sample with a larger sample size.'. D. 'A stratified random sample provides information about each stratum in the population as well as an estimate for the population as a whole, and a simple random sample does not.'

=====

Answer Option 1:

The answer is D. A stratified random sample provides information about each stratum in the population as well as an estimate for the population as a whole, and a simple random sample does not. This statement is true and is indeed an advantage of stratified random sampling over simple random sampling. Therefore, it is not a disadvantage of stratified random sampling. The question asks for the option that is not an advantage, so the correct answer is D. Final answer: D

\*\*\*\*\*

Answer Option 2:

The statement that is not an advantage of stratified random sampling over simple random sampling is option D. A stratified random sample provides information about each stratum in the population as well as an estimate for the population as a whole, and a simple random sample does not. This is actually an advantage of stratified random sampling as it allows for more detailed analysis and insights into different subgroups within the population. Therefore, the correct answer is option D. Final Answer: D

\*\*\*\*\*

Answer Option 3:

The correct answer is A. 'When done correctly, a stratified random sample is less biased than a simple random sample.' This statement is incorrect because both stratified random sampling and simple random sampling, when done correctly, have the potential to produce unbiased estimates of population parameters. The advantage of stratified random sampling lies in its ability to provide more precise estimates and reduce variability (option B), increase efficiency by using a smaller sample size (option C), and provide information about each stratum in the population (option D). Therefore, the statement that stratified random sampling is less biased than simple random sampling is not a valid advantage and is the answer that is not an advantage. Final Answer: A

\*\*\*\*\*

Answer Option 4:

The correct answer is B. 'When done correctly, a stratified random sampling process has less variability from sample to sample than a simple random sample.' This statement is actually an advantage of stratified random sampling over simple random sampling. By dividing the population into strata and sampling proportionally from each stratum, stratified random sampling tends to reduce the variability in the sample compared to a simple random sample which can have more variability from sample to sample.

Options A, C, and D are advantages of stratified random sampling. A stratified random sample is less biased than a simple random sample because it ensures representation from each stratum. It can provide reliable estimates with a smaller sample size because it considers the variability within each stratum and can be more efficient. Additionally, stratified random sampling provides information about each stratum, which can be useful for making comparisons and drawing conclusions about specific subgroups within the population.

Therefore, the answer is B.

\*\*\*\*\*

Answer Option 5:

The correct answer is B. 'When done correctly, a stratified random sampling process has less variability from sample to sample than a simple random sample.'

Stratified random sampling is a sampling technique where the population is divided into distinct subgroups or strata, and a simple random sample is taken from each stratum. This method ensures that each subgroup is represented in the sample, which can improve the precision of estimates for each subgroup. This is an advantage of stratified random sampling.

Option A is true because stratified random sampling reduces biases by ensuring representation from each subgroup.

Option C is also true because stratified random sampling can provide reliable estimates even with a smaller sample size compared to simple random sampling.

Option D is true because stratified random sampling allows for separate analysis and estimation for each stratum, providing more detailed information about the population.

Option B is not an advantage of stratified random sampling. In fact, stratified random sampling tends to have more variability from sample to sample compared to simple random sampling. This is because stratification introduces additional complexity and potential sources of variability in the sampling process.

Therefore, the correct answer is B. 'When done correctly, a stratified random sampling process has less variability from sample to sample than a simple random sample.'

\*\*\*\*\*

Answer Option 6:

The correct answer is C. 'When done correctly, a stratified random sample can provide, with a smaller sample size, an estimate that is just as reliable as that of a simple random sample with a larger sample size.'

Stratified random sampling is designed to ensure that the sample represents the different subgroups or strata of the population. This helps to reduce bias by ensuring that each stratum is adequately represented in the sample. It also allows for estimates and inferences to be made for each stratum separately, which can be advantageous in certain situations.

Stratified random sampling can also help to reduce variability from sample to sample, as it ensures that each stratum is represented proportionally in the sample. This can help to increase the precision and accuracy of the estimates.

Additionally, stratified random sampling can sometimes provide estimates that are just as reliable as those from a simple random sample, but with a smaller sample size. This is because stratifying the population allows for more targeted sampling within each stratum, which can lead to more efficient use of the sample.

Therefore, all of the other options listed are advantages of stratified random sampling over simple random sampling.

\*\*\*\*\*



Answer Option 7:

The correct answer is A. 'When done correctly, a stratified random sample is less biased than a simple random sample.'

Stratified random sampling is designed to reduce bias by ensuring that each stratum of the population is properly represented in the sample. By dividing the population into homogeneous groups (strata) and randomly selecting samples from each stratum, stratified random sampling helps to reduce the potential bias that can occur in simple random sampling.

Option B is an advantage of stratified random sampling. By ensuring that each stratum is represented in the sample, stratified random sampling can reduce the variability from sample to sample compared to simple random sampling. This is because stratified sampling allows for a more accurate representation of the population, which in turn leads to less variability in the estimates obtained from different samples.

Option C is also an advantage of stratified random sampling. By targeting specific subgroups within the population, stratified random sampling allows for a smaller sample size to achieve the same level of precision as a larger simple random sample. This is because stratified sampling focuses resources on the subgroups that may have higher variability or interest, while still obtaining reliable estimates for the entire population.

Option D is another advantage of stratified random sampling. By dividing the population into strata and selecting samples from each stratum, stratified random sampling provides information about each stratum in the population as well as an estimate for the population as a whole. This can provide a more comprehensive understanding of the population compared to simple random sampling, which only provides an estimate for the entire population.

Therefore, the correct answer is A. Final Answer: A

\*\*\*\*\*

Answer Option 8:

The correct answer is B. 'When done correctly, a stratified random sampling process has less variability from sample to sample than a simple random sample.' is not an advantage of stratified random sampling over simple random sampling. In fact, stratified random sampling can have more variability from sample to sample compared to simple random sampling. This is because in stratified random sampling, different strata may have different characteristics and variability, and the variability within each stratum may not be fully represented in the overall sample. Final Answer: B

\*\*\*\*\*

Answer Option 9:

The statement that is not an advantage of stratified random sampling over simple random sampling is D: "A stratified random sample provides information about each stratum in the population as well as an estimate for the population as a whole, and a simple random sample does not."

In fact, stratified random sampling does provide information about each stratum in the population, but it also provides an estimate for the population as a whole. This is because the sample is designed to be representative of the population, and by stratifying the population into different groups, the sample can capture the variability within each group and provide a more accurate estimate for the entire population.

Therefore, the correct answer is D. Final Answer: D

\*\*\*\*\*

Question 189:

'In the casino game of roulette, there are 38 slots for a ball to drop into when it is rolled around the rim of a revolving wheel: 18 red, 18 black, and 2 green. What is the probability that the first time a ball drops into the red slot is on the 8th trial (in other words, suppose you are betting on red every time-what is the probability of losing 7 straight times before you win the first time)?', A. '0.0278', B. '0.0112', C. '0.0053', D. '0.0101'

=====

Answer Option 1:

The probability of losing on any given trial is  $\frac{20}{38} = \frac{10}{19}$  since there are 20 non-red slots out of 38 total. Thus, the probability of losing 7 straight times before winning on the 8th trial is  $\left(\frac{10}{19}\right)^7 \cdot \frac{18}{38} = \frac{720}{19^7 \cdot 2} = \frac{720}{13032161 \cdot 2} = \frac{360}{13032161} \approx \boxed{0.0278}$ . Final answer: A

\*\*\*\*\*

Answer Option 2:

The probability of losing on a single trial is  $\frac{20}{38} = \frac{10}{19}$ . The probability of losing 7 straight times is  $\left(\frac{10}{19}\right)^7$ . The probability of winning on the 8th trial is  $\frac{18}{38} = \frac{9}{19}$ . Therefore, the probability of losing 7 straight times and winning on the 8th trial is  $\left(\frac{10}{19}\right)^7 \cdot \frac{9}{19} = \frac{7290}{130321} \approx \boxed{0.0559}$ . Final answer: None of the above.

\*\*\*\*\*

Answer Option 3:

The probability of losing a bet on red in one trial is  $\frac{18}{38}$ , and the probability of losing 7 straight times is  $\left(\frac{18}{38}\right)^7$ . The probability of winning on the 8th trial is  $\frac{18}{38}$ . Therefore, the probability of losing 7 times and then winning on the 8th trial is  $\left(\frac{18}{38}\right)^7 \cdot \frac{18}{38} = \frac{18^8}{38^8}$ . Simplifying the fraction gives  $\frac{9^8}{19^8} = \frac{81}{19^8}$ . The answer is  $\boxed{0.0112}$ . Final answer: B

\*\*\*\*\*

Answer Option 4:

The probability of losing on each individual trial is  $\frac{20}{38}$  because there are 20 non-red slots out of 38 total slots. Therefore, the probability of losing 7 times in a row is  $\left(\frac{20}{38}\right)^7$ . On the 8th trial, the ball must land in the red slot, which has a probability of  $\frac{18}{38}$ . Thus, the probability of losing 7 times and then winning on the 8th trial is  $\left(\frac{20}{38}\right)^7 \cdot \frac{18}{38} = \frac{324}{548903680} = \frac{1}{1696035} \approx \boxed{0.000589}$ . Final Answer: B

\*\*\*\*\*

Answer Option 5:

The probability of the ball landing on red on any given trial is  $\frac{18}{38}$ . Thus, the probability of the ball not landing on red on any given trial (landing on black or green) is  $\frac{20}{38}$ .

The probability of the ball not landing on red on all 7 trials is  $\left(\frac{20}{38}\right)^7$ .

The probability of the ball landing on red on the 8th trial is  $\frac{18}{38}$ .

The probability of losing 7 straight times before winning the first time (the ball lands on red on the 8th trial) is  $\left(\frac{20}{38}\right)^7 \cdot \frac{18}{38} = \frac{20^7 \cdot 18}{38^8} \approx 0.0112$ .

Therefore, the correct answer is  $\boxed{\text{B. } 0.0112}$ .

\*\*\*\*\*

Answer Option 6:

The probability of losing on any given trial is  $\frac{20}{38}$  (since there are 18 black slots and 2 green slots out of a total of 38 slots). The probability of losing 7 times in a row is  $\left(\frac{20}{38}\right)^7$ .

On the 8th trial, the ball must drop into the red slot for the first time. Therefore, the probability of winning on the 8th trial is  $\frac{18}{38}$ .

The probability of losing 7 times and then winning on the 8th trial is  $\left(\frac{20}{38}\right)^7 \cdot \frac{18}{38} = \frac{810}{130321}$ .

Simplifying, we get  $\frac{810}{130321} = \frac{90}{14583} = \frac{10}{1453}$ .

Therefore, the answer is  $\boxed{\text{0.0069}}$ .

\*\*\*\*\*

Answer Option 7:

The probability of losing on any trial is  $\frac{20}{38}$  because there are 20 slots that are not red out of the total 38 slots.

Therefore, the probability of losing 7 straight times is  $\left(\frac{20}{38}\right)^7 = \frac{4000000}{23191740069}$ .

The probability of winning on the 8th trial is  $\frac{18}{38}$ .

Thus, the probability of losing 7 straight times before winning on the 8th trial is  $\frac{4000000}{23191740069} \cdot \frac{18}{38} = \frac{100000}{579793501}$ .

To simplify this fraction, we can divide both the numerator and denominator by 1000:  $\frac{100}{5797935}$ .

Therefore, the answer is  $100 + 5797935 = \boxed{5798035}$ . Final Answer: C

\*\*\*\*\*

Answer Option 8:

The probability of losing on any given trial is  $\frac{20}{38}$  (since there are 20 non-red slots out of 38 total slots). Therefore, the probability of losing 7 times in a row is  $\left(\frac{20}{38}\right)^7 = \frac{200}{814^2} = \frac{25}{101^2}$ .

The probability of winning on the 8th trial is then  $\frac{18}{38}$  (since there are 18 red slots out of 38 total slots). Therefore, the probability of losing 7 straight times before winning on the 8th trial is  $\frac{25}{101^2} \cdot \frac{18}{38} = \frac{450}{401^2 \cdot 38} = \frac{450}{38438} = \frac{225}{19219}$ .

Since 225 and 19219 do not have any common factors, the fraction is already in simplest form. Therefore, the probability is  $\frac{225}{19219} \approx \boxed{0.0117}$ . Final Answer: B

\*\*\*\*\*

Answer Option 9:

The probability of losing on each trial is  $\frac{18+2}{38} = \frac{10}{19}$ , since there are 18 non-red slots and 2 green slots out of 38 total slots.

The probability of losing 7 straight times before winning is  $\left(\frac{10}{19}\right)^7$ .

Thus, the probability of winning on the 8th trial is  $1 - \left(\frac{10}{19}\right)^7 = \frac{204533}{191943} \approx 0.0101$ .

Therefore, the answer is  $\boxed{\text{(D) } 0.0101}$ . Final Answer: D

\*\*\*\*\*

Question 190:

'A talk show host recently reported that in response to his on-air question, 82% of the more than 2500 e-mail messages received through his publicized address supported the death penalty for anyone convicted of selling drugs to children. What does this show?', A. 'The survey is meaningless because of voluntary response bias.', B. 'No meaningful conclusion is possible without knowing something more about the characteristics of his listeners.', C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.', D. 'The survey would have been more meaningful if he had used a control group.'

=====

Answer Option 1:

The correct answer is C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.'

This is because the survey in question suffers from voluntary response bias. The host received responses only from those who chose to email him, which means that the sample is not representative of the entire population of listeners. People who feel more strongly about the topic are more likely to respond, skewing the results. To obtain a more accurate representation, a random sample should have been used to ensure that all listeners had an equal chance of being selected. Final answer: C

\*\*\*\*\*

Answer Option 2:

The correct answer is C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.'

The survey suffers from voluntary response bias because the respondents are self-selected and not randomly chosen. This means that the results cannot be generalized to the entire population or even to the listeners of the show. The survey only reflects the opinions of the individuals who chose to respond, which may not accurately represent the overall population of listeners. To obtain more meaningful and representative results, a random sample should have been selected from the 2500 listeners who responded. Final Answer: C

\*\*\*\*\*

Answer Option 3:

The correct answer is C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.'

This is because the survey suffers from voluntary response bias. Voluntary response bias occurs when individuals self-select into a survey or study, leading to a non-random sample. In this case, the survey only includes individuals who chose to respond to the host's question. This means that the results may not be representative of the entire population or the listeners' views as a whole. To obtain more meaningful results, it would be better to randomly select a sample from the 2500 listeners who responded. Final answer: C.

\*\*\*\*\*

Answer Option 4:

The correct answer is C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.'

This is because the survey conducted by the talk show host suffers from voluntary response bias. The individuals who chose to respond to the survey may not be representative of the population as a whole. Therefore, the survey results may not accurately reflect the opinions of all the listeners. To obtain more meaningful results, a random sample should have been chosen from the 2500 listeners who responded.

Final Answer: C

\*\*\*\*\*

Answer Option 5:

The correct answer is A. 'The survey is meaningless because of voluntary response bias.'

The survey suffers from voluntary response bias because individuals who choose to respond to the question are self-selected. This can lead to a skewed representation of the population as it only includes those who have a strong opinion on the topic and are motivated to respond. The results from this survey cannot be generalized to the entire population or even the listeners of the talk show, as it does not represent a random sample. Therefore, the survey is not meaningful in drawing conclusions about the opinions of the listeners or the general population. Final Answer: A

\*\*\*\*\*

Answer Option 6:

The correct answer is C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.'

This is because the survey suffers from voluntary response bias. People who feel strongly about the topic are more likely to respond to the question, while those who do not have a strong opinion may choose not to respond. Therefore, the survey results may not be representative of the entire population of listeners. To obtain more accurate results, a random sample should have been selected from the respondents. Final answer: C

\*\*\*\*\*

Answer Option 7:

The correct answer is C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.'

This is because the sample of listeners who responded to the e-mail question is likely not representative of the entire population. People who feel strongly about the death penalty for selling drugs to children may be more motivated to respond to the question, leading to a voluntary response bias. By selecting a random sample from the 2500 listeners who responded, the survey would have been more representative and the results would have been more meaningful. Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct answer is C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.'

The survey conducted by the talk show host suffers from voluntary response bias. This occurs when individuals self-select into a survey or study, meaning that they have chosen whether or not to participate. In this case, individuals who feel strongly about the death penalty for selling drugs to children are more likely to respond to the survey.

Therefore, the survey results may not accurately represent the views of the entire audience or population. To obtain more meaningful results, it would be better to select a random sample from the 2500 listeners who responded. This would help to reduce bias and provide a more representative view of the overall population's opinions.

\*\*\*\*\*

Answer Option 9:

The correct answer is C. 'The survey would have been more meaningful if he had picked a random sample of the 2500 listeners who responded.'

The survey suffers from a selection bias known as voluntary response bias. This means that the respondents are self-selected and may not be representative of the entire population. In this case, only those who felt strongly about the issue and were motivated to respond would have participated. Therefore, the survey results cannot be generalized to the entire population, and a random sample would have provided a more accurate representation. Final Answer: C

\*\*\*\*\*

Question 191:

'To determine the average number of children living in single-family homes, a researcher picks a simple random sample of 50 such homes. However, even after one follow-up visit the interviewer is unable to make contact with anyone in 8 of these homes. Concerned about nonresponse bias, the researcher picks another simple random sample and instructs the interviewer to keep trying until contact is made with someone in a total of 50 homes. The average number of children is determined to be 1.73. Is this estimate probably too low or too high?', A. 'Too low, because of undercoverage bias.', B. 'Too low, because convenience samples overestimate average results.', C. 'Too high, because of undercoverage bias.', D. 'Too high, because convenience samples overestimate average results.'

=====

Answer Option 1:

The estimate is probably too high. This is because nonresponse bias can lead to an underestimation of the average number of children living in single-family homes. In this case, the researcher was unable to make contact with 8 out of the 50 homes initially selected, which could potentially result in a lower average number of children if those homes had a different distribution of children than the ones that were successfully contacted. Therefore, the estimate of 1.73 children is likely to be inflated, and the correct answer is  $\boxed{\text{C. 'Too high, because of undercoverage bias.'}}$

\*\*\*\*\*

Answer Option 2:

The estimate is probably too high. This is because of nonresponse bias. The researcher was unable to make contact with 8 out of the 50 homes in the first sample. These 8 homes may have different characteristics from the homes that were successfully contacted. If the homes with a larger number of children were less likely to respond, then the average number of children in the contacted homes would be lower than the true average of all single-family homes. Therefore, the estimate of 1.73 is likely too high. The correct answer is  $\boxed{\text{C. 'Too high, because of undercoverage bias.'}}$

\*\*\*\*\*

Answer Option 3:

The estimate of the average number of children living in single-family homes is probably too low. This is because the researcher was unable to make contact with anyone in 8 out of the 50 homes initially sampled. This nonresponse bias means that the researcher may have missed households with higher average number of children. The second simple random sample, in which the researcher kept trying until contact was made with a total of 50 homes, may not fully compensate for this bias. Therefore, the estimate is likely to be lower than the true average number of children. The correct answer is  $\boxed{\text{A. 'Too low, because of undercoverage bias.'}}$

\*\*\*\*\*

Answer Option 4:

The estimate of 1.73 is probably too high, because of nonresponse bias. The researcher was unable to make contact with anyone in 8 of the homes in the initial sample. This nonresponse could potentially bias the results if the homes with nonresponse have a different average number of children compared to the homes that did respond. Since the researcher had to keep trying until contact was made with someone in a total of 50 homes, it is possible that the homes that did not respond had a lower average number of children. Therefore, the estimate of 1.73 is likely too high. The correct answer is: C. 'Too high, because of undercoverage bias.'

\*\*\*\*\*

Answer Option 5:

The estimate is probably too low because of nonresponse bias. In the first sample, the researcher was unable to make contact with 8 out of 50 homes. This means that the researcher only obtained information from 42 out of 50 homes, which may not be representative of the entire population of single-family homes. The fact that the researcher had to make additional attempts to reach a total of 50 homes suggests that there was a higher nonresponse rate in the first sample. This nonresponse bias can lead to an underestimation of the true average number of children in single-family homes. Therefore, the correct answer is A. 'Too low, because of undercoverage bias.' Final Answer: A

\*\*\*\*\*

Answer Option 6:

The estimate is probably too low due to nonresponse bias. Nonresponse bias occurs when the individuals who do not respond to a survey have different characteristics or values compared to those who do respond. In this case, the researcher was unable to make contact with 8 out of 50 homes in the first sample. The fact that the researcher had to make additional attempts to reach a total of 50 homes suggests that those who did not respond may have different characteristics or values compared to those who did respond. Therefore, the estimate based on the sample is likely to be biased and may not accurately represent the true average number of children living in single-family homes. Hence, the answer is A. 'Too low, because of undercoverage bias.'

\*\*\*\*\*

Answer Option 7:

The estimate is probably too high because of nonresponse bias. In the first sample, the researcher was unable to make contact with 8 out of 50 homes, which means that these homes are not represented in the estimate. This can lead to undercoverage bias, where certain groups or types of homes are systematically excluded from the sample. Therefore, the estimate of 1.73 is probably too high. The correct answer is  $\boxed{\text{C. 'Too high, because of undercoverage bias.'}}$

\*\*\*\*\*

Answer Option 8:

The estimate of 1.73 children is probably too high. This is because the researcher may be experiencing nonresponse bias, where the homes that did not respond may have different characteristics from the homes that did respond. In this case, the researcher was unable to make contact with 8 out of the 50 homes in the initial sample. The fact that these homes did not respond could indicate that they have different characteristics, such as a lower average number of children. Therefore, the estimate based on the homes that did respond may be higher than the true average number of children in all single-family homes. Thus, the correct answer is  $\boxed{\text{C. Too high, because of undercoverage bias.'}}$

\*\*\*\*\*

Answer Option 9:

The estimate of 1.73 is probably too high. This is because the researcher had difficulties making contact with people in 8 out of 50 homes. These non-responses introduce a form of nonresponse bias, as there may be a systematic difference between the homes that were successfully contacted and the homes that were not. It is possible that the homes that were not contacted have a different average number of children compared to the homes that were contacted. Since the researcher was unable to make contact with 8 homes, the estimate is likely biased towards the homes that were successfully contacted, which could result in an overestimation of the average number of children. Therefore, the estimate is probably too high. The correct answer is  $\boxed{\text{Too high, because of undercoverage bias.}}$

\*\*\*\*\*

Question 192:

'If all the values of a data set are the same, all of the following must equal zero except for which one?', A. 'Mean', B. 'Standard deviation', C. 'Variance', D. 'Range'

=====

Answer Option 1:

If all the values of a data set are the same, then the mean is equal to that value, so it is not zero. The standard deviation, variance, and range are all measures of dispersion or spread, and since there is no spread when all the values are the same, they will all be equal to zero. Therefore, the answer is  $\boxed{\text{A. 'Mean'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 2:

If all the values of a data set are the same, then the mean will be equal to that value, so option A is not equal to zero. The standard deviation, variance, and range all involve the differences between the values in the data set, so if all the values are the same, these measures will also be equal to zero. Therefore, the answer is  $\boxed{\text{A. 'Mean'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 3:

If all the values of a data set are the same, the only measure that will not equal zero is the 'Range'. The range is the difference between the maximum value and the minimum value in the data set. Since all the values are the same, the maximum value and minimum value are equal, and the range will be nonzero. Therefore, the answer is  $\boxed{\text{D. 'Range'}}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

If all the values in a data set are the same, then the mean of the data set will be equal to that value. Therefore, the mean is not equal to zero. However, since the standard deviation, variance, and range all involve calculations based on the differences between the values in the data set, they will all be equal to zero. Therefore, the only option that does not equal zero is the mean, so the answer is  $\boxed{\text{A. 'Mean'}}$ .

\*\*\*\*\*

Answer Option 5:

If all the values in a data set are the same, then the mean (average) of the data set will also be the same value. Therefore, the mean will not equal zero and is the exception among the given options. The standard deviation, variance, and range will all be equal to zero since there is no variation in the data set. Therefore, the answer is  $\boxed{\text{A. 'Mean'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 6:

If all the values of a data set are the same, then the mean, variance, and standard deviation must all be zero. This is because the mean is the average of all the values, and if all the values are the same, then their average is the same value. Similarly, the variance and standard deviation measure the spread or variability of the data, and if all the values are the same, there is no spread or variability. However, the range is defined as the difference between the maximum and minimum values in the data set. If all the values are the same, then the maximum and minimum values are the same, so the range is also zero.

Therefore, the correct answer is  $\boxed{\text{D. 'Range'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

If all the values of a data set are the same, the mean, variance, and standard deviation will all be zero. The mean is the average value of the data set, so if all the values are the same, the mean will be equal to that value. The variance and standard deviation measure the dispersion or spread of the data set around the mean, but in this case, since all the values are the same, there is no spread and therefore the variance and standard deviation will be zero. However, the range is defined as the difference between the maximum and minimum values in the data set. Since all the values are the same, the maximum and minimum values will also be the same, and therefore the range will be equal to zero. Thus, the correct answer is  $\boxed{\text{D. 'Range'}}$ .

\*\*\*\*\*

Answer Option 8:

If all the values of a data set are the same, then the mean, variance, and standard deviation will all be zero. The mean is the average of the values, so if they are all the same, the mean will also be that same value. The variance and standard deviation both measure the spread or variability of the data, but if all the values are the same, there is no variability and these measures will be zero.

However, the range is the difference between the maximum and minimum values in the data set. If all the values are the same, the maximum and minimum values will also be the same, so the range will be zero as well. Therefore, the correct answer is  $\boxed{\text{D. 'Range'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

If all the values of a data set are the same, then the mean of the data set is equal to each value in the data set. Therefore, the mean is not equal to zero. However, the standard deviation, variance, and range of the data set will all equal zero.

The range is the difference between the maximum and minimum values in the data set. Since all the values are the same, the maximum and minimum values are equal, and the range is zero.

The variance is the average of the squared differences between each value and the mean. Since all the values are the same, the squared differences will all be zero, and the variance is zero.

The standard deviation is the square root of the variance. Since the variance is zero, the standard deviation is also zero.

Therefore, the only measure that does not equal zero is the mean, so the answer is  $\boxed{\text{A. 'Mean'}}$ .

\*\*\*\*\*



Question 193:

'A university sent out a survey to a random sample of 120 recent graduates to see how those graduates feel about the education they received. Thirty-two of the graduates did not respond to the survey. Which of these would be the best way to proceed?', A. 'Use the 88 who did respond, using 88 as the sample size in the analysis.', B. 'Use the 88 who did respond, using 120 as the sample size in the analysis.', C. 'Attempt to contact the 32 who did not respond and get the information from them.', D. 'Select 32 more graduates at random and use them to replace the nonresponders.'

=====

Answer Option 1:

The best way to proceed in this situation is to use the 88 respondents who did respond, using 120 as the sample size in the analysis. This is because the nonresponse in the survey introduces potential bias into the sample. If we were to only use the 88 respondents, we would introduce further bias by not accounting for the nonrespondents. By using the entire sample of 120 graduates, we ensure that all graduates have an equal chance of being included in the analysis, regardless of whether they responded to the survey or not. Therefore, the correct answer is  $\boxed{\text{B. 'Use the 88 who did respond, using 120 as the sample size in the analysis.'}}$ . Final answer: B

=====

Answer Option 2:

The best way to proceed in this situation would be to use the 88 graduates who did respond, using 120 as the sample size in the analysis. This is because the 88 respondents are the only available data for analysis. The 32 non-respondents cannot be included in the analysis because their views and opinions are unknown. Including them in the analysis or attempting to contact them may introduce bias or may not accurately reflect the views of the entire population of recent graduates. Thus, the correct answer is  $\boxed{\text{B. 'Use the 88 who did respond, using 120 as the sample size in the analysis.'}}$ . Final Answer: B

=====

Answer Option 3:

The best way to proceed in this situation is to use the 88 graduates who did respond, using 120 as the sample size in the analysis. This is because the 88 respondents represent a random sample of the 120 recent graduates who were initially surveyed. By using the 88 respondents as the sample, we are still taking into account the responses of the non-respondents, as they are considered part of the original sample. This method ensures that the sample is representative of the entire group of recent graduates. Therefore, the correct answer is  $\boxed{\text{B. 'Use the 88 who did respond, using 120 as the sample size in the analysis.'}}$

=====

Answer Option 4:

The best way to proceed in this situation is to use option B, 'Use the 88 who did respond, using 120 as the sample size in the analysis.' This is because the survey was sent out to a random sample of 120 recent graduates, and the 88 who responded represent the individuals who took the time to provide their feedback. By using the full sample size of 120 in the analysis, we can maintain the randomness and representativeness of the original sample. Option A would not accurately reflect the total sample size, while options C and D would introduce potential bias by either attempting to contact the non-responders or selecting additional graduates at random. Therefore, option B is the most appropriate way to proceed. Final answer: B.

=====

Answer Option 5:

The best way to proceed in this situation would be option C, 'Attempt to contact the 32 who did not respond and get the information from them.' This is because the non-responders may have different opinions and experiences compared to the responders, and including their responses in the analysis would provide a more representative sample of the population of recent graduates. Final Answer: C

=====

Answer Option 6:

The best way to proceed in this situation is to use the 88 graduates who did respond, using 88 as the sample size in the analysis. This is because the non-responders may have different opinions or characteristics compared to the responders, and including them in the analysis without any information could introduce bias. Therefore, it is best to work with the available data and analyze the responses from the 88 graduates who did respond. Final answer: A

=====

Answer Option 7:

The best way to proceed in this situation would be to use the 88 graduates who did respond and use 120 as the sample size in the analysis, as stated in option B. This is because the initial survey was sent out as a random sample, and the 88 respondents who did respond can be considered as representative of the entire sample of 120 recent graduates. It is important to use the original sample size in the analysis to maintain the integrity and representativeness of the data. Final answer: B

=====

Answer Option 8:

The best way to proceed in this situation would be to use the 88 graduates who responded, but still consider the original sample size of 120 in the analysis. This is because the non-responders may have different opinions or characteristics compared to the responders, which could introduce bias if they are ignored completely. By using the entire sample size of 120, we can still account for the non-responders and have a more accurate representation of the entire population of recent graduates. Therefore, the best option is  $\boxed{\text{B. 'Use the 88 who did respond, using 120 as the sample size in the analysis.'}}$

=====

Answer Option 9:

The best way to proceed in this situation is to use the 88 graduates who did respond, using 88 as the sample size in the analysis. This is because the survey was sent out to a random sample of 120 recent graduates, and the 88 graduates who responded are representative of the overall population of recent graduates. By using the 88 respondents, we can still make statistically valid inferences about the larger population of recent graduates. Options C and D would introduce bias into the analysis by either attempting to contact the nonresponders or selecting additional graduates at random. Option B would not be

appropriate because the sample size should reflect the number of respondents, not the number of individuals who were initially contacted. Therefore, the best way to proceed is to use the 88 who did respond, using 88 as the sample size in the analysis. Final Answer: A

\*\*\*\*\*

Question 194:

'Given the probabilities  $P(A) = 0.3$  and  $P(B) = 0.2$ , what is the probability of the union  $P(A \cup B)$  if A and B are mutually exclusive? If A and B are independent? If B is a subset of A?' A. '0.44, 0.5, 0.2'; B. '.44, 0.5, 0.3'; C. '.5, 0.44, 0.2'; D. '.5, 0.44, 0.3'

=====

Answer Option 1:

If A and B are mutually exclusive, meaning they cannot occur at the same time, then  $P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$ .

If A and B are independent, meaning the occurrence of one event does not affect the occurrence of the other, then  $P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B) = 0.3 + 0.2 - (0.3 \cdot 0.2) = 0.44$ .

If B is a subset of A, meaning every outcome in B is also in A, then  $P(A \cup B) = P(A) = 0.3$ .

Therefore, the correct answer is  $\boxed{\text{(C) } 0.5, 0.44, 0.2}$ .

\*\*\*\*\*

Answer Option 2:

If A and B are mutually exclusive, then  $P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$ . So the answer is C.

If A and B are independent, then  $P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.3 + 0.2 - 0.3 \cdot 0.2 = 0.44$ . So the answer is also C.

If B is a subset of A, then  $P(A \cup B) = P(A) = 0.3$ . So the answer is C.

Therefore, the answer is C.

\*\*\*\*\*

Answer Option 3:

If A and B are mutually exclusive, then  $P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$ . So the answer is A.

If A and B are independent, then  $P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.3 + 0.2 - (0.3)(0.2) = 0.44$ . So the answer is A.

If B is a subset of A, then  $P(A \cup B) = P(A) = 0.3$ . So the answer is A.

Therefore, the answer is A.

\*\*\*\*\*

Answer Option 4:

If A and B are mutually exclusive, then  $P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$ . Therefore, we can eliminate options A and B.

If A and B are independent, then  $P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.3 + 0.2 - (0.3)(0.2) = 0.44$ . Therefore, the answer is option C.

If B is a subset of A, then  $P(A \cup B) = P(A) = 0.3$ . Therefore, we can eliminate option D.

The correct answer is option C. Final answer: C

\*\*\*\*\*

Answer Option 5:

If A and B are mutually exclusive, then they have no intersection. Therefore, the probability of their union is simply the sum of their individual probabilities:  $P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$ .

If A and B are independent, then the probability of their union is given by the formula:  $P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.3 + 0.2 - (0.3)(0.2) = 0.44$ .

If B is a subset of A, then A and B have an intersection. Therefore, the probability of their union is simply equal to P(A):  $P(A \cup B) = P(A) = 0.3$ .

Thus, the correct choice is C. '0.5, 0.44, 0.2'.

\*\*\*\*\*

Answer Option 6:

If A and B are mutually exclusive, then  $P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$ . Therefore, the answer is .5.

If A and B are independent, then  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Since A and B are independent,  $P(A \cap B) = P(A)P(B) = 0.3 \cdot 0.2 = 0.06$ . Therefore,  $P(A \cup B) = 0.3 + 0.2 - 0.06 = 0.44$ . Therefore, the answer is 0.44.

If B is a subset of A, then  $A \cup B = A$ . Therefore,  $P(A \cup B) = P(A) = 0.3$ . Therefore, the answer is 0.3.

The answer is therefore A. '.44, 0.5, 0.2'.

\*\*\*\*\*

Answer Option 7:

If A and B are mutually exclusive, then  $P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$ . Therefore, the probability of the union is 0.5.

If A and B are independent, then  $P(A \cap B) = P(A) \times P(B) = 0.3 \times 0.2 = 0.06$ . Since A and B are independent,  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.2 - 0.06 = 0.44$ . Therefore, the probability of the union is 0.44.

If B is a subset of A, then  $P(A \cup B) = P(A)$ . Since B is a subset of A,  $A \cup B = A$ . Therefore, the probability of the union is equal to P(A), which is 0.3. Therefore, the probability of the union is 0.3.

Therefore, the correct answer is A. 0.5, 0.44, 0.2. Final Answer: A

\*\*\*\*\*

Answer Option 8:

If A and B are mutually exclusive, then they cannot occur at the same time. This means that  $P(A \cup B)$  is simply the sum of the probabilities of A and B, since there is no overlap.

$P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$

If A and B are independent, then the probability of their union is the product of their probabilities, minus the probability of their intersection (since the intersection is counted twice).

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.2 - (P(A) \cdot P(B)) = 0.3 + 0.2 - (0.3 \cdot 0.2) = 0.44$

If B is a subset of A, then B is contained within A and therefore the union of A and B is simply A.

$$P(A \cup B) = P(A)$$

Thus, the probabilities are:

$$\text{If A and B are mutually exclusive: } P(A \cup B) = 0.5$$

$$\text{If A and B are independent: } P(A \cup B) = 0.44$$

$$\text{If B is a subset of A: } P(A \cup B) = 0.3$$

Therefore, the correct answer is: C. '.5, 0.44, 0.2'

\*\*\*\*\*

Answer Option 9:

If A and B are mutually exclusive, then they cannot occur at the same time. Therefore,  $P(A \cup B) = P(A) + P(B) = 0.3 + 0.2 = 0.5$ .

If A and B are independent, then  $P(A \cap B) = P(A) \cdot P(B) = 0.3 \cdot 0.2 = 0.06$ . Since A and B are independent,  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.2 - 0.06 = 0.44$ .

If B is a subset of A, then B can only occur if A occurs. Therefore,  $P(A \cup B) = P(A) = 0.3$ .

The correct answer is C. '.5, 0.44, 0.2'. Final Answer: C

\*\*\*\*\*

Question 195:

'Which of the following is true about any discrete probability distribution of a random variable  $X$ ?', A. 'The expected value of  $X = np$ .', B. 'The sum of all possible values of  $X$  is equal to one.', C. 'The probabilities of all possible values of  $X$  must add up to one.', D. 'The probability distribution is bell-shaped and symmetric.'

=====

Answer Option 1:

The correct statement about any discrete probability distribution of a random variable  $X$  is C. 'The probabilities of all possible values of  $X$  must add up to one.' This is a fundamental property of probability distributions, as the total probability of any outcome must be equal to one. Therefore, the answer is C. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The correct statement about any discrete probability distribution of a random variable  $X$  is that 'The probabilities of all possible values of  $X$  must add up to one.' This is known as the probability axiom and it is a fundamental property of probability. Therefore, the correct answer is  $\boxed{\text{C. 'The probabilities of all possible values of } X \text{ must add up to one.'}}$

\*\*\*\*\*

Answer Option 3:

The correct statement about any discrete probability distribution of a random variable  $X$  is C. 'The probabilities of all possible values of  $X$  must add up to one.' This is a fundamental property of probability distributions. The sum of the probabilities of all possible values of  $X$  must always equal one, ensuring that the distribution accounts for all possible outcomes. Thus, the correct answer is C. Final answer: C

\*\*\*\*\*

Answer Option 4:

The correct statement about any discrete probability distribution of a random variable  $X$  is that 'The probabilities of all possible values of  $X$  must add up to one.' This is known as the probability mass function (PMF) property, which states that the sum of the probabilities of all possible outcomes must equal 1. Therefore, the correct answer is C. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The correct statement about any discrete probability distribution of a random variable  $X$  is C. 'The probabilities of all possible values of  $X$  must add up to one.'

A discrete probability distribution represents the probabilities of different outcomes of a discrete random variable. The probabilities assigned to all possible values of  $X$  must add up to one, as the sum of all probabilities in a probability distribution should equal one.

Option A is incorrect because the expected value of  $X$  is not necessarily equal to  $np$ , where  $n$  is the number of trials and  $p$  is the probability of success in each trial. This formula holds true only for specific distributions such as the binomial distribution.

Option B is incorrect because the sum of all possible values of  $X$  does not have to equal one. The sum of the probabilities equals one, not the sum of the values of  $X$ .

Option D is incorrect because the shape of the probability distribution depends on the specific distribution and is not necessarily bell-shaped and symmetric. It could be skewed, multimodal, or have other shapes depending on the situation.

\*\*\*\*\*

Answer Option 6:

The correct answer is C. 'The probabilities of all possible values of  $X$  must add up to one.'

A probability distribution of a random variable  $X$  assigns probabilities to each possible value of  $X$ . Since the sum of the probabilities of all possible outcomes must equal 1, this implies that the probabilities of all possible values of  $X$  must add up to one.

Option A is incorrect because the expected value of  $X$  is equal to the sum of all possible values of  $X$  multiplied by their respective probabilities, which is not necessarily equal to  $np$ .

Option B is incorrect because the sum of all possible values of  $X$  does not have to be equal to one. It could be any real number depending on the values and probabilities of the outcomes.

Option D is incorrect because the shape of the probability distribution depends on the specific probabilities assigned to each value of  $X$  and is not necessarily bell-shaped or symmetric.

Therefore, the correct answer is C. 'The probabilities of all possible values of  $X$  must add up to one.'

\*\*\*\*\*

Answer Option 7:

The correct answer is C. 'The probabilities of all possible values of  $X$  must add up to one.'

A discrete probability distribution assigns probabilities to each possible value of a random variable. The sum of all these probabilities must be equal to one. This property ensures that the probabilities cover all possible outcomes and that the total probability is accounted for.

Option A, 'The expected value of  $X = np$ ,' is specific to a binomial distribution, where  $n$  is the number of trials and  $p$  is the probability of success in each trial.

Option B, 'The sum of all possible values of  $X$  is equal to one,' is incorrect and does not hold true for a probability distribution. The sum of all possible values of  $X$  depends on the range and values of the random variable and can take any real number, not necessarily one.

Option D, 'The probability distribution is bell-shaped and symmetric,' is specific to a normal distribution, which is not always the case for all probability distributions.

Therefore, option C is the correct statement for any discrete probability distribution. Final Answer: C

\*\*\*\*\*

Answer Option 8:

The correct statement is C. 'The probabilities of all possible values of  $X$  must add up to one.' In any probability distribution, the sum of the probabilities of all possible values of the random variable must be equal to one. This ensures that the distribution represents all possible outcomes and that the total probability is accounted for. Final Answer: C

\*\*\*\*\*

Answer Option 9:

The correct answer is C. 'The probabilities of all possible values of  $X$  must add up to one.' In a discrete probability distribution, the probabilities assigned to each possible value of the random variable must sum to one. This is a fundamental property of probability distributions. Option A is incorrect because the expected value of  $X$  is equal to  $n$  multiplied by the probability of  $X$  taking each value, not  $np$ . Option B is incorrect because the sum of all possible values of  $X$  is not necessarily equal to one; it depends on the specific values and their corresponding probabilities. Option D is incorrect because a probability distribution can take various shapes and does not have to be bell-shaped or symmetric. Final Answer: C

\*\*\*\*\*

Question 196:

'Which among the following would result in the narrowest confidence interval?', A. 'Small sample size and 95% confidence', B. 'Small sample size and 99% confidence', C. 'Large sample size and 95% confidence', D. 'Large sample size and 99% confidence'

=====

Answer Option 1:

The width of a confidence interval depends on the sample size and the confidence level. Generally, as the sample size increases, the confidence interval becomes narrower, and as the confidence level increases, the confidence interval becomes wider.

Therefore, the narrowest confidence interval would result from having a large sample size and a high confidence level. Among the given options, this corresponds to option C: 'Large sample size and 95% confidence'. The other options either have a small sample size or a higher confidence level, both of which would result in wider confidence intervals.

Therefore, the answer is  $\boxed{\text{C. Large sample size and 95% confidence}}$ . Final answer: C

\*\*\*\*\*

Answer Option 2:

A narrower confidence interval corresponds to a smaller margin of error, which in turn depends on the sample size and the confidence level.

A larger sample size tends to result in a smaller margin of error because it provides more data points and reduces the uncertainty. Therefore, "Large sample size" is likely to result in a narrower confidence interval.

A higher confidence level leads to a wider confidence interval because it requires a higher level of certainty, which in turn requires a larger margin of error. Therefore, "95% confidence" is likely to result in a narrower confidence interval compared to "99% confidence."

Combining these observations, the option that would result in the narrowest confidence interval is "Large sample size and 95% confidence," which corresponds to answer choice C.

\*\*\*\*\*

Answer Option 3:

A confidence interval is a range of values that is likely to contain the true value of a population parameter. The width of the confidence interval is affected by two factors: sample size and confidence level.

A larger sample size generally results in a narrower confidence interval because it provides more information about the population. More data points decrease the variability of the estimate, resulting in a more precise interval.

A higher confidence level, such as 99%, requires a wider interval because we want to be more confident that the interval contains the true value. This wider interval allows for more potential values and accommodates for greater uncertainty.

Therefore, the narrowest confidence interval would result from a combination of a large sample size and a lower confidence level. In this case, the correct answer is  $\boxed{\text{C. 'Large sample size and 95% confidence'}}$ . Final answer: C

\*\*\*\*\*

Answer Option 4:

The width of a confidence interval is influenced by two main factors: the sample size and the confidence level.

When the sample size is small, the variability of the data points is higher, leading to a wider confidence interval. Therefore, options A and B, which both mention small sample size, would result in wider confidence intervals compared to options C and D.

When it comes to the confidence level, a higher confidence level implies a wider confidence interval. This is because a higher confidence level requires a larger range of values to be included in the interval, making it wider. Therefore, option D, which mentions a higher confidence level of 99%, would result in a wider confidence interval compared to option C.

Thus, the answer is option C: 'Large sample size and 95% confidence', which would result in the narrowest confidence interval. Final Answer: C

\*\*\*\*\*

Answer Option 5:

The width of a confidence interval is determined by two factors: the sample size and the confidence level. A larger sample size generally leads to a narrower confidence interval because it provides more information and reduces the uncertainty in the estimate. Similarly, a higher confidence level leads to a wider confidence interval because it requires a higher level of certainty.

Therefore, the option that would result in the narrowest confidence interval is 'C. Large sample size and 95% confidence'. This is because a large sample size reduces the uncertainty, resulting in a narrower interval, while a 95% confidence level is still reasonably high but narrower than a 99% confidence level.

Final Answer: C

\*\*\*\*\*

Answer Option 6:

The width of a confidence interval is influenced by two factors: the sample size and the confidence level. A larger sample size will generally result in a narrower confidence interval because it provides more data and reduces the uncertainty. Therefore, the options "Large sample size and 95% confidence" and "Large sample size and 99% confidence" are likely to result in narrower confidence intervals compared to options with a small sample size.

On the other hand, the confidence level measures the level of certainty that the interval contains the true population parameter. A higher confidence level will result in a wider interval because it requires a higher degree of certainty. Therefore, the options with a 99% confidence level are likely to result in wider intervals compared to options with a 95% confidence level.

Based on these considerations, the narrowest confidence interval would be achieved with the option "Large sample size and 95% confidence". Thus, the answer is  $\boxed{\text{A. Small sample size and 95% confidence}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 7:

The width of a confidence interval is determined by two factors: the sample size and the chosen level of confidence.

A smaller sample size generally leads to a wider confidence interval because there is more uncertainty associated with estimating the population parameter with limited data.

On the other hand, a higher level of confidence leads to a wider confidence interval because we are more confident in capturing the true population parameter, and thus, we need to allow for a larger range.

Therefore, the narrowest confidence interval would result from having a large sample size and a lower level of confidence.

Among the given options, the combination that satisfies this condition is:

C. Large sample size and 95% confidence.

Thus, the answer is C. Final answer: C

\*\*\*\*\*

Answer Option 8:

The confidence interval gives a range of values within which the true population parameter is estimated to lie. A narrower confidence interval indicates a more precise estimate.

In general, increasing the sample size tends to result in a narrower confidence interval because a larger sample provides more information about the population. Therefore, between options A and C, we would expect the confidence interval to be narrower for a large sample size.

Similarly, a higher confidence level corresponds to a wider confidence interval. A 95% confidence level implies that we are 95% confident the true parameter falls within the interval, while a 99% confidence level implies we are 99% confident. Therefore, between options A and B, we would expect the confidence interval to be narrower for a 95% confidence level.

Based on these considerations, the option that would result in the narrowest confidence interval is option C, 'Large sample size and 95% confidence'. Therefore, the answer is C.

\*\*\*\*\*

Answer Option 9:

A confidence interval is a range of values that is likely to contain the true population parameter with a certain level of confidence. The width of the confidence interval depends on two factors: the sample size and the confidence level.

A smaller sample size generally leads to a wider confidence interval, as there is less data to estimate the population parameter. So, options A and B (small sample size) would result in wider confidence intervals compared to options C and D (large sample size).

The confidence level determines the probability that the true population parameter falls within the confidence interval. A higher confidence level requires a wider interval to be more certain about capturing the true parameter. Therefore, option B (99% confidence) would result in a wider interval compared to option A (95% confidence).

Combining these factors, the option that would result in the narrowest confidence interval is option C, with a large sample size and 95% confidence. This is because a large sample size allows for more precise estimation of the population parameter, while a 95% confidence level still provides a reasonable level of certainty.

Therefore, the answer is option C.

\*\*\*\*\*



Question 197:

The correlation between height in inches and weight in pounds for a particular class is 0.65. If the heights are converted from inches to centimeters, what will the correlation be? (1 in. = 2.54 cm), A. '-0.65', B. '-0.26', C. '0.1', D. '0.65'

=====

Answer Option 1:

The correlation coefficient is unaffected by changes in scale or units of measurement. Therefore, the correlation between height in centimeters and weight in pounds will still be  $\boxed{0.65}$ . Final answer:

D

\*\*\*\*\*

Answer Option 2:

The correlation coefficient is a measure of the linear relationship between two variables. It is unaffected by changes in units or scaling of the variables. Therefore, converting the heights from inches to centimeters will not change the correlation coefficient.

Hence, the correlation between height in centimeters and weight in pounds will still be  $\boxed{0.65}$ . Final answer: D

\*\*\*\*\*

Answer Option 3:

Converting the heights from inches to centimeters will not change the correlation between height and weight. The correlation coefficient is a measure of the linear relationship between two variables and is unaffected by changes in units of measurement. Therefore, the correlation after converting the heights to centimeters will still be  $\boxed{0.65}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

The correlation coefficient is a measure of the linear relationship between two variables. Converting the measurements of height from inches to centimeters does not change the linear relationship between height and weight; it only changes the units of measurement. Therefore, the correlation coefficient will remain the same. Thus, the correlation between height in centimeters and weight in pounds will also be 0.65. The answer is  $\boxed{\text{(D) } 0.65}$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

The correlation coefficient is a measure of the linear relationship between two variables. It is not affected by changes in scale or units. Therefore, converting the heights from inches to centimeters will not change the correlation coefficient.

Thus, the correlation between height in centimeters and weight in pounds will still be 0.65. Therefore, the answer is  $\boxed{\text{(D) } 0.65}$ .

\*\*\*\*\*

Answer Option 6:

The correlation coefficient is a measure of the linear relationship between two variables. Let's call the height variable  $X$  and the weight variable  $Y$ . The correlation coefficient is given by the formula:

$$r = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

where  $\text{Cov}(X,Y)$  is the covariance between  $X$  and  $Y$ , and  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of  $X$  and  $Y$  respectively.

If we convert the heights from inches to centimeters, the units of the height variable will change, but the relationship between height and weight will remain the same. Therefore, the correlation between the converted heights and weight will also remain the same. In other words, the correlation will still be 0.65.

Therefore, the answer is  $\boxed{\text{(D) } 0.65}$ .

\*\*\*\*\*

Answer Option 7:

We know that the correlation coefficient is a measure of the linear relationship between two variables. Let's assume that  $X$  represents height in inches and  $Y$  represents weight in pounds. The given correlation coefficient between  $X$  and  $Y$  is 0.65.

To convert height from inches to centimeters, we multiply the values in inches by 2.54. Therefore, we now have a new variable  $X'$  representing height in centimeters.

If we have a linear relationship between  $X$  and  $Y$ , then this relationship will also hold true between  $X'$  and  $Y$ . However, the correlation coefficient will change due to the change in units.

The correlation coefficient is unaffected by scaling or translations, but it is affected by changes of units. Since we have only changed units (from inches to centimeters), the linear relationship between  $X'$  and  $Y$  remains the same. Therefore, the correlation between  $X'$  and  $Y$  will also be  $\boxed{0.65}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

Let  $X$  be the random variable representing height in inches and  $Y$  be the random variable representing weight in pounds. The given correlation coefficient between  $X$  and  $Y$  is 0.65.

Let  $X'$  be the random variable representing height in centimeters. We have  $X' = 2.54X$ .

The correlation coefficient between  $X'$  and  $Y$  can be found using the formula:

$$\text{cor}(X', Y) = \frac{\text{cov}(X', Y)}{\sqrt{\text{var}(X')\text{var}(Y)}}$$

The variance of  $X'$  can be found as follows:

$$\text{var}(X') = \text{var}(2.54X) = (2.54)^2 \text{var}(X) = (2.54)^2 \sigma_X^2$$

where  $\sigma_X^2$  is the variance of  $X$ .

Since the correlation coefficient is scale-invariant, the correlation coefficient between  $X$  and  $Y$  remains the same. Therefore,  $\text{cor}(X', Y) = 0.65$ .

Converting the correlation coefficient back to its original units, we find that the correlation between heights in centimeters and weights in pounds is  $\boxed{0.65}$ . Final answer: D

\*\*\*\*\*

Answer Option 9:

To convert the heights from inches to centimeters, we multiply each height by 2.54. Let  $X$  represent the heights in inches,  $Y$  represent the weights in pounds,  $X'$  represent the heights in centimeters, and  $Y'$  represent the weights in pounds. The correlation coefficient is defined as

$$r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}.$$

We can see that  $\text{Cov}(X', Y') = \text{Cov}(2.54X, Y) = 2.54 \cdot \text{Cov}(X, Y)$  and  $\text{Var}(X') = (2.54)^2 \cdot \text{Var}(X)$ . Substituting into the correlation coefficient formula, we get

$$r = \frac{\text{Cov}(X', Y')}{\sqrt{\text{Var}(X') \cdot \text{Var}(Y')}} = \frac{2.54 \cdot \text{Cov}(X, Y)}{\sqrt{(2.54)^2 \cdot \text{Var}(X) \cdot \text{Var}(Y)}} = \frac{2.54}{2.54} \cdot r = \boxed{0.65}.$$

Final answer: D

\*\*\*\*\*

Question 198:

The number of days it takes to build a new house has a variance of 386. A sample of 40 new homes shows an average building time of 83 days. With what confidence can we assert that the average building time for a new house is between 80 and 90 days?, A. '15.4%', B. '17.8%', C. '20.0%', D. '82.1%'

=====

Answer Option 1:

To determine the confidence interval for the average building time, we can use the formula:

Confidence interval = sample mean  $\pm$  (critical value) \* (standard deviation /  $\sqrt{\text{sample size}}$ )

Since we don't know the population standard deviation, we'll use the sample standard deviation as an estimate. The sample variance is given as 386, so the sample standard deviation is  $\sqrt{386}$ .

The critical value can be obtained from a standard normal distribution table or calculator. For a 95% confidence interval, the critical value is approximately 1.96.

Plugging in the values, the confidence interval is:

$$83 \pm 1.96 * (\sqrt{386} / \sqrt{40})$$

$$= 83 \pm 1.96 * (\sqrt{9.65})$$

$$= 83 \pm 1.96 * 3.11$$

$$= 83 \pm 6.09$$

$$= (76.91, 89.09)$$

Therefore, we can assert with 95% confidence that the average building time for a new house is between 76.91 and 89.09 days.

The range between 80 and 90 days falls within this confidence interval, so we can assert with 95% confidence that the average building time is between 80 and 90 days.

Therefore, the answer is  $\boxed{\text{D. '82.1\%'}}$ .

\*\*\*\*\*

Answer Option 2:

Since we are given the sample mean and sample size, we can use the Central Limit Theorem to approximate the sampling distribution of the mean. Under the assumption that the population is normally distributed, the distribution of the sample mean will also be approximately normal for large sample sizes.

The standard deviation of the sample mean, also known as the standard error, can be calculated as the population standard deviation divided by the square root of the sample size. In this case, the standard deviation is the square root of the variance, so the standard error is  $\sqrt{\frac{386}{40}}$   $\approx 3.1$ .

We can then calculate the z-score for each of the lower and upper bounds:

$$\text{Lower z-score: } \frac{80 - 83}{3.1} \approx -0.97$$

$$\text{Upper z-score: } \frac{90 - 83}{3.1} \approx 2.26$$

Using a standard normal distribution table or calculator, we can find the probabilities associated with these z-scores:

Lower probability: 0.166

Upper probability: 0.988

The confidence interval is then given by the sum of the upper probability and the complement of the lower probability:

$$\text{Confidence interval: } 0.988 - 0.166 = 0.822$$

To express this as a percentage, we multiply by 100:

$$\text{Confidence: } 0.822 \times 100 \approx \boxed{82.1\%}$$

Final Answer: D

\*\*\*\*\*

Answer Option 3:

To determine the confidence level, we need to calculate the margin of error and then determine the corresponding t-score for a sample size of 40.

The margin of error, E, is given by:

$$E = t \cdot \frac{s}{\sqrt{n}}$$

where t is the t-score, s is the standard deviation of the sample, and n is the sample size.

Since the standard deviation is the square root of the variance, we have:

$$s = \sqrt{386} = 19.64$$

and

$$n = 40$$

To find the t-score, we need to determine the degrees of freedom, which is given by:

$$df = n - 1 = 40 - 1 = 39$$

Using a t-table or calculator, we find that for a 95% confidence level and 39 degrees of freedom, the t-score is approximately 2.022.

Substituting these values into the margin of error formula, we have:

$$E = 2.022 \cdot \frac{19.64}{\sqrt{40}} \approx 6.36$$

Since the confidence interval is given as 80 to 90 days, with a margin of error of 6.36, we can assert with a confidence of approximately 95% that the true average building time falls between  $80 - 6.36 = 73.64$  days and  $90 + 6.36 = 96.36$  days.

Thus, the answer is  $\boxed{\text{D. '82.1\%'}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

To determine the confidence interval for the average building time, we can use the formula:

Confidence Interval = Average Building Time  $\pm$  Margin of Error

The margin of error is calculated using the formula:

Margin of Error = Critical Value \* (Standard Deviation /  $\sqrt{\text{Sample Size}}$ )

Since we are given the variance (386), we can find the standard deviation by taking the square root of the variance, which is  $\sqrt{386} = 19.65$ .

To calculate the critical value, we need to determine the level of confidence. Let's assume a 95% confidence level, which corresponds to a critical value of 1.96.

Substituting the values into the formula, we get:

Margin of Error =  $1.96 * (19.65 / \sqrt{40}) = 11.00$

Therefore, the confidence interval for the average building time is:

$83 \pm 11$ , or (72, 94)

To determine the confidence level, we can calculate the proportion of the interval width to the average building time:

Interval Width / Average Building Time =  $(94 - 72) / 83 = 0.265$

Multiplying by 100% to obtain the percentage, we find that the confidence level is approximately 26.5%.

Since we want to assert that the average building time is between 80 and 90 days, which is within the confidence interval, the confidence we can assert is 26.5%.

So the answer is approximately 26.5%, which is closest to A. '15.4%'. Final answer: A

\*\*\*\*\*

Answer Option 5:

To determine the confidence level, we will use the concept of a confidence interval. The formula for constructing a confidence interval for the population mean, given the sample mean, sample size, and variance, is:

$$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

Where:

-  $\bar{x}$  is the sample mean (83 days),

-  $z$  is the z-score corresponding to the desired confidence level,

-  $\sigma$  is the population standard deviation (square root of the variance, which is  $\sqrt{386}$ ),

-  $n$  is the sample size (40 homes).

To find the z-score corresponding to the desired confidence level, we need to determine the area under the normal distribution curve. Since we want to find the confidence level for the average building time to be between 80 and 90 days, we need to find the area between the z-scores corresponding to 80 and 90 days.

Using a standard normal distribution table or calculator, we find that the z-score for 80 days is approximately -1.56 and the z-score for 90 days is approximately 1.56.

Substituting these values into the confidence interval formula, we have:

$$83 \pm 1.56 \cdot \frac{\sqrt{386}}{\sqrt{40}}$$

Calculating this expression, we get the confidence interval:

$$83 \pm 4.976$$

This means that we can assert with 95% confidence that the average building time for a new house is between  $83 - 4.976 = 78.024$  and  $83 + 4.976 = 87.976$  days.

Since the interval includes the range of 80 to 90 days, we can assert with 95% confidence that the average building time for a new house is between 80 and 90 days.

Thus, the answer is  $\boxed{\text{C. '20.0\%'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 6:

To determine the confidence level, we can use the formula for the confidence interval for the mean. The formula is:

$$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

where  $\bar{x}$  is the sample mean,  $\sigma$  is the population standard deviation,  $n$  is the sample size, and  $z$  is the z-score corresponding to the desired level of confidence.

In this case, we have  $\bar{x} = 83$ ,  $\sigma^2 = 386$ ,  $n = 40$ , and we want to find the confidence level when the average building time is between 80 and 90 days.

First, we find the population standard deviation:

$$\sigma = \sqrt{386} \approx 19.65$$

Next, we find the z-score corresponding to the desired level of confidence. Since we want to find the confidence level, we need to find the z-score that leaves a tail probability of  $\frac{1 - \text{Confidence Level}}{2}$ .

Using a standard normal distribution table or calculator, we can find that the z-score corresponding to a tail probability of  $\frac{1 - \text{Confidence Level}}{2} = \frac{1 - 0.95}{2} = 0.025$  is approximately 1.96.

Now, we can calculate the confidence interval:

$$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}} = 83 \pm 1.96 \cdot \frac{19.65}{\sqrt{40}} \approx 83 \pm 6.14$$

Therefore, we can assert with approximately 95% confidence that the average building time for a new house is between  $83 - 6.14 = 76.86$  and  $83 + 6.14 = 89.14$  days.

The range of 80 to 90 days is within this confidence interval, so we can assert with at least 95% confidence that the average building time for a new house is between 80 and 90 days.

The closest answer choice is  $\boxed{\text{D. '82.1\%'}}$ .

\*\*\*\*\*

Answer Option 7:

To determine the confidence interval for the average building time, we will use the concept of a t-distribution. Since the variance of the population is not known, we need to estimate it using the sample variance.

The standard error of the mean can be calculated as the square root of the sample variance divided by the square root of the sample size. In this case, the standard error is  $\sqrt{\frac{386}{40}}$ , which is approximately 3.08.

Next, we need to determine the critical value to use for the confidence interval. Since the sample size is greater than 30, we can use the standard normal distribution. For a 95% confidence level, the critical value is approximately 1.96.

Finally, we can calculate the margin of error by multiplying the standard error by the critical value:  $3.08 \times 1.96$ , which is approximately 6.04.

Therefore, the confidence interval for the average building time is  $\$83 \pm 6.04$ , or  $(\$76.96, \$89.04)$ .

To determine the confidence level, we can calculate the proportion of the total area under the curve within this interval. Since the total area under the curve is 1, and the interval represents approximately 95% of this area, the confidence level is approximately 95%.

Thus, we can assert with approximately 95% confidence that the average building time for a new house is between 80 and 90 days. The correct answer is A. '15.4%'.

\*\*\*\*\*

Answer Option 8:

To determine the confidence level, we can use the concept of the confidence interval. Given that the sample size is large (40), we can assume that the sampling distribution of the sample mean is approximately normally distributed.

The standard deviation of the sample mean is given by the formula:  $\frac{\sigma}{\sqrt{n}}$ , where  $\sigma$  is the population standard deviation and  $n$  is the sample size. In this case, the population standard deviation is the square root of the variance, which is  $\sqrt{386} \approx 19.6$ .

To calculate the confidence interval, we need to determine the margin of error. The margin of error is given by the formula:  $z \cdot \frac{\sigma}{\sqrt{n}}$ , where  $z$  is the z-score corresponding to the desired confidence level. Since we want to find the confidence interval for an average building time between 80 and 90 days, we can use a 95% confidence level, which corresponds to a z-score of approximately 1.96.

Plugging in the values, the margin of error is approximately  $1.96 \cdot \frac{19.6}{\sqrt{40}} \approx 5.47$ .

The confidence interval is given by the mean  $\pm$  the margin of error. Therefore, the confidence interval is  $\$83 \pm 5.47$ , which is approximately  $(\$77.53, \$88.47)$ .

Since the average building time for a new house falls within this confidence interval, we can assert with approximately 95% confidence that the average building time for a new house is between 80 and 90 days.

Therefore, the answer is  $\boxed{\text{D. '82.1\%'}}$ .

\*\*\*\*\*

Answer Option 9:

To determine the confidence interval for the average building time, we can use the formula:

$$\text{Confidence Interval} = \text{Sample Mean} \pm \text{Margin of Error}$$

The margin of error depends on the sample size, the standard deviation, and the desired level of confidence. Since the sample size is 40 and the variance is given as 386, we can calculate the standard deviation as  $\sqrt{386} \approx 19.65$ .

To find the margin of error, we need to determine the critical value corresponding to the desired level of confidence. Since the sample size is large (greater than 30), we can use a Z-distribution. For a 95% confidence level, the critical value is approximately 1.96.

The margin of error is then calculated as:

$$\text{Margin of Error} = \frac{\text{Standard Deviation}}{\sqrt{\text{Sample Size}}} \times \text{Critical Value}$$

Substituting the values, we get:

$$\text{Margin of Error} = \frac{19.65}{\sqrt{40}} \times 1.96 \approx 6.9$$

Therefore, the confidence interval is  $\$83 \pm 6.9$ , which gives us  $\$76.1$  to  $\$89.9$ .

Since the range  $\$80$  to  $\$90$  falls within this confidence interval, we can assert with 95% confidence that the average building time for a new house is between  $\$80$  and  $\$90$  days. This corresponds to option  $\boxed{\text{D. '82.1\%'}}$ .

\*\*\*\*\*

Question 199:

'A significance test of the hypothesis  $H_0: p = 0.3$  against the alternative  $H_A: p > 0.3$  found a value of  $\hat{p} = 0.35$  for a random sample of size 95. What is the P-value of this test?', A. '1.06', B. '0.1446', C. '0.2275', D. '0.8554'

=====

Answer Option 1:

To calculate the p-value, we need to find the probability of obtaining a test statistic as extreme as the one observed, assuming the null hypothesis is true.

In this case, we are conducting a one-tailed test with the alternative hypothesis  $H_A: p > 0.3$ . We have a sample proportion of 0.35 and a sample size of 95.

To calculate the test statistic, we can use the formula:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

where

-  $\hat{p}$  is the sample proportion (0.35)

-  $p$  is the hypothesized population proportion under the null hypothesis (0.3)

-  $n$  is the sample size (95)

Substituting the values, we get:

$$Z = \frac{0.35 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{95}}}$$

$$Z \approx 1.215$$

To find the p-value, we need to calculate the probability of obtaining a z-value greater than 1.215. Using a standard normal distribution table or a calculator, we find that the p-value is approximately 0.1446.

Therefore, the correct answer is:  $\boxed{\text{B. '0.1446'}}$ .

=====

Answer Option 2:

To find the P-value, we need to determine the probability of obtaining a test statistic as extreme as or more extreme than the observed value, assuming that the null hypothesis is true. In this case, we are testing whether the true proportion,  $p$ , is greater than 0.3.

Since the alternative hypothesis is  $p > 0.3$ , the P-value corresponds to the probability of observing a sample proportion greater than or equal to 0.35, given that  $p = 0.3$ .

We can use a normal approximation to the binomial distribution when the sample size is large ( $n = 95$  in this case). The test statistic for a one-proportion test is given by:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

where  $\hat{p}$  is the observed sample proportion,  $p$  is the hypothesized proportion under the null hypothesis, and  $n$  is the sample size.

In our case, the observed sample proportion is 0.35, the hypothesized proportion under the null hypothesis is 0.3, and the sample size is 95. Plugging these values into the formula, we get:

$$Z = \frac{0.35 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{95}}}$$

Calculating this expression, we find that  $Z = 1.06$ .

To find the P-value, we need to find the probability that  $Z$  is greater than or equal to 1.06. Using a standard normal distribution table or a calculator, we find that the area to the right of  $Z = 1.06$  is approximately 0.1446.

Therefore, the P-value of this test is approximately 0.1446. Thus, the correct answer is B. Final Answer: B

=====

Answer Option 3:

To find the P-value, we need to calculate the probability of observing a test statistic as extreme as the one obtained, assuming that the null hypothesis is true.

We can use the normal approximation to the binomial distribution to calculate the P-value. Under the null hypothesis, the test statistic follows a normal distribution with mean  $\mu = 0.3$  and standard deviation  $\sigma = \sqrt{(0.3)(0.7)/95} = 0.0477$ .

The observed test statistic is  $z = (0.35 - 0.3) / 0.0477 = 1.05$ .

To find the P-value, we calculate the probability that a standard normal random variable is greater than 1.05.

Using a standard normal table or a calculator, we find that  $P(Z > 1.05) = 0.1446$ .

Therefore, the P-value of this test is approximately 0.1446.

The correct answer is B. Final answer: B.

=====

Answer Option 4:

To find the p-value, we need to calculate the probability of observing a test statistic as extreme as 0.35 or more extreme, assuming the null hypothesis is true.

Since we are conducting a one-sided test and the alternative hypothesis states that  $p > 0.3$ , we are interested in the upper tail of the distribution.

We can use the normal approximation to the binomial distribution to find the p-value. Under the null hypothesis, the test statistic follows a normal distribution with mean  $np$  and standard deviation  $\sqrt{np(1-p)}$ ,

where  $n$  is the sample size and  $p$  is the null hypothesis proportion.

In this case, the test statistic is  $(0.35 - 0.3) / \sqrt{(0.3)(0.7)/95} = 1.166$ .

Using a standard normal distribution table or calculator, the probability of observing a test statistic as extreme as 1.166 or more extreme is approximately 0.1446.

Therefore, the p-value for this test is approximately 0.1446. The correct answer is B. Final Answer: B.

=====

Answer Option 5:

To find the P-value, we need to calculate the probability of observing a test statistic as extreme as the one obtained, assuming the null hypothesis is true.

Given that the alternative hypothesis is  $p > 0.3$ , we have a one-tailed test. The test statistic we are interested in is the sample proportion  $\hat{p} = \frac{x}{n}$ , where  $x$  is the number of successes (in this case, the number of individuals in the sample that have the characteristic of interest) and  $n$  is the sample size.

In this case,  $\hat{p} = \frac{0.35 \times 95}{95} = 0.35$ .

To calculate the P-value, we need to find the probability of observing a test statistic as extreme as 0.35, assuming the null hypothesis is true. This can be done using the binomial distribution.

Using a calculator or statistical software, we find that the P-value is approximately 0.2275.

Therefore, the correct answer is  $\boxed{\text{C. '0.2275'}}$ . Final Answer: C

\*\*\*\*\*

Answer Option 6:

To find the p-value of this test, we need to determine the probability of observing a test statistic as extreme as the one we obtained, assuming the null hypothesis is true. In this case, our test statistic is the sample proportion,  $\hat{p} = 0.35$ , and we are testing against the alternative hypothesis  $H_A: p > 0.3$ .

To calculate the p-value, we can use the normal approximation to the binomial distribution, since the sample size is large ( $n = 95$ ). Under the null hypothesis, the sampling distribution of  $\hat{p}$  is approximately normal with mean  $p_0 = 0.3$  and standard deviation  $\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.3(1-0.3)}{95}} \approx 0.0494$ .

The z-score corresponding to the observed sample proportion is  $z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.35 - 0.3}{0.0494} \approx 1.01$ .

The p-value can be calculated as the probability of observing a z-score greater than or equal to 1.01, which can be found using a standard normal distribution table or calculator. The p-value is approximately 0.1562.

Therefore, the correct answer is not listed in the options provided. The p-value of this test is approximately 0.1562.

\*\*\*\*\*

Answer Option 7:

To find the p-value, we need to calculate the probability of observing a test statistic as extreme as the one obtained under the null hypothesis. In this case, since the alternative hypothesis is one-sided ( $p > 0.3$ ), we are interested in the probability of observing a value of 0.35 or greater.

To calculate this probability, we can use the binomial distribution. The test statistic is the proportion of successes ( $\hat{p}$ ) in the sample, which follows a binomial distribution with parameters  $n = 95$  (sample size) and  $p = 0.3$  (under the null hypothesis).

Let  $X$  denote the number of successes (number of observations with  $\hat{p}$  greater than or equal to 0.35) in the sample. We want to calculate  $P(X \geq 35)$ .

Using the normal approximation to the binomial distribution, we can approximate this probability as  $P(X \geq 35) = 1 - P(X \leq 34)$ , where the distribution of  $X$  can be approximated by a normal distribution with mean  $np$  and variance  $np(1-p)$ .

Using  $np = 95 \times 0.3 = 28.5$  and  $np(1-p) = 95 \times 0.3 \times (1 - 0.3) = 19.95$ , we can calculate the z-score for  $X = 34$ :

$$z = \frac{34 - 28.5}{\sqrt{19.95}} = 1.378$$

Looking up the z-score in the standard normal distribution table or using a calculator, we find that the probability of observing a z-score of 1.378 or less is approximately 0.9154.

Therefore, the p-value for this test is  $1 - 0.9154 = 0.0846$ , which is closest to 0.085. Thus, the correct answer is A. 0.085.

Final Answer: A

\*\*\*\*\*

Answer Option 8:

To find the p-value, we need to calculate the probability of observing a test statistic as extreme as the one obtained, assuming that the null hypothesis is true. In this case, the test is one-tailed because we are testing if  $p$  is greater than 0.3.

We can use the normal approximation to the binomial distribution since the sample size is large. The test statistic is given by:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

where  $\hat{p}$  is the sample proportion,  $p$  is the null hypothesis value, and  $n$  is the sample size.

Plugging in the values, we have:

$$Z = \frac{0.35 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{95}}} = \frac{0.05}{0.0492} \approx 1.016$$

To find the p-value, we need to find the area under the standard normal curve to the right of  $Z = 1.016$ . Using a standard normal table or a calculator, we find that the p-value is approximately 0.1471.

Therefore, the correct answer is B. '0.1446'. Final answer: B

\*\*\*\*\*

Answer Option 9:

To find the P-value of this test, we need to calculate the probability of observing a test statistic as extreme or more extreme than the one obtained, assuming that the null hypothesis is true. In this case, the test statistic is the sample proportion  $\hat{p}$ .

Under the null hypothesis  $H_0: p=0.3$ , the test statistic follows an approximate normal distribution. We can calculate the z-score as:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

where  $\hat{p} = 0.35$  is the observed sample proportion,  $p = 0.3$  is the hypothesized population proportion, and  $n = 95$  is the sample size.

Plugging in the values, we get:

$$Z = \frac{0.35 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{95}}} \approx 1.06$$

To find the P-value, we need to find the probability of observing a z-score greater than 1.06. Consulting a standard normal distribution table or using a calculator, we find that the corresponding area under the standard normal curve is approximately 0.1446.

Therefore, the P-value of this test is approximately 0.1446. The correct answer choice is  $\boxed{\text{B. } 0.1446}$ . Final Answer: B

\*\*\*\*\*



Question 200:

Two events A and B each have a nonzero probability. If A and B are independent, which of the following statements is true?, A. ' $P(A \text{ and } B) = P(A) \cdot P(B)$ ', B. 'A and B may or may not be mutually exclusive.', C. 'A and B must be mutually exclusive.', D. ' $P(A|B) = P(B|A)$ '

=====

Answer Option 1:

If events A and B are independent, it means that the occurrence of one event does not affect the probability of the occurrence of the other event. In this case, the correct statement is A. ' $P(A \text{ and } B) = P(A) \cdot P(B)$ '. This is the definition of independence - the probability of both events A and B happening is equal to the product of the individual probabilities of A and B. Option B is incorrect because events A and B can be both mutually exclusive or not mutually exclusive in the case of independent events. Option C is incorrect because events A and B can be both mutually exclusive or not mutually exclusive in the case of independent events. Option D is incorrect because independence does not imply that the conditional probabilities are equal. Therefore, the correct answer is A. Final Answer: A.

\*\*\*\*\*

Answer Option 2:

If events A and B are independent, it means that the occurrence of one event does not affect the probability of the other event occurring.

Statement A, ' $P(A \text{ and } B) = P(A) \cdot P(B)$ ', is true when events A and B are independent. Therefore, statement A is true.

Statement B, 'A and B may or may not be mutually exclusive', is unrelated to the independence of events A and B. Events can be independent and mutually exclusive at the same time, or they can be independent and not mutually exclusive. Therefore, statement B may or may not be true.

Statement C, 'A and B must be mutually exclusive', is not necessarily true. Independent events can occur at the same time, so they can be mutually exclusive or not. Therefore, statement C is false.

Statement D, ' $P(A|B) = P(B|A)$ ', is not necessarily true for independent events. The probability of A given B and the probability of B given A are not affected by the independence of the events. Therefore, statement D is false.

Therefore, the correct statement is A. Final answer: A

\*\*\*\*\*

Answer Option 3:

If events A and B are independent, then the occurrence (or non-occurrence) of one event does not affect the probability of the other event. In other words, the probability of both A and B occurring is equal to the product of the probabilities of each event occurring individually. Therefore, the correct statement is: ' $P(A \text{ and } B) = P(A) \cdot P(B)$ ' (option A). Final answer: A

\*\*\*\*\*

Answer Option 4:

If events A and B are independent, then the occurrence or non-occurrence of one event does not affect the occurrence or non-occurrence of the other event. In other words, the probability of both events A and B occurring is simply the product of their individual probabilities. Therefore, the correct statement is A. ' $P(A \text{ and } B) = P(A) \cdot P(B)$ '. Final Answer: A

\*\*\*\*\*

Answer Option 5:

If events A and B are independent, it means that the occurrence of one event does not affect the probability of the other event happening. In this case, statement A, ' $P(A \text{ and } B) = P(A) \cdot P(B)$ ', is true. This is the definition of independent events. Statement B, 'A and B may or may not be mutually exclusive', is not necessarily true. Events A and B can be mutually exclusive or they can overlap - it depends on the specific events. Statement C, 'A and B must be mutually exclusive,' is not true because independent events can occur at the same time. Statement D, ' $P(A|B) = P(B|A)$ ', is not necessarily true. The probability of A given B is not necessarily the same as the probability of B given A. Therefore, the correct statement is A. Final answer: A

\*\*\*\*\*

Answer Option 6:

If events A and B are independent, it means that the occurrence (or non-occurrence) of one event does not affect the occurrence (or non-occurrence) of the other event. In this case, the correct statement is A. ' $P(A \text{ and } B) = P(A) \cdot P(B)$ '. This is the definition of independence - the probability of both events A and B occurring is equal to the product of the probabilities of each individual event occurring.

Option B is incorrect because independence does not imply that the events A and B are mutually exclusive. It is possible for both events to occur simultaneously.

Option C is incorrect because if events A and B are independent, they cannot be mutually exclusive.

Option D is incorrect because independence does not imply that the conditional probabilities  $P(A|B)$  and  $P(B|A)$  are equal.

\*\*\*\*\*

Answer Option 7:

If events A and B are independent, it means that the occurrence or non-occurrence of one event does not affect the probability of the other event. In other words, the probability of A happening given that B has happened is the same as the probability of A happening without considering whether B has happened or not.

Statement A, ' $P(A \text{ and } B) = P(A) \cdot P(B)$ ', is true when events A and B are independent. This is one of the key properties of independent events.

Statement B, 'A and B may or may not be mutually exclusive', is not necessarily true or false. Two events can be independent and still have some outcomes in common.

Statement C, 'A and B must be mutually exclusive', is false. Independent events can have outcomes in common.

Statement D, ' $P(A|B) = P(B|A)$ ', is not necessarily true or false. The probability of A given that B has occurred may or may not be the same as the probability of B given that A has occurred. It depends on the specific probabilities of A and B and their relationship.

Therefore, the correct statement is A. ' $P(A \text{ and } B) = P(A) \cdot P(B)$ '. Final answer: A

\*\*\*\*\*

Answer Option 8:

If events A and B are independent, then the occurrence of event A does not affect the probability of event B, and vice versa.

The correct statement is A:  $P(A \text{ and } B) = P(A) \cdot P(B)$ . This is the definition of independent events. If A and B are independent, then the probability of both events happening is equal to the product of their individual probabilities.

Statement B is not necessarily true. Events A and B can be independent even if they are not mutually exclusive. For example, if A is the event of rolling a 1 on a fair six-sided die, and B is the event of rolling an even number, A and B are not mutually exclusive (A and B can both occur if a 2 is rolled), but they can still be independent.

Statement C is false. If events A and B are independent, then they cannot be mutually exclusive. If they were mutually exclusive, then the occurrence of one event would guarantee that the other event does not happen, which would imply dependence.

Statement D is not necessarily true. The conditional probabilities  $P(A|B)$  and  $P(B|A)$  are not necessarily equal if A and B are independent. The fact that A and B are independent means that the probability of A occurring is not affected by whether B has occurred, and vice versa. Therefore, it does not imply that the conditional probabilities are equal.

Therefore, the correct statement is A. Final Answer: A

\*\*\*\*\*

Answer Option 9:

If events A and B are independent, that means that the occurrence (or non-occurrence) of event A does not affect the probability of event B, and vice versa.

Statement A, ' $P(A \text{ and } B) = P(A) \cdot P(B)$ ', is true for independent events. The probability of both events A and B occurring is equal to the product of the probabilities of each event occurring individually.

Statement B, 'A and B may or may not be mutually exclusive', is also true for independent events. Two events can be independent without being mutually exclusive. For example, if event A is rolling an even number on a fair six-sided die, and event B is rolling a number greater than 4 on the same die, these events are independent but not mutually exclusive.

Statement C, 'A and B must be mutually exclusive', is not necessarily true for independent events. As stated above, independent events can still overlap or have elements in common.

Statement D, ' $P(A|B) = P(B|A)$ ', is not necessarily true for independent events. If events A and B are independent, the occurrence or non-occurrence of one event does not affect the probability of the other event, so the conditional probabilities are not necessarily equal.

Therefore, the correct statement is A. ' $P(A \text{ and } B) = P(A) \cdot P(B)$ '. Final answer: A

\*\*\*\*\*

Question 201:

'A teacher believes that giving her students a practice quiz every week will motivate them to study harder, leading to a greater overall understanding of the course material. She tries this technique for a year, and everyone in the class achieves a grade of at least C. Is this an experiment or an observational study?', A. 'An experiment, but with no reasonable conclusion possible about cause and effect', B. 'An experiment, thus making cause and effect a reasonable conclusion', C. 'An observational study, because there was no use of a control group', D. 'An observational study, but a poorly designed one because randomization was not used'

=====

Answer Option 1:

The scenario described is an experiment. The teacher implemented a specific intervention (giving students a practice quiz every week) and observed its effect on the students' grades. Therefore, cause and effect can be reasonably concluded from this experiment. The correct answer is B. Final Answer: B

\*\*\*\*\*

Answer Option 2:

The given scenario is an example of an experiment. The teacher intentionally implements a practice quiz every week with the purpose of motivating the students to study harder. The outcome of the experiment, which is that everyone in the class achieves a grade of at least C, can reasonably be attributed to the implementation of the practice quizzes. Therefore, the answer is B. 'An experiment, thus making cause and effect a reasonable conclusion'. Final Answer: B

\*\*\*\*\*

Answer Option 3:

The scenario described is an example of an experiment. The teacher implemented a specific intervention, which was giving the students a practice quiz every week, and then observed the outcomes, which were the grades of the students. Therefore, cause and effect can be reasonably concluded. However, the given information does not provide any evidence or reasoning about the effectiveness or impact of the practice quizzes on the students' grades. So the correct answer is  $\boxed{\text{A. 'An experiment, but with no reasonable conclusion possible about cause and effect'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 4:

The scenario described is an experiment. The teacher implemented a specific intervention, which was giving the students a practice quiz every week, and observed the outcome, which was that everyone in the class achieved a grade of at least C. This experimental design allows for a reasonable conclusion about cause and effect, as the intervention was deliberately implemented to observe its impact on the students' grades. Therefore, the correct answer is B. 'An experiment, thus making cause and effect a reasonable conclusion.' Final Answer: B

\*\*\*\*\*

Answer Option 5:

The given scenario is an experiment. The teacher implemented a specific intervention, which is giving practice quizzes every week, with the purpose of motivating the students to study harder. The outcome of this intervention was that everyone in the class achieved a grade of at least C.

However, the conclusion about cause and effect is not reasonable because there are other factors that could have influenced the students' performance, such as individual motivation, external resources, or teaching methods. Therefore, the correct answer is  $\boxed{\text{A. 'An experiment, but with no reasonable conclusion possible about cause and effect'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 6:

The scenario described is an experiment. The teacher deliberately implemented a practice quiz every week with the intention of motivating the students to study harder and improve their understanding of the course material. This intervention was under the control of the teacher, making it an experimental study.

However, the correct choice is A. 'An experiment, but with no reasonable conclusion possible about cause and effect.' This is because even though the teacher implemented the intervention, there is no control group to compare the results to. Without a control group, it is difficult to confidently attribute the improvement in grades solely to the practice quizzes. Other factors, such as changes in teaching methods or student motivation, could have also influenced the outcomes. Therefore, while it is an experiment, there is no reasonable conclusion that can be made about cause and effect. Final Answer: A

\*\*\*\*\*

Answer Option 7:

The given scenario is an experiment, as the teacher deliberately implements a specific intervention (giving practice quizzes) to observe its effect on the students' performance. However, based on the information provided, we cannot draw a reasonable conclusion about cause and effect. Thus, the correct answer is  $\boxed{\text{A. 'An experiment, but with no reasonable conclusion possible about cause and effect'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 8:

The given scenario is an 'An experiment, but with no reasonable conclusion possible about cause and effect'. This is because the teacher implemented the practice quiz technique and observed that everyone in the class achieved a grade of at least C. However, there could be other factors at play that could have contributed to the students' improved grades, such as studying outside of class or increased effort. Therefore, it is not possible to conclusively attribute the improved grades solely to the practice quizzes. Final Answer: A

\*\*\*\*\*

Answer Option 9:

The given scenario is an experiment because the teacher implemented a specific intervention (giving practice quizzes every week) with the intention of observing its impact on the students' understanding of the course material. However, the conclusion about cause and effect cannot be reached definitively because there may be other factors at play that contributed to the students' improved grades. Therefore, the correct answer is  $\boxed{\text{A. 'An experiment, but with no reasonable conclusion possible about cause and effect'}}$ . Final Answer: A

\*\*\*\*\*



Question 202:

'Which of these is the best description of a P-value?', A. 'The probability of making a Type I error.', B. 'The probability of making a Type II error.', C. 'The probability of rejecting the null hypothesis if it is, in fact, false.', D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.'

=====

Answer Option 1:

The best description of a P-value is D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.' This definition captures the essence of what a P-value represents in hypothesis testing. It quantifies the strength of evidence against the null hypothesis based on the observed data. Final answer: D

\*\*\*\*\*

Answer Option 2:

The best description of a P-value is D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.' This definition captures the essence of the P-value as a measure of evidence against the null hypothesis. It quantifies how likely we would observe the data we have or more extreme data if the null hypothesis were true. Final answer: D

\*\*\*\*\*

Answer Option 3:

The best description of a P-value is D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.' The P-value is a measure of the strength of evidence against the null hypothesis. It quantifies the probability of observing data as extreme or more extreme than what was observed, assuming that the null hypothesis is true. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The best description of a P-value is option D: "The probability of getting a test statistic at least as extreme as the observed test statistic if the null hypothesis is true." The P-value is a measure of the evidence against the null hypothesis and is used in hypothesis testing. It quantifies the strength of the observed data in favor of or against the null hypothesis. If the P-value is small, it suggests that the observed data is unlikely to occur under the assumption that the null hypothesis is true, providing evidence to reject the null hypothesis. Final answer: D

\*\*\*\*\*

Answer Option 5:

The best description of a P-value is D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.' The P-value is used in hypothesis testing to determine the strength of evidence against the null hypothesis. It represents the probability of observing a test statistic as extreme as, or more extreme than, the observed test statistic, assuming the null hypothesis is true. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The best description of a P-value is D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.'

A P-value represents the strength of evidence against the null hypothesis. It is a measure of how likely we would observe the given data or more extreme data, assuming the null hypothesis is true. If the P-value is low, it suggests that the observed data is unlikely to have occurred by chance alone, leading us to reject the null hypothesis. Final Answer: D

\*\*\*\*\*

Answer Option 7:

The best description of a P-value is D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.'

The P-value is a measure of the strength of evidence against the null hypothesis. It quantifies the probability of observing a test statistic as extreme as the one obtained in the sample, assuming that the null hypothesis is true. If the P-value is small, it suggests that the observed data is unlikely under the null hypothesis, and therefore provides evidence in favor of rejecting the null hypothesis. Final answer: D

\*\*\*\*\*

Answer Option 8:

The best description of a p-value is D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.' The p-value measures the strength of evidence against the null hypothesis. It quantifies how likely the observed data would be if the null hypothesis were true. A small p-value suggests that the observed data is unlikely under the null hypothesis, leading to the rejection of the null hypothesis in favor of the alternative hypothesis. Therefore, option D is the most accurate description of a p-value. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The best description of a P-value is D. 'The probability of getting a test statistic at least as extreme as the observed test statistic, if the null hypothesis is true.' The P-value is used in hypothesis testing to determine the strength of evidence against the null hypothesis. It represents the probability of obtaining the observed data, or data more extreme, under the assumption that the null hypothesis is true. Therefore, option D accurately describes the concept of a P-value. Final answer: D.

\*\*\*\*\*

Question 203:

The financial aid office at a state university conducts a study to determine the total student costs per semester. All students are charged \$4500 for tuition. The mean cost for books is \$350 with a standard deviation of \$65. The mean outlay for room and board is \$2800 with a standard deviation of \$380. The mean personal expenditure is \$675 with a standard deviation of \$125. Assuming independence among categories, what is the standard deviation of the total student costs?, A. '\$24', B. '\$91', C. '\$190', D. '\$405'

=====

Answer Option 1:

Since the costs for tuition, books, room and board, and personal expenditure are independent, the variances simply add up. Therefore, the variance of the total student costs is  $\$(0^2) + (65^2) + (380^2) + (125^2) = 65^2 + 380^2 + 125^2$ . Taking the square root of this variance gives us the standard deviation of the total student costs, which is  $\sqrt{65^2 + 380^2 + 125^2} \approx \sqrt{5830 + 144400 + 15625} = \sqrt{178855} \approx \boxed{424}$ . Final Answer: D

\*\*\*\*\*

Answer Option 2:

To find the standard deviation of the total student costs, we need to add up the variances of each category and take the square root. Since the categories are independent, the variances can be added.

The variance in the cost of books is  $\$65^2 = 4225$ .

The variance in the cost of room and board is  $\$380^2 = 144400$ .

The variance in the cost of personal expenditure is  $\$125^2 = 15625$ .

The total variance is  $\$4225 + 144400 + 15625 = 162250$ .

Taking the square root, the standard deviation of the total student costs is  $\sqrt{162250} \approx 403.05$ .

Rounded to the nearest dollar, the standard deviation is  $\boxed{\$405}$ . Final Answer: D.

\*\*\*\*\*

Answer Option 3:

The standard deviation of the total student costs can be calculated by using the formula for the standard deviation of the sum of independent random variables.

The standard deviation of the total student costs is given by:

$$\sqrt{(\text{standard deviation of tuition})^2 + (\text{standard deviation of books})^2 + (\text{standard deviation of room and board})^2 + (\text{standard deviation of personal expenditure})^2}$$

Plugging in the given values, we have:

$$\sqrt{0^2 + 65^2 + 380^2 + 125^2} = \sqrt{4225 + 144400 + 15625} = \sqrt{299250} \approx \boxed{\$547}. \text{ Final Answer: None of the above.}$$

\*\*\*\*\*

Answer Option 4:

The total student cost is the sum of the costs for tuition, books, room and board, and personal expenditure. Since the costs for each category are independent, the variances add up. The variance of the total student cost is given by:

$$\begin{aligned} \end{aligned}$$

$$\text{Var}(X) = \text{Var}(\text{Tuition}) + \text{Var}(\text{Books}) + \text{Var}(\text{Room \& Board}) + \text{Var}(\text{Personal})$$

$$= 0 + (65^2) + (380^2) + (125^2)$$

$$= 4225 + 144400 + 15625$$

$$= 162250$$

$$\end{aligned}$$

The standard deviation is the square root of the variance, so the standard deviation of the total student costs is  $\sqrt{162250} \approx \boxed{403}$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

To find the standard deviation of the total student costs, we need to consider the variances of each category and add them up. Since the categories are independent, we can simply add the variances.

The variance of the tuition is 0, because it is a fixed cost.

The variance of the books cost is  $\$65^2 = 4225$ .

The variance of the room and board cost is  $\$380^2 = 144400$ .

The variance of the personal expenditure cost is  $\$125^2 = 15625$ .

Adding up the variances, we have  $\$0 + 4225 + 144400 + 15625 = 163250$ .

Taking the square root of the variance, we find that the standard deviation of the total student costs is  $\sqrt{163250} \approx \boxed{\$405}$ . Final answer: D

\*\*\*\*\*

Answer Option 6:

To find the standard deviation of the total student costs, we need to sum up the variances of each category and then take the square root. The variance of the tuition cost is \$0\$ because it is a fixed amount.

For the books cost, the variance is  $\$65^2 = 4225$ .

For the room and board cost, the variance is  $\$380^2 = 144400$ .

For the personal expenditure, the variance is  $\$125^2 = 15625$ .

Since the categories are independent, we can add up the variances to get the variance of the total student costs:  $\$0 + 4225 + 144400 + 15625 = 162250$ .

Finally, we take the square root of the variance to get the standard deviation:  $\sqrt{162250} \approx 402.78$ .

Therefore, the standard deviation of the total student costs is approximately  $\boxed{\$405}$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

The total student cost per semester is the sum of the costs for tuition, books, room and board, and personal expenditure. Since the costs in each category are independent, the variances add up. Therefore, the variance of the total student cost is:

$$\text{Var}[\text{total cost}] = \text{Var}[\text{tuition}] + \text{Var}[\text{books}] + \text{Var}[\text{room and board}] + \text{Var}[\text{personal expenditure}]$$

$$\text{Var}[\text{total cost}] = 0 + 65^2 + 380^2 + 125^2$$

$$\text{Var}[\text{total cost}] = 4225 + 144400 + 15625$$

$$\text{Var}[\text{total cost}] = 161250$$

The standard deviation is the square root of the variance, so:

$$\text{SD}[\text{total cost}] = \sqrt{161250} \approx \$401.56$$

Therefore, the standard deviation of the total student costs is  $\boxed{\$405}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

Since the cost for each category is independent, the variances of each category can be added to obtain the variance of the total cost. The variance of the total cost is given by:

$$\text{Var}(\text{Total Cost}) = \text{Var}(\text{Tuition}) + \text{Var}(\text{Books}) + \text{Var}(\text{Room and Board}) + \text{Var}(\text{Personal Expenditure})$$

The variance of tuition is  $\$0$  since it is a fixed cost. The variances of books, room and board, and personal expenditure are  $\$65^2$ ,  $\$380^2$ , and  $\$125^2$  respectively. Therefore, the variance of the total cost is:

$$\text{Var}(\text{Total Cost}) = 65^2 + 380^2 + 125^2$$

The standard deviation is the square root of the variance, so the standard deviation of the total cost is:

$$\sqrt{65^2 + 380^2 + 125^2} \approx \sqrt{17125 + 144400 + 15625} \approx \sqrt{177150} \approx 421$$

Therefore, the standard deviation of the total student costs is approximately  $\boxed{\$405}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

To find the standard deviation of the total student costs, we need to add up the variances of each category and then take the square root.

The variance of the tuition cost is  $\$0$  since it is a fixed amount.

The variance of the books cost is  $\$65^2 = 4225$ .

The variance of the room and board cost is  $\$380^2 = 144400$ .

The variance of the personal expenditure cost is  $\$125^2 = 15625$ .

Since the categories are independent, we can simply add up the variances to find the variance of the total student costs:  $\$0 + 4225 + 144400 + 15625 = 162250$ .

Finally, we take the square root of the variance to find the standard deviation:  $\sqrt{162250} \approx \boxed{\$403}$ . Final Answer: D

\*\*\*\*\*

Question 204:

'Self-efficacy (the belief that one has control over one's situation) as it related to job satisfaction was studied. When a group of teachers rated their ability to control their situation and their satisfaction with their job, the two variables had a correlation of 0.30. Which statement follows from this correlation?', A. 'If you want teachers to be happy with their job, give them more control over their situation.', B. 'If you want teachers to take more control over their situation, make them happier at their jobs.', C. 'Teachers in the study who were more satisfied with their job were less confident in their ability to control their situation.', D. '9% of the variability in job satisfaction can be explained by the linear model with self-efficacy as a predictor.'

=====

Answer Option 1:

The given correlation of 0.30 between the variables indicates a positive relationship, but it does not tell us about causation or the direction of the relationship. Therefore, we cannot conclude statement A or statement B.

The correlation coefficient does not provide information about the confidence level or the direction of the relationship. Therefore, we cannot conclude statement C.

The correlation coefficient does not provide information about the variability in job satisfaction that can be explained by the linear model. Therefore, we cannot conclude statement D.

Thus, none of the given statements can be directly inferred from the correlation of 0.30. Final answer: None of the above.

\*\*\*\*\*

Answer Option 2:

The given correlation of 0.30 between self-efficacy and job satisfaction does not allow us to make any causal statements. Correlation does not imply causation. Therefore, we cannot conclude options A or B.

The correlation coefficient only measures the strength and direction of the linear relationship between two variables. It does not provide information about the causality or the direction of causality. Therefore, we cannot conclude option C.

Option D makes a statement about the variability in job satisfaction that can be explained by the linear model with self-efficacy as a predictor. However, the correlation coefficient does not directly give us this information. Therefore, we cannot conclude option D.

In conclusion, none of the given statements can be inferred from the correlation coefficient of 0.30.

\*\*\*\*\*

Answer Option 3:

The given correlation of 0.30 indicates a positive relationship between self-efficacy and job satisfaction. However, correlation does not imply causation. Therefore, we cannot conclude that giving teachers more control over their situation will make them happier at their job (option A) or that making teachers happier at their jobs will make them take more control over their situation (option B).

Similarly, we cannot conclude that teachers who were more satisfied with their job were less confident in their ability to control their situation (option C). The correlation coefficient only measures the strength and direction of the linear relationship between the two variables and does not provide information about the direction of causality or the magnitude of the relationship.

Lastly, the correlation coefficient of 0.30 does not provide information about the variability in job satisfaction that can be explained by the linear model with self-efficacy as a predictor (option D). The coefficient of determination (r-squared) would be needed to determine the proportion of variability in job satisfaction that can be explained by self-efficacy.

Therefore, none of the given statements can be definitively concluded from the correlation of 0.30.

\*\*\*\*\*

Answer Option 4:

The correlation coefficient of 0.30 indicates a positive, but moderate, linear relationship between the two variables. However, it does not provide information about causation or the direction of the relationship. Therefore, we cannot conclude that either statement A or B is true based solely on the correlation coefficient.

Additionally, the correlation coefficient does not provide information about the relationship between job satisfaction and confidence in controlling one's situation for individual teachers. Therefore, statement C cannot be concluded based solely on the correlation coefficient.

Lastly, the correlation coefficient of 0.30 does not imply that 9% of the variability in job satisfaction can be explained by the linear model with self-efficacy as a predictor. The square of the correlation coefficient (r-squared) represents the proportion of the variability in one variable that can be explained by the linear relationship with the other variable. In this case, r-squared would be 0.30 squared, which is 0.09 or 9%. However, this does not imply that 9% of the variability in job satisfaction can be explained by the linear model with self-efficacy as a predictor.

Therefore, none of the given statements can be concluded based solely on the correlation coefficient of 0.30.

\*\*\*\*\*

Answer Option 5:

The correlation of 0.30 indicates a positive relationship between the two variables, but it does not imply causation. Therefore, statements A and B, which imply a causal relationship, cannot be concluded from the correlation alone.

Statement C cannot be concluded from the correlation either. The correlation coefficient of 0.30 only indicates the strength and direction of the linear relationship between the variables, not the specific relationship between satisfaction and confidence.

Statement D also cannot be concluded from the correlation. The correlation coefficient only measures the strength of the linear relationship between the variables, but it does not provide information about the variability in job satisfaction that can be explained by self-efficacy as a predictor.

Therefore, none of the statements can be concluded from the given correlation.

\*\*\*\*\*

Answer Option 6:

The correlation coefficient of 0.30 indicates a positive relationship between self-efficacy and job satisfaction. However, correlation does not imply causation. Therefore, we cannot conclude that giving teachers more control over their situation will make them happier with their job (option A) or that making teachers happier at their jobs will make them take more control over their situation (option B).



Furthermore, the correlation coefficient does not provide information about the direction of causation or the causal relationship between the variables. Therefore, we cannot conclude that teachers who were more satisfied with their job were less confident in their ability to control their situation (option C).

Finally, the correlation coefficient does not provide information about the amount of variability explained by the linear model. Therefore, we cannot conclude that 9% of the variability in job satisfaction can be explained by the linear model with self-efficacy as a predictor (option D).

Based on the given information, none of the statements can be concluded.

\*\*\*\*\*

Answer Option 7:

The correlation of 0.30 only tells us that there is a positive relationship between the two variables (self-efficacy and job satisfaction), but it does not provide information about causation or the direction of the relationship. Therefore, we cannot conclude that giving teachers more control over their situation will make them happier with their job (option A) or that making teachers happier at their jobs will make them take more control over their situation (option B).

Similarly, we cannot conclude that teachers who were more satisfied with their job were less confident in their ability to control their situation (option C), as the correlation coefficient does not indicate the direction of the relationship between the variables.

Furthermore, we cannot determine the amount of variability in job satisfaction that can be explained by the linear model with self-efficacy as a predictor based solely on the correlation coefficient (option D).

Therefore, none of the statements necessarily follow from the given correlation coefficient.

\*\*\*\*\*

Answer Option 8:

The given correlation of 0.30 implies a positive relationship between self-efficacy and job satisfaction, but it does not provide information about causation. Therefore, neither statement A nor statement B can be inferred from the correlation.

Statement C contradicts the positive correlation between self-efficacy and job satisfaction. The correlation of 0.30 suggests that teachers who rated higher in self-efficacy were more likely to be satisfied with their job, not less confident in their ability to control their situation.

Statement D is also incorrect. The correlation coefficient of 0.30 does not provide information about the percentage of variability in job satisfaction that can be explained by the linear model with self-efficacy as a predictor. It only indicates the strength and direction of the linear relationship between the two variables.

Therefore, none of the statements can be inferred from the given correlation, and the correct answer is: None of the above.

\*\*\*\*\*

Answer Option 9:

The given correlation of 0.30 indicates a positive relationship between self-efficacy and job satisfaction. However, correlation does not imply causation, so we cannot conclude that one variable causes the other.

Option A suggests a causal relationship where giving teachers more control over their situation will make them happier with their job. This cannot be conclusively determined based on the correlation coefficient alone.

Option B suggests a reverse causal relationship where making teachers happier at their jobs will lead them to take more control over their situation. Again, this cannot be determined solely from the given correlation.

Option C suggests a negative relationship between job satisfaction and self-efficacy, which contradicts the given correlation of 0.30, indicating a positive relationship.

Option D is unrelated to the given correlation. It introduces the concept of variability explained by a linear model, which is not mentioned in the question.

Therefore, none of the given options can be conclusively determined based on the given correlation coefficient.

\*\*\*\*\*

Question 205:

'Suppose X and Y are random variables with  $E(X) = 312$ ,  $\text{var}(X) = 6$ ,  $E(Y) = 307$ , and  $\text{var}(Y) = 8$ . What are the expected value and variance of the random variable  $X + Y$ ?', A. ' $E(X + Y) = 619$ ,  $\text{var}(X + Y) = 7$ ', B. ' $E(X + Y) = 619$ ,  $\text{var}(X + Y) = 10$ ', C. ' $E(X + Y) = 619$ ,  $\text{var}(X + Y) = 14$ ', D. 'There is insufficient information to answer this question.'

=====

Answer Option 1:

The expected value of the sum of two random variables is the sum of their individual expected values. Therefore,  $E(X + Y) = E(X) + E(Y) = 312 + 307 = 619$ .

The variance of the sum of two independent random variables is the sum of their individual variances. Therefore,  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14$ .

Therefore, the expected value of  $X + Y$  is 619 and the variance of  $X + Y$  is 14. The correct answer is C. Final answer: C.

\*\*\*\*\*

Answer Option 2:

We know that for any two random variables X and Y, the expected value of their sum is equal to the sum of their expected values, and the variance of their sum is equal to the sum of their variances.

Therefore,

$\begin{aligned} & \\ & \end{aligned}$

$$E(X + Y) = E(X) + E(Y) = 312 + 307 = 619$$

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14$$

$\end{aligned}$

Therefore, the expected value of  $X + Y$  is 619 and the variance of  $X + Y$  is 14. Thus, the correct answer is  $\boxed{\text{C. 'E(X + Y) = 619, var(X + Y) = 14'}}$ .

\*\*\*\*\*

Answer Option 3:

The expected value of the sum of two random variables is equal to the sum of their individual expected values. Thus,  $E(X + Y) = E(X) + E(Y) = 312 + 307 = 619$ .

The variance of the sum of two independent random variables is equal to the sum of their individual variances. Thus,  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14$ .

Therefore, the expected value and variance of the random variable  $X + Y$  are  $E(X + Y) = 619$  and  $\text{var}(X + Y) = 14$ .

The correct answer is C. Final Answer: C

\*\*\*\*\*

Answer Option 4:

The expected value of a sum of random variables is the sum of their individual expected values. Therefore, we have:

$$E(X + Y) = E(X) + E(Y) = 312 + 307 = 619.$$

The variance of a sum of independent random variables is the sum of their individual variances. Therefore, we have:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14.$$

Thus, the correct answer is  $E(X + Y) = 619$ ,  $\text{var}(X + Y) = 14$ , which corresponds to choice C.

\*\*\*\*\*

Answer Option 5:

The expected value of a sum of random variables is the sum of the expected values, so we have

$$E(X + Y) = E(X) + E(Y) = 312 + 307 = 619.$$

The variance of a sum of independent random variables is the sum of the variances, so we have

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14.$$

Therefore, the expected value of  $X + Y$  is 619 and the variance of  $X + Y$  is 14. So the correct answer is  $\boxed{\text{C} } E(X + Y) = 619, \text{ var}(X + Y) = 14$ .

\*\*\*\*\*

Answer Option 6:

The expected value of a sum of random variables is equal to the sum of their individual expected values, so we have  $E(X + Y) = E(X) + E(Y) = 312 + 307 = 619$ .

The variance of a sum of independent random variables is equal to the sum of their individual variances, so we have  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14$ .

Therefore, the expected value of  $X + Y$  is 619 and the variance of  $X + Y$  is 14. The correct answer is  $\boxed{\text{C} } E(X + Y) = 619, \text{ var}(X + Y) = 14$ .

\*\*\*\*\*

Answer Option 7:

We can use the properties of expected value and variance to find the expected value and variance of the sum of X and Y.

First, we know that the expected value of the sum of two random variables is equal to the sum of their individual expected values. Therefore,  $E(X + Y) = E(X) + E(Y) = 312 + 307 = 619$ .

Next, the variance of the sum of two independent random variables is equal to the sum of their individual variances. Therefore,  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14$ .

Thus, the expected value of  $X + Y$  is 619 and the variance of  $X + Y$  is 14. The correct answer is C. ' $E(X + Y) = 619$ ,  $\text{var}(X + Y) = 14$ '.

\*\*\*\*\*

Answer Option 8:

We know that the expected value of a sum of random variables is the sum of their expected values. Therefore, we have

$$E(X + Y) = E(X) + E(Y) = 312 + 307 = 619.$$

Similarly, the variance of a sum of independent random variables is the sum of their variances. Since X and Y are independent, we have

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14.$$

Therefore, the expected value of  $X + Y$  is 619 and the variance of  $X + Y$  is 14, so the answer is  $\boxed{\text{C}}$   $E(X + Y) = 619, \text{var}(X + Y) = 14$ .

\*\*\*\*\*

Answer Option 9:

The expected value of the sum of two random variables is equal to the sum of their individual expected values. Therefore,  $E(X + Y) = E(X) + E(Y) = 312 + 307 = 619$ .

The variance of the sum of two independent random variables is equal to the sum of their individual variances. Therefore,  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) = 6 + 8 = 14$ .

Therefore, the expected value of  $X + Y$  is 619 and the variance of  $X + Y$  is 14. Thus, the correct answer is C.  $E(X + Y) = 619, \text{var}(X + Y) = 14$ .

\*\*\*\*\*

Question 206:

'Consider the two sets  $X = \{10, 30, 45, 50, 55, 70, 90\}$  and  $Y = \{10, 30, 35, 50, 65, 70, 90\}$ . Which of the following is false?', A. 'The sets have identical medians.', B. 'The sets have identical means.', C. 'The sets have identical ranges.', D. 'None of the above are false.'

=====

Answer Option 1:

To determine which statement is false, we can compare the given sets X and Y.

The median is the middle value when the numbers are arranged in ascending order. For both sets X and Y, the middle value is 50. Therefore, statement A is true.

The mean is calculated by adding up all the values in a set and dividing by the number of elements. For set X, the mean is  $\frac{10+30+45+50+55+70+90}{7} = \frac{350}{7} = 50$ . For set Y, the mean is  $\frac{10+30+35+50+65+70+90}{7} = \frac{350}{7} = 50$ . Therefore, statement B is true.

The range is the difference between the largest and smallest values in a set. For set X, the range is  $90-10 = 80$ . For set Y, the range is  $90-10 = 80$ . Therefore, statement C is true.

Since all the statements are true, the answer is  $\boxed{\text{D. 'None of the above are false.'}}$

\*\*\*\*\*

Answer Option 2:

To determine which statement is false, we can compare the given sets X and Y.

The median is the middle value when the numbers are arranged in ascending order. X and Y have 7 elements each, so the median is the 4th element. In both sets, the 4th element is 50. Therefore, statement A is true.

The mean is calculated by summing all the elements and dividing by the number of elements. For set X, the mean is  $(10+30+45+50+55+70+90)/7 = 350/7 = 50$ . For set Y, the mean is  $(10+30+35+50+65+70+90)/7 = 350/7 = 50$ . Therefore, statement B is true.

The range is the difference between the maximum and minimum values. For set X, the minimum is 10 and the maximum is 90, so the range is  $90-10=80$ . For set Y, the minimum is 10 and the maximum is 90, so the range is  $90-10=80$ . Therefore, statement C is true.

Since all the statements are true, the correct answer is  $\boxed{\text{D. 'None of the above are false.'}}$ .

\*\*\*\*\*

Answer Option 3:

To determine which statement is false, we can compare the medians, means, and ranges of the two sets:

The median of a set is the middle value when the numbers are arranged in increasing order. For set X, the median is 50, and for set Y, the median is also 50. Therefore, the statement "The sets have identical medians" is true.

The mean of a set is calculated by summing all the numbers in the set and dividing by the number of elements. For set X, the mean is  $(10 + 30 + 45 + 50 + 55 + 70 + 90)/7 = 50$ . For set Y, the mean is  $(10 + 30 + 35 + 50 + 65 + 70 + 90)/7 = 50$ . Therefore, the statement "The sets have identical means" is true.

The range of a set is the difference between the largest and smallest values. For set X, the range is  $90 - 10 = 80$ . For set Y, the range is  $90 - 10 = 80$ . Therefore, the statement "The sets have identical ranges" is true.

Since all the statements are true, the correct answer is  $\boxed{\text{D. None of the above are false.'}}$

\*\*\*\*\*

Answer Option 4:

To determine whether the given statements are true or false, we can compare the medians, means, and ranges of the two sets X and Y.

The median of a set is the value in the middle when all the numbers are arranged in ascending order. For set X, the median is 50, and for set Y, the median is also 50. Therefore, statement A, "The sets have identical medians," is true.

The mean of a set is the sum of all the numbers divided by the count of numbers. For set X, the mean is  $(10 + 30 + 45 + 50 + 55 + 70 + 90) / 7 = 50$ , and for set Y, the mean is  $(10 + 30 + 35 + 50 + 65 + 70 + 90) / 7 = 51.43$ . Therefore, statement B, "The sets have identical means," is false.

The range of a set is the difference between the largest and smallest values in the set. For set X, the range is  $90 - 10 = 80$ , and for set Y, the range is also  $90 - 10 = 80$ . Therefore, statement C, "The sets have identical ranges," is true.

Since statement B is false, the correct answer is  $\boxed{\text{B. 'The sets have identical means.'}}$

\*\*\*\*\*

Answer Option 5:

To determine if the given statements are true or false, we can compare the properties of the two sets.

The median of a set is the middle value when the elements are arranged in increasing order. For both sets X and Y, the medians are both 50. Therefore, statement A is true.

The mean of a set is the sum of all the elements divided by the number of elements. For set X, the mean is  $\frac{10 + 30 + 45 + 50 + 55 + 70 + 90}{7} = \frac{350}{7} = 50$ . For set Y, the mean is  $\frac{10 + 30 + 35 + 50 + 65 + 70 + 90}{7} = \frac{350}{7} = 50$ . Therefore, statement B is true.

The range of a set is the difference between the maximum and minimum values. For set X, the maximum value is 90 and the minimum value is 10, so the range is  $90 - 10 = 80$ . For set Y, the maximum value is 90 and the minimum value is 10, so the range is  $90 - 10 = 80$ . Therefore, statement C is true.

Since all of the given statements are true, the answer is  $\boxed{\text{D. None of the above are false.'}}$

\*\*\*\*\*

Answer Option 6:

To determine which statement is false, we need to compare the medians, means, and ranges of the sets X and Y.

The median of a set is the middle value when the numbers are arranged in increasing order. The medians of both sets X and Y are 50, so statement A is true.

The mean of a set is the average of all the numbers in the set. The means of both sets X and Y can be calculated as  $(10+30+45+50+55+70+90)/7 = 50$ , so statement B is true.

The range of a set is the difference between the maximum and minimum values. The range of set X is  $90 - 10 = 80$ , and the range of set Y is  $90 - 10 = 80$ . Therefore, statement C is true.

Since all the statements are true, the answer is  $\boxed{\text{D. None of the above are false.}}$

\*\*\*\*\*

Answer Option 7:

The median of a set is the middle value when the values are arranged in ascending order. For both sets X and Y, the middle value is 50. Therefore, statement A is true.

The mean of a set is the average value of all the values in the set. For both sets X and Y, the mean is equal to  $(10+30+45+50+55+70+90)/7 = (10+30+35+50+65+70+90)/7 = 50$ . Therefore, statement B is true.

The range of a set is the difference between the largest and smallest values in the set. For set X, the range is  $90-10 = 80$ . For set Y, the range is  $90-10 = 80$ . Therefore, statement C is true.

Since all the statements are true, the answer is  $\boxed{\text{D. None of the above are false.}}$

\*\*\*\*\*

Answer Option 8:

To determine if the statements are true or false, we need to compare the medians, means, and ranges of the two sets X and Y.

The median is the middle value of a set when it is arranged in ascending order. For both sets X and Y, the medians are both 50. Therefore, statement A is true.

The mean is the average of the values in a set. For both sets X and Y, the means can be calculated as follows:

Mean of X =  $(10 + 30 + 45 + 50 + 55 + 70 + 90) / 7 = 350 / 7 = 50$

Mean of Y =  $(10 + 30 + 35 + 50 + 65 + 70 + 90) / 7 = 350 / 7 = 50$

Therefore, statement B is true.

The range is the difference between the largest and smallest values in a set. For set X, the range is  $90 - 10 = 80$ . For set Y, the range is  $90 - 10 = 80$ . Therefore, statement C is true.

Since statements A, B, and C are all true, the correct answer is  $\boxed{\text{D. None of the above are false.}}$

\*\*\*\*\*

Answer Option 9:

To determine which statement is false, we compare the given characteristics of the sets:

A. The sets have identical medians.

To find the median, we arrange the numbers in ascending order:

X: 10, 30, 45, 50, 55, 70, 90

Y: 10, 30, 35, 50, 65, 70, 90

Both sets have the same number of elements, so the medians are the middle values:

X median: 50

Y median: 50

A is true.

B. The sets have identical means.

To find the mean, we sum all the numbers in each set and divide by the number of elements:

X mean:  $(10 + 30 + 45 + 50 + 55 + 70 + 90)/7 = 50$

Y mean:  $(10 + 30 + 35 + 50 + 65 + 70 + 90)/7 = 50$

B is true.

C. The sets have identical ranges.

The range is the difference between the maximum and minimum values in each set:

X range:  $90 - 10 = 80$

Y range:  $90 - 10 = 80$

C is true.

Since all the statements are true, the answer is  $\boxed{\text{D. None of the above are false.}}$

\*\*\*\*\*

Question 207:

'Which of the following is a true statement?', A. 'In a well-designed, well-conducted sample survey, sampling error is effectively eliminated.', B. 'In a well-designed observational study, responses are influenced through an orderly, carefully planned procedure during the collection of data.', C. 'In a well-designed experiment, the treatments are carefully planned to result in responses that are as similar as possible.', D. 'None of the above is a true statement.'

=====

Answer Option 1:

The correct answer is D. 'None of the above is a true statement.'

Explanation:

A. In a well-designed, well-conducted sample survey, sampling error is not eliminated. Sampling error refers to the variability that occurs due to sampling, and it cannot be completely eliminated. However, proper sampling techniques can help minimize sampling error.

B. In a well-designed observational study, responses are not influenced through an orderly, carefully planned procedure during the collection of data. In an observational study, the researcher observes and collects data without applying any treatments or interventions. Hence, the responses are not influenced in a controlled manner.

C. In a well-designed experiment, the treatments are not planned to result in responses that are as similar as possible. Rather, in a well-designed experiment, treatments are deliberately varied to assess their impact on the response variable. The purpose is to compare the effects of different treatments, not to make the responses as similar as possible.

Therefore, the correct answer is D. 'None of the above is a true statement.'

\*\*\*\*\*

Answer Option 2:

The correct answer is D. 'None of the above is a true statement.'

Explanation:

A. In a well-designed, well-conducted sample survey, sampling error is not eliminated. Sampling error refers to the variability that occurs when a sample is selected from a population rather than the entire population being surveyed. It is impossible to completely eliminate sampling error, although it can be minimized through proper sampling techniques.

B. In a well-designed observational study, responses are not influenced through an orderly, carefully planned procedure during the collection of data. In an observational study, the researcher only observes and collects data without intervening or manipulating any variables. The responses are not influenced by the researcher in any way.

C. In a well-designed experiment, the treatments are not planned to result in responses that are as similar as possible. In fact, the goal of a well-designed experiment is to carefully plan and manipulate the treatments in order to observe the effects on the responses which are expected to vary based on the treatments applied.

Therefore, none of the statements are true, making D the correct answer. Final answer: D

\*\*\*\*\*

Answer Option 3:

The correct answer is D. 'None of the above is a true statement.'

A. 'In a well-designed, well-conducted sample survey, sampling error is effectively eliminated.'

This statement is false. Sampling error is a natural part of any sampling process and cannot be completely eliminated. It refers to the difference between the characteristics of the sample and the characteristics of the population being studied.

B. 'In a well-designed observational study, responses are influenced through an orderly, carefully planned procedure during the collection of data.'

This statement is false. In an observational study, the researcher does not have control over the variables being studied. The responses are not influenced by any planned procedure.

C. 'In a well-designed experiment, the treatments are carefully planned to result in responses that are as similar as possible.'

This statement is false. In a well-designed experiment, the treatments are carefully planned to result in responses that are different from each other. This allows for the comparison of the effects of different treatments.

Therefore, the correct answer is D. 'None of the above is a true statement.'

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'None of the above is a true statement.'

A. 'In a well-designed, well-conducted sample survey, sampling error is effectively eliminated.' Sampling error is inherent in any sample survey and cannot be completely eliminated, even with a well-designed and well-conducted survey. Sampling error refers to the difference between the characteristics of the sample and the characteristics of the population being studied.

B. 'In a well-designed observational study, responses are influenced through an orderly, carefully planned procedure during the collection of data.' In an observational study, the researcher does not actively influence or manipulate the responses of the individuals being observed. The responses occur naturally, without any planned intervention or influence from the researcher.

C. 'In a well-designed experiment, the treatments are carefully planned to result in responses that are as similar as possible.' In a well-designed experiment, the treatments are carefully planned to result in responses that are different, not similar. The goal of an experiment is to compare the effects of different treatments or conditions on a response variable.

Therefore, none of the statements given is true. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The correct answer is D. 'None of the above is a true statement.'

Explanation:

- Option A is incorrect because even in a well-designed and well-conducted sample survey, sampling error cannot be completely eliminated. Sampling error refers to the natural variability that occurs when a sample is used to estimate characteristics of a larger population. While a well-designed survey can minimize sampling error, it cannot eliminate it entirely.

- Option B is incorrect because in an observational study, responses are not influenced through an orderly, carefully planned procedure. In an observational study, the researcher observes and collects data without intervening or influencing the subjects or variables being studied. The goal is to observe and analyze existing conditions or relationships without any manipulation or intervention.

- Option C is incorrect because in a well-designed experiment, the treatments are carefully planned to create differences between groups or conditions. The goal of an experiment is to investigate cause-and-effect relationships by manipulating one or more variables. The treatments are intentionally designed to create variations or differences in the independent variable(s) to observe the corresponding effects on the dependent variable(s).

Therefore, none of the statements are true, making option D the correct answer. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The correct answer is D. 'None of the above is a true statement.'

Explanation:

A. 'In a well-designed, well-conducted sample survey, sampling error is effectively eliminated.'

Sampling error refers to the difference between the characteristics of the sample and the characteristics of the population being studied. While a well-designed and well-conducted sample survey can minimize sampling error, it is impossible to completely eliminate it. Sampling error is inherent in any form of sampling.

B. 'In a well-designed observational study, responses are influenced through an orderly, carefully planned procedure during the collection of data.'

Observational studies involve observing and collecting data on subjects without any intervention or influence. In a well-designed observational study, responses are not influenced by any planned procedure. The goal is to simply observe and record data without any deliberate manipulation.

C. 'In a well-designed experiment, the treatments are carefully planned to result in responses that are as similar as possible.'

In a well-designed experiment, the treatments are actually carefully planned to result in responses that are as different as possible. The purpose of an experiment is to compare the effects of different treatments on the response variable. By intentionally manipulating the treatments to be different, researchers can assess the impact of these differences on the responses.

Therefore, none of the statements are true, and the correct answer is D. 'None of the above is a true statement.'

\*\*\*\*\*

Answer Option 7:

The correct statement is D. 'None of the above is a true statement.'

Explanation:

A. 'In a well-designed, well-conducted sample survey, sampling error is effectively eliminated.'

Sampling error refers to the difference between the characteristics of the sample and the characteristics of the population being studied. In any sample survey, there will always be some level of sampling error. A well-designed and well-conducted sample survey can minimize sampling error, but it cannot eliminate it entirely. Therefore, statement A is not true.

B. 'In a well-designed observational study, responses are influenced through an orderly, carefully planned procedure during the collection of data.'

In an observational study, the researcher observes and collects data without intervening or manipulating any variables. The responses are not influenced by a planned procedure, as the researcher is simply observing and recording what is happening naturally. Therefore, statement B is not true.

C. 'In a well-designed experiment, the treatments are carefully planned to result in responses that are as similar as possible.'

In a well-designed experiment, the treatments are carefully planned to result in responses that are different, not similar. The purpose of an experiment is to test the effect of different treatments on the response variable. Therefore, statement C is not true.

Since none of the statements are true, the correct answer is D. 'None of the above is a true statement.' Final Answer: D

\*\*\*\*\*

Answer Option 8:

The true statement among the options is D. 'None of the above is a true statement.'

Explanation:

A. 'In a well-designed, well-conducted sample survey, sampling error is effectively eliminated.' - This statement is not true. Sampling error refers to the natural variation that can occur when taking a sample from a larger population. While a well-designed and well-conducted sample survey can help reduce sampling error, it cannot be completely eliminated.

B. 'In a well-designed observational study, responses are influenced through an orderly, carefully planned procedure during the collection of data.' - This statement is also not true. In an observational study, responses are observed or recorded as they naturally occur, without any intervention or influence from the researcher. The goal is to observe and gather data on existing behaviors or characteristics.

C. 'In a well-designed experiment, the treatments are carefully planned to result in responses that are as similar as possible.' - This statement is not true. In a well-designed experiment, the treatments are carefully planned to manipulate or influence certain variables to observe their effects on the response variable. The goal is to create differences between the treatments to understand their impact, not to make them as similar as possible.

Therefore, the correct answer is D. 'None of the above is a true statement.' Final Answer: D

\*\*\*\*\*

Answer Option 9:

The correct answer is D. 'None of the above is a true statement.'

A. 'In a well-designed, well-conducted sample survey, sampling error is effectively eliminated.' - This statement is not true. Sampling error is inherent in any sample survey, no matter how well-designed and well-conducted it is. Sampling error refers to the difference between the characteristics of the sample and the characteristics of the population being studied. It is impossible to completely eliminate sampling

error, although it can be minimized through proper sampling techniques.

B. 'In a well-designed observational study, responses are influenced through an orderly, carefully planned procedure during the collection of data.' - This statement is not true. In an observational study, the researcher does not have control over the factors being studied. The responses are not influenced by any planned procedure during data collection. Observational studies aim to observe and analyze relationships between variables without any intervention or manipulation by the researcher.

C. 'In a well-designed experiment, the treatments are carefully planned to result in responses that are as similar as possible.' - This statement is not true. In a well-designed experiment, the treatments are carefully planned to be different from each other. The goal is to compare the effects of different treatments on the response variable. The treatments are intentionally varied to determine the impact of the different factors on the response variable.

Therefore, the correct answer is D. 'None of the above is a true statement.' Final Answer: D

\*\*\*\*\*



Question 208:

A researcher is hoping to find a predictive linear relationship between the explanatory and response variables in her study. Accordingly, as part of her analysis she plans to generate a 95% confidence interval for the slope of the regression line for the two variables. The interval is determined to be (0.45, 0.80). Which of the following is (are) true? (Assume conditions for inference are met.) I. She has good evidence of a linear relationship between the variables. II. It is likely that there is a non-zero correlation ( $r$ ) between the two variables. III. It is likely that the true slope of the regression line is 0. A. 'I and II only', B. 'I and III only', C. 'II and III only', D. 'I only'

=====

Answer Option 1:

Based on the 95% confidence interval for the slope of the regression line, which is (0.45, 0.80), we can conclude that the researcher has good evidence of a linear relationship between the variables (I). This is because the confidence interval does not include zero.

However, the confidence interval does not give us any information about the correlation ( $r$ ) between the two variables (II). The correlation can be non-zero even if the confidence interval includes zero.

Therefore, statement II cannot be determined from the given information.

Lastly, the confidence interval does not include zero, which means it is unlikely that the true slope of the regression line is zero (III). Therefore, statement III is not true.

In conclusion, the correct answer is  $\boxed{\text{A. 'I and II only'}}$ . Final answer: A

\*\*\*\*\*

Answer Option 2:

The researcher has determined a 95% confidence interval for the slope of the regression line to be (0.45, 0.80). This means that she is 95% confident that the true slope of the regression line falls within this interval.

I. She has good evidence of a linear relationship between the variables.

This statement is true because the confidence interval for the slope does not contain zero. Since the interval is entirely positive, it suggests that there is a positive linear relationship between the variables.

II. It is likely that there is a non-zero correlation ( $r$ ) between the two variables.

This statement is also true because a non-zero slope indicates that there is a non-zero correlation between the two variables. A positive slope suggests a positive correlation, while a negative slope suggests a negative correlation.

III. It is likely that the true slope of the regression line is 0.

This statement is false because the confidence interval does not include 0. Therefore, it is unlikely that the true slope of the regression line is 0.

Based on the analysis, the correct answer is  $\boxed{\text{A. 'I and II only'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 3:

A 95% confidence interval for the slope of the regression line of the two variables does not include zero, so we can conclude that it is likely that the true slope of the regression line is not zero. This means that there is evidence of a linear relationship between the variables (I is true). Since there is evidence of a linear relationship, it is likely that there is a non-zero correlation ( $r$ ) between the two variables (II is true).

However, we cannot conclude whether the true slope of the regression line is within the given confidence interval or not, so we cannot say whether it is likely that the true slope of the regression line is zero or not (III is not necessarily true). Therefore, the correct answer is  $\boxed{\text{A. 'I and II only'}}$ .

\*\*\*\*\*

Answer Option 4:

In this case, since the confidence interval for the slope of the regression line does not include 0, we can conclude that the true slope of the regression line is likely not 0. Therefore, we can say that statement III is false.

However, the fact that the confidence interval does not include 0 provides evidence that there is a linear relationship between the variables. Thus, statement I is true.

The confidence interval for the slope does not provide information about the correlation between the variables. The correlation coefficient is a measure of the strength and direction of the linear relationship between the variables, and it is not directly determined by the interval estimate for the slope of the regression line. Therefore, statement II is not necessarily true.

Based on the above analysis, the correct option is  $\boxed{\text{A. 'I and II only'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 5:

For a 95% confidence interval for the slope of the regression line, if the interval does not contain the value 0, then we have evidence of a linear relationship between the variables. In this case, the interval is (0.45, 0.80), which does not include 0. Therefore, we have evidence of a linear relationship between the variables, so statement I is true.

Since the interval does not include 0, it is likely that the true slope of the regression line is not 0. Therefore, statement III is false.

The presence of a non-zero correlation ( $r$ ) between the two variables is not directly determined by the confidence interval for the slope of the regression line. Therefore, we cannot conclude whether statement II is true or false.

Thus, the correct answer is  $\boxed{\text{D. 'I only'}}$

\*\*\*\*\*

Answer Option 6:

In this case, since the confidence interval for the slope of the regression line does not include 0, we can conclude that the researcher has good evidence of a linear relationship between the variables.

Therefore, statement I is true.

However, the fact that the confidence interval does not include 0 does not provide information about the correlation ( $r$ ) between the two variables. The correlation coefficient measures the strength and direction of the linear relationship, whereas the confidence interval for the slope provides information about the statistical significance of the slope. Therefore, statement II is not necessarily true.

Similarly, the fact that the confidence interval does not include 0 does not imply that the true slope of the regression line is exactly 0. It only means that there is evidence of a non-zero slope. Therefore, statement III is not necessarily true.

Hence, the correct answer is  $\boxed{\text{D. 'I only'}}$ .

\*\*\*\*\*

Answer Option 7:

The researcher has obtained a 95% confidence interval for the slope of the regression line for the two variables, which is (0.45, 0.80). This means that, based on the sample data, the researcher is 95% confident that the true slope of the regression line falls within this interval.

I. The fact that the confidence interval does not include 0 indicates that there is evidence of a linear relationship between the variables. If the true slope of the regression line was 0, it would imply that there is no linear relationship between the variables. Since the confidence interval does not include 0, we can say that there is evidence of a linear relationship.

II. The fact that the confidence interval does not include 0 also indicates that it is likely that there is a non-zero correlation ( $r$ ) between the two variables. The slope of the regression line is directly related to the correlation between the variables. A non-zero slope indicates a non-zero correlation.

III. The confidence interval does not contain the value 0, so it is not likely that the true slope of the regression line is 0.

Therefore, the correct option is A. 'I and II only'. Final Answer: A

\*\*\*\*\*

Answer Option 8:

In this case, we have a 95% confidence interval for the slope of the regression line, which is (0.45, 0.80). This means that we can be 95% confident that the true slope falls within this interval.

I. She has good evidence of a linear relationship between the variables.

Since the confidence interval does not contain zero, we can conclude that there is evidence of a linear relationship between the variables. This is because a slope of zero would indicate no linear relationship.

II. It is likely that there is a non-zero correlation ( $r$ ) between the two variables.

Since the slope of the regression line is not zero, we can conclude that there is likely a non-zero correlation between the variables. This is because the slope represents the change in the response variable for a one-unit increase in the explanatory variable.

III. It is likely that the true slope of the regression line is 0.

This statement is not true, as the confidence interval does not include 0. Therefore, it is not likely that the true slope of the regression line is 0.

Based on the explanations above, we can conclude that the correct answer is  $\boxed{\text{A. 'I and II only'}}$ . Final Answer: A

\*\*\*\*\*

Answer Option 9:

The fact that the researcher has generated a confidence interval for the slope of the regression line indicates that she has found evidence of a linear relationship between the variables (I). This is because the confidence interval does not include zero, suggesting that the slope is significantly different from zero.

Since the confidence interval does not include zero, it is also likely that there is a non-zero correlation between the two variables (II). This is because the slope of the regression line is related to the correlation coefficient.

However, we cannot conclude that the true slope of the regression line is exactly zero because the confidence interval does not include zero. Therefore, statement III is incorrect.

In conclusion, statements I and II are true, so the answer is  $\boxed{\text{A. 'I and II only'}}$ . Final answer: A

\*\*\*\*\*

Question 209:

'A student organization at a university is interested in estimating the proportion of students in favor of showing movies biweekly instead of monthly. How many students should be sampled to get a 90 percent confidence interval with a width of at most 0.08?', A. '27', B. '64', C. '106', D. '423'

=====

Answer Option 1:

To determine the sample size needed, we must use the formula for sample size in a proportion estimation problem. The formula is given by

$$n = \frac{z^2 \cdot p \cdot (1-p)}{E^2}$$

where  $n$  is the required sample size,  $z$  is the z-score corresponding to the desired level of confidence (90% in this case),  $p$  is the estimated proportion, and  $E$  is the desired margin of error.

Since we do not have an estimate of the proportion, we can use  $p = 0.5$  as a conservative estimate since this will yield the largest sample size. Plugging in the values, we have

$$n = \frac{z^2 \cdot p \cdot (1-p)}{E^2} = \frac{1.645^2 \cdot 0.5 \cdot 0.5}{(0.08)^2} \approx 64$$

Therefore, the required sample size is approximately 64. Final answer: B

\*\*\*\*\*

Answer Option 2:

To determine the sample size needed, we use the formula

$$n = \left( \frac{z \cdot p \cdot (1-p)}{E} \right)^2$$

where  $n$  is the sample size,  $z$  is the z-score corresponding to the desired confidence level (in this case, 90% confidence level corresponds to a z-score of approximately 1.645),  $p$  is the estimated proportion, and  $E$  is the maximum error or width of the confidence interval.

Since we do not have an initial estimate of the proportion, we can use  $p = 0.5$  which gives the maximum sample size. Substituting the given values into the formula, we have

$$n = \left( \frac{1.645 \cdot 0.5 \cdot (1-0.5)}{0.08} \right)^2 \approx 106.859$$

Since we cannot have a fraction of a student, we round up the sample size to the next whole number, giving us  $n = 107$ .

Therefore, the answer is 106. Final Answer: C

\*\*\*\*\*

Answer Option 3:

To determine the sample size needed to estimate a proportion with a desired margin of error, we need to use the formula:

$$n = \frac{z^2 \cdot p(1-p)}{E^2}$$

where:

- $n$  is the sample size
- $z$  is the z-score corresponding to the desired level of confidence (in this case, 90% confidence corresponds to a z-score of approximately 1.645)
- $p$  is the estimated proportion (since we don't have an estimate, we can use 0.5, which maximizes the required sample size)
- $E$  is the desired margin of error (0.08 in this case)

Plugging in the values, we get:

$$n = \frac{1.645^2 \cdot 0.5(1-0.5)}{(0.08)^2} = 106.4$$

Rounding up to the nearest whole number, the required sample size is 107. Therefore, the answer is 106. Final answer: C

\*\*\*\*\*

Answer Option 4:

To determine the sample size needed, we need to use the formula for sample size calculation for estimating proportions. The formula is given by:

$$n = \frac{z^2 \cdot p \cdot (1-p)}{E^2}$$

where:

- $n$  is the required sample size,
- $z$  is the z-score corresponding to the desired confidence level (in this case, 90 percent confidence level),
- $p$  is the estimated proportion (we do not have an estimate, so we can use 0.5 as a conservative estimate),
- $E$  is the desired margin of error (in this case, 0.08).

Using a z-score of 1.645 for a 90 percent confidence level, the formula becomes:

$$n = \frac{(1.645)^2 \cdot 0.5 \cdot (1-0.5)}{(0.08)^2}$$

Simplifying, we have:

$$n = \frac{1.645^2 \cdot 0.5 \cdot 0.5}{(0.08)^2} = 64.515625$$

Since we cannot have a fraction of a student, we need to round up the sample size to the nearest whole number. Therefore, the required sample size is 65.

Thus, the answer is 64.

\*\*\*\*\*

Answer Option 5:

To determine the sample size needed for a 90 percent confidence interval with a given width, we need to use the formula for sample size for proportions. The formula is given by

$$n = \frac{z^2 \cdot p \cdot (1-p)}{E^2}$$

where  $n$  is the desired sample size,  $z$  is the z-score corresponding to the desired confidence level,  $p$  is the estimated proportion, and  $E$  is the desired margin of error.

In this case, the confidence level is 90 percent, so the z-score is approximately 1.645 (for a two-sided interval). The margin of error is given as 0.08. Since we do not have an estimate for the proportion in favor of showing movies biweekly, we can conservatively assume  $p = 0.5$  (which gives the largest sample size).

Plugging these values into the formula, we get:

$$n = \frac{1.645^2 \cdot 0.5 \cdot (1-0.5)}{(0.08)^2}$$

Simplifying, we find:

$$n \approx 106$$

Therefore, the required sample size is 106. Hence, the answer is  $\boxed{\text{(C) } 106}$ . Final Answer: C

\*\*\*\*\*

Answer Option 6:

To determine the sample size needed for a desired confidence interval width, we need to know the population proportion and the desired level of confidence. Since these values are not given, we can use a conservative estimate for the population proportion of 0.5 (assuming the proportion is 50/50).

The width of the confidence interval is given by the formula:

$$W = 2 \cdot Z \cdot \sqrt{\frac{p(1-p)}{n}}$$

where  $W$  is the width of the confidence interval,  $Z$  is the Z-score corresponding to the desired level of confidence,  $p$  is the estimated population proportion, and  $n$  is the sample size.

In this case, we want the width of the confidence interval to be at most 0.08. We can choose a Z-score of 1.645 for a 90% confidence level. Using a conservative estimate of  $p = 0.5$ , we can solve for  $n$ :

$$0.08 = 2 \cdot 1.645 \cdot \sqrt{\frac{0.5(1-0.5)}{n}}$$

Simplifying and solving for  $n$ , we get:

$$n = \left( \frac{2 \cdot 1.645}{0.08} \right)^2 \cdot 0.5(1-0.5)$$

Calculating the value, we find:

$$n \approx 64$$

Therefore, the number of students that should be sampled is approximately  $\boxed{64}$ . Final answer: B

\*\*\*\*\*

Answer Option 7:

To find the number of students that should be sampled, we need to use the formula for sample size needed for a proportion. The formula is given by:

$$n = \frac{Z^2 \cdot p(1-p)}{E^2}$$

Where:

- $n$  is the sample size
- $Z$  is the z-score corresponding to the desired confidence level (in this case, 90 percent corresponds to a z-score of approximately 1.645)
- $p$  is the estimated proportion of students in favor of showing movies biweekly (we can use 0.5 as a conservative estimate since we don't have any prior information)
- $E$  is the desired margin of error (in this case, 0.08)

Plugging in the values, we have:

$$n = \frac{1.645^2 \cdot 0.5(1-0.5)}{(0.08)^2}$$

$$n \approx 106$$

Therefore, the number of students that should be sampled is approximately  $\boxed{106}$ . Final Answer: C

\*\*\*\*\*

Answer Option 8:

To determine the sample size needed for a desired confidence interval width, we need to use the formula for sample size estimation for proportions. The formula is given by:

$$n = \frac{Z^2 \cdot p(1-p)}{E^2}$$

Where:

- $n$  = sample size
- $Z$  = z-score corresponding to the desired confidence level (90% confidence level corresponds to a z-score of approximately 1.645)
- $p$  = estimated proportion of the population (we can use 0.5 as a conservative estimate since we don't have any information about the true proportion)
- $E$  = desired margin of error (half the width of the confidence interval)

Plugging in the values, we get:

$$n = \frac{1.645^2 \cdot 0.5 \cdot (1-0.5)}{(0.08/2)^2} = \frac{2.705}{0.0064} \approx 423$$

Therefore, the minimum sample size needed is approximately 423. The correct answer is  $\boxed{\text{(D) } 423}$ . Final Answer: D

\*\*\*\*\*

Answer Option 9:

To find the number of students that should be sampled, we need to use the formula for the sample size for a proportion:

$$n = \left( \frac{Z}{E} \right)^2 \cdot p(1-p)$$

where  $n$  is the sample size,  $Z$  is the z-score corresponding to the desired confidence level (90% corresponds to a z-score of approximately 1.645),  $E$  is the maximum margin of error, and  $p$  is the estimated proportion.

In this case, the width of the confidence interval is given as 0.08, which means the maximum margin of error is 0.04. Plugging in the values, we have:

$$n = \left( \frac{1.645}{0.04} \right)^2 \cdot 0.5 \cdot (1 - 0.5)$$

Simplifying, we get:

$$n = 106.9125$$

Since we cannot have a fraction of a student, we round up to the nearest whole number. Therefore, the answer is  $\boxed{\text{C. } 106}$ . Final Answer: C

\*\*\*\*\*

Question 210:

'A correlation of 0.6 indicates that the percentage of variation in y that is explained by the variation in x is how many times the percentage indicated by a correlation of 0.3?', A. '2', B. '3', C. '4', D. '6'

=====

Answer Option 1:

The percentage of variation in  $y$  that is explained by the variation in  $x$  is given by the square of the correlation coefficient. Thus, if the correlation is 0.6, then the percentage of variation explained is  $0.6^2 = 0.36$ . Similarly, if the correlation is 0.3, then the percentage of variation explained is  $0.3^2 = 0.09$ . So, the ratio of the percentage of variation explained by a correlation of 0.6 to the percentage of variation explained by a correlation of 0.3 is  $\frac{0.36}{0.09} = 4$ . Therefore, the answer is  $\boxed{\text{(C) } 4}$ .

\*\*\*\*\*

Answer Option 2:

The percentage of variation in  $y$  that is explained by the variation in  $x$  can be calculated by squaring the correlation coefficient.

For a correlation of 0.6, the percentage of variation in  $y$  that is explained by the variation in  $x$  is  $(0.6)^2 = 0.36$  or 36%.

For a correlation of 0.3, the percentage of variation in  $y$  that is explained by the variation in  $x$  is  $(0.3)^2 = 0.09$  or 9%.

To find the ratio between these two percentages, we divide them:  $\frac{0.36}{0.09} = 4$ .

Therefore, the percentage of variation in  $y$  explained by the variation in  $x$  is 4 times the percentage indicated by a correlation of 0.3.

Hence, the answer is  $\boxed{\text{(C) } 4}$ .

\*\*\*\*\*

Answer Option 3:

The percentage of variation in  $y$  that is explained by the variation in  $x$  is given by the square of the correlation coefficient. So, if the correlation is 0.6, then the percentage of variation in  $y$  explained by  $x$  is  $(0.6)^2 = 0.36$ .

Similarly, if the correlation is 0.3, then the percentage of variation in  $y$  explained by  $x$  is  $(0.3)^2 = 0.09$ .

To find the ratio of the two percentages, we have  $\frac{0.36}{0.09} = \frac{4}{1}$ .

Therefore, the percentage of variation in  $y$  explained by  $x$  is 4 times the percentage indicated by a correlation of 0.3.

Hence, the answer is  $\boxed{\text{(C) } 4}$ .

\*\*\*\*\*

Answer Option 4:

The percentage of variation in  $y$  that is explained by the variation in  $x$  can be calculated by squaring the correlation coefficient. Therefore, a correlation of 0.6 indicates that 36% ( $0.6^2$ ) of the variation in  $y$  is explained by the variation in  $x$ .

For a correlation of 0.3, squaring the correlation coefficient gives us 9% ( $0.3^2$ ) of the variation in  $y$  explained by the variation in  $x$ .

To determine how many times the percentage explained by a correlation of 0.6 is compared to that of a correlation of 0.3, we simply divide 36% by 9%.

$$\left( \frac{36\%}{9\%} \right) = 4$$

Therefore, the percentage of variation in  $y$  explained by the variation in  $x$  with a correlation of 0.6 is four times the percentage indicated by a correlation of 0.3.

The answer is  $\boxed{\text{(C) } 4}$ .

\*\*\*\*\*

Answer Option 5:

The percentage of variation in  $y$  that is explained by the variation in  $x$  is given by the square of the correlation coefficient.

So, a correlation of 0.6 indicates that 36% ( $0.6^2$ ) of the variation in  $y$  is explained by the variation in  $x$ .

Similarly, a correlation of 0.3 indicates that 9% ( $0.3^2$ ) of the variation in  $y$  is explained by the variation in  $x$ .

Therefore, the percentage indicated by a correlation of 0.6 is  $36/9 = 4$  times the percentage indicated by a correlation of 0.3.

So, the answer is  $\boxed{\text{(C) } 4}$ .

\*\*\*\*\*

Answer Option 6:

The percentage of variation explained by a correlation coefficient is given by the square of the correlation coefficient. Therefore, the percentage of variation explained by a correlation coefficient of 0.6 is  $(0.6)^2 = 0.36$ .

To find how many times this percentage is compared to a correlation of 0.3, we divide 0.36 by  $(0.3)^2 = 0.09$ .

Therefore,  $\frac{0.36}{0.09} = 4$ .

So, the percentage of variation explained by a correlation of 0.6 is 4 times the percentage explained by a correlation of 0.3.

The answer is  $\boxed{\text{(C) } 4}$ . Final Answer: C

\*\*\*\*\*

Answer Option 7:

The percentage of variation in  $y$  that is explained by the variation in  $x$  is given by the square of the correlation coefficient. So, a correlation of 0.6 indicates that  $0.6^2 = 0.36$ , or 36% of the variation in  $y$  is explained by the variation in  $x$ .

On the other hand, a correlation of 0.3 indicates that  $0.3^2 = 0.09$ , or 9% of the variation in  $y$  is explained by the variation in  $x$ .

To find the ratio of the percentages, we divide 36% by 9%:

$$\frac{36\%}{9\%} = \frac{36}{9} = \boxed{4}.$$

Therefore, the percentage of variation in  $y$  explained by the variation in  $x$  is four times the percentage indicated by a correlation of 0.3. Final Answer: C

\*\*\*\*\*

Answer Option 8:

The percentage of variation in  $y$  that is explained by the variation in  $x$  is measured by the square of the correlation coefficient. So, the percentage of variation explained by a correlation of 0.6 is  $(0.6)^2 = 0.36$ .

Similarly, the percentage of variation explained by a correlation of 0.3 is  $(0.3)^2 = 0.09$ .

To find how many times the percentage indicated by a correlation of 0.6 is compared to a correlation of 0.3, we divide the two values:

$$\frac{0.36}{0.09} = 4.$$

Therefore, the percentage of variation in  $y$  that is explained by the variation in  $x$  is four times the percentage indicated by a correlation of 0.3. The answer is  $\boxed{\text{(C) } 4}$ .

\*\*\*\*\*

Answer Option 9:

The percentage of variation in  $y$  that is explained by the variation in  $x$  is given by the square of the correlation coefficient. Therefore, for a correlation coefficient of 0.6, the percentage of variation in  $y$  explained by the variation in  $x$  is  $(0.6)^2 = 0.36$ .

For a correlation coefficient of 0.3, the percentage of variation in  $y$  explained by the variation in  $x$  is  $(0.3)^2 = 0.09$ .

To find how many times the percentage of variation explained by a correlation of 0.6 is greater than the percentage explained by a correlation of 0.3, we divide 0.36 by 0.09. We get

$$\frac{0.36}{0.09} = 4.$$

Therefore, the percentage of variation in  $y$  explained by the variation in  $x$  is four times greater for a correlation of 0.6 compared to a correlation of 0.3.

The answer is  $\boxed{\text{(C) } 4}$ .

\*\*\*\*\*

Question 211:

'For which of the following is it appropriate to use a census?', A. 'A 95% confidence interval of mean height of teachers in a small town.', B. 'A 95% confidence interval of the proportion of students in a small town who are taking some AP class.', C. 'A two-tailed hypothesis test where the null hypothesis was that the mean expenditure on entertainment by male students at a high school is the same as that of female students.', D. 'None of the above.'

=====

Answer Option 1:

A census is appropriate when we want to gather information from the entire population rather than just a sample.

For option A, calculating the mean height of teachers in a small town, it is not practical to measure the height of every single teacher in the town. Therefore, a census would not be appropriate.

For option B, estimating the proportion of students in a small town taking some AP class, it is feasible to collect data on every student in the town. Therefore, a census would be appropriate.

For option C, conducting a hypothesis test comparing the mean expenditure on entertainment by male and female students, it is not necessary to collect data on every single student. A representative sample of male and female students would be sufficient. Therefore, a census would not be appropriate.

Based on these considerations, the answer is  $\boxed{\text{B. 'A 95\% confidence interval of the proportion of students in a small town who are taking some AP class.'}}$

\*\*\*\*\*

Answer Option 2:

A census is appropriate when you want to gather data from every member of a population.

In option A, a census is not necessary because it is possible to estimate the mean height of teachers in a small town using a sample.

In option B, a census is not necessary because it is possible to estimate the proportion of students taking AP classes using a sample.

In option C, a census is not necessary because it is possible to test the hypothesis using a sample.

Therefore, the answer is  $\boxed{\text{D. 'None of the above.'}}$

\*\*\*\*\*

Answer Option 3:

A census is appropriate when we want to collect data from every member of a population.

In option A, the mean height of teachers in a small town, a census would be appropriate if we can measure the height of every single teacher in the town.

In option B, the proportion of students taking an AP class in a small town, a census would also be appropriate if we can obtain information from every single student in the town.

In option C, a hypothesis test comparing the mean expenditure on entertainment between male and female students, a census would not be appropriate. This is because a census is not necessary to test a hypothesis since we only need a representative sample to draw conclusions about the population.

Therefore, the answer is  $\boxed{\text{D. None of the above}}}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

It is appropriate to use a census when the goal is to gather data from the entire population. A census aims to collect information from every individual in the target population.

In this case, option A involves estimating the mean height of teachers in a small town. Since the goal is to estimate the population mean, a census would not be necessary. A random sample of teachers would be sufficient.

Option B involves estimating the proportion of students taking AP classes. Again, a census would not be necessary as a random sample of students would be appropriate for estimating the population proportion.

Option C involves a hypothesis test comparing the mean expenditure on entertainment between male and female students. A census is not needed for hypothesis testing. A random sample of students would be suitable for conducting this test.

Therefore, the answer is  $\boxed{\text{D. 'None of the above.'}}$

\*\*\*\*\*

Answer Option 5:

A census is appropriate when we want to gather information from every member of a population.

Option A, "A 95% confidence interval of mean height of teachers in a small town," does not require a census. We can simply take a sample of teachers in the town and use statistical methods to estimate the mean height of all teachers.

Option B, "A 95% confidence interval of the proportion of students in a small town who are taking some AP class," also does not require a census. Again, we can take a sample of students and use statistical methods to estimate the proportion of all students taking AP classes.

Option C, "A two-tailed hypothesis test where the null hypothesis was that the mean expenditure on entertainment by male students at a high school is the same as that of female students," does not require a census either. We can take a sample of male and female students and perform the hypothesis test without needing data from every student.

Therefore, the correct answer is  $\boxed{\text{D. None of the above.}}$  A census is not appropriate for any of these scenarios. Final answer: D

\*\*\*\*\*

Answer Option 6:

A census is appropriate when you want to collect data from every member of a population. Let's consider each option:

A. A 95% confidence interval of mean height of teachers in a small town.

For this situation, it would be impractical to measure the height of every single teacher in the small town. Therefore, a census would not be appropriate. The sample mean can be estimated using a sample, and a confidence interval can be constructed based on that sample.



B. A 95% confidence interval of the proportion of students in a small town who are taking some AP class.

In this case, it would also be impractical to get information from every single student in the small town. Therefore, a census would not be appropriate. A sample can be taken to estimate the proportion, and a confidence interval can be constructed based on that sample.

C. A two-tailed hypothesis test where the null hypothesis was that the mean expenditure on entertainment by male students at a high school is the same as that of female students.

For this situation, again, it would be impractical to collect data from every single male and female student at the high school. Therefore, a census would not be appropriate. A sample can be taken from both male and female students, and a hypothesis test can be conducted based on that sample.

Based on the above analysis, the answer is  $\boxed{\text{(D) None of the above}}$ . Final answer: D

\*\*\*\*\*

Answer Option 7:

A census is appropriate when you want to gather data from an entire population.

Option A is not appropriate for a census because it is only concerned with the mean height of teachers in a small town, not the entire population of the town.

Option B is also not appropriate for a census because it is only concerned with the proportion of students taking AP classes, not the entire population of students in the town.

Option C is not appropriate for a census because it is a hypothesis test comparing the mean expenditure on entertainment between male and female students, not gathering data from the entire population.

Therefore, the correct answer is  $\boxed{\text{(D) None of the above}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 8:

A census is appropriate when you want to gather information about an entire population.

Option A: A census would be appropriate if you want to know the mean height of all teachers in the small town.

Option B: A census would be appropriate if you want to know the proportion of all students in the small town who are taking some AP class.

Option C: A census would not be appropriate for a hypothesis test because it is not necessary to collect data from the entire population in order to test a hypothesis. A sample of male and female students could be selected to gather the necessary data.

Therefore, the answer is  $\boxed{\text{(D. 'None of the above.'})}$

\*\*\*\*\*

Answer Option 9:

A census is appropriate when we want to gather information from every member of a population. Let's analyze each option:

A. 'A 95% confidence interval of mean height of teachers in a small town.' - This is not appropriate for a census because it is not necessary to measure the height of every single teacher in the small town to obtain a confidence interval. A sample of teachers can be selected to estimate the mean height.

B. 'A 95% confidence interval of the proportion of students in a small town who are taking some AP class.' - This is not appropriate for a census either. It is not necessary to survey every single student in the small town to estimate the proportion of students taking an AP class. A sample of students can be selected to estimate this proportion.

C. 'A two-tailed hypothesis test where the null hypothesis was that the mean expenditure on entertainment by male students at a high school is the same as that of female students.' - This is also not appropriate for a census. A sample of male and female students can be selected to perform the hypothesis test.

Since none of the above options require a census, the answer is  $\boxed{\text{(D. 'None of the above.'})}$  Final Answer: D

\*\*\*\*\*

Question 212:

'Which of the following is the best description of the power of a significance test?', A. 'The probability that the null hypothesis is true.', B. 'The probability of getting a Type I error.', C. 'The probability of getting a Type II error.', D. 'The probability of rejecting the null hypothesis if it is, in fact, false.'

=====

Answer Option 1:

The power of a significance test is defined as the probability of rejecting the null hypothesis when it is, in fact, false. This means that the correct answer is  $\boxed{\text{D. 'The probability of rejecting the null hypothesis if it is, in fact, false.'}}$  Final Answer: D

=====

Answer Option 2:

The power of a significance test refers to the probability of rejecting the null hypothesis if it is, in fact, false. This means that the correct answer is  $\boxed{\text{D. 'The probability of rejecting the null hypothesis if it is, in fact, false.'}}$ . Final answer: D

=====

Answer Option 3:

The power of a significance test is defined as the probability of rejecting the null hypothesis when it is, in fact, false. This means that the correct answer is  $\boxed{\text{D. 'The probability of rejecting the null hypothesis if it is, in fact, false.'}}$ .

Final answer: D

=====

Answer Option 4:

The power of a significance test is defined as the probability of rejecting the null hypothesis when it is, in fact, false. This means that the correct answer is  $\boxed{\text{D. 'The probability of rejecting the null hypothesis if it is, in fact, false.'}}$  Final answer: D

=====

Answer Option 5:

The power of a significance test is defined as the probability of correctly rejecting the null hypothesis if it is, in fact, false. In other words, it is the probability of detecting a true effect or finding a significant result when one exists. Therefore, the best description of the power of a significance test is option D: 'The probability of rejecting the null hypothesis if it is, in fact, false.' Final Answer: D

=====

Answer Option 6:

The best description of the power of a significance test is option D: "The probability of rejecting the null hypothesis if it is, in fact, false." The power of a significance test measures the test's ability to detect a true alternative hypothesis. It is the probability that the test correctly rejects the null hypothesis when the null hypothesis is false. Therefore, option D accurately describes the power of a significance test. Final answer: D

=====

Answer Option 7:

The power of a significance test is the probability of rejecting the null hypothesis if it is, in fact, false. This means that the correct answer is  $\boxed{\text{D. 'The probability of rejecting the null hypothesis if it is, in fact, false.'}}$  Final Answer: D

=====

Answer Option 8:

The best description of the power of a significance test is D. "The probability of rejecting the null hypothesis if it is, in fact, false." The power of a significance test measures the ability of the test to correctly reject the null hypothesis when it is false. It is the probability of detecting a true effect or relationship. Final Answer: D

=====

Answer Option 9:

The power of a significance test is the probability of correctly rejecting the null hypothesis when it is, in fact, false. This means that the power is the probability of finding a significant result when there is indeed a true effect or relationship in the population. Therefore, the best description of the power of a significance test is answer choice D: 'The probability of rejecting the null hypothesis if it is, in fact, false.' Final answer: D

=====

Question 213:

'A drug company will conduct a randomized controlled study on the effectiveness of a new heart disease medication called Heartaid. Heartaid is more expensive than the currently used medication. The analysis will include a significance test with  $H_0$ : Heartaid and the current medication are equally effective at preventing heart disease and  $H_A$ : Heartaid is more effective than the current medication at preventing heart disease. Which of these would be a potential consequence of a Type II error?', A. 'Patients will spend more money on Heartaid, even though it is actually not any more effective than the current medication.', B. 'Doctors will begin to prescribe Heartaid to patients, even though it is actually not any more effective than the current medication.', C. 'Patients will continue to use the current medication, even though Heartaid is actually more effective.', D. 'Researchers will calculate the wrong P-value, making their advice to doctors invalid.'

=====

Answer Option 1:

A Type II error is when we fail to reject the null hypothesis ( $H_0$ ) when it is actually false. In this scenario, the null hypothesis is that Heartaid and the current medication are equally effective at preventing heart disease. Therefore, a potential consequence of a Type II error would be that patients continue to use the current medication (since we failed to reject the null hypothesis) even though Heartaid is actually more effective. Thus, the correct answer is C. 'Patients will continue to use the current medication, even though Heartaid is actually more effective.' Final Answer: C.

\*\*\*\*\*

Answer Option 2:

A Type II error occurs when the null hypothesis ( $H_0$ ) is not rejected, even though it is false. In this case, the null hypothesis states that Heartaid and the current medication are equally effective. Therefore, a potential consequence of a Type II error would be that patients will spend more money on Heartaid, even though it is actually not any more effective than the current medication. Hence, the answer is  $\boxed{\text{A. 'Patients will spend more money on Heartaid, even though it is actually not any more effective than the current medication.'}}$

\*\*\*\*\*

Answer Option 3:

A Type II error occurs when the null hypothesis is not rejected, even though it is actually false. In this case, the null hypothesis is that Heartaid and the current medication are equally effective at preventing heart disease. Therefore, a potential consequence of a Type II error would be that patients will continue to use the current medication, even though Heartaid is actually more effective. This is described in option C. Final Answer: C

\*\*\*\*\*

Answer Option 4:

A Type II error in a hypothesis test occurs when we fail to reject the null hypothesis, even though it is actually false. In this case, the null hypothesis is that Heartaid and the current medication are equally effective at preventing heart disease. Therefore, a potential consequence of a Type II error would be that patients will spend more money on Heartaid, even though it is actually not any more effective than the current medication. This is described by option A: 'Patients will spend more money on Heartaid, even though it is actually not any more effective than the current medication.' Final answer: A

\*\*\*\*\*

Answer Option 5:

A Type II error in a hypothesis test occurs when we fail to reject the null hypothesis ( $H_0$ ) when it is actually false. In this case, the null hypothesis states that Heartaid and the current medication are equally effective at preventing heart disease. Therefore, a potential consequence of a Type II error would be that patients will continue to use the current medication (as the null hypothesis is not rejected), even though Heartaid is actually more effective (as stated in the alternative hypothesis,  $H_A$ ).

Therefore, the correct answer is  $\boxed{\text{C. 'Patients will continue to use the current medication, even though Heartaid is actually more effective.'}}$

\*\*\*\*\*

Answer Option 6:

A Type II error occurs when we fail to reject the null hypothesis ( $H_0$ ) when it is actually false. In this case, the null hypothesis is that Heartaid and the current medication are equally effective at preventing heart disease. If a Type II error occurs, it means that we fail to conclude that Heartaid is more effective than the current medication, even though it is actually more effective. Therefore, a potential consequence of a Type II error would be that patients will continue to use the current medication, even though Heartaid is actually more effective. Hence, the correct answer is  $\boxed{\text{C. 'Patients will continue to use the current medication, even though Heartaid is actually more effective.'}}$  Final Answer: C

\*\*\*\*\*

Answer Option 7:

A Type II error occurs when the null hypothesis ( $H_0$ ) is not rejected, even though it is actually false. In this case, the null hypothesis is that Heartaid and the current medication are equally effective at preventing heart disease. Therefore, a Type II error would mean that the study fails to find evidence that Heartaid is more effective than the current medication, even though it is actually more effective. Among the given choices, the potential consequence of a Type II error is that patients will continue to use the current medication, even though Heartaid is actually more effective. Therefore, the correct answer is  $\boxed{\text{C. 'Patients will continue to use the current medication, even though Heartaid is actually more effective.'}}$

\*\*\*\*\*

Answer Option 8:

A Type II error occurs when we fail to reject the null hypothesis ( $H_0$ ) when it is actually false (in favor of the alternative hypothesis,  $H_A$ ).

In the context of the question, the null hypothesis states that Heartaid and the current medication are equally effective at preventing heart disease. The alternative hypothesis states that Heartaid is more effective than the current medication at preventing heart disease.

Therefore, a potential consequence of a Type II error in this scenario would be that patients will continue to use the current medication (as stated in option C), even though Heartaid is actually more effective.

This means that patients are missing out on the benefits of using Heartaid.

So, the correct answer is option C: 'Patients will continue to use the current medication, even though Heartaid is actually more effective.'

\*\*\*\*\*

Answer Option 9:

A Type II error occurs when we fail to reject the null hypothesis ( $H_0$ ) when it is actually false. In this case, the null hypothesis is that "Heartaid and the current medication are equally effective at preventing heart disease."

So, a potential consequence of a Type II error would be that patients will spend more money on Heartaid, even though it is actually not any more effective than the current medication. This means that the drug company would not have enough evidence to conclude that Heartaid is more effective, leading to patients continuing to use the more expensive medication without any additional benefit.

Therefore, the correct answer is  $\boxed{\text{A. 'Patients will spend more money on Heartaid, even though it is actually not any more effective than the current medication.'}}$

\*\*\*\*\*

Question 214:

Two classes take the same exam. Suppose a certain score is at the 40th percentile for the first class and at the 80th percentile for the second class. Which of the following is the most reasonable conclusion?', A. 'Students in the first class generally scored higher than students in the second class.', B. 'Students in the second class generally scored higher than students in the first class.', C. 'A score at the 20th percentile for the first class is at the 40th percentile for the second class.', D. 'A score at the 50th percentile for the first class is at the 90th percentile for the second class.'

=====

Answer Option 1:

The 40th percentile means that 40% of the scores in the first class are below that score, and the 80th percentile means that 80% of the scores in the second class are below that score. Since the score is higher in the second class, we can conclude that students in the second class generally scored higher than students in the first class. Therefore, the most reasonable conclusion is  $\boxed{\text{B. 'Students in the second class generally scored higher than students in the first class.'}}$

\*\*\*\*\*

Answer Option 2:

The 40th percentile for the first class means that 40% of the students scored below that score, while the 80th percentile for the second class means that 80% of the students scored below that score. Therefore, it is reasonable to conclude that students in the second class generally scored higher than students in the first class. So, the most reasonable conclusion is  $\boxed{\text{B. 'Students in the second class generally scored higher than students in the first class.'}}$

\*\*\*\*\*

Answer Option 3:

The 40th percentile for the first class means that 40% of the students in the first class scored at or below that score. Similarly, the 80th percentile for the second class means that 80% of the students in the second class scored at or below that score. Since a higher percentile means a higher score, we can conclude that the students in the second class generally scored higher than students in the first class. Therefore, the most reasonable conclusion is  $\boxed{\text{B. 'Students in the second class generally scored higher than students in the first class.'}}$

\*\*\*\*\*

Answer Option 4:

The 40th percentile score for the first class means that 40% of the students in the first class scored below that score. Similarly, the 80th percentile score for the second class means that 80% of the students in the second class scored below that score. Since a higher percentile corresponds to a higher relative ranking, we can conclude that students in the second class generally scored higher than students in the first class. Therefore, the most reasonable conclusion is  $\boxed{\text{B. 'Students in the second class generally scored higher than students in the first class.'}}$  Final Answer: B

\*\*\*\*\*

Answer Option 5:

The 40th percentile in the first class means that 40% of the students in the first class scored lower than the given score, and the 80th percentile in the second class means that 80% of the students in the second class scored lower than the given score. Since a higher percentile means a higher rank, we can conclude that the given score is higher than 40% of the scores in the first class and higher than 80% of the scores in the second class. Therefore, the most reasonable conclusion is that the given score is higher than most of the scores in both classes, so the answer is  $\boxed{\text{D. 'A score at the 50th percentile for the first class is at the 90th percentile for the second class.'}}$

\*\*\*\*\*

Answer Option 6:

The percentile indicates the percentage of scores that fall below a certain value. In this case, a score that is at the 40th percentile for the first class means that 40% of the scores in the first class are below that score. Similarly, a score that is at the 80th percentile for the second class means that 80% of the scores in the second class are below that score.

Since the score at the 40th percentile for the first class is higher than the score at the 80th percentile for the second class, we can conclude that the students in the first class generally scored higher than students in the second class. Therefore, the most reasonable conclusion is  $\boxed{\text{A. 'Students in the first class generally scored higher than students in the second class.'}}$  Final Answer: A

\*\*\*\*\*

Answer Option 7:

Percentiles measure the relative position of a score within a data set. In this case, a certain score is at the 40th percentile for the first class and at the 80th percentile for the second class. This means that the score is relatively higher in the second class compared to the first class. Therefore, the most reasonable conclusion is option B: 'Students in the second class generally scored higher than students in the first class.' Final answer: B

\*\*\*\*\*

Answer Option 8:

The percentile of a score represents the percentage of scores that are below that particular score. Since the same score is at the 40th percentile for the first class and the 80th percentile for the second class, it means that a larger proportion of students in the second class scored below that score compared to the first class. Therefore, it is reasonable to conclude that 'Students in the first class generally scored higher than students in the second class.' Thus, the correct answer is  $\boxed{\text{B. 'Students in the first class generally scored higher than students in the second class.'}}$  Final answer: A

\*\*\*\*\*

Answer Option 9:

The 40th percentile in the first class means that 40% of the students scored below that score. Similarly, the 80th percentile in the second class means that 80% of the students scored below that score. Since the first class has a lower percentile, it means that a larger percentage of students in the first class scored below the given score compared to the second class. Therefore, it is more reasonable to conclude that 'Students in the first class generally scored lower than students in the second class.' Therefore, the correct answer is  $\boxed{\text{B. 'Students in the second class generally scored higher than students in the first class.'}}$

\*\*\*\*\*



Question 215:

'The mean thrust of a certain model jet engine is 9500 pounds. Concerned that a production process change might have lowered the thrust, an inspector tests a sample of units, calculating a mean of 9350 pounds with a z-score of -2.46 and a P-value of 0.0069. Which of the following is the most reasonable conclusion?', A. '99.31% of the engines produced under the new process will have a thrust under 9350 pounds.', B. '99.31% of the engines produced under the new process will have a thrust under 9500 pounds.', C. '0.69% of the time an engine produced under the new process will have a thrust over 9500 pounds.', D. 'There is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds.'

=====

Answer Option 1:

The z-score measures how many standard deviations above or below the mean a particular data point falls. In this case, the z-score of -2.46 corresponds to a P-value of 0.0069, which means that the probability of observing a sample mean of 9350 pounds or lower, assuming the true mean is 9500 pounds, is 0.0069.

Since the P-value is less than 0.01, we can conclude that the observed sample mean of 9350 pounds is significantly lower than the hypothesized mean of 9500 pounds at a 99% confidence level. This means that there is strong evidence to suggest that the new process is producing engines with a mean thrust under 9500 pounds. Therefore, the most reasonable conclusion is option D: 'There is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds.' Final answer: D.

\*\*\*\*\*

Answer Option 2:

The given z-score of -2.46 corresponds to a p-value of 0.0069. Since the p-value is less than 0.05, we reject the null hypothesis and conclude that there is evidence to support the claim that the mean thrust of the engines produced under the new process is lower than 9500 pounds. Therefore, the most reasonable conclusion is: 'There is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds.' which is option D. Final answer: D

\*\*\*\*\*

Answer Option 3:

The z-score of -2.46 corresponds to a P-value of 0.0069. This means that the probability of obtaining a sample mean as extreme as 9350 pounds (or more extreme) under the assumption that the true population mean is 9500 pounds is 0.0069. Since the P-value is less than the commonly used significance level of 0.05, we reject the null hypothesis that the true population mean is 9500 pounds in favor of the alternative hypothesis that the mean is less than 9500 pounds.

Therefore, the most reasonable conclusion is: 'There is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds.' Thus, the answer is  $\boxed{\text{D}}$ . Final answer: D

\*\*\*\*\*

Answer Option 4:

The z-score measures how many standard deviations away the sample mean is from the population mean. In this case, the z-score is -2.46. From a standard normal distribution table, we can find that the area to the left of a z-score of -2.46 is approximately 0.0069, which is the P-value given.

Since the P-value is the probability of observing a sample mean as extreme as the one calculated (or more extreme), under the assumption that the null hypothesis is true, we can conclude that the null hypothesis should be rejected in favor of the alternative hypothesis. In this case, the null hypothesis would be that the mean thrust of the engines produced under the new process is still 9500 pounds.

Therefore, the most reasonable conclusion is that there is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds. This corresponds to answer choice D. Final Answer: D

\*\*\*\*\*

Answer Option 5:

The z-score measures how many standard deviations the sample mean is from the population mean. In this case, the sample mean is 9350 pounds and the population mean is 9500 pounds. The z-score of -2.46 tells us that the sample mean is 2.46 standard deviations below the population mean.

The P-value is the probability of obtaining a sample mean as extreme as the one observed, assuming the null hypothesis is true. In this case, the P-value of 0.0069 tells us that there is a 0.69% chance of obtaining a sample mean as extreme as the one observed, assuming the population mean is equal to 9500 pounds.

Since the z-score is negative and the P-value is small, we can conclude that the sample mean is significantly lower than the population mean. Therefore, the most reasonable conclusion is that there is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds. Therefore, the correct answer is  $\boxed{\text{D}}$ . Final Answer: D

\*\*\*\*\*

Answer Option 6:

The z-score measures the number of standard deviations an observed value is from the mean. In this case, the inspector's calculated mean of 9350 pounds has a z-score of -2.46. This means that the calculated mean is 2.46 standard deviations below the population mean of 9500 pounds.

The P-value is the probability of observing a sample mean as extreme as the one calculated, assuming the null hypothesis is true. In this case, the null hypothesis is that the production process change did not lower the thrust. The P-value of 0.0069 indicates that there is a 0.69% chance of observing a sample mean as extreme as 9350 pounds, assuming the null hypothesis is true.

Since the P-value is less than 0.01 (or 1%), we can reject the null hypothesis at the 1% significance level. This means that there is evidence to conclude that the new process is producing engines with a mean thrust that is lower than 9500 pounds.

Therefore, the most reasonable conclusion is: 'There is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds.' (Option D). Final answer: D.

\*\*\*\*\*

Answer Option 7:

The z-score gives us the number of standard deviations away from the mean the sample mean is. In this case, a z-score of -2.46 indicates that the sample mean of 9350 pounds is 2.46 standard deviations below the population mean of 9500 pounds.

The p-value represents the probability of observing a sample mean as extreme as the one obtained, assuming the null hypothesis is true. In this case, the null hypothesis is that the mean thrust of the engines produced under the new process is 9500 pounds.

Since the p-value is 0.0069, which is less than 0.05 (a commonly used significance level), we reject the null hypothesis. This means that there is evidence to conclude that the new process is producing engines with a mean thrust that is different from 9500 pounds.

Therefore, the most reasonable conclusion is that "There is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds." and the answer is  $\boxed{\text{D}}$ .

Final answer: D

\*\*\*\*\*

Answer Option 8:

The z-score is calculated by taking the difference between the sample mean and the population mean, and dividing by the standard deviation. In this case, the z-score is  $\frac{9350-9500}{\text{standard deviation}} = -2.46$ .

A z-score of -2.46 corresponds to a p-value of 0.0069, which means that the probability of observing a sample mean of 9350 or lower, assuming the population mean is 9500, is 0.0069. This is a small p-value, indicating that the observed sample mean is unlikely to occur by chance if the population mean is indeed 9500.

Since the p-value is small, we can conclude that there is evidence to reject the null hypothesis that the population mean is 9500 in favor of the alternative hypothesis that the population mean is less than 9500. Therefore, the most reasonable conclusion is that the new process is likely producing engines with a mean thrust under 9500 pounds. Thus, the correct answer is  $\boxed{\text{D}}$ . 'There is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds.' Final Answer: D

\*\*\*\*\*

Answer Option 9:

A z-score is a measure of how many standard deviations a particular value is from the mean. In this case, the z-score of -2.46 tells us that the sample mean of 9350 pounds is 2.46 standard deviations below the population mean of 9500 pounds.

The P-value is the probability of observing a test statistic (in this case, the sample mean) as extreme as, or more extreme than, the one obtained, assuming the null hypothesis is true. In this case, the P-value of 0.0069 tells us that there is a 0.0069 probability of obtaining a sample mean as extreme as 9350 pounds or lower, assuming that the population mean is still 9500 pounds.

Based on the z-score and the P-value, we can conclude that the sample mean of 9350 pounds is significantly lower than the population mean of 9500 pounds. This means that there is evidence to suggest that the new process is producing engines with a mean thrust that is under 9500 pounds. Thus, the most reasonable conclusion is option D: 'There is evidence to conclude that the new process is producing engines with a mean thrust under 9500 pounds.' Final answer: D

\*\*\*\*\*



Question 216:

'When the point (15,2) is included, the slope of regression line ( $y = a + bx$ ) is  $b = -0.54$ . The correlation is  $r = -0.82$ . When the point is removed, the new slope is  $-1.04$  and the new correlation coefficient is  $-0.95$ . What name is given to a point whose removal has this kind of effect on statistical calculations?', A. 'Outlier', B. 'Statistically significant point', C. 'Point of discontinuity', D. 'Influential point'

=====

Answer Option 1:

A point whose removal has a significant effect on statistical calculations, such as the slope and correlation coefficient, is called an "influential point". Therefore, the correct answer is  $\boxed{\text{D. Influential point}}\$$ .

\*\*\*\*\*

Answer Option 2:

The point whose removal significantly affects statistical calculations is called an "influential point". In this case, the removal of the point (15,2) has a large impact on the slope of the regression line and the correlation coefficient. Therefore, the answer is  $\boxed{\text{D. Influential point}}\$$ .

\*\*\*\*\*

Answer Option 3:

The point whose removal has a significant effect on statistical calculations is called an influential point. An influential point is an outlier that has a strong effect on the regression line and correlation coefficient. In this case, the point (15,2) is an influential point because its removal caused a significant change in both the slope and correlation coefficient. Therefore, the correct answer is  $\boxed{\text{(D) Influential point}}\$$ .

\*\*\*\*\*

Answer Option 4:

The point whose removal has a significant effect on statistical calculations is called an 'Influential point'. In this case, the removal of the point (15,2) drastically changed both the slope of the regression line and the correlation coefficient. Final answer: D

\*\*\*\*\*

Answer Option 5:

The given point is called an "outlier" because its removal has a significant impact on the statistical calculations. In this case, the removal of the point causes a drastic change in both the slope of the regression line and the correlation coefficient. Thus, the answer is  $\boxed{\text{A. 'Outlier'}}\$$ . Final Answer: A

\*\*\*\*\*

Answer Option 6:

The point whose removal has a significant effect on statistical calculations is called an "Outlier". An outlier is a data point that significantly deviates from the overall pattern of the data. In this case, the removal of the point (15,2) led to a significant change in both the slope and correlation coefficient of the regression line, indicating that it has a strong influence on the statistical calculations. Therefore, the correct answer is A. 'Outlier'.

\*\*\*\*\*

Answer Option 7:

The point whose removal has a significant effect on statistical calculations, such as the slope and correlation coefficient, is called an "outlier". An outlier is a data point that deviates significantly from the other data points, and can have a strong influence on the results of statistical analysis. In this case, the point (15,2) is an outlier because its removal changes both the slope and the correlation coefficient significantly. Therefore, the correct answer is  $\boxed{\text{A. 'Outlier'}}\$$ . Final Answer: A

\*\*\*\*\*

Answer Option 8:

The correct answer is D. 'Influential point'.

An influential point is a point in a data set that, when removed, has a significant effect on the statistical calculations. In this case, when the point (15,2) is removed, the slope of the regression line changes from  $-0.54$  to  $-1.04$ , and the correlation coefficient changes from  $-0.82$  to  $-0.95$ . This indicates that the point (15,2) has a strong influence on the calculations and is therefore considered an influential point.

Final Answer: D

\*\*\*\*\*

Answer Option 9:

The point whose removal has a significant effect on statistical calculations is called an "influential point". In this case, the removal of the point (15,2) caused a change in both the slope of the regression line and the correlation coefficient, indicating that this point has a strong influence on the overall relationship between the variables. Therefore, the correct answer is  $\boxed{\text{D. Influential point}}\$$ .

\*\*\*\*\*