

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:

The P-value of the hypothesis test would be represented by '2P(t > 1.54) with df = 6' since we are looking at the probability of a result as extreme as the one obtained in either direction (hence the multiplication by 2) in a t-distribution with 6 degrees of freedom (since degrees of freedom is n-1 for a sample size n=7). Final Answer: B

\*\*\*\*\*

Answer Option 2:

The P-value is calculated by comparing the test statistic to a t-distribution with degrees of freedom equal to the sample size minus 1. So the degrees of freedom should be 7-1=6. The test is two-sided, because the investigator wants to know if the true mean weight is different from 300 mg, not specifically whether it is larger or smaller. So the P-value is the probability of getting a test statistic as extreme as 1.54 in either direction. Therefore, the P-value is given by B. '2P(t > 1.54) with df = 6'. Final Answer: B

\*\*\*\*\*

Answer Option 3:

In statistics, a one-sample t-test checks whether a sample mean differs from the population mean. The number of degrees of freedom is given by the sample size minus 1. So, since we have a sample of seven tablets, we have 7 minus 1 equals 6 degrees of freedom. The hypothesis test checks if the weight is less than or more than 300mg. This is a two-tailed test because we are testing the possibility of the relationship in both directions, so we want the P-value for both sides of the distribution. Therefore, the P-value of this test is given by '2P(t > 1.54) with df = 6'. Final Answer: B.

\*\*\*\*\*

Answer Option 4:

The P-value of a hypothesis test for a mean is calculated using a t-distribution. The degrees of freedom (df) for the t-distribution used in this scenario is n-1, where n is the sample size (7 in this case). So the degrees of freedom is 7-1=6. The P-value is the probability of observing a t-statistic as extreme as the one calculated from the sample data, given that the null hypothesis is true. This is a two-tailed test because the sample mean could be either greater than or less than the hypothesized population mean. So the P-value is the probability of observing a t-statistic greater than the absolute value of the calculated t-statistic or less than the negative of this value. Hence the P-value is '2P(t > 1.54) with df = 6'. Final Answer: B

\*\*\*\*\*

Answer Option 5:

The P-value is the probability, under the null hypothesis, of getting a test statistic as extreme or more extreme than what was observed. In this case, the null hypothesis is that the average weight is 300 mg. Since we're testing whether the mean is different from 300, not specifically whether it's larger or smaller, we use a two-tailed test. Our degrees of freedom are 7-1=6, because we have a sample size of 7 and we're estimating one parameter (the mean). Therefore, the P-value is '2P(t > 1.54) with df = 6'. Final answer: B.

\*\*\*\*\*

Answer Option 6:

The sample mean is  $\frac{299+300+305+302+299+301+303}{7}=429.857$ . The sample standard deviation is  $\sqrt{\frac{(299-429.857)^2+(300-429.857)^2+(305-429.857)^2+(302-429.857)^2+(299-429.857)^2+(301-429.857)^2+(303-429.857)^2}{6}}\approx 2.76887$ . The test statistic is  $t=\frac{429.857-300}{2.76887/\sqrt{7}}\approx 1.54$ . The null hypothesis is that the manufacturer's claim is correct, and the alternative hypothesis is that the manufacturer's claim is not correct, which is a two-tailed test. Therefore, the P-value is the probability that a t-distributed random variable with the appropriate degrees of freedom falls in the tail(s) corresponding to the t statistic calculated from the sample data. The degrees of freedom are 7-1=6. Therefore, the P-value is  $2P(t > 1.54)$  with df = 6. Final Answer: B.

\*\*\*\*\*

Answer Option 7:

The hypothesis test is for the mean, and the investigator will use a t-test since the population standard deviation is unknown. The t statistic will be based on the sample mean, which from the data given is approximately 301.14, and the sample standard deviation, which from the data given is approximately 2.33. The t statistic will be approximately 1.54 (calculated from the sample mean, population mean, sample standard deviation, and sample size).

Since the manufacturer's claim is that the weight of an aspirin tablet is 300 milligrams, this is a two-sided test (the actual weight could be greater than or less than 300 milligrams). The degrees of freedom for the t distribution in this case is the sample size minus 1, which is 6.

Therefore, the P-value of this test is given by "2P(t > 1.54) with df = 6". Final answer: B.

\*\*\*\*\*

Answer Option 8:

The P-value of a hypothesis test is the probability, assuming the null hypothesis is true, of obtaining a result as extreme or more extreme as the actual observed result. In this case, the null hypothesis is that the average weight of the tablets is 300 mg.

The average weight of the seven tablets is \$301\$, and the standard deviation is approximately \$2.16\$. The formula for the test statistic in this case is  $t = \frac{\overline{x} - \mu}{s / \sqrt{n}}$ , where  $\overline{x}$  is the sample mean,  $\mu$  is the hypothesized population mean,  $s$  is the sample standard deviation, and  $n$  is the sample size. Substituting the given values gives  $t = \frac{301 - 300}{2.16 / \sqrt{7}} \approx 1.54$ .

Since the sample mean is greater than the hypothesized population mean, the test is one-sided and we are interested in the probability of obtaining a test statistic greater than the observed value. However, the number of degrees of freedom is  $n - 1 = 7 - 1 = 6$ , not \$7\$. Therefore, the P-value is given by  $P(t > 1.54)$  with df = 6.

Final Answer: A

\*\*\*\*\*

Answer Option 9:

A t-test is used to test a claim about a population mean when the population standard deviation is unknown. The null hypothesis is that the mean weight of an aspirin tablet is 300 milligrams, and the alternative hypothesis is that the mean weight is not 300 milligrams. This is a two-tailed test because the alternative hypothesis does not specify a direction.

The test statistic is computed from the sample data. In this case, the sample size is 7, so the degrees of freedom is  $7 - 1 = 6$ . Based on the given information, the test statistic is approximately 1.54.

The P-value is the probability of obtaining a test statistic as extreme as, or more extreme than, the observed test statistic, under the assumption that the null hypothesis is true. Because this is a two-tailed test, the P-value is the probability of obtaining a test statistic greater than 1.54 or less than -1.54. Thus, the P-value is represented by  $2P(t > 1.54)$  with  $df = 6$ .

Final answer: B

\*\*\*\*\*

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 question asks about the shortest time interval that includes 100,000 of the 150,000 riders. This corresponds to the middle 2/3 of the distribution. Since the waiting times are normally distributed, this corresponds to within one standard deviation of the mean. The mean is 35 minutes and the standard deviation is 10 minutes, so this interval is  $35 - 10 = 25$  minutes to  $35 + 10 = 45$  minutes. The choice that corresponds to this interval is '25.3 to 44.7 minutes'.

Final Answer: C

=====

Answer Option 2:

In the standard normal distribution, approximately 68% of all values lie within one standard deviation from the mean. For the given distribution, 68% of all values lie within 10 minutes of 35 minutes, that is, between 25 minutes and 45 minutes. This is equivalent to 102,000 riders. Since we want the shortest time interval for 100,000 riders, we should consider the time interval that is symmetric about the mean and that contains approximately 67% of all values. This interval is '31.7 to 39.3 minutes'. So, the answer is B.

=====

Answer Option 3:

Since the distribution is normal, about 68% of the riders would be within one standard deviation of the mean. This means that about 102,000 riders would wait between 25 and 45 minutes. So we need a time interval that is a little bit smaller. The answer choice that makes the most sense is from 31.7 minutes to 39.3 minutes, which is within half a standard deviation of the mean. This interval is smaller than one standard deviation and is centered around the mean, so it is likely to contain fewer riders than an interval of one standard deviation.

Final Answer: B

=====

Answer Option 4:

For a normal distribution, about 68% of the data falls within one standard deviation of the mean. This means that approximately 68% of riders will have a wait time between 25 minutes ( $35 - 10$ ) and 45 minutes ( $35 + 10$ ). For 150,000 riders, this corresponds to 102,000 riders. Since we want the shortest time interval associated with 100,000 riders, we need to remove some riders from this group.

Since the distribution is symmetric, we can remove an equal number of riders from the beginning and the end of the distribution to preserve the centrality of the mean. Removing 2000 riders from each end means we are excluding riders who waited less than 25.3 minutes and those who waited more than 44.7 minutes. So, the shortest time interval associated with 100,000 riders is from 25.3 to 44.7 minutes.

Final Answer: C. '25.3 to 44.7 minutes'

=====

Answer Option 5:

If 100,000 out of 150,000 riders fall within a certain time interval, then that time interval includes approximately 2/3 of the total. In a normal distribution, about 68% of data falls within one standard deviation of the mean. Therefore, the time interval that includes the middle 2/3 of the data is from one standard deviation below the mean to one standard deviation above the mean. With a mean of 35 minutes and a standard deviation of 10 minutes, this interval is from 25 minutes to 45 minutes.

However, none of the answer choices exactly match this interval. Instead, we look for the interval that is closest to this range. The interval '25.3 to 44.7 minutes' is very close to our computed range, so the answer is C. '25.3 to 44.7 minutes'. Final answer: C

=====

Answer Option 6:

If the waiting times are normally distributed, then about 68% of the data lies within one standard deviation of the mean, 95% lies within two standard deviations, and 99.7% lies within three standard deviations.

Since there are 150,000 riders and we are looking for the shortest time interval associated with 100,000 riders, we want the time interval that includes the middle 2/3 (or approximately 67%) of the data.

This corresponds to the interval from one standard deviation below the mean to one standard deviation above the mean, or from  $35 - 10 = 25$  minutes to  $35 + 10 = 45$  minutes.

However, the options are not given in this form. They are given in decimal approximations. We approximate 25 minutes and 45 minutes to the nearest tenth to get 25.3 and 44.7, respectively.

Therefore, the correct answer is '25.3 to 44.7 minutes', or answer choice C.

=====

Answer Option 7:

If the waiting times are normally distributed, about 68% of the data falls within one standard deviation from the mean, about 95% falls within two standard deviations from the mean, and about 99.7% falls within three standard deviations from the mean. This is known as the empirical rule or the 68-95-99.7 rule.

In this case, 100,000 riders out of 150,000 is approximately 67% of all riders. Therefore, the shortest time interval associated with 100,000 riders lies within one standard deviation from the mean.

The mean waiting time is 35 minutes and the standard deviation is 10 minutes, so the interval is from  $35 - 10 = 25$  minutes to  $35 + 10 = 45$  minutes.

Hence, the shortest time interval associated with 100,000 riders is '25 to 45 minutes'. However, this exact option is not available. The closest available option that covers the mean and lies within one standard deviation is '25.3 to 44.7 minutes' or C. Final answer: C.

=====

Answer Option 8:

The shortest time interval associated with 100,000 riders would be the 100,000 riders who had the shortest wait times. To find this interval, we need to find where the cut-off is in the normal distribution for the bottom 100,000 out of 150,000 riders, or the bottom 2/3 of riders.

Looking up in the standard normal distribution table, the z-score associated with the bottom 2/3 of data is approximately 0.43.

Using the z-score formula,  $z = (X - \mu) / \sigma$ , where X is the cut-off time,  $\mu$  is the mean, and  $\sigma$  is the standard deviation, we solve to find X:

$$0.43 = (X - 35) / 10$$

$$X = 0.43 * 10 + 35 = 38.3$$

Therefore, the shortest time interval associated with 100,000 riders is 0 to 38.3 minutes. However, this option is not listed in the provided multiple choices. Considering the approximations usually involved in these calculations and the options provided, the closest option is '0 to 31.7 minutes'.

Final Answer: A

\*\*\*\*\*

Answer Option 9:

The shortest time interval associated with 100,000 riders means that we are interested in the time interval that includes the middle 100,000 riders out of 150,000 total riders. This means that we want to exclude the lowest 25,000 riders and the highest 25,000 riders, leaving us with the middle 100,000 riders.

This corresponds to a percentile range from the  $25,000/150,000 = 1/6$ th percentile (or approximately the 17th percentile) to the  $125,000/150,000 = 5/6$ th percentile (or approximately the 83rd percentile).

Using a standard normal distribution table or calculator, we find that the 17th percentile corresponds to a z-score of approximately -0.93 while the 83rd percentile corresponds to a z-score of approximately 0.93.

Recall that the z-score is calculated as  $(X - \text{mean})/\text{standard deviation}$ . Solving for X in each case, we find that the 17th percentile corresponds to a time of  $(-0.93*10) + 35 = 25.3$  minutes and the 83rd percentile corresponds to a time of  $(0.93*10) + 35 = 44.7$  minutes.

Therefore, the shortest time interval associated with 100,000 riders is from 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 correct answer is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This is a characteristic of normal distribution and not applicable to all discrete random variables.

=====

Answer Option 2:

The false statement is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This is the empirical rule, which applies to continuous random variables that are normally distributed, not to all discrete random variables. Final Answer: D

=====

Answer Option 3:

The correct answer is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This statement is part of the empirical rule, which is only applicable to normal distributions and not necessarily to all discrete random variables. The other options A, B, and C are true for all discrete random variables. Final Answer: D

=====

Answer Option 4:

The statement that does not necessarily hold true for all discrete random variables is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This is a characteristic of normal distribution, not all discrete random variables. Other distributions may not have this property. Final Answer: D.

=====

Answer Option 5:

The correct answer is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This is a characteristic of continuous random variables that follow a normal distribution, not a property of all discrete random variables.

=====

Answer Option 6:

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

This statement is not always true for all discrete random variables. It's an application of the empirical rule, which applies specifically to normal distributions, not all discrete random variables. Some discrete random variables may follow this pattern, but it is not a rule that applies to all of them.

=====

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 not necessarily true for all discrete random variables. It is part of the Empirical Rule which applies to normal distributions, not to all discrete random variables. There can be discrete random variables where the outcomes and their probabilities are distributed in such a way that more or less than 95% of the outcomes are within two standard deviations of the mean.

=====

Answer Option 8:

The correct answer is D. 'Approximately 95% of the outcomes will be within two standard deviations of the mean.' This statement is part of the empirical rule, which applies to normal distributions, not to all discrete random variables. While it may turn out to be true for some discrete random variables, it is not generally applicable to all discrete random variables. The other three options (A, B, and C) are all true for any discrete random variable. Therefore, the answer is D.

=====

Answer Option 9:

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

This statement is a general rule but it is based on the assumption that the distribution of the data is normal (or nearly normal). This is known as the Empirical Rule or 68-95-99.7 rule. However, not all distributions follow this rule. For example, a uniform distribution or a skewed distribution will not necessarily have 95% of outcomes within two standard deviations of the mean. Therefore, this statement is not always true for all discrete random variables.

=====

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:

D. 'The power of a test concerns its ability to detect an alternative hypothesis.'

\*\*\*\*\*

Answer Option 2:

D. 'The power of a test concerns its ability to detect an alternative hypothesis.'

\*\*\*\*\*

Answer Option 3:

D. 'The power of a test concerns its ability to detect an alternative hypothesis.'

\*\*\*\*\*

Answer Option 4:

The correct answer is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a test is the probability that the test correctly rejects the null hypothesis when the alternative hypothesis is true.

\*\*\*\*\*

Answer Option 5:

The correct answer is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a test is its ability to correctly reject the null hypothesis when the alternative hypothesis is true. That is, it is the probability that the test will lead to a correct conclusion.

\*\*\*\*\*

Answer Option 6:

The correct statement 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 a null hypothesis when a specific alternative hypothesis is true. It measures the test's ability to detect a difference from the null hypothesis, if such a difference really exists.

\*\*\*\*\*

Answer Option 7:

The correct answer is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' A test's power is the probability that it correctly rejects a null hypothesis when it is false and accepts the alternative hypothesis. It is a measure of the test's ability to detect a significant effect or difference if one actually exists. Therefore, statement D is the only one that is correct in the context of hypothesis testing.

\*\*\*\*\*

Answer Option 8:

The correct answer is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a statistical test measures the test's ability to reject the null hypothesis when the alternative hypothesis is true - in other words, to detect an effect if there is one.

\*\*\*\*\*

Answer Option 9:

The correct statement is D. 'The power of a test concerns its ability to detect an alternative hypothesis.' The power of a statistical test measures the test's ability to reject the null hypothesis when the alternative hypothesis is true. It is the probability that the test correctly rejects the null hypothesis. The other options are incorrect. Option A is incorrect because if there is sufficient evidence to reject the null hypothesis at the 10% level, it doesn't necessarily mean it can be rejected at the 5% level, which is a stricter threshold. Option B is incorrect because the decision to use a one- or two-sided test must be made before data collection, based on the research question. Option C is also incorrect because the level of the test (1% in this case) is the probability of rejecting the null hypothesis when it is true (Type I error), not the probability of rejecting the null hypothesis in general.

\*\*\*\*\*

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:

D. 'is significantly different from the other observations in the data set'

\*\*\*\*\*

Answer Option 2:

D. 'is significantly different from the other observations in the data set'

\*\*\*\*\*

Answer Option 3:

D. 'is significantly different from the other observations in the data set'

\*\*\*\*\*

Answer Option 4:

D. 'is significantly different from the other observations in the data set'

\*\*\*\*\*

Answer Option 5:

D. 'is significantly different from the other observations in the data set'

\*\*\*\*\*

Answer Option 6:

D. 'is significantly different from the other observations in the data set'

\*\*\*\*\*

Answer Option 7:

D. 'is significantly different from the other observations in the data set'

\*\*\*\*\*

Answer Option 8:

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'. 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:

D. 'There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005.'

\*\*\*\*\*

Answer Option 2:

D. 'There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005.'

\*\*\*\*\*

Answer Option 3:

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

\*\*\*\*\*

Answer Option 4:

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

\*\*\*\*\*

Answer Option 5:

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

\*\*\*\*\*

Answer Option 6:

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

\*\*\*\*\*

Answer Option 7:

The correct interpretation is 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 an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. In this case, it's the mean income per family in the school.

\*\*\*\*\*

Answer Option 8:

The correct interpretation of a confidence interval is D: 'There is a 95% probability that the mean income per family in the school is between \$35,095 and \$45,005.' This is because a confidence interval provides an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

\*\*\*\*\*

Answer Option 9:

The correct interpretation is 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 provides an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. In this case, the parameter is the mean income per family in the school.

\*\*\*\*\*



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:

The probability of a Type I error is the significance level, which is set by the researchers, so it would not change. Increasing the number of experimental units for each treatment would increase the power of the test (i.e., the probability of correctly rejecting the null hypothesis when it is false). Therefore, the correct answer is 'The probability of a Type I error would stay the same and the power would increase.'

Final Answer: D

\*\*\*\*\*

Answer Option 2:

The probability of a Type I error is determined by the chosen significance level, which has not changed in this case. Therefore, the probability of a Type I error remains the same. However, increasing the sample size generally increases the power of the test (the ability to detect a false null hypothesis), thus reducing the probability of a Type II error. Therefore, the correct answer is 'The probability of a Type I error would stay the same and the power would increase.' Final answer: D

\*\*\*\*\*

Answer Option 3:

Increasing the sample size in an experiment generally does not affect the probability of a Type I error, which is determined by the significance level. However, it does reduce the probability of a Type II error, thus increasing the power of the test (the ability to correctly reject a false null hypothesis). So the correct answer is: D. 'The probability of a Type I error would stay the same and the power would increase.'

\*\*\*\*\*

Answer Option 4:

Increasing the sample size usually decreases the standard error of the estimate and thus increases the power of the statistical test (i.e., the probability of correctly rejecting the null hypothesis when it is false). This is because a larger sample size allows the experiment to detect smaller differences more reliably. The probability of a Type I error, however, is determined solely by the significance level (0.05 in this case) and is not affected by the sample size. Therefore, the probability of a Type I error would stay the same and the power would increase. Final Answer: D.

\*\*\*\*\*

Answer Option 5:

Increasing the sample size generally reduces the standard error and widens the margin of error. Hence, it makes it more likely that the researchers will reject the null hypothesis when it is false. That is, the power of the test (which is the probability of correctly rejecting a false null hypothesis) would increase. The probability of a Type I error is determined by the significance level, which is not changed, so it remains the same. Therefore, 'The probability of a Type I error would stay the same and the power would increase.' is the correct answer. Final answer: D

\*\*\*\*\*

Answer Option 6:

The probability of a Type I error is given by the significance level, which is not changed by the sample size, so the probability of a Type I error would stay the same. Increasing the sample size generally makes it easier to detect a false null hypothesis, if one exists, so the power of the test (probability of correctly rejecting a false null hypothesis) would increase. Therefore, the correct answer is 'The probability of a Type I error would stay the same and the power would increase.' Final answer: D.

\*\*\*\*\*

Answer Option 7:

The probability of a Type I error, also known as the significance level, is set by the researchers and does not depend on the number of experimental units. In this case, it is set at 0.05 and would remain the same. Therefore, options A and B cannot be correct.

Increasing the number of experimental units increases the sample size and can make it easier to detect an effect if one truly exists. This means the power of the test, or the probability of correctly rejecting the null hypothesis when it is false, would increase.

Option D correctly states that the probability of a Type I error would stay the same and the power would increase. Option C cannot be correct because it incorrectly states that the probability of a Type II error would increase - in fact, the probability of a Type II error (failing to reject a false null hypothesis) decreases as power increases. Therefore, the correct answer is D. Final Answer: D

\*\*\*\*\*

Answer Option 8:

The probability of a Type I error is determined by the significance level, so doubling the number of experimental units does not change the probability of a Type I error. However, increasing the sample size generally increases the power of a test (the probability of correctly rejecting a false null hypothesis), which decreases the probability of a Type II error. So, 'The probability of a Type I error would stay the same and the power would increase.' is correct. Final answer: D

\*\*\*\*\*

Answer Option 9:

The probability of a Type I error is the significance level, which is fixed at 0.05 and does not change with the number of experimental units. So, the probability of a Type I error stays the same. Increasing the number of experimental units increases the power of the test (the ability to detect an effect if there is one), which means the probability of a Type II error (failing to reject a false null hypothesis) would decrease. So, the correct answer is '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:

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 2:

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 3:

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 4:

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 5:

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 6:

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 7:

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 8:

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 9:

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.'

\*\*\*\*\*

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:

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

=====

Answer Option 2:

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

=====

Answer Option 3:

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

=====

Answer Option 4:

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

=====

Answer Option 5:

Option C is not true. While t-distributions do have a mean of 0, they do not always have a standard deviation of 1. The standard deviation of a t-distribution varies depending on the degrees of freedom.

Therefore, the correct answer is C.

=====

Answer Option 6:

The statement C. 't-distributions always have mean 0 and standard deviation 1.' is not true about t-distributions. While t-distributions do have a mean of 0, the standard deviation depends on the degrees of freedom and is not always 1. Final answer: C

=====

Answer Option 7:

C. 't-distributions always have mean 0 and standard deviation 1.' is not true. All t-distributions are symmetric with a mean of 0, but they do not always have a standard deviation of 1. The standard deviation depends on the degrees of freedom and is generally greater than 1, which is why the t-distribution is more spread out than the normal distribution.

=====

Answer Option 8:

The statement C: 't-distributions always have mean 0 and standard deviation 1.' is not true about t-distributions. While t-distributions do indeed have a mean of 0, they do not always have a standard deviation of 1. Rather, the standard deviation (or scale parameter) of a t-distribution depends on its degrees of freedom; the fewer the degrees of freedom, the greater the standard deviation. Final Answer: C

=====

Answer Option 9:

The statement that is not true about t-distributions is C. 't-distributions always have mean 0 and standard deviation 1.' While it's true that t-distributions have a mean of 0, they do not always have a standard deviation of 1. Instead, the standard deviation varies depending on the degrees of freedom. Final Answer: C.

=====

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:

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

\*\*\*\*\*

Answer Option 2:

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

\*\*\*\*\*

Answer Option 3:

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

\*\*\*\*\*

Answer Option 4:

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

\*\*\*\*\*

Answer Option 5:

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

\*\*\*\*\*

Answer Option 6:

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

\*\*\*\*\*

Answer Option 7:

Since the subjects were given a treatment (the beer) and then observed, this study is not merely observational. They were used as their own controls, since their reaction times were measured before and after the treatment. Therefore, the correct statement is: 'This study was an experiment in which the subjects were used as their own controls.' Final Answer: D

\*\*\*\*\*

Answer Option 8:

This is an experimental design study. The subjects are used as their own controls because their reaction time is measured before and after the alcohol consumption to see how they change. Therefore, the correct answer is D. 'This study was an experiment in which the subjects were used as their own controls.'

\*\*\*\*\*

Answer Option 9:

The correct statement is: 'This study was an experiment in which the subjects were used as their own controls.' This is because each subject was measured before and after drinking the alcohol, effectively serving as their own control for comparison. This kind of design is often used in studies where individual variance is high. 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:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

Answer Option 2:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

Answer Option 3:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

Answer Option 4:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

Answer Option 5:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

Answer Option 6:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

Answer Option 7:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

Answer Option 8:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

Answer Option 9:

A. 'It describes a process that creates estimates that are too high on average or too low on average.'

\*\*\*\*\*

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 true Null Hypothesis. In this scenario, the Null Hypothesis would be that the tissue strength is within specifications. Therefore, a Type I error would result in 'Halting the production process when the tissue strength is within specifications.' Final answer: C

\*\*\*\*\*

Answer Option 2:

A Type I error occurs when a true null hypothesis is rejected. Here, the null hypothesis would be that the mean strength is within specifications. Therefore, a Type I error would result in "Halting the production process when the tissue strength is within specifications." Final Answer: C

\*\*\*\*\*

Answer Option 3:

A Type I error, in hypothesis testing, is the incorrect rejection of a true null hypothesis (a "false positive"). In this context, it would mean halting the production process when there's actually no problem, so the correct answer is C. 'Halting the production process when the tissue strength is within specifications.'

\*\*\*\*\*

Answer Option 4:

A Type I error is the incorrect rejection of a true null hypothesis (also known as a "false positive" finding or conclusion). In this context, it would be halting the production process when the tissue strength is actually within specifications. Hence, the answer is C. 'Halting the production process when the tissue strength is within specifications.'

\*\*\*\*\*

Answer Option 5:

Type I error is the incorrect rejection of a true null hypothesis, also known as a "false positive". In this context, a Type I error would occur if the company incorrectly concludes that the tissue strength has dropped below the specified level when it's actually within the specifications. So, the correct answer is C. 'Halting the production process when the tissue strength is within specifications.'

\*\*\*\*\*

Answer Option 6:

A Type I error in statistical terms is when we reject the null hypothesis when it is actually true. In this scenario, the null hypothesis is that the tissue strength is good (within specifications). So, a Type I error would be halting production when the tissue strength is in fact within specifications. Thus, the correct option is C: 'Halting the production process when the tissue strength is within specifications.'

\*\*\*\*\*

Answer Option 7:

A Type I error occurs when we reject a true null hypothesis, meaning we decide something is wrong when it actually isn't.

In this context, the null hypothesis is that the tissue strength is within specifications. If we reject this when it's actually true, that means we're halting the production process when the tissue strength is within specifications.

So, the correct answer is C. 'Halting the production process when the tissue strength is within specifications.'

\*\*\*\*\*

Answer Option 8:

A Type I error refers to the incorrect rejection of a true null hypothesis. In this context, it would mean incorrectly concluding that the tissue strength is below the specified level, when it is in fact within specifications. Therefore, the result of a Type I error would be 'Halting the production process when the tissue strength is within specifications.'. Final Answer: C.

\*\*\*\*\*

Answer Option 9:

A Type I error occurs when we reject a true null hypothesis. In this context, the null hypothesis would be that the tissue strength is within specifications. Therefore, a Type I error would result in halting the production process when the tissue strength is actually within specifications. So, the correct answer is C. 'Halting the production process when the tissue strength is within specifications.'

\*\*\*\*\*

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:

In a normal distribution, 30% of the data is approximately 0.5244 standard deviations below the mean. Thus, if 30% of the people live longer than 75 years, then the mean longevity would be approximately 75 + 0.5244(14). Therefore, the correct answer is D. '75 + 0.5244(14)'.

=====

Answer Option 2:

Using the standard normal table, a cumulative probability of 0.30 corresponds to a Z-score of approximately -0.5244. Knowing that  $Z = (X - \mu)/\sigma$ , we can substitute the given values and solve.  $-0.5244 = (75 - \mu)/14$ . Solving for  $\mu$ , we get  $\mu = 75 + 0.5244 \cdot 14$ . Therefore the correct answer is B. '75 + 0.30(14)'.

=====

Answer Option 3:

In a normal distribution, if 30% of the people live longer than 75 years, then 70% live less than 75 years. Looking up 0.70 in a standard normal distribution table, we find a z-score of approximately 0.5244. The z-score is calculated as  $\frac{x - \mu}{\sigma}$ , where  $x$  is the value of the point,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Setting the z-score equal to 0.5244 and solving for the mean longevity gives  $\mu = x - z \cdot \sigma = 75 - 0.5244(14)$ . Therefore the mean longevity is '75 - 0.5244(14)' which is answer choice D.

=====

Answer Option 4:

In a normal distribution, approximately 50% of data is above the mean and 50% is below. If 30% of people live longer than 75 years, then 70% live 75 years or less. Therefore, 75 years is at the 70th percentile, which is 0.5244 standard deviations above the mean. Using the formula for normal distributions, Mean = Value - (Z-score)(Standard deviation), we get Mean = 75 - 0.5244(14). So, the correct answer is D. '75 - 0.5244(14)'.

=====

Answer Option 5:

In a normal distribution, approximately 50% of values are below the mean and 50% are above. If 30% of people live longer than 75 years, then 70% live 75 years or less. This means that 75 years is above the mean, and is higher than 70% of all values. In terms of standard deviations, this corresponds approximately to 0.5244 standard deviations above the mean. Therefore, the mean is  $75 - 0.5244(14) = 75 - 7.3416$ . So, the final answer is D. '75 - 0.5244(14)'.

=====

Answer Option 6:

For a normal distribution, approximately 50% of observations lie below the mean and 50% lie above. If 30% of people live longer than 75 years, then 70% live less than 75 years. This means that 75 years is more than the mean lifespan.

The normal distribution table shows that approximately 70% of observations lie within one standard deviation ( $\pm 1\sigma$ ) from the mean. Therefore, 75 years corresponds to the mean plus 0.5244 standard deviations (since the normal distribution table value for 0.30 is approximately 0.5244).

Therefore, the mean longevity is  $75 - 0.5244(14)$ . So the correct answer is D. '75 - 0.5244(14)'.

=====

Answer Option 7:

The question states that 30% of people live longer than 75 years. In a standard normal distribution, 50% of the data is above the mean and 50% is below. Since 30% of people live longer than 75 years, this means that 70% of people live less than or equal to 75 years. Looking up 0.70 on the standard normal distribution table gives us a z-score of approximately 0.5244. The z-score is defined as  $(x - \text{mean}) / \text{standard deviation}$ . Rearranging and solving for the mean gives us  $\text{mean} = x - z \cdot \text{standard deviation}$ . Plugging in the numbers gives us  $75 - 0.5244 \cdot 14$ . Final Answer: D

=====

Answer Option 8:

From a standard normal distribution table, we find that about 30% of data lies to the right of approximately 0.5244 standard deviations above the mean. In this context, 75 years is 0.5244 standard deviations above the mean. So if we let M be the mean longevity, we can set up the equation  $M + 0.5244(14) = 75$ . Solving for M gives  $M = 75 - 0.5244(14)$ . So the mean longevity is '75 - 0.5244(14)'. Final answer: D

=====

Answer Option 9:

In a normal distribution, approximately 50% of people live longer than the mean longevity and 50% live less than the mean longevity. If 30% of people live longer than 75 years, this means that 70% of people live less than or equal to 75 years. Since 70% is greater than 50%, this means the mean longevity is less than 75 years.

The standard deviation (14 years) gives us an idea of how spread out the data is. Because we know that 70% live less than or equal to 75 years, and 50% live less than the mean, this places the mean at roughly 0.2 standard deviations less than 75 years.

Hence, the mean longevity would be  $75 - 0.20(14)$ , which falls within the options in 'A'. Thus, the answer is A. '75 - 0.30(14)'.

=====

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:

A. 'increases the interval size by 9%.'

=====

Answer Option 2:

A. 'increases the interval size by 9%.'

=====

Answer Option 3:

A. 'increases the interval size by 9%.'

=====

Answer Option 4:

C. 'increases the interval size by 57%.'

=====

Answer Option 5:

C. 'increases the interval size by 57%.'

=====

Answer Option 6:

A. 'increases the interval size by 9%.'

=====

Answer Option 7:

A higher confidence level results in a wider confidence interval, all other things being equal. So, the 99% confidence interval will be larger than the 90% confidence interval. But the increase is not by a specific percentage. Therefore, the answer is: C. 'increases the interval size by 57%.'

=====

Answer Option 8:

The confidence interval is wider for a 99% confidence interval than a 90% confidence interval because a 99% confidence interval captures the population parameter with greater certainty. Therefore, changing from a 90% confidence interval estimate to a 99% confidence interval estimate, with all other things being equal, A. 'increases the interval size'. However, it does not increase the interval size by a specific percentage such as 9% or 57%. The percentage increase is dependent on the standard deviation and sample size. So there is no correct answer among the provided choices.

=====

Answer Option 9:

A. 'increases the interval size by 9%.'

In statistics, a confidence interval (CI) is a type of interval estimate, computed from the statistics of the observed data, that might contain the true value of an unknown population parameter. The interval has an associated confidence level that, loosely speaking, quantifies the level of confidence that the parameter lies in the interval.

Increasing the confidence level from 90% to 99% would increase the size of the interval because you're more certain (99% confident) that the true population parameter lies within the range. So the answer is C. 'increases the interval size by 57%.' The exact percentage increase would depend on the standard deviation and the sample size, but we're looking for a general trend here, not an exact value.

=====



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 defined by the number of "successes" in a fixed number of trials, where each trial is independent and has the same probability of success. The only scenario that fits this definition is D: 'The number of motorists not wearing seat belts in a random sample of five drivers.' Each driver is an independent event with two outcomes: they are either wearing a seat belt (failure) or not (success). There is a fixed number of trials (five drivers) and each trial (driver) has the same probability of success (not wearing a seat belt). Final Answer: D

\*\*\*\*\*

Answer Option 2:

A binomial random variable is defined by the number of "successes" in a fixed number of independent and identically distributed Bernoulli trials. A Bernoulli trial is an experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted.

Here, option D. 'The number of motorists not wearing seat belts in a random sample of five drivers.' is a binomial random variable. This is because each driver either is or is not wearing a seatbelt (two outcomes), the drivers are sampled independently, and the probability a given driver is not wearing a seatbelt is the same for each driver. Final answer: D.

\*\*\*\*\*

Answer Option 3:

A binomial random variable is one that counts the number of successes in a fixed number of trials, where each trial is independent, and the probability of success is the same for each trial.

Option A is not a binomial random variable because the number of trials is not fixed, it continues until a "5" appears.

Option B is not a binomial random variable because there are three possible outcomes for each game (win, tie, loss), not just success or failure.

Option C is not a binomial random variable because the trials are not independent; the outcome of one card draw affects the probabilities for the next draw.

Option D is a binomial random variable because it counts the number of successes (not wearing a seat belt) in a fixed number of trials (five drivers), where each trial is independent (one driver's seat belt use does not affect another's), and the probability of success is the same for each trial (assuming each driver is equally likely to not wear a seat belt).

Final Answer: D

\*\*\*\*\*

Answer Option 4:

A binomial random variable is defined by the number of "successes" (a certain outcome happening) in a set number of repeated, independent trials, each of which results in a "success" with a certain probability or a "failure" with the complementary probability.

Option D. 'The number of motorists not wearing seat belts in a random sample of five drivers.' is a binomial random variable. Here, the "success" is a motorist not wearing a seat belt, the number of trials is 5 (the number of drivers sampled), and each trial is independent (the state of one driver does not affect others) and has two outcomes (either the driver is wearing a seat belt or not). So, the correct answer is D.

\*\*\*\*\*

Answer Option 5:

A binomial random variable is defined by a fixed number of Bernoulli trials (a trial with two possible outcomes), each of which is independent of the others and has the same probability of success. From the given options only D. 'The number of motorists not wearing seat belts in a random sample of five drivers.' fits this definition. Each of the five drivers is a trial, each trial is independent, and each has two outcomes: the driver is either wearing a seatbelt or not. Final Answer: D

\*\*\*\*\*

Answer Option 6:

A binomial random variable has two possible outcomes. It counts the number of "successes" in a fixed number of independent Bernoulli trials (where a Bernoulli trial is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted).

Option D, 'The number of motorists not wearing seat belts in a random sample of five drivers.', is a binomial random variable. In this case, the two outcomes are a motorist wearing a seat belt (failure) and a motorist not wearing a seat belt (success). There are exactly five trials (five drivers being sampled), and the outcome of each trial is independent of the others.

The other options do not satisfy the conditions necessary to be considered a binomial random variable.

Final Answer: D

\*\*\*\*\*

Answer Option 7:

A binomial random variable represents the number of successes in a series of independent and identically distributed Bernoulli trials. A Bernoulli trial is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted.

Option A is not a binomial random variable because the number of trials (the tosses of the die) is not fixed in advance.

Option B is not a binomial random variable because there are three possible outcomes (win, tie, loss), not two.

Option C is not a binomial random variable because the draws are not independent (once a card is drawn, it changes the probabilities for the subsequent draws).

Option D is a binomial random variable. There are a fixed number of trials (five drivers), there are two outcomes (wearing a seat belt or not), and each trial is independent (the seat belt status of one driver does not affect the status of another). Hence the final answer is D.

\*\*\*\*\*

Answer Option 8:

A binomial random variable refers to the number of successes in a fixed number of independent and identically distributed Bernoulli trials. A Bernoulli trial is a random experiment in which there are only two possible outcomes - success or failure.

In this case, the only option that refers to a binomial random variable is D. 'The number of motorists not wearing seat belts in a random sample of five drivers.' This is because each driver (trial) is independent, there are a fixed number of trials (five drivers), and there are only two outcomes for each driver - either they are wearing a seat belt (failure) or they are not (success). Final Answer: D.

\*\*\*\*\*

Answer Option 9:

A binomial random variable is one that arises from a series of Bernoulli trials, which are identical, independent trials that have two possible outcomes: success and failure. This requires a fixed number of trials, each with a constant probability of success.

Option A does not have a fixed number of trials. Option B does not have exactly two outcomes for each trial. Option C does not have a constant probability of success since the cards are drawn without replacement.

Option D, however, does meet all the characteristics of a binomial random variable. There are a fixed number of trials (five drivers), each trial has two possible outcomes (wearing a seat belt or not), and the probability of success (a driver not wearing a seat belt) is constant for each trial (assuming the sample is drawn from a large enough population). Therefore, option D is the correct answer. Final Answer: 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:

D. 'I and II'

\*\*\*\*\*

Answer Option 2:

The power of a test is the probability that it correctly rejects the null hypothesis when the null hypothesis is false. This is the opposite of making a Type II error, so increasing the probability of a Type II error will decrease the power of the test, not increase it. Increasing the sample size and increasing the significance level will both increase the power of the test, so the correct answer is 'I and II'. Final

Answer: D

\*\*\*\*\*

Answer Option 3:

The power of a hypothesis test is the probability that the test correctly rejects the null hypothesis when the alternative hypothesis is true. It's influenced by several factors.

I. Increasing the sample size: This will increase the power of a hypothesis test. The larger the sample size, the greater the chance of detecting a true effect if one exists.

II. Increasing the significance level: A higher significance level (alpha) means a greater probability of rejecting the null hypothesis, therefore increasing the power of the test.

III. Increasing the probability of a Type II error: A Type II error occurs when a true alternative hypothesis is not detected, so increasing the probability of a Type II error would actually decrease the power of the test.

So, the correct answer is D. 'I and II'

\*\*\*\*\*

Answer Option 4:

The power of a hypothesis test is the probability that it correctly rejects a false null hypothesis.

I. Increasing the sample size generally increases the power of a hypothesis test, because a larger sample size gives us more information and reduces the impact of random variation.

II. Increasing the significance level (i.e., the cut-off for rejecting the null hypothesis) increases power by making it easier to reject the null hypothesis, but at the cost of increasing the chance of a Type I error (rejecting a true null hypothesis).

III. Increasing the probability of a Type II error (failing to reject a false null hypothesis) actually decreases the power of the test, since power is defined as 1 minus the probability of a Type II error.

So, the correct answer is D. 'I and II'.

\*\*\*\*\*

Answer Option 5:

The power of a hypothesis test is the probability that it correctly rejects a false null hypothesis.

I. Increasing the sample size increases the power of a hypothesis test because it reduces the standard error, making the test statistic larger and increasing the chances of rejecting the null hypothesis when it is false.

II. Increasing the significance level also increases the power of a hypothesis test. The significance level is the probability of rejecting the null hypothesis when it is true (Type I error). A larger significance level means a larger critical region and a higher chance of rejecting the null hypothesis when it is false.

III. Increasing the probability of a Type II error (failing to reject a false null hypothesis) actually decreases the power of a test. The power of a test is 1 minus the probability of a Type II error, so as the probability of a Type II error increases, power decreases.

So, options I and II increase the power of a hypothesis test. Final Answer: D

\*\*\*\*\*

Answer Option 6:

The power of a hypothesis test is the probability that the test correctly rejects the null hypothesis. Therefore, it can be influenced by the following factors:

I. Increasing the sample size will increase the power of a hypothesis test. This is because larger samples provide more information and lead to more precise estimates of the population parameters, which in turn makes it easier to detect a true effect if one exists.

II. Increasing the significance level (i.e., the probability of rejecting the null hypothesis when it is true, also known as the Type I error rate) will also increase the power of a test. This is because a larger significance level will make it easier to reject the null hypothesis, thereby increasing the power of the test.

III. Increasing the probability of a Type II error (i.e., the probability of failing to reject the null hypothesis when it is false) will decrease, not increase, the power of the test. This is because the power of a test is equal to 1 minus the probability of a Type II error, so increasing the probability of a Type II error will decrease the power.

Therefore, the correct answer is "I and II" (option D).

\*\*\*\*\*

Answer Option 7:

The power of a hypothesis test is the probability that it correctly rejects the null hypothesis when the alternative hypothesis is true.

I. Increasing the sample size. This will increase the power of the test, because the larger the sample size, the more evidence we have for the alternative hypothesis.

II. Increasing the significance level. This will also increase the power of the test. The significance level is the probability of rejecting the null hypothesis when it is true (a Type I error). If we increase this probability, we are more likely to reject the null hypothesis and therefore more likely to have a powerful test.

III. Increasing the probability of a Type II error. This will decrease the power of the test. A Type II error occurs when we fail to reject the null hypothesis when the alternative hypothesis is true. If we increase this probability, we are less likely to reject the null hypothesis and therefore less likely to have a powerful test.

Therefore, the correct answer is D. 'I and II'.

\*\*\*\*\*

Answer Option 8:

The power of a hypothesis test is the probability that the test correctly rejects the null hypothesis when the alternative hypothesis is true. It is generally increased by the following factors:

I. Increasing the sample size: This allows for more data to be considered, reducing the chance of random errors and thus increasing the power of the test.

II. Increasing the significance level: This makes it easier to reject the null hypothesis, but at the expense of potentially increasing the chance of a Type I error (rejecting the null hypothesis when it's actually true). This, however, does increase the power of the test.

III. Increasing the probability of a Type II error (failing to reject the null hypothesis when the alternative is true) actually decreases the power of a test, not increases it.

So the correct answer is D. 'I and II'

\*\*\*\*\*

Answer Option 9:

The power of a hypothesis test is defined as the probability of rejecting the null hypothesis when the alternative hypothesis is true. Increasing the sample size allows for a more accurate representation of the population, thus increasing the power of the test. Increasing the significance level (alpha) increases the likelihood of rejecting the null hypothesis, thus increasing power. Increasing the probability of a Type II error (failing to reject the null hypothesis when the alternative is true) would actually decrease the power of the test. Therefore, the correct answer is D. 'I and II'.

\*\*\*\*\*

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. This is a common rule of thumb for using chi-square tests - that each cell should have an expected frequency of at least 5.

Statement II is not necessarily true. You can have a chi-square test with just 1 degree of freedom. A chi-square test with 1 degree of freedom is known as a chi-square test for goodness of fit.

Statement III is not true. In a  $4 \times 2$  two-way table, the degrees of freedom would be  $(4-1)(2-1) = 3 \times 1 = 3$ .

Therefore, the correct answer is A. 'I only'

\*\*\*\*\*

Answer Option 2:

I. This statement is true. In order to use a  $\chi^2$  procedure, the expected count for each cell of a one- or two-way table should be at least 5.

II. This statement is false. It's possible to use  $\chi^2$  procedures with 1 degree of freedom.

III. This statement is false. In a  $4 \times 2$  two-way table, the number of degrees of freedom would be not 3, but  $(4 - 1)(2 - 1) = 3$ .

So, the correct answer is A. 'I only'.

\*\*\*\*\*

Answer Option 3:

The correct answer is B. 'I and III only'.

Statement I: This is a commonly used rule of thumb for applying a  $\chi^2$  test. It is not a strict requirement, but if the expected count in each cell is less than 5, the  $\chi^2$  test can be less accurate.

Statement II: This is not true. A  $\chi^2$  test can be applied even with one degree of freedom. For example, a  $\chi^2$  test can be used to compare observed and expected frequencies for a single categorical variable with two categories.

Statement III: In a two-way table, the degrees of freedom are calculated as (number of rows - 1) \* (number of columns - 1). So for a  $4 \times 2$  table, the degrees of freedom would indeed be 3.

\*\*\*\*\*

Answer Option 4:

Statement I is true. The  $\chi^2$  test requires that the expected count for each cell in the table be at least 5 for the approximation to be valid.

Statement II is not necessarily true. It is possible to perform a  $\chi^2$  test with only 1 degree of freedom.

Statement III is true. The degrees of freedom for a two-way table is calculated as (number of rows - 1) \* (number of columns - 1). For a  $4 \times 2$  table, this would be  $(4-1) * (2-1) = 3$ .

So, the correct answer is: B. 'I and III only'

\*\*\*\*\*

Answer Option 5:

Statement I is true. The expected value (or count) for each cell in a chi-square test should be at least 5 for the validity of the approximation to the chi-square distribution.

Statement II is not necessarily true. Chi-square procedures can be used with 1 degree of freedom. For example, a chi-square test for goodness of fit can have 1 degree of freedom.

Statement III is true. In a  $4 \times 2$  two-way table, the number of degrees of freedom would be (number of rows - 1) \* (number of columns - 1) =  $3 * 1 = 3$ .

So, the correct answer is B. 'I and III only'.

\*\*\*\*\*

Answer Option 6:

The correct answer is A. 'I only'.

Explanation:

I. In order to use a chi-square ( $\chi^2$ ) procedure, the expected count for each cell in a contingency table should be at least 5. This is a common rule of thumb for applying the  $\chi^2$  test, as this helps ensure that the  $\chi^2$  approximation to the distribution of the test statistic is valid.

II. The statement is incorrect. The degrees of freedom for a  $\chi^2$  test depends on the number of categories or groups being compared (or the dimensions of the contingency table, in the case of a two-way  $\chi^2$  test). It's not necessary to have at least 2 degrees of freedom.

III. The number of degrees of freedom for a two-way  $\chi^2$  test is calculated as (rows - 1) \* (cols - 1). Therefore, for a  $4 \times 2$  table, the degrees of freedom would be  $(4-1)(2-1) = 3$ , making the statement true.

However, as statement II is false, this option is not correct.

So, the correct answer is 'I only'.

\*\*\*\*\*

Answer Option 7:

Statement I is correct. The  $\chi^2$  procedure requires that the expected counts for each cell in the table be at least 5 for the approximation to be valid.

Statement II is incorrect. A  $\chi^2$  procedure can be used with 1 degree of freedom. For example, a  $\chi^2$  test for goodness of fit with two categories only has 1 degree of freedom.

Statement III is correct. In a two-way table with r rows and c columns, the number of degrees of freedom is (r-1)(c-1). So, in a  $4 \times 2$  two-way table, the number of degrees of freedom is  $(4-1)(2-1) = 3$ .

Therefore, the correct answer is B. 'I and III only'

\*\*\*\*\*

Answer Option 8:

Statement I: This is a general guideline but not a rule. The correct requirement for expected cell frequencies in a chi-square test is that no more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater.

Statement II: You can have less than 2 degrees of freedom while using  $\chi^2$  procedures. For example, in a one-way table with three categories, there are 2 degrees of freedom. But for a one-way table with two categories, there is only 1 degree of freedom.

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

So, the correct answer is D. 'III only'.

\*\*\*\*\*

Answer Option 9:

Statement I: This statement is generally true; it is a common rule of thumb for using  $\chi^2$  procedures that each expected cell count should be at least 5. However, it is not absolutely necessary, especially for large sample sizes. Some statisticians use a guideline of at least 1 or at least 5 in all cells.

Statement II: This statement is not necessarily true. You can use  $\chi^2$  procedures with only 1 degree of freedom (for instance, in a  $2 \times 2$  contingency table).

Statement III: This statement is not true. For a  $r \times c$  (row x column) contingency table, the degrees of freedom are equal to  $(r-1)(c-1)$ . Plugging in  $r=4$  and  $c=2$ , we get  $(4-1)(2-1)=3$ , which would make this statement true.

Therefore, the correct response is B. 'I and III only'. Final answer: B

\*\*\*\*\*

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:

The definition of independent events is that the occurrence of one does not affect the probability of the occurrence of the other. However, mutually exclusive events cannot both occur. Therefore, two events cannot be both independent and mutually exclusive if they both have non-zero probabilities. So, the correct answer is D.

\*\*\*\*\*

Answer Option 2:

D. 'E and F are both independent and mutually exclusive.'

Two events cannot be both independent and mutually exclusive. If two events are mutually exclusive, then the occurrence of one event means the other cannot happen. On the other hand, two events are independent if the occurrence of one event does not affect the probability of the occurrence of the other event. Hence, two events cannot be both independent and mutually exclusive.

\*\*\*\*\*

Answer Option 3:

The correct answer is D. Two events cannot be both independent and mutually exclusive. Independent events have no impact on each other's probability, while mutually exclusive events cannot occur at the same time. If E and F are mutually exclusive, the occurrence of E means that F cannot occur, and vice versa, which contradicts the definition of independent events.

\*\*\*\*\*

Answer Option 4:

D. 'E and F are both independent and mutually exclusive.'

Two events cannot be both independent and mutually exclusive. If two events are independent, then the occurrence of one event does not affect the probability of the other event. If two events are mutually exclusive, then the occurrence of one event means the other cannot occur. Therefore, mutually exclusive events are dependent, not independent.

\*\*\*\*\*

Answer Option 5:

D. 'E and F are both independent and mutually exclusive.'

Two events are independent if the probability of both events occurring is the product of the probabilities of each event. Two events are mutually exclusive if they cannot both occur. Therefore, if E and F are both independent and mutually exclusive, that would mean  $p \cdot q = 0$ , but we are given that both p and q are nonzero. Therefore, option D is impossible.

\*\*\*\*\*

Answer Option 6:

The correct answer is D. 'E and F are both independent and mutually exclusive.'

For two events to be independent, the occurrence of one event does not affect the probability of the occurrence of the other event. For two events to be mutually exclusive, they cannot both occur at the same time.

If two events are mutually exclusive, then they cannot be independent because the occurrence of one event means the other cannot occur, thereby affecting its probability. So it is impossible for two events to be both independent and mutually exclusive.

\*\*\*\*\*

Answer Option 7:

The correct answer is D. 'E and F are both independent and mutually exclusive.'

In probability theory, two events are said to be independent if the occurrence of one does not affect the probability of the occurrence of the other. On the other hand, two events are mutually exclusive if the occurrence of one event excludes the occurrence of the other.

So, it is impossible for two events to be both independent and mutually exclusive at the same time. If two events are mutually exclusive, then the occurrence of one event definitely affects the probability of the occurrence of the other event, making them dependent. Therefore, they cannot be independent.

\*\*\*\*\*

Answer Option 8:

The correct answer is D. 'E and F are both independent and mutually exclusive.'

By definition, two events are independent if the occurrence of one event does not affect the probability of the occurrence of the other event. This means that  $P(E \cap F) = P(E)P(F)$ .

On the other hand, two events are mutually exclusive if they cannot both occur at the same time. This means that  $P(E \cap F) = 0$ .

Clearly, these two definitions cannot both be true unless  $P(E) = 0$  or  $P(F) = 0$ . But the problem states that E and F have nonzero probabilities, so D is impossible.

\*\*\*\*\*

Answer Option 9:

D. 'E and F are both independent and mutually exclusive.'

Explanation: Two events cannot be both independent and mutually exclusive. If E and F are mutually exclusive (i.e., they cannot both occur), then the occurrence of E necessarily implies the nonoccurrence of F, and thus E and F are dependent. For two events to be independent, the occurrence of one event must not affect the probability of the occurrence of the other event. For E and F to be mutually exclusive, the occurrence of one event must eliminate all possibility of the occurrence of the other event. Therefore, it is impossible for events E and F to be both independent and mutually exclusive.

\*\*\*\*\*





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:

By the law of total probability,

$$P(\text{accident}) = P(\text{accident} | \text{wet})P(\text{wet}) + P(\text{accident} | \text{dry})P(\text{dry})$$

$$= 0.01 * 0.2 + 0.002 * 0.8$$

$$= 0.002 + 0.0016$$

$$= 0.0036$$

So, the probability of there being an accident on Highway 48 today is 0.0036.

Final answer: D.

\*\*\*\*\*

Answer Option 2:

The total probability of an accident is the sum of the probability of an accident given dry weather times the probability of dry weather and the probability of an accident given wet weather times the probability of wet weather. This is  $(0.002)(0.8) + (0.01)(0.2) = 0.0016 + 0.002 = 0.0036$ . Final Answer: D. 0.0036

\*\*\*\*\*

Answer Option 3:

The probability that there will be an accident on Highway 48 today is the sum of the probability that there will be an accident given it is wet and the probability that there will be an accident given it is dry, each multiplied by the probability of that weather condition. This gives a total probability of  $\$(0.2/100)(0.8) + (1.0/100)(0.2) = 0.0016 + 0.002 = \boxed{0.0036}\$$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

The probability that there will be an accident on Highway 48 today is the sum of the probability that there will be an accident given that it is wet and the probability that there will be an accident given that it is dry. This is given by:

$$P(\text{accident}) = P(\text{accident} | \text{wet})P(\text{wet}) + P(\text{accident} | \text{dry})P(\text{dry})$$

$$= 0.01 * 0.2 + 0.002 * 0.8 = \boxed{0.0024}$$

\*\*\*\*\*

Answer Option 5:

We can solve this by using the total probability theorem. The total probability of an accident is the sum of the probability of an accident given dry weather times the probability of dry weather, and the probability of an accident given wet weather times the probability of wet weather.

$$P(\text{Accident}) = P(\text{Accident} | \text{Dry})P(\text{Dry}) + P(\text{Accident} | \text{Wet})P(\text{Wet})$$

$$= (0.002)(0.8) + (0.01)(0.2)$$

$$= 0.0016 + 0.002$$

$$= 0.0036$$

So, the probability of an accident on Highway 48 today is 0.0036. Final Answer: D

\*\*\*\*\*

Answer Option 6:

This is a classic problem of total probability. The total probability of an accident is the sum of the probability of an accident given the weather is dry and the probability of an accident given the weather is wet, each weighted by the probability of that weather condition. That is,

$$P(\text{Accident}) = P(\text{Dry}) * P(\text{Accident} | \text{Dry}) + P(\text{Wet}) * P(\text{Accident} | \text{Wet})$$

$$= 0.80 * 0.002 + 0.20 * 0.01$$

$$= 0.0016 + 0.002$$

$$= 0.0036$$

So the probability that there will be an accident on Highway 48 today is 0.0036. Final Answer: D.

\*\*\*\*\*

Answer Option 7:

We need to add the probability of an accident when the weather is dry to the probability of an accident when the weather is wet. The probability that the weather is dry is 80%, or 0.8, so the probability of an accident when the weather is dry is  $(0.8)(0.002) = 0.0016$ . The probability of an accident when the weather is wet is  $(0.2)(0.01) = 0.002$ . Adding these together gives  $0.0016 + 0.002 = 0.0036$ . So the probability that there will be an accident on Highway 48 today is 0.0036. Final answer: D

\*\*\*\*\*

Answer Option 8:

We can solve this problem using conditional probability. The probability of an accident happening is the sum of the probability of an accident happening given dry weather times the probability of dry weather and the probability of an accident happening given wet weather times the probability of wet weather.

In mathematical terms, this means the probability of an accident happening is:

$$P(\text{Accident}) = P(\text{Accident} \mid \text{Dry}) * P(\text{Dry}) + P(\text{Accident} \mid \text{Wet}) * P(\text{Wet}).$$

Given in the problem, we know that  $P(\text{Accident} \mid \text{Dry}) = 0.002$ ,  $P(\text{Accident} \mid \text{Wet}) = 0.01$ ,  $P(\text{Wet}) = 0.2$ . Therefore,  $P(\text{Dry}) = 1 - P(\text{Wet}) = 0.8$ . Plugging these values into our equation gives:

$$P(\text{Accident}) = 0.002 * 0.8 + 0.01 * 0.2 = 0.0016 + 0.002 = 0.0036.$$

So, there is a 0.0036 probability of an accident on Highway 48 today. Final Answer: D.

\*\*\*\*\*

Answer Option 9:

We can solve this problem using the total probability theorem. The total probability of an event A is the sum of the probabilities of A happening given each separate event in a partition of the sample space. In this case, our partition is the event that the weather is wet and the event that the weather is dry.

The probability of an accident given dry weather,  $P(A \mid D)$ , is 0.2% or 0.002. The probability of dry weather,  $P(D)$ , is 80% or 0.8, because if there's a 20% chance of wet weather, there's an 80% chance of dry weather.

The probability of an accident given wet weather,  $P(A \mid W)$ , is 1% or 0.01. The probability of wet weather,  $P(W)$ , is 20% or 0.2.

So, the total probability of an accident is  $P(A) = P(D)P(A \mid D) + P(W)P(A \mid W) = 0.8 * 0.002 + 0.2 * 0.01 = 0.0016 + 0.002 = 0.0036$ .

Final answer: D. '0.0036'

\*\*\*\*\*

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:

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.'

\*\*\*\*\*

Answer Option 2:

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.'

\*\*\*\*\*

Answer Option 3:

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.'

\*\*\*\*\*

Answer Option 4:

The binomial distribution models the number of successes in a fixed number of independent Bernoulli trials with the same probability of success. The best fit for this description among the given choices is: 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.' Here, each city (Bernoulli trial) independently has the weather being cloudy (success) or not cloudy (failure). Hence, the correct answer is B.

\*\*\*\*\*

Answer Option 5:

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.'

The binomial distribution is used when there are exactly two mutually exclusive outcomes of a trial (often referred to as success and failure). These outcomes need not be equally likely, and each trial is assumed to be independent of the others. The option B fulfills these requirements: for each city, it's either cloudy for most of the day (success) or not (failure), and the weather in each city can be considered as independent trials.

\*\*\*\*\*

Answer Option 6:

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.'

The binomial distribution model applies to a series of independent trials with two possible outcomes (success and failure), where the probability of success is the same for each trial. In this case, each city is an independent trial, and the two outcomes are 'cloudy' and 'not cloudy'. Hence, option B is the correct choice.

\*\*\*\*\*

Answer Option 7:

The binomial distribution model is appropriate for a random experiment where each trial has only two possible outcomes, and the trials are independent. Among the given options, the option 'The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock.' fits the criterion for a binomial distribution model as there are only two outcomes (either the driver is wearing a seat belt or not) and each trial (each driver being stopped) is independent. Hence, the answer is C.

\*\*\*\*\*

Answer Option 8:

The binomial distribution is appropriate for a statistical experiment where there are only two possible outcomes (success or failure) for each trial, the trials are independent, and the probability of success remains the same for each trial.

Option C, 'The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock.' fits these conditions. The two outcomes are wearing a seat belt (success) or not wearing a seat belt (failure). Each driver stopped is an independent trial, and the probability that a driver is wearing a seat belt should be approximately constant.

So, the correct final answer is C.

\*\*\*\*\*

Answer Option 9:

The binomial distribution describes the number of successes in a sequence of  $n$  independent experiments, each asking a yes-no question, and each with its own boolean-valued outcome: success (with probability  $p$ ) or failure (with probability  $q = 1 - p$ ).

Option C, 'The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock.' is a situation that fits this description. Here, the 'success' is a driver wearing a seat belt, and each driver is an independent experiment with a yes-no outcome (they're either wearing a seat belt or they're not). Therefore, the correct answer is 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 residual for a point in a regression analysis is the observed value minus the predicted value. In this case, the observed value of y is 6. The predicted value of y, based on the regression equation, is  $-2.29 + 1.70(5) = 5.21$ . Thus, the residual is  $6 - 5.21 = 0.79$ . However, none of the provided options match this result. Please check the problem statement or the options.

\*\*\*\*\*

Answer Option 2:

The residual for a point in a regression analysis is the observed y-value minus the predicted y-value. The predicted y-value is found by plugging the x-value into the regression equation. So the predicted y-value is  $-2.29 + 1.70(5) = 6.21$ . The observed y-value is 6. The residual is  $6 - 6.21 = -0.21$ . Thus the residual for the point (5, 6) is  $-0.21$ . Final Answer: D

\*\*\*\*\*

Answer Option 3:

The residual for a point in a least squares regression is given by the actual value minus the predicted value.

The predicted value of y given  $x = 5$  is:

$$y = -2.29 + 1.70(5) = 5.21$$

So, the residual is  $6 - 5.21 = 0.79$

However, none of the options match this value. It seems like there is a mistake in the options or the question. The correct answer should be '0.79'. There might be a typo in the question or the options.

\*\*\*\*\*

Answer Option 4:

The residual is the observed y-value minus the predicted y-value. The predicted y-value can be found by substituting  $x = 5$  into the regression equation:  $y = -2.29 + 1.70*5 = -2.29 + 8.5 = 6.21$ . Thus, the residual is  $6 - 6.21 = -0.21$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

The residual is the observed value minus the predicted value. The predicted value of y when  $x = 5$  is  $y = -2.29 + 1.70*5 = 5.21$ . Thus, the residual for the point (5,6) is  $6 - 5.21 = 0.79$ . None of the provided options match this value. There may be a mistake in the question or the options provided.

\*\*\*\*\*

Answer Option 6:

The residual of a point from the regression line is the observed value minus the predicted value. The observed value of y (the actual value from the data) is 6. The predicted value of y (the value that the regression equation predicts) is  $-2.29 + 1.70(5) = 6.21$ . The residual is then  $6 - 6.21 = -0.21$ . So the residual for the point (5, 6) is  $-0.21$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

The residual is the actual y value minus the y value predicted by the regression line. The y value predicted by the regression line when  $x = 5$  is  $-2.29 + 1.70*5 = 5.21$ . Thus, the residual is  $6 - 5.21 = 0.79$ .

However, none of the options match this value.

Let's double-check our computation.

We calculate the predicted y value as  $-2.29 + 1.70*5 = 5.21$ . The actual y value is 6. Then, the residual is  $6 - 5.21 = 0.79$ .

It seems there might be a mistake in the options provided as none of them match the calculated residual value.

\*\*\*\*\*

Answer Option 8:

The residual is the difference between the observed and predicted values of the dependent variable. For the point (5, 6), the predicted value of y (denoted as ■) is given by plugging  $x=5$  into the regression equation: ■ =  $-2.29 + 1.70*5 = 5.21$ . Therefore, the residual is  $y - \blacksquare = 6 - 5.21 = 0.79$ . None of the options match this value so there might be a mistake in the problem or the options. However, if we consider the question to ask for the error, which is the absolute value of the residual, then we get  $|-0.79| = 0.79$  which is still not in the options. The problem seems to have a mistake.

\*\*\*\*\*

Answer Option 9:

The residual for a data point in a regression analysis is the difference between the observed and predicted value of the dependent variable. In the given regression equation,  $y = -2.29 + 1.70x$ , we can substitute  $x = 5$  to predict the value of y:

$$\text{Predicted } y = -2.29 + 1.70*5 = 5.21$$

The observed value of y is given as 6. Therefore, the residual is:

$$\text{Residual} = \text{observed } y - \text{predicted } y = 6 - 5.21 = 0.79. \text{ This value is not in the given options, so let's check the calculation again.}$$

$$\text{Predicted } y = -2.29 + 1.70*5 = 5.21$$

$$\text{Residual} = \text{observed } y - \text{predicted } y = 6 - 5.21 = 0.79$$

It seems there was a miscalculation in the question options. The correct residual is 0.79, which is not in the provided answer choices. The closest available answer is C. '0.21'. However, this answer isn't accurate based on the calculation.

It's also possible to check other answer choices. If we choose D. '-0.21' as a residual, it means the predicted y value should be 6.21, not 5.21. And, indeed, if we calculate predicted y value using x=5 in the regression equation, we get:

$$\text{Predicted } y = -2.29 + 1.70 \cdot 5 = 6.21$$

In this case, the residual is:

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

So, the correct answer is D. '-0.21'. 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:

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 2:

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 3:

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 4:

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 5:

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 6:

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 7:

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 8:

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 9:

D. 'When using the sample standard deviation to estimate the population standard deviation, more variability is introduced into the sampling distribution of the statistic.'  
.....

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:

A. 'Following up with those who did not respond to the survey the first time'

\*\*\*\*\*

Answer Option 2:

A. 'Following up with those who did not respond to the survey the first time'

\*\*\*\*\*

Answer Option 3:

A. 'Following up with those who did not respond to the survey the first time'

\*\*\*\*\*

Answer Option 4:

A. 'Following up with those who did not respond to the survey the first time'. This might not reduce bias, as the results would still be skewed towards those who were either more eager to respond, or could respond more easily. Therefore the correct answer is A.

\*\*\*\*\*

Answer Option 5:

C. 'Using stratified random sampling rather than simple random sampling'

Stratified random sampling can actually reduce bias by ensuring representation from different subgroups within a population. However, this method also has the potential to introduce bias if the strata (or groups) are not chosen correctly. For instance, if certain groups are over-represented or under-represented, this could lead to a biased sample. Thus, while stratified random sampling can be a useful tool for reducing bias, it is not guaranteed to do so and may not be as effective as the other methods listed in reducing bias. Therefore, the answer is C.

\*\*\*\*\*

Answer Option 6:

A. 'Following up with those who did not respond to the survey the first time' - This reduces bias by ensuring that non-respondents, who may have differing opinions from respondents, are included in the survey results.

B. 'Asking questions in a neutral manner to avoid influencing the responses' - This reduces bias by avoiding leading questions or phrasing that may influence respondents' answers.

C. 'Using stratified random sampling rather than simple random sampling' - Stratified random sampling can help decrease bias by ensuring that subgroups of the population are adequately represented. However, it may not necessarily always reduce bias compared to simple random sampling, especially if the stratification criteria are not related to the survey topic.

D. 'Selecting samples randomly' - This reduces bias by ensuring each member of the population has an equal chance of being selected.

Thus, the answer is C. 'Using stratified random sampling rather than simple random sampling'. Final Answer: C.

\*\*\*\*\*

Answer Option 7:

A. 'Following up with those who did not respond to the survey the first time' - This can reduce non-response bias.

B. 'Asking questions in a neutral manner to avoid influencing the responses' - This can reduce response bias.

C. 'Using stratified random sampling rather than simple random sampling' - This can reduce bias by ensuring all groups are adequately represented in the sample.

D. 'Selecting samples randomly' - This should reduce bias by giving all members of the population an equal chance of being selected.

From the given options, 'Using stratified random sampling rather than simple random sampling' is the least likely to reduce bias as it might not be necessary in all contexts. While stratified random sampling can ensure all groups are adequately represented, it might introduce bias if the strata are not defined correctly. Also, in a population that is homogeneous or when the strata are not related to the measured attribute, simple random sampling might be as good or better.

Final Answer: C

\*\*\*\*\*

Answer Option 8:

A. 'Following up with those who did not respond to the survey the first time' - This reduces non-response bias, so it is likely to reduce bias in a sample survey.

B. 'Asking questions in a neutral manner to avoid influencing the responses' - This reduces response bias, so it is likely to reduce bias in a sample survey.

C. 'Using stratified random sampling rather than simple random sampling' - Stratified random sampling can reduce sampling bias, particularly when the population is heterogeneous, so it is likely to reduce bias in a sample survey.

D. 'Selecting samples randomly' - This is a fundamental part of reducing sampling bias, so it is likely to reduce bias in a sample survey.

Therefore, the correct answer is C. 'Using stratified random sampling rather than simple random sampling'. While stratified random sampling can sometimes reduce bias, it is not always the case and depends on the specifics of the population being surveyed. In particular, if the strata are not chosen appropriately, it could even increase bias. Therefore, it is the least likely of the options to reduce bias in a sample survey.

\*\*\*\*\*

Answer Option 9:

A. 'Following up with those who did not respond to the survey the first time' - This can help to reduce non-response bias, a problem that can occur when those who do not respond to a survey have different opinions or characteristics than those who do.

B. 'Asking questions in a neutral manner to avoid influencing the responses' - This can help to reduce response bias, which can occur when the way a question is asked influences the responses.

C. 'Using stratified random sampling rather than simple random sampling' - This can help to reduce sampling bias. Stratified random sampling can ensure that all groups within a population are adequately represented.

D. 'Selecting samples randomly' - Selecting samples randomly can help to reduce selection bias, which can occur when some members of the population are more likely to be selected for the sample than others.

All these options can potentially reduce bias in a survey. However, the option least likely to reduce bias is C: 'Using stratified random sampling rather than simple random sampling'. Although stratified random sampling can reduce bias if there are groups within a population that have different responses, if there is no substantial difference between these subgroups, it won't reduce bias any more than simple random sampling. 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:

Since for the second exam, every student studied for exactly 10 hours, there is no variability at all in the number of hours students studied for the second exam. Therefore, the two sets of scores are uncorrelated, and the correlation is  $\boxed{0}$ . Final answer: B

\*\*\*\*\*

Answer Option 2:

Because all the students studied the exact same number of hours for the second exam, there is no variation in the number of hours studied for the second exam. As a consequence of there being no variability in the second variable, there is no linear association between the two variables. Hence, the correlation between the two variables is  $\boxed{0}$ . Final answer: B.

\*\*\*\*\*

Answer Option 3:

Correlation measures the linear dependence between two variables. In this case, the number of hours each student studied for the first exam has no impact on the number of hours they studied for the second exam because every student studied the same amount, 10 hours, for the second exam. Therefore, the correlation between the hours studied for each exam is  $\boxed{0}$ . Final answer: B

\*\*\*\*\*

Answer Option 4:

Since the amount of time each student studied for the second exam is a constant (10 hours), there is no linear relationship between the amount of time the students studied for the first exam and the amount of time they studied for the second exam. Therefore, the correlation is  $\boxed{0}$ . Final Answer: B

\*\*\*\*\*

Answer Option 5:

The correlation between two variables measures the strength and direction of the relationship between them. However, for the second exam, every student studied exactly 10 hours. This means there is no variability in the number of hours studied for the second exam, and therefore, it is impossible to establish any kind of relationship, or correlation, with the number of hours studied for the first exam. Therefore, the correlation is  $\boxed{0}$ . Final Answer: B

\*\*\*\*\*

Answer Option 6:

The standard deviation of the number of hours the students studied for the second exam is 0, since everyone studied exactly 10 hours. Because the standard deviation is 0 (everyone studied the exact same amount), there can be no linear relationship, positive or negative, between the number of hours studied for each exam, so the correlation is  $\boxed{0}$ . Final Answer: B

\*\*\*\*\*

Answer Option 7:

The correlation between the numbers of hours students studied for each exam is 0. This is because correlation measures the linear relationship between two variables, and in this case, there is no linear relationship between the hours studied for the first exam and the hours studied for the second exam - the number of hours studied for the second exam is the same for all students, regardless of how much they studied for the first exam. Therefore, the correlation is  $\boxed{0}$ . Final Answer: B

\*\*\*\*\*

Answer Option 8:

The correlation between two variables measures the extent to which they linearly relate. In this case, all the students studied the same amount for the second exam, regardless of how much they studied for the first exam. This indicates no linear relationship between the number of hours studied for the first and second exams. Therefore, the correlation is  $\boxed{0}$ .

Final Answer: B

\*\*\*\*\*

Answer Option 9:

The correlation coefficient measures the strength and direction of a linear relationship between two variables. In this case, the number of hours studied for the second exam has no linear relationship with the number of hours studied for the first exam -- it doesn't matter how long a student studied for the first exam, they all studied the same amount for the second exam. This means that the correlation coefficient is  $\boxed{0}$ . Final Answer: B

\*\*\*\*\*

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:

Adding a constant to each data point of a data set does not change the spread of the data, hence the new standard deviation will remain \$23,000. Final answer: A.

\*\*\*\*\*

Answer Option 2:

Adding a constant to all data values increases their mean but doesn't affect the spread of the data. Therefore, the standard deviation remains the same, i.e., \$23000. So, the correct answer is A. '\$23,000'.

\*\*\*\*\*

Answer Option 3:

Since all the house values are increased by the same amount, the differences between the values (which is what standard deviation measures) will remain the same. Therefore, the new standard deviation will still be \$23,000. Final Answer: A

\*\*\*\*\*

Answer Option 4:

Changing all scores in a data set by adding a constant increases the mean by the constant, but it does not change the standard deviation. Therefore, the standard deviation of the new appraised values is the same as the old standard deviation, or  $\boxed{\$23,000}$ . Final answer: A

\*\*\*\*\*

Answer Option 5:

When all values in a data set are increased by a constant, the mean increases by that constant but the standard deviation does not change. So, the new standard deviation of the appraised values of houses in the city will still be  $\boxed{\$23,000}$ . Final answer: A

\*\*\*\*\*

Answer Option 6:

Increasing every value in a data set by a certain number (in this case, \$5,000) will increase the mean by that number, but the standard deviation will remain the same. The standard deviation measures how spread out the data is, and adding the same amount to every data point doesn't change the spread, just the location.

So, the new standard deviation of the appraised values will still be  $\boxed{\$23,000}$ .

Final answer: A

\*\*\*\*\*

Answer Option 7:

In a data set, adding a constant to each value increases the mean by the constant but does not change the standard deviation. This is because standard deviation is a measure of dispersion, and adding a constant value to each data point does not change the relative dispersion of the data. So, the new standard deviation of the appraised values of houses in the city will be  $\boxed{\$23,000}$ . Final answer: A

\*\*\*\*\*

Answer Option 8:

When all values in a data set are increased by the same amount, the standard deviation stays the same. Thus, the new standard deviation of the appraised values of houses is  $\boxed{23,000}$ . Final

Answer: A

\*\*\*\*\*

Answer Option 9:

The standard deviation is a measure of dispersion. When you add a constant to each data point, the dispersion of the data does not change. Therefore, the new standard deviation of the appraised values of the houses in the city 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:

The total salary paid to office assistants in New Jersey is  $\$28,500 \times 128 = \$3,648,000$ .

The total salary paid to office assistants in Utah is  $\$22,500 \times 32 = \$720,000$ .

The total number of office assistants is  $128 + 32 = 160$ .

The total salary paid to all office assistants is  $\$3,648,000 + \$720,000 = \$4,368,000$ .

The mean salary is therefore  $\$4,368,000 / 160 = \$27,300$ . Final Answer: D. '\$27,300 '

\*\*\*\*\*

Answer Option 2:

The total amount of salaries for the New Jersey office is  $\$28,500 \times 128 = \$3,648,000$ . For the Utah office, it is  $\$22,500 \times 32 = \$720,000$ . Therefore, the total amount of salaries for both offices is  $\$3,648,000 + \$720,000 = \$4,368,000$ . As there are  $128 + 32 = 160$  office assistants in total, the mean salary is  $\$4,368,000 / 160 = \$27,300$ . So, the correct answer is D. '\$27,300 '.

\*\*\*\*\*

Answer Option 3:

The total salary paid to the New Jersey office assistants is  $\$28,500 \times 128 = \$3,648,000$ . The total salary paid to the Utah office assistants is  $\$22,500 \times 32 = \$720,000$ . The total salary paid to all the office assistants is  $\$3,648,000 + \$720,000 = \$4,368,000$ . The total number of office assistants is  $128 + 32 = 160$ . The mean salary paid to the office assistants is  $\$4,368,000 \div 160 = \$27,300$ . Final answer: D. '\$27,300 '

\*\*\*\*\*

Answer Option 4:

The total paid to office assistants at the New Jersey office is  $\$28,500 \times 128 = \$3,648,000\$$ .

The total paid to office assistants at the Utah office is  $\$22,500 \times 32 = \$720,000\$$ .

The total number of office assistants is  $128 + 32 = 160\$$ .

So, the mean salary paid to the office assistants in the company is  $(\$3,648,000 + \$720,000) / 160 = \$27,300\$$ .

The final answer is  $\boxed{\$27,300}$  (Option D).

\*\*\*\*\*

Answer Option 5:

To find the overall mean salary of office assistants, we need to find the total salary paid to office assistants and divide it by the total number of office assistants. The total salary paid to New Jersey office assistants is  $\$28,500 \times 128 = \$3,648,000$ , and the total salary paid to Utah office assistants is  $\$22,500 \times 32 = \$720,000$ . The total salary paid to all office assistants is  $\$3,648,000 + \$720,000 = \$4,368,000$ .

The total number of office assistants is  $128 + 32 = 160$ . Therefore, the mean salary is  $\$4,368,000 / 160 = \$27,300$ . So, the correct answer is D. '\$27,300 '.

\*\*\*\*\*

Answer Option 6:

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  $\$22,500 \times 32 = \$720,000$ . So, the total salary paid to office assistants in the company is  $\$3,648,000 + \$720,000 = \$4,368,000$ . The company has a total of  $128 + 32 = 160$  office assistants. So, the mean salary paid to the office assistants in the company is  $\$4,368,000 / 160 = \$27,300$ . Final Answer: D

\*\*\*\*\*

Answer Option 7:

The total salary paid to the office assistants in the New Jersey office is  $\$28,500 \times 128 = \$3,648,000\$$  and 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  $128 + 32 = 160\$$  and the total salary paid to all the office assistants is  $\$3,648,000 + \$720,000 = \$4,368,000\$$ . Therefore, the mean salary paid to the office assistants in this company is  $\$4,368,000 \div 160 = \boxed{\$27,300}$  \$.

Final answer: D

\*\*\*\*\*

Answer Option 8:

The total salary paid to the office assistants in the New Jersey office is  $\$28,500 \times 128 = \$3,648,000$  and the total salary paid to the office assistants in the Utah office is  $\$22,500 \times 32 = \$720,000$ . So, the total salary paid to all the office assistants is  $\$3,648,000 + \$720,000 = \$4,368,000$ . The total number of office assistants is  $128 + 32 = 160$ . Therefore, the mean salary paid to the office assistants in this company is  $\$4,368,000 / 160 = \$27,300$ . So, the final answer is D. '\$27,300 '.

\*\*\*\*\*

Answer Option 9:

To find the mean salary, we need to find the total salary paid to all of the office assistants and then divide by the number of office assistants.

The total salary paid to the New Jersey office assistants is  $128 \times \$28,500 = \$3,648,000$  and the total salary paid to the Utah office assistants is  $32 \times \$22,500 = \$720,000$ . The total salary paid to all office assistants is  $\$3,648,000 + \$720,000 = \$4,368,000$ .

The total number of office assistants is  $128 + 32 = 160$ .

The mean salary paid to the office assistants in this company is  $\$4,368,000 / 160 = \$27,300$ .

So, the correct answer is D. '\$27,300'.

\*\*\*\*\*

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:

A. 'A sample must be reasonably large to be properly considered a simple random sample.'

\*\*\*\*\*

Answer Option 2:

A. 'A sample must be reasonably large to be properly considered a simple random sample.'

\*\*\*\*\*

Answer Option 3:

Option A is false. A simple random sample doesn't have to be large. The main characteristic of a simple random sample is that every member of the population has an equal chance of being chosen, not the size of the sample. Final answer: A.

\*\*\*\*\*

Answer Option 4:

The false statement is A. 'A sample must be reasonably large to be properly considered a simple random sample.' The size of a simple random sample does not determine whether it is a simple random sample or not. What matters is that every member of the population has an equal chance of being included in the sample, not the size of the sample.

\*\*\*\*\*

Answer Option 5:

A. 'A sample must be reasonably large to be properly considered a simple random sample.' This is a false statement. A simple random sample is defined by every member of the set being equally likely to be chosen, not by the size of the sample. Even a sample size of 1 could technically be considered a simple random sample. Therefore, the correct answer is A.

\*\*\*\*\*

Answer Option 6:

A. 'A sample must be reasonably large to be properly considered a simple random sample.' is a false statement. The size of the sample doesn't determine whether it's a simple random sample or not. A sample is a simple random sample if every member of the population has an equal chance of being selected, regardless of the sample size. Final answer: A.

\*\*\*\*\*

Answer Option 7:

A. 'A sample must be reasonably large to be properly considered a simple random sample.'

This statement is false. A simple random sample does not necessarily have to be large. It just needs to be random, meaning every subset of the population has an equal chance of being selected, regardless of its size. The other choices are characteristics of simple random samples. So, the final answer is A.

\*\*\*\*\*

Answer Option 8:

A. 'A sample must be reasonably large to be properly considered a simple random sample.' is a false statement. The size of a sample does not determine whether it is a simple random sample or not. A simple random sample is one in which every member of the population has an equal chance of being selected, regardless of the sample size. Final Answer: A

\*\*\*\*\*

Answer Option 9:

The answer is A. 'A sample must be reasonably large to be properly considered a simple random sample.' This statement is false. The size of a sample isn't what characterizes it as a simple random sample. A sample is a simple random sample if every member of the population has an equal chance of being included in the sample, regardless of the sample's size. Even a small sample can be a simple random sample if it's selected correctly.

\*\*\*\*\*

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:

A. 'Her procedure makes use of chance.'

=====

Answer Option 2:

A. 'Her procedure makes use of chance.'

=====

Answer Option 3:

A. 'Her procedure makes use of chance.'

=====

Answer Option 4:

A. 'Her procedure makes use of chance.'

=====

Answer Option 5:

A. 'Her procedure makes use of chance.'

=====

Answer Option 6:

A. 'Her procedure makes use of chance.'

=====

Answer Option 7:

A. 'Her procedure makes use of chance.'

=====

Answer Option 8:

A. 'Her procedure makes use of chance.'

=====

Answer Option 9:

A. 'Her procedure makes use of chance.'

=====

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:

The expected return for Option II is  $0.5 * \$50,000 + 0.5 * \$10,000 = \$30,000$ , which is larger than the return for Option I. Thus, Option II maximizes the expected return.

If the company needs at least \$20,000 to pay off an overdue loan, Option I is the only choice that guarantees at least \$20,000 return.

Option III only has a 5% chance of returning \$100,000 and no chance of returning more than \$80,000. Therefore, if the company needs at least \$80,000 to pay off an overdue loan, Option III is not a suitable choice.

Therefore, the correct answer is D. 'All of the above answers are correct.'

\*\*\*\*\*

Answer Option 2:

The expected return of option II is  $\$(0.5 * 50000) + (0.5 * 10000) = 30000\$$ , which is higher than the certain return of option I. So, option II would be preferred if the company wants to maximize expected return.

Option I ensures a return of \$25000, which is higher than the minimum amount of return the company needs to pay off its loan in the second statement. So, option I is preferred in this case.

Since option III only gives a 5% chance of returning \$100000 and no return in other cases, it is not preferred over option I or II, by the company that needs to pay off an overdue loan of at least \$80000.

So, the correct answer is A and B, hence the final answer: D. 'All of the above answers are correct.'

\*\*\*\*\*

Answer Option 3:

The expected return of Option I is \$25,000.

The expected return of Option II is  $\$(0.5)*50,000 + (0.5)*10,000 = \$30,000$ .

The expected return of Option III is  $\$(0.05)*100,000 + (0.95)*0 = \$5,000$ .

Therefore, Option II has the highest expected return.

If the company has an overdue loan of \$20,000, it should choose Option I since it is the only one that guarantees at least \$20,000.

If the company has an overdue loan of \$80,000, none of the options can guarantee this amount, so none of the options would be appropriate to choose.

Thus, the correct answer is 'Option II if it wants to maximize expected return' and 'Option I if it needs at least \$20,000 to pay off an overdue loan'. Therefore, the final answer is D. 'All of the above answers are correct.'

\*\*\*\*\*

Answer Option 4:

The expected return of each option is as follows:

Option I: \$25,000.

Option II:  $\$0.5 * \$50,000 + 0.5 * \$10,000 = \$30,000$ .

Option III:  $\$0.05 * \$100,000 + 0.95 * \$0 = \$5,000$ .

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 a loan, it should not choose Option III because there's a 95% chance it will return nothing. Option I is the only option that guarantees a return of at least \$20,000, so the company should choose Option I in this case.

Option III only has a 5% chance of returning \$100,000 and cannot guarantee a return of \$80,000 to pay off a loan. Therefore, it is not the correct choice if the company needs at least \$80,000 to pay off a loan.

So, the correct answer is D. 'All of the above answers are correct.'

\*\*\*\*\*

Answer Option 5:

The expected return for each option can be calculated as follows:

Option I: The expected return is a guaranteed \$25,000.

Option II: The expected return is  $\$50,000*0.5 + \$10,000*0.5 = \$30,000$ .

Option III: The expected return is  $\$100,000*0.05 + \$0*0.95 = \$5,000$ .

A. 'Option II if it wants to maximize expected return' - This is correct, as Option II has the highest expected return of \$30,000.

B. 'Option I if it needs at least \$20,000 to pay off an overdue loan' - This is correct, as Option I is the only one that guarantees a return of at least \$20,000.

C. 'Option III if it needs at least \$80,000 to pay off an overdue loan' - This is incorrect, as there is only a 5% chance of getting a return of \$100,000 and a 95% chance of getting nothing.

Therefore, the final answer is D. 'All of the above answers are correct.'

\*\*\*\*\*

Answer Option 6:

First, we calculate the expected return for each option. The expected return is the sum of the possible returns, each multiplied by its probability of occurring.

Option I: The expected return is \$25,000 because this is the only possible outcome.

Option II: The expected return is  $\$0.5 \times 50,000 + 0.5 \times 10,000 = \$30,000$ .

Option III: The expected return is  $\$0.05 \times 100,000 + 0.95 \times 0 = \$5,000$ .

So, 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 can't risk getting nothing or only \$10,000, so it should choose Option I.

If the company needs at least \$80,000 to pay off an overdue loan, none of the options can guarantee this amount.

So, the correct answer is 'Option II if it wants to maximize expected return' and 'Option I if it needs at least \$20,000 to pay off an overdue loan'. Therefore, the final answer is B. 'Option I if it needs at least \$20,000 to pay off an overdue loan'.

\*\*\*\*\*

Answer Option 7:

The expected return of Option II is  $(0.5 \times 50000) + (0.5 \times 10000) = 30000$ . The expected return of Option III is  $(0.05 \times 100000) + (0.95 \times 0) = 5000$ . Thus, if the company wants to maximize expected return, it should choose Option II. If the company needs at least \$20000 to pay off an overdue loan, it should choose Option I to guarantee a return of \$25000. Option III only gives a 5% chance of returning \$100000 and a 95% chance of returning nothing, so it's not a good choice if the company needs at least \$80000. Therefore, the correct answer is 'Option II if it wants to maximize expected return' and 'Option I if it needs at least \$20000 to pay off an overdue loan', so the final answer is D. 'All of the above answers are correct.'

\*\*\*\*\*

Answer Option 8:

The expected return of an investment option is a measure of its average outcome and is calculated by multiplying each possible outcome by its probability and adding these values.

Option I: The expected return is \$25,000 because it's a sure return.

Option II: The expected return is  $\$(0.5 \times \$50,000) + (0.5 \times \$10,000) = \$30,000$ .

Option III: The expected return is  $\$(0.05 \times \$100,000) + (0.95 \times \$0) = \$5,000$ .

If the company wants to maximize its expected return, it should choose Option II.

If the company needs at least \$20,000 to pay off an overdue loan and wants to be sure of having this amount, it should choose Option I.

Option III is risky because it has a 95% chance of returning nothing. So, if the company needs at least \$80,000, it cannot count on Option III.

Hence, the correct answer is: 'Option II if it wants to maximize expected return' and 'Option I if it needs at least \$20,000 to pay off an overdue loan'. The final answer is A and B. But since the options do not provide such a choice, the correct answer is 'None of the above'. We cannot choose any of the given options A, B, C, or D.

\*\*\*\*\*

Answer Option 9:

The expected return for Option II is  $\$(0.5 \times 50000) + (0.5 \times 10000) = 30000\$$ . The expected return for Option III is  $\$(0.05 \times 100000) + (0.95 \times 0) = 5000\$$ . 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 a loan, it should choose Option I because it guarantees a return of \$25,000, unlike the other two options.

If the company needs at least \$80,000 to pay off a loan, none of the options are guaranteed to return that much. Even though Option III offers the possibility of \$100,000, it only has a 5% chance of happening.

Therefore, the company should not choose any of the options if it needs at least \$80,000.

So, the correct answer is 'Option II if it wants to maximize expected return' and 'Option I if it needs at least \$20,000 to pay off an overdue loan'. Therefore, the final answer is 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:

For a one-sided test with a sample size of 14 (meaning the degrees of freedom is 13), a P-value of 0.075 corresponds to a t-value of approximately 1.771 (from t-distribution tables). Any t-value larger than this would yield a smaller P-value. So, the range of t-values that would have given a P-value of 0.075 is C. ' $1.771 < t < 2.160$ '.

\*\*\*\*\*

Answer Option 2:

The t-value corresponding to a P-value of 0.075 with degrees of freedom 13 (which is  $14-1$ ) is approximately 1.771. Since this is a one-sided test and the P-value is the probability of observing a test statistic as extreme or more extreme than the observed one, the t-values that would have given this P-value are greater than 1.771. So, the best range of t values would be ' $1.771 < t < 2.160$ '. Final Answer: C.

\*\*\*\*\*

Answer Option 3:

From a t-table or t-distribution calculator, we can find a P-value of 0.075 corresponds to a t-value of approximately 1.771 for a sample size of 14. Because it's a one-sided test, the t-values that would result in a P-value of 0.075 or less fall on one end of the distribution. Therefore, the correct range of t-values is anything greater than this critical t-value. Thus, the range would be C. ' $1.771 < t < 2.160$ '.

\*\*\*\*\*

Answer Option 4:

The p-value is the probability that a random variable T with a student's t-distribution, with the given degrees of freedom (here,  $14-1=13$ ), has  $T > t$ . So we need to find the t-value such that  $\Pr(T > t) = 0.075$ . The best tool to use for this is a t-distribution table. Looking at the row with 13 degrees of freedom, we find that this value lies between 1.771 (which gives a p-value of 0.05) and 1.350 (which gives a p-value of 0.1). So the answer is D. Final Answer: D.

\*\*\*\*\*

Answer Option 5:

If the P-value is 0.075, this means that 7.5% of values lie to the right of the observed test statistic in a one-tailed test. Looking up 0.075 in the t-table for 13 degrees of freedom ( $14-1$ ), we see a t-value of approximately 1.771. Because we only consider values to the right of this, the range should start with this value. The t-table doesn't give us a maximum t-value, but in reality, the t-distribution extends indefinitely. So, the best possible range is ' $1.771 < t < \infty$ '. However, since this option is not given, the closest provided option is ' $1.771 < t < 2.160$ '.

\*\*\*\*\*

Answer Option 6:

Since it's a one-sided test and using the t-distribution table, the P-value of 0.075 corresponds to a t-value of approximately 1.771 for a sample size of 14 (degrees of freedom are 13). This is the critical t-value that separates values that would lead to rejecting the null hypothesis from those that do not.

Therefore, any observed t-value higher than this would lead to rejecting the null hypothesis at this significance level. Thus, the range of t-values that would have given this P-value is any value greater than 1.771.

So, the correct answer is C. ' $1.771 < t < 2.160$ '.

\*\*\*\*\*

Answer Option 7:

A one-sided test with a P-value of 0.075 means that 7.5% of the t-distribution is above the observed t-value (since it's one-sided). For a t-distribution with 13 degrees of freedom ( $n-1 = 14-1 = 13$ ), a P-value of 0.075 corresponds to a t-value of approximately 1.771 (using a t-table or statistical software). Therefore, t-values that would yield this P-value in a one-sided test would need to be greater than this. The only range that fits this criterion is C. ' $1.771 < t < 2.160$ '. Final answer: C

\*\*\*\*\*

Answer Option 8:

The P value is the probability that a particular statistical test will produce a result as extreme as the one you observed, assuming the null hypothesis is true. In this case, the P value is 0.075, which indicates that there's a 7.5% chance the results could be due to random chance (assuming the null hypothesis is true).

The t-value is a standardized statistic that measures the degree of difference relative to the variation in your data sets. It's a way to quantify the difference between the two groups and express it in terms of standard deviation.

In a one-sided test, we're looking at whether the population mean is greater than or less than a certain value, not simply different.

For a one-tailed test, with a significance level of 0.075 and degrees of freedom of 13 ( $14-1$ ), the critical t-value is approximately 1.771.

So, the range of t-values that would have given a P-value of 0.075 in a one-sided test is any t-value greater than 1.771. Thus, the correct answer is C. ' $1.771 < t < 2.160$ '.

\*\*\*\*\*

Answer Option 9:

A P-value of 0.075 indicates that the observed test statistic is in the tail of the distribution, suggesting that the null hypothesis would be rejected. For a t-distribution with 13 degrees of freedom (sample size 14 minus 1), a one-tailed test with a P-value of 0.075 corresponds to a t-value of approximately 1.771 (you can look this up in a t-distribution table or use a statistical calculator).

Since it's a one-sided test, we're interested in the tail of the distribution--the range of t-values greater than this critical value. Therefore, the correct range of t-values would be ' $1.771 < t < \infty$ '. However, since this option is not available, the closest available option is ' $1.771 < t < 2.160$ '.

Final Answer: C

\*\*\*\*\*

Question 31:

'The president of an online music streaming service whose customers pay a fee wants to gather additional information about customers who have joined in the past 12 months. The company plans to send out an e-mail survey to a sample of current customers with a link that gives participants a month of streaming service for free once the survey has been completed. They know that musical tastes vary by geographical region. Which of the following sample plans would produce the most representative sample of its customers?', A. 'Choose all of the customers who joined in the last month.', B. 'Make a list of all the customers who joined in the last 12 months and choose a random sample of customers on this list.', C. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 10 customers from each state.', D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

=====

Answer Option 1:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

Answer Option 2:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

Answer Option 3:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

Answer Option 4:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

Answer Option 5:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

Answer Option 6:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

Answer Option 7:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

Answer Option 8:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

Answer Option 9:

D. 'From the list of all customers who joined in the last 12 months, classify customers by the state in which they live, then choose 3% of the customers from each state.'

\*\*\*\*\*

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:

The Z score for a 94% confidence interval is 1.88. Final answer: D

=====

Answer Option 2:

A 94% confidence interval corresponds to an upper critical value of z equal to 1.88. This is because, in a standard normal distribution, approximately 94% of the values lie within 1.88 standard deviations of the mean. Final answer: D.

=====

Answer Option 3:

For a 94% confidence interval, the remaining 6% is split equally on both ends of the distribution. This means you are looking for a z-score that cuts off the top 3% of the distribution. The z-score that does this is '1.88'. Final Answer: D

=====

Answer Option 4:

The Z-score for a 94% confidence interval can be found using the standard normal distribution table or a Z-score calculator. The Z-score that corresponds to a 94% confidence interval is approximately 1.88.

Therefore, the upper critical value of z to be used in constructing this interval is 1.88. Final answer: D.

=====

Answer Option 5:

A 94% confidence interval corresponds to an alpha level of 0.06 ( $100\% - 94\% = 6\%$ ), which leaves 3% in each tail of the distribution (since confidence intervals are two-tailed). Looking this up in a standard Z-table, or using an inverse normal calculator, we find the critical value of z to be approximately 1.88. Final answer: D.

=====

Answer Option 6:

The critical value for a 94% confidence interval is found by determining the Z-score that leaves 3% in each tail of the standard normal distribution (because  $100\% - 94\% = 6\%$  and this is split evenly in the two tails). The Z-score that leaves 3% in the tail is approximately 1.88. So the correct answer is D. '1.88'.

=====

Answer Option 7:

For a 94% confidence interval, the critical value of z corresponds to having 3% in each tail of the standard normal distribution (because  $100\% - 94\% = 6\%$  and this is split across the two tails). This is equivalent to having 97% of the data to the left, because  $100\% - 3\% = 97\%$ . The z-score that corresponds to this percentile is approximately '1.88'. Final Answer: D

=====

Answer Option 8:

A 94% confidence interval is equivalent to a significance level of 0.06 ( $1 - 0.94 = 0.06$ ). Since the confidence interval is two-tailed, the critical value of z corresponds to the 97th percentile ( $100 - 0.06/2 * 100 = 97$ ). The z-score corresponding to the 97th percentile is approximately 1.88. Hence, the upper critical value of z to be used in constructing this interval is 1.88. Final answer: D

=====

Answer Option 9:

For a 94% confidence interval, the area in each tail is  $(100\% - 94\%)/2 = 3\%$ . Using a Z-table, a tail area of 3% corresponds to a Z-value of approximately 1.88. So, the upper critical value of Z for a 94% confidence interval is '1.88'. Final answer: D

=====

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:

B. 'A two-sample t-test of population means'

=====

Answer Option 2:

B. 'A two-sample t-test of population means'

=====

Answer Option 3:

B. 'A two-sample t-test of population means'

=====

Answer Option 4:

B. 'A two-sample t-test of population means'

=====

Answer Option 5:

B. 'A two-sample t-test of population means'

=====

Answer Option 6:

B. 'A two-sample t-test of population means'

=====

Answer Option 7:

B. 'A two-sample t-test of population means'

=====

Answer Option 8:

B. 'A two-sample t-test of population means'

=====

Answer Option 9:

B. 'A two-sample t-test of population means'

=====

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:

By adding 50 cents to each item, the mean price will increase by 50 cents (since the average of a set of numbers is changed by the same amount as each individual number). The range will also increase by 50 cents because both the maximum and minimum prices were increased by 50 cents. The standard deviation will not be affected because standard deviation measures the spread of the data, and adding a constant to each data point does not change the spread of data. Thus, 'The mean price and the range of prices will increase by 50 cents.' is correct. Final answer: A

\*\*\*\*\*

Answer Option 2:

Since all items' prices increase by the same amount (50 cents), the mean price will increase by 50 cents as well (from \$14.75 to \$15.25). The range of the prices will also increase by 50 cents (from \$29.99 - \$3.99 = \$26 to \$30.49 - \$4.49 = \$26.50). However, since each price increases by the same constant amount, the standard deviation, which is a measure of the spread of the prices, will remain the same.

Thus, the correct answer is 'The mean price and the range of prices will increase by 50 cents.'. Final Answer: A

\*\*\*\*\*

Answer Option 3:

By increasing the price of each item by 50 cents, the mean price will also increase by 50 cents. The range of prices will also increase by 50 cents (from \$29.99-\$3.99 = \$26.00 to \$30.49-\$4.49 = \$26.50). But adding a constant to each data point in a set does not change the standard deviation of the set, because standard deviation measures dispersion and the dispersion here does not change. The correct answer is A. 'The mean price and the range of prices will increase by 50 cents.'

\*\*\*\*\*

Answer Option 4:

The mean price will increase by 50 cents because each item's price increased by that amount, increasing the total price of all items (which is divided by the number of items to get the mean) by a fixed amount. The range of prices will also increase by 50 cents, since both the minimum and maximum prices increased by this amount. However, the standard deviation will not change, as this measures the spread of the prices from the mean, which will not be affected by a uniform price increase. Hence, the correct answer is: A. 'The mean price and the range of prices will increase by 50 cents.'

\*\*\*\*\*

Answer Option 5:

Since the price of every item is increasing by fifty cents, the mean price (average) will increase by fifty cents. The range of the prices (which is the difference between the highest and lowest prices) will also increase by fifty cents, since both the highest and lowest prices are increasing by this amount. However, the standard deviation (which measures the dispersion of the prices) will not change, because the prices are all shifting by the same amount. Therefore, the correct answer is 'The mean price and the range of prices will increase by 50 cents.' Final answer: A.

\*\*\*\*\*

Answer Option 6:

When the kiosk owner decides to increase the prices of all items by 50 cents, each price increases by the same amount. Hence, the mean price also increases by 50 cents because the mean is the sum of all prices divided by the total number of items. The range of prices, defined as the difference between the maximum and minimum prices, remains the same because both minimum and maximum prices increase by 50 cents. The standard deviation, a measure of the spread of data, remains the same because the relative positions of the prices do not change; they all shift up by the same amount. Therefore, the correct answer is 'The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.' Final answer: D.

\*\*\*\*\*

Answer Option 7:

Adding a constant of 50 cents to each price will cause the mean price and the range of prices to increase by 50 cents, but the standard deviation will not be affected because the distances between each price and the mean price will remain the same. Therefore, the correct answer is A. 'The mean price and the range of prices will increase by 50 cents.'

\*\*\*\*\*

Answer Option 8:

The answer is D. 'The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.'

When the owner increases the prices of all items by 50 cents, the mean price of the items will also increase by 50 cents since the mean is the sum of all prices divided by the number of items. Therefore, the mean price will increase from \$14.75 to \$15.25.

However, the standard deviation, which measures the dispersion or spread of the data values from the mean, will not change. This is because all prices are increased by the same amount, preserving their relative positions to each other and the mean. Thus, the dispersion of the data is not affected.

\*\*\*\*\*

Answer Option 9:

D. 'The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.'

Explanation: Since every single item's price has increased by 50 cents, this will also increase the mean price by 50 cents. However, the standard deviation of prices (a measure of the spread of the prices) will remain the same because the increase is uniform and doesn't change the relative distance between prices. The range of prices will also increase by 50 cents.

\*\*\*\*\*



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:

A. 'This November, the average donation is \$155.'

=====

Answer Option 2:

A. 'This November, the average donation is \$155.'

=====

Answer Option 3:

A. 'This November, the average donation is \$155.'

=====

Answer Option 4:

A. 'This November, the average donation is \$155.'

=====

Answer Option 5:

The correct statement is A. 'This November, the average donation is \$155.' The other options involve misinterpretations of the average and standard deviation. Option B is incorrect because it confuses the average with the median. Options C and D are incorrect because they apply a confidence interval to the data, which is something that is used for samples to estimate population parameters, not for a complete set of data.

Final Answer: A

=====

Answer Option 6:

The appropriate statement is A. 'This November, the average donation is \$155.' The other options involve misconceptions about the interpretation of standard deviation and confidence intervals. Option B is incorrect because the average does not necessarily represent the median (the point at which 50% of donations are more or less). Options C and D are incorrect because the standard deviation does not provide a confidence interval for the average donation. Confidence intervals are more complex statistical constructs that require additional assumptions and calculations.

=====

Answer Option 7:

The statement 'This November, the average donation is \$155.' is the most appropriate one. The average donation is a factual statement and does not involve any uncertainty or prediction. The other options involve misconceptions about what the average and standard deviation represent. The average donation is not necessarily the median donation (making option B incorrect), and the range of one or two standard deviations from the mean (options C and D) refers to the spread of individual data points, not the possible range of the average itself. Final Answer: A

=====

Answer Option 8:

The statement that is most appropriate is A. 'This November, the average donation is \$155.' The other statements make assumptions about the distribution of the donations that are not supported by the data given. The average is simply the total amount of donations divided by the number of donations, and does not tell us anything about how those donations were distributed. Therefore, we can't say that 50% of donations were more than \$155 (as in option B) or make any statement about a confidence interval (as in options C and D) without further information. Final answer: A

=====

Answer Option 9:

While all the statements may seem reasonable, the most appropriate statement is A. 'This November, the average donation is \$155.'

The standard deviation gives us a measure of the variability in the data, it does not allow us to make precise statements about the confidence interval for the average donation. Statement B is incorrect as the average does not necessarily represent the midpoint of all donations. Statements C and D are incorrect as they assume a normal distribution and apply the concept of a confidence interval incorrectly.

Confidence intervals are used to estimate population parameters, not to make claims about individual observations or specific samples. The given data does not provide sufficient information to make these claims. Therefore, the correct answer is A.

=====



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 Central Limit Theorem states all three of these statements. Therefore, the correct answer is not listed.

\*\*\*\*\*

Answer Option 2:

The Central Limit Theorem (CLT) states that the distribution of sample means will approximate a normal distribution as the sample size becomes larger, regardless of the shape of the population distribution.

This is stated in statement III. Statement I is also correct; the mean of a sampling distribution is indeed equal to the mean of the population from which the samples were taken. Statement II accurately describes the standard deviation of the sampling distribution. Therefore, all three statements correctly describe the Central Limit Theorem. However, none of the options matches this description. It seems there's a mistake in the provided options. The correct answer would be 'I, II & III'.

\*\*\*\*\*

Answer Option 3:

The central limit theorem suggests three things:

I. The mean of the sampling distribution ( $\mu_{\bar{x}}$ ) is equal to the mean of the population ( $\mu$ ). This is correct.

II. The standard deviation of the sampling distribution ( $\sigma_{\bar{x}}$ ) is equal to the standard deviation of the population ( $\sigma$ ) divided by the square root of the sample size ( $n$ ). This is also correct.

III. If the sample size is large (typically  $n > 30$ ), the sampling distribution of  $\bar{x}$  is approximately normally distributed, regardless of the shape of the population distribution. This statement is also correct.

So, all the three statements I, II and III are correct descriptions of the Central Limit Theorem. However, none of the given options include all three statements. There seems to be an error in the question or the options.

\*\*\*\*\*

Answer Option 4:

The Central Limit Theorem states all three conditions:

I. The mean of the sampling distribution is the same as the mean of the population.

II. The standard deviation of the sampling distribution is 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 is approximately normal.

So none of the options A, B, C, or D are correct.

\*\*\*\*\*

Answer Option 5:

The Central Limit Theorem states all three of these points. So the answer is not mentioned in the given options. The Central Limit Theorem states that the sum of a large number of independent and identically distributed measurements will have an approximately normal distribution, regardless of the shape of the original measurement distribution. This is assuming that the mean and the variance are finite. The theorem also indicates that the mean of this distribution will be equal to the mean of the original population (I), the standard deviation will be equal to the standard deviation of the original distribution divided by the square root of the sample size (II), and the distribution will approach a normal distribution as the sample size increases (III).

\*\*\*\*\*

Answer Option 6:

The Central Limit Theorem primarily states that if you have a population with mean  $\mu$  and standard deviation  $\sigma$  and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed. This will hold true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large (usually  $n > 30$ ). If the population is normal, then the theorem holds true even for samples smaller than 30.

In other words, statement III correctly describes the Central Limit Theorem.

Statement I is also correct, as the expected value (mean) of the sample means will be the population mean.

Statement II correctly describes the standard deviation of the sampling distribution (known as the standard error), which is indeed the population standard deviation divided by the square root of the sample size.

So, all the statements correctly describe aspects of the Central Limit Theorem. However, if we need to choose the one that most directly describes the theorem itself, it would be statement III - which states the theorem's main point about the distribution of sample means.

So, the correct answer is D. 'III only'.

\*\*\*\*\*

Answer Option 7:

The Central Limit Theorem (CLT) has three main parts:

I. The mean of the sampling distribution of  $\bar{x}$  (the sample mean) is the same as the mean of the population. This is true, as the CLT states that the expectation (mean) of the sampling distribution of  $\bar{x}$  is the population mean.

II. The standard deviation of the sampling distribution of  $\bar{x}$  is the standard deviation of the population divided by the square root of the sample size. This is also true, as according to the CLT, the standard deviation (also known as the standard error) of the sampling distribution is the population standard deviation 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. This is the main point of the CLT, which states that if the sample size is large enough, the sampling distribution of  $\bar{x}$  will be approximately normally distributed, regardless of the shape of the population distribution.

Therefore, all three statements I, II, and III are part of the Central Limit Theorem. However, the answer choices do not provide this option. The closest correct answer would be B. 'I & II only'. Unfortunately, none of the options accurately represent the Central Limit Theorem.

\*\*\*\*\*

Answer Option 8:

The Central Limit Theorem (CLT) states that:

(I) The mean of the sampling distribution of the sample mean ( $\bar{x}$ ) is the same as the mean of the population.

(II) The standard deviation of the sampling distribution of the sample mean ( $\bar{x}$ ) is the standard deviation of the population divided by the square root of the sample size.

(III) If the sample size is large enough, the shape of the sampling distribution of the sample mean ( $\bar{x}$ ) is approximately normal, regardless of the shape of the population distribution.

So all three statements (I, II, III) describe the Central Limit Theorem. However, none of the provided options (A, B, C, D) reflect this. This question seems to be flawed as it does not provide a correct answer choice. It might be a good idea to consult the source of this question for clarification or possible correction.

\*\*\*\*\*

Answer Option 9:

The central limit theorem describes the distribution of sample means. It states that:

I. The mean of the sampling distribution of  $\bar{x}$  (the sample mean) is the same as the mean of the population. This is accurate.

II. The standard deviation of the sampling distribution of  $\bar{x}$  is the standard deviation of the population divided by the square root of the sample size. This is also accurate.

III. If the sample size is large (usually considered to be greater than or equal to 30), the shape of the sampling distribution of  $\bar{x}$  is approximately normal, regardless of the shape of the population distribution. This is also true.

So, all three statements are correct descriptions of the central limit theorem. However, the options do not include an 'All of the above' choice, so the options may be incorrect. But if we are to choose from the given options, the closest would be B. 'I & II only', although it should also include III.

\*\*\*\*\*

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:

C. 'It is the difference between the same statistics computed from two different samples.'

\*\*\*\*\*

Answer Option 2:

C. 'It is the difference between the same statistics computed from two different samples.'

\*\*\*\*\*

Answer Option 3:

C. 'It is the difference between the same statistics computed from two different samples.'

\*\*\*\*\*

Answer Option 4:

C. 'It is the difference between the same statistics computed from two different samples.'

\*\*\*\*\*

Answer Option 5:

C. 'It is the difference between the same statistics computed from two different samples.'

\*\*\*\*\*

Answer Option 6:

C. 'It is the difference between the same statistics computed from two different samples.'

\*\*\*\*\*

Answer Option 7:

B. 'It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.'

\*\*\*\*\*

Answer Option 8:

B. 'It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.'

\*\*\*\*\*

Answer Option 9:

B. 'It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.'

\*\*\*\*\*

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 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.'

\*\*\*\*\*

Answer Option 2:

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.'

\*\*\*\*\*

Answer Option 3:

The correct interpretation 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.' This is because the interval is negative, suggesting that the proportion of men interested in shopping on Sundays is greater than the proportion of women.

\*\*\*\*\*

Answer Option 4:

The correct interpretation 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.' The confidence interval of  $(-0.23, -0.18)$  means that the proportion of men interested in shopping on Sundays is between 0.18 and 0.23 greater than the proportion of women, with a 95 percent level of confidence.

\*\*\*\*\*

Answer Option 5:

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.'

The interval  $(-0.23, -0.18)$  is negative, which means the proportion of men interested in shopping on Sundays is higher than the proportion of women.

\*\*\*\*\*

Answer Option 6:

The 95% confidence interval for the difference in proportions is given as  $(-0.23, -0.18)$ . The fact that the interval is negative indicates that the proportion of female customers who answered "yes" is less than the proportion of male customers who answered "yes". 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.' Final Answer: B

\*\*\*\*\*

Answer Option 7:

The interval  $(-0.23, -0.18)$  represents the range in which the difference between the proportion of men and women interested in shopping on Sundays lies, with a 95% certainty. Since the interval is negative, we can infer that the proportion of women who are interested is less than the proportion of men. 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.' Final Answer: B

\*\*\*\*\*

Answer Option 8:

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.'

The negative values in the interval indicate that the proportion of male customers in favor of shopping on Sundays is greater than the proportion of female customers.

\*\*\*\*\*

Answer Option 9:

The confidence interval is  $(-0.23, -0.18)$  which indicates that the proportion of male customers interested in shopping on Sundays is higher than the proportion of female customers. This is because the interval is negative, meaning the proportion of female customers (which we are subtracting) is smaller. 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.' 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:

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 2:

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 3:

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 4:

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 5:

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 6:

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:

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:

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 9:

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. ' $27.70 + 22.30 = 1.24$ ', D. ' $2.00(4.00) = 8.00$ '

=====

Answer Option 1:

The interquartile range is the range between the first quartile (the 25th percentile) and the third quartile (the 75th percentile). For a normal distribution, the first quartile is approximately 0.67 standard deviations below the mean, and the third quartile is approximately 0.67 standard deviations above the mean. Therefore, the interquartile range is approximately  $2 * 0.67 * 4 = 5.36$ . The closest answer choice to this is B. ' $27.70 - 22.30 = 5.40$ '. Final Answer: B

\*\*\*\*\*

Answer Option 2:

The interquartile range for a normally distributed random variable is approximately equal to 1.35 times the standard deviation. So, the interquartile range for the given distribution is  $1.35 * 4 = 5.40$ . Therefore, the correct answer is B. ' $27.70 - 22.30 = 5.40$ '.

\*\*\*\*\*

Answer Option 3:

The interquartile range (IQR) is a measure of statistical dispersion, and for a normal distribution, about 50% of values are within the IQR. The IQR for a normal distribution is approximately 1.35 times the standard deviation. So the IQR is  $1.35 * 4 = 5.40$ . Final Answer: B

\*\*\*\*\*

Answer Option 4:

The interquartile range (IQR) is the range between the first quartile (the 25th percentile) and the third quartile (the 75th percentile) of a distribution. For a normal distribution, the first quartile is approximately 0.67 standard deviations below the mean and the third quartile is approximately 0.67 standard deviations above the mean. Therefore, the interquartile range is approximately  $2 * 0.67$  standard deviations, or 1.34 standard deviations. With a standard deviation of 4, the interquartile range is approximately  $1.34 * 4 = 5.36$ . The closest answer is B. ' $27.70 - 22.30 = 5.40$ '. Final Answer: B

\*\*\*\*\*

Answer Option 5:

The interquartile range is the range between the first quartile (25th percentile) and the third quartile (75th percentile). For a normal distribution, the first quartile is approximately 0.675 standard deviations below the mean and the third quartile is approximately 0.675 standard deviations above the mean. So, the interquartile range is approximately  $2 * 0.675$  standard deviations, or 1.35 standard deviations. Given a standard deviation of 4, the interquartile range is approximately  $1.35 * 4 = 5.40$ . So, the correct answer is B. ' $27.70 - 22.30 = 5.40$ '.

\*\*\*\*\*

Answer Option 6:

The interquartile range is defined as the difference between the upper quartile and the lower quartile. For a normal distribution, the upper quartile corresponds approximately to 0.67 standard deviations above the mean, and the lower quartile corresponds approximately to 0.67 standard deviations below the mean. Therefore, the interquartile range is approximately  $2 * 0.67$  standard deviations, or 1.34 standard deviations. Given that the standard deviation is 4, the interquartile range is approximately  $1.34 * 4 = 5.36$ , which is closest to 5.40. Therefore, the correct answer is B. ' $27.70 - 22.30 = 5.40$ '

\*\*\*\*\*

Answer Option 7:

The interquartile range is the range between the first quartile (the 25th percentile) and the third quartile (the 75th percentile) of the data. For a normal distribution, the first quartile is approximately 0.67 standard deviations below the mean and the third quartile is approximately 0.67 standard deviations above the mean. So, the interquartile range is approximately 1.34 times the standard deviation.

For this distribution, with a standard deviation of 4, the interquartile range is approximately  $1.34 * 4 = 5.36$ . The closest option to this is ' $27.70 - 22.30 = 5.40$ '. So, the answer is B. ' $27.70 - 22.30 = 5.40$ '.

\*\*\*\*\*

Answer Option 8:

The interquartile range of a normally distributed random variable is approximately the difference between the 75th percentile and the 25th percentile. These percentiles are approximately 0.675 standard deviations above and below the mean, respectively. So the interquartile range is approximately  $2 * 0.675$  standard deviations, or 1.35 standard deviations. Given a standard deviation of 4, the interquartile range is approximately  $1.35 * 4 = 5.4$ . Therefore, the correct answer is B. ' $27.70 - 22.30 = 5.40$ '.

\*\*\*\*\*

Answer Option 9:

The interquartile range for a normally distributed random variable is the difference between the 75th percentile (Q3) and 25th percentile (Q1). For a normal distribution, Q3 corresponds approximately to the value that is 0.675 standard deviations above the mean, and Q1 corresponds to the value that is 0.675 standard deviations below the mean. Thus, Q3 is equal to  $25 + 0.675(4) = 27.70$  and Q1 is equal to  $25 - 0.675(4) = 22.30$ . So, the interquartile range is approximately ' $27.70 - 22.30 = 5.40$ '. Final Answer: B.

\*\*\*\*\*

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:

Statement D is true. When  $P(A \cap B) = 0$ , it implies that events A and B are mutually exclusive meaning they cannot occur at the same time. So the correct answer is D.

\*\*\*\*\*

Answer Option 2:

$P(A \cap B) = 0$  implies events A and B are mutually exclusive. This is because the intersection of A and B being 0 means that the two events cannot occur at the same time, which is the definition of mutually exclusive events. So the correct answer is D.

\*\*\*\*\*

Answer Option 3:

The statement ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' is true. If the probability of A and B occurring together (their intersection) is 0, this implies that the events cannot occur at the same time, which is the definition of mutually exclusive events. Final Answer: D

\*\*\*\*\*

Answer Option 4:

The correct answer is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' This is because by definition, two events are mutually exclusive if they cannot both occur at the same time - in other words, the probability of both events occurring (their intersection) is zero.

\*\*\*\*\*

Answer Option 5:

The correct answer is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' This means that the intersection of events A and B is empty, or in other words, A and B cannot occur at the same time. Thus, they are mutually exclusive.

\*\*\*\*\*

Answer Option 6:

The correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' This is because two events are considered mutually exclusive if they cannot both occur at the same time, i.e., their intersection is empty. Thus, if the probability of their intersection is zero, it means that both events cannot occur simultaneously. Hence, the two events must be mutually exclusive.

\*\*\*\*\*

Answer Option 7:

If  $P(A \cap B) = 0$ , it means that the two events A and B cannot occur simultaneously. This is the definition of mutually exclusive events. So, ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' is the correct statement. Final answer: D

\*\*\*\*\*

Answer Option 8:

The correct statement is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' This is because two events are considered mutually exclusive if they cannot occur at the same time—an intersection of the two events would thus be empty, leading to a probability of zero. The other options can lead to incorrect conclusions about the relationship between events A and B. For example, just because the union of two events has a certain probability, it doesn't tell us whether the events are independent or mutually exclusive, and similarly, just because the intersection of two events is zero, it doesn't necessarily mean the events are independent. Final Answer: D

\*\*\*\*\*

Answer Option 9:

The answer is D. ' $P(A \cap B) = 0$  implies events A and B are mutually exclusive.' If the probability of both A and B occurring is zero, then events A and B must be mutually exclusive. This means that if one event occurs, the other cannot occur. The other options do not necessarily hold true for all events A and B.

\*\*\*\*\*

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:

The number of standard deviations away from the mean that \$13,000 and \$28,000 are equal to  $\frac{13000-20500}{3750}=-2$  and  $\frac{28000-20500}{3750}=2$ , respectively. Based on the empirical rule for normal distribution (68-95-99.7 rule), approximately 95% of the data lies within 2 standard deviations of the mean. Therefore, we expect 95% of the 400 employees, or  $0.95 \times 400 = 380$  employees, to have an income between \$13,000 and \$28,000. Final answer: D

\*\*\*\*\*

Answer Option 2:

First, we need to convert the income range to a z-score range. The z-score is calculated by subtracting the mean and dividing by the standard deviation.

For \$13,000, the z-score is:  $(13,000-20,500)/3750 = -2$ .

For \$28,000, the z-score is:  $(28,000-20,500)/3750 = 2$ .

This means we are looking for the proportion of employees within 2 standard deviations of the mean. According to the empirical rule (or 68-95-99.7 rule) for normal distributions, about 95% of the values lie within 2 standard deviations of the mean.

Therefore, approximately  $0.95 \times 400 = 380$  employees are expected to have an income between \$13,000 and \$28,000.

Final answer: D. '390'.

\*\*\*\*\*

Answer Option 3:

The standard deviation of \$3750 is used to measure the extent of variation or dispersion of a set of values. \$13,000 is \$7500 below the mean and \$28,000 is \$7500 above. Each of these is 2 standard deviations away from the mean. According to the empirical rule, or the 68-95-99.7 rule, for a normal distribution, 95% of the data falls within two standard deviations from the mean. Therefore, we would expect 95% of the 400 employees to have an income between \$13,000 and \$28,000. 95% of 400 is 380. So the answer is  $\boxed{390}$ . Final Answer: D

\*\*\*\*\*

Answer Option 4:

The Z-scores corresponding to \$13,000 and \$28,000 are  $Z_{13000} = \frac{13000-20500}{3750} = -2$  and  $Z_{28000} = \frac{28000-20500}{3750} = 2$ . In a standard normal distribution, about 95.45% of all values are within 2 standard deviations of the mean (between -2 and 2). So, we expect about  $0.9545 \times 400 = 381.8$  employees to have an income between \$13,000 and \$28,000. Since we can't have 0.8 of an employee, we round this to the nearest whole number, 382. Our closest answer choice is  $\boxed{390}$ . Final Answer: D

\*\*\*\*\*

Answer Option 5:

For 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.

$\$20,500 - \$13,000 = \$7,500$  which is 2 standard deviations below the mean.

$\$28,000 - \$20,500 = \$7,500$  which is also 2 standard deviations above the mean.

Therefore, the range \$13,000 - \$28,000 covers approximately 95% of the employees.

Since there are 400 employees, we expect about  $0.95 \times 400 = 380$  employees to have an income between \$13,000 and \$28,000.

This is closest to the choice D, '390'. So, the final answer is D.

\*\*\*\*\*

Answer Option 6:

The number of standard deviations away from the mean that \$13,000 and \$28,000 are can be calculated as follows:

For \$13,000:

Number of SDs =  $(\$20,500 - \$13,000) / \$3,750 = 2$  SD

For \$28,000:

Number of SDs =  $(\$28,000 - \$20,500) / \$3,750 = 2$  SD

In a normal distribution, about 95% of the data falls within 2 standard deviations of the mean. Therefore, we would expect about 95% of the 400 employees to have an income between \$13,000 and \$28,000.

95% of 400 =  $0.95 \times 400 = 380$ . Therefore, we would expect about 380 employees to have an income between \$13,000 and \$28,000.

Final Answer: D. '390'

\*\*\*\*\*

Answer Option 7:

We first need to standardize the income values to z-scores. The z-score is calculated as follows:  $z = (X - \mu) / \sigma$ , where X is the value from the dataset,  $\mu$  is the mean and  $\sigma$  is the standard deviation.

For \$13,000, the z-score is  $(13000-20500)/3750 = -2$ .

For \$28,000, the z-score is  $(28000-20500)/3750 = 2$ .

Looking these values up in the z-table, we find that about 95.4% of values in a normal distribution fall within this range. Therefore, we would expect about  $0.954 \times 400 = 381.6$ , which rounds to about 382 employees.



However, as the options provided don't include 382, it's possible the question anticipates a different method of rounding. Normally, we would round the figure down as you can't have a fraction of an employee. However, considering the options provided, the closest option would be 390.

Final Answer: D. '390'.

\*\*\*\*\*

Answer Option 8:

The z-score of \$13,000 is  $(\$13,000 - 20,500) / 3,750 = -2\$$ . The z-score of \$28,000 is  $(\$28,000 - 20,500) / 3,750 = 2\$$ .

Looking up these z-scores in the z-score table, we find that about 0.0228 (or 2.28%) of the distribution is below a z-score of -2 and about 0.9772 (or 97.72%) is below a z-score of 2.

So, the proportion of the distribution between these two z-scores is  $0.9772 - 0.0228 = 0.9544$  (or 95.44%).

Applying this proportion to the 400 employees, we find that about  $400 * 0.9544 = 381.76$  employees are expected to have an income between \$13,000 and \$28,000. Because we can't have a fraction of an employee, we round this to 382.

The closest answer to 382 is 390. So, Final Answer: D.

\*\*\*\*\*

Answer Option 9:

First, we need to find how many standard deviations from the mean the incomes \$13,000 and \$28,000 are. The mean is \$20,500 and the standard deviation is \$3,750, so \$13,000 is  $(\$20,500 - 13,000)/3,750 = 2\$$  standard deviations below the mean, and \$28,000 is  $(\$28,000 - 20,500)/3,750 = 2\$$  standard deviations above the mean.

Looking at a standard normal distribution table, we know that about 95.4% of values lie within 2 standard deviations of the mean. Therefore, we would expect about  $0.954 * 400 = 381.6$  employees to have incomes between \$13,000 and \$28,000.

Since we can't have a fraction of an employee, we round to the nearest whole number to get 382 employees. However, this option isn't available, with the closest option being 390. It's important to note that this is an estimate, as the distribution may not perfectly follow the normal distribution and the exact number of employees in this range could be slightly higher or lower.

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. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.'

\*\*\*\*\*

Answer Option 2:

A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.'

\*\*\*\*\*

Answer Option 3:

A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.'

\*\*\*\*\*

Answer Option 4:

A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.'

\*\*\*\*\*

Answer Option 5:

A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.'

\*\*\*\*\*

Answer Option 6:

A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.'

\*\*\*\*\*

Answer Option 7:

A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.'

\*\*\*\*\*

Answer Option 8:

The margin of error for a poll reflects the uncertainty inherent in sampling: it is a measure of the likely difference between the results from a sample of the population and the results from the full population.

Thus, a margin of error of 5% indicates that the true proportion in the full population is likely to be within 5% of the result obtained from the sample. Hence, the correct answer is: A. 'It is likely that the true proportion of high school students afraid to go to school is between 38% and 48%.'

\*\*\*\*\*

Answer Option 9:

The margin of error for a survey is the range in which the true value (in this case, the true percentage of high school students afraid to go to school) is likely to fall, given the survey results. So if the poll reports a 43% result with a 5% margin of error, that means the true value is likely to be between  $43\% - 5\% = 38\%$  and  $43\% + 5\% = 48\%$ .

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:

A. 'Census'

\*\*\*\*\*

Answer Option 2:

A. 'Census'

\*\*\*\*\*

Answer Option 3:

B. 'Sample survey'

\*\*\*\*\*

Answer Option 4:

A. 'Census'

\*\*\*\*\*

Answer Option 5:

A. 'Census'

\*\*\*\*\*

Answer Option 6:

D. 'Observational study'

\*\*\*\*\*

Answer Option 7:

A. 'Census'

\*\*\*\*\*

Answer Option 8:

A. 'Census'

\*\*\*\*\*

Answer Option 9:

A. 'Census'

\*\*\*\*\*

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:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

Answer Option 2:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

Answer Option 3:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

Answer Option 4:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

Answer Option 5:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

Answer Option 6:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

Answer Option 7:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

Answer Option 8:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

Answer Option 9:

A. 'The first study was an experiment without a control group, while the second was an observational study.'

.....

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:

B. 'Plan II'

=====

Answer Option 2:

B. 'Plan II'

=====

Answer Option 3:

B. 'Plan II'

=====

Answer Option 4:

B. 'Plan II'

Plan II is better since it targets the group of students who actually have investments. Plan I includes students who don't have investments in the stock market, which can result in a lower average and a less accurate estimate.

=====

Answer Option 5:

B. 'Plan II'

Plan I would result in a large number of students in the sample who have zero investment in the stock market, which would skew the estimate towards zero. Plan II, on the other hand, specifically targets the population of interest (students with investments), leading to a more accurate and useful estimate.

=====

Answer Option 6:

The correct answer is B. 'Plan II'.

Plan I is flawed because it includes students who have no investments in the stock market. The result would be an underestimation of the average investment among the students who do invest, because the zeros from the non-investing students would pull down the average.

Plan II, on the other hand, only includes students who have investments in the stock market. Therefore, the resulting estimate of the average investment would be more accurate. When this is multiplied by 132 to estimate the total, the result would be a more accurate estimate of the total investment by students in this school.

=====

Answer Option 7:

B. 'Plan II'

=====

Answer Option 8:

B. 'Plan II'

=====

Answer Option 9:

B. 'Plan II'

Plan I assumes that every student in the school has a personal investment in the stock market, which is not true. The sample taken in Plan I might not even include a student with personal investments. Plan II, on the other hand, only considers the students that actually have investments. Therefore, Plan II is more likely to give a better estimate of the total investment.

=====

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:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

Answer Option 2:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

Answer Option 3:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

Answer Option 4:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

Answer Option 5:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

Answer Option 6:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

Answer Option 7:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

Answer Option 8:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

Answer Option 9:

B. 'because a sample statistic is used to estimate a population parameter.'

\*\*\*\*\*

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:

An R-squared value of 98.1% indicates that 98.1% of the variation in the response variable can be explained by the regression on the explanatory variable, which means that there is a strong linear relationship between the explanatory and response variables. However, R-squared does not indicate the sign of the relationship, so it could either be positive or negative. Therefore, the correct answer is 'I or II (but not both) and III'. Final Answer: C

\*\*\*\*\*

Answer Option 2:

The R-squared value (also known as the coefficient of determination) measures the proportion of the variance in the dependent variable that is predictable from the independent variable. It does not, however, indicate the direction of the relationship (whether it's positive or negative). Thus, the given R-sq value of 98.1% suggests that about 98.1% of the variation in the response variable can be explained by the regression on the explanatory variable. This suggests a strong relationship between the variables. However, we cannot determine from this whether the relationship is positive or negative. So, the correct answer is 'I or II (but not both) and III'. Final Answer: C.

\*\*\*\*\*

Answer Option 3:

The statement "R-sq = 98.1%" refers to the coefficient of determination, a measure of how well the regression line predicts the actual values. An R-sq value close to 100% means that the regression line closely fits the data. However, the R-sq value itself does not tell us the direction of the relationship (positive or negative) - it only tells us the strength of the linear relationship. Therefore, we cannot determine from the given information whether there is a strong positive or negative linear relationship. However, we can conclude that about 98% of the variation in the response variable can be explained by the regression on the explanatory variable. Final Answer: C. 'I or II (but not both) and III'.

\*\*\*\*\*

Answer Option 4:

The R-squared value (also known as the coefficient of determination) measures the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. An R-squared of 98.1% means that all but about 2% of the variability of the response data is accounted for by the explanatory variable. However, R-squared doesn't tell us about the direction of the relationship between the variables, only the strength. Therefore, option III is correct.

As for options I and II, it's not clear from the R-squared value alone whether the relationship is positive or negative. That would be determined by the sign of the correlation coefficient or the slope of the regression line. So, either I or II could be true, but not both at the same time because the relationships they describe are opposites.

So, the correct answer is 'I or II (but not both) and III'. Final answer: C

\*\*\*\*\*

Answer Option 5:

The R-squared value (R-sq) measures the proportion of the variation in the response variable that can be explained by the regression on the explanatory variable. Therefore, statement III is true. However, R-squared does not indicate the sign (positive or negative) of the relationship between the variables, it simply denotes the strength of the linear relationship. So, the relationship could be either positive or negative (but not both), as stated in statements I and II. Thus, the correct answer is C. 'I or II (but not both) and III'. Final answer: C

\*\*\*\*\*

Answer Option 6:

The statement "R-sq = 98.1%" implies that about 98% of the variation in the response variable can be explained by the regression on the explanatory variable, which makes Statement III true. However, the R-squared value does not provide information about the direction (positive or negative) of the relationship between the explanatory and response variables. Therefore, we can't determine if Statement I and II are true or false. Thus, the correct answer is 'III only', and none of the provided options (A, B, C, D) is correct. There seems to be an error in the question or options provided.

\*\*\*\*\*

Answer Option 7:

The R-squared value ranges from 0 to 1 and reflects the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. As it is not negative here, it can be a strong positive linear relationship (I) or a strong negative linear relationship (II). It does not indicate the sign of the relationship, just the strength. So either I or II could be true, but not both at the same time. Statement III is also correct as it correctly interprets the meaning of the R-sq = 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, the correct answer is 'I or II (but not both) and III', so final answer: C

\*\*\*\*\*

Answer Option 8:

The R-squared value, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variable(s). An R-squared of 98.1% reveals that 98.1% of the data fit the regression model. However, R-squared does not tell us about the nature of the relationship, whether it is positive or negative. Therefore, the correct answer is 'III only'. Since it's not one of the options, the statement that comes closest is 'I or II (but not both) and III'. Therefore, the final answer is C.

\*\*\*\*\*

Answer Option 9:

R-squared is a statistical measure that's used to assess the goodness of fit of a regression model. In the context of predictive modeling, R-squared is a measure of how well the model's predictions approximate the actual data. An R-squared of 98.1% reveals that all movements of a variable are completely explained by movements in another variable minus 1.9%. However, R-squared cannot determine whether the coefficient estimates and predictions are biased, which is why you must assess the residual plots. It does not indicate the strength of the relationship between the explanatory and response variables (not necessarily positive or negative), only the proportion of variance explained by the model. The answer is 'III only' but none of the options given fit this, so there seems to be an error.

\*\*\*\*\*



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:

C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

\*\*\*\*\*

Answer Option 2:

C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

\*\*\*\*\*

Answer Option 3:

C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

\*\*\*\*\*

Answer Option 4:

C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

\*\*\*\*\*

Answer Option 5:

C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

\*\*\*\*\*

Answer Option 6:

C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

\*\*\*\*\*

Answer Option 7:

C. 'This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.'

\*\*\*\*\*

Answer Option 8:

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.' The patients cannot be blinded to their treatment, because they will know whether they are taking medication or using a blindfold and listening to soft music. However, the doctor examining the data from the breathing patterns can be blinded to which treatment the patient received, making the experiment single-blind.

\*\*\*\*\*

Answer Option 9:

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.' In a single-blind experiment, either the subject does not know which group they are in, or the experimenter does not know which treatment the subject is receiving. In this case, the patients would obviously know whether they are taking medication or wearing a blindfold and listening to soft music, but the examining doctor, who is analyzing the breathing patterns, might not know which treatment each patient received. This is assuming that the doctor is not informed about the treatment each patient received, making the experiment single-blind. 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 P-value is the probability of observing a result at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. Therefore, 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.'

\*\*\*\*\*

Answer Option 2:

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.'

\*\*\*\*\*

Answer Option 3:

The correct interpretation of this 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.'

\*\*\*\*\*

Answer Option 4:

The correct interpretation of this 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.'

\*\*\*\*\*

Answer Option 5:

The P-value is the probability of observing a result as extreme as, or more extreme than, the observed data, under the assumption that the null hypothesis is true. So the correct interpretation of this 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.'

\*\*\*\*\*

Answer Option 6:

The correct interpretation of this 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.'

\*\*\*\*\*

Answer Option 7:

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 is the probability of observing a result as extreme as the one in the study, assuming that the null hypothesis (the new keyboard is no better at preventing injuries than the traditional one) is true. It is not a probability that the null hypothesis is correct or not.

\*\*\*\*\*

Answer Option 8:

The P-value represents the probability of obtaining a result as extreme as, or more extreme than, the observed data, assuming the null hypothesis is true. Therefore, the correct interpretation of a 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.' Final Answer: D

\*\*\*\*\*

Answer Option 9:

The P-value is the probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct. Therefore, a P-value of 0.07 indicates that, under the null hypothesis (that there is no difference between the keyboards), there is a 7% chance that we would observe a difference as extreme (or more) as the one we observed. So, the correct interpretation would be: 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.'

\*\*\*\*\*