Part 1 Quiz 1 solutions

February 14, 2025

1 Solutions/answers

Question 1 A team of researchers want to study the average height of students in a high school. They record the heights of 200 students. After collecting the data, they calculate the mean, median, and standard deviation of the heights to summarize the findings.

solution Descriptive statistics. (since there is no goal mentioned of using the data for some generalization/inference)

Question 2 A pharmaceutical company develops a new drug to lower blood pressure. To test its effectiveness, they conduct a clinical trial with a sample of 500 patients. After analyzing the results, they would like to know whether the drud is likely to reduce blood pressure in the entire polulation of patients with high blood pressure, not just those in the sample.

solution Inferential statistics

Question 3 Match the following variables with the corresponding measurement scales:

solution

- Eye color- nominal
- Education level - Ordinal
- IQ scores - Interval
- Weight - ratio

Question 4 What are the most suitable ways to represent the following dataset?

Favorite Ice Cream Flavors among 100 People

Vanilla: 30 Chocolate: 25 Strawberry: 15 Mint: 20 Other: 10

solution Bar chart and Pie chart

Question 5 Given that the 50th percentile and the trimean of a certain data set are 90 and 100 respectively. If the interquartile range is 160, determine the 25th and the 75th percentile.

Express your answers in integer numbers only (without any space, full stop or any characters other than numbers).

25th percentile = [1]

75th percentile = [2]

solution We have that $P_{50} = 90$ and $\frac{P_{25}+2P_{50}+P_{75}}{4} = 100$. Therefore,

$$P_{25} + P_{75} = 400 - 2P_{50} = 400 - 2 \cdot 90 = 220$$

In addition,

$$IQR = P_{75} - P_{25} = 160$$

Adding the above, we get:

$$2P_{75} = 380$$

$$P_{75} = 190$$

Thus,

$$P_{25} = P_{75} - 160 = 190 - 160 = 30$$

Question 6 The distance between the median and the mean in a negatively skewed distribution is [x]. If the mean is [y], what is the median value of the distribution?

solution In a negatively skewed distribution, we know that the median is larger than the mean. Therefore, the median is equal to $y + x$.

Question 7 Calculate the mean and (population) variance of the following dataset: 28, 58, 43, 49, 57, 80

solution The mean is:

$$\frac{28 + 58 + 43 + 49 + 57 + 80}{6} = \frac{315}{6} = 52.5$$

The population variance is given by the formula:

$$Var(X) = E[X^2] - (E[X])^2 = \frac{28^2 + 58^2 + 43^2 + 49^2 + 57^2 + 80^2}{6} - 52.5^2 = 251.583$$

Question 8 Two separate studies measure the same variable, and the sample variances and sample sizes for each group are as follows:

study	Sample size	Sample mean	Sample Variance
1	6	10	0.8
2	7	11	1

After combining data from both studies, what is the sample variance of the combined data set?

solution Let $X_1 = \{x_{1,1}, \ldots, x_{1,n}\}, X_2 = \{x_{2,1}, \ldots, x_{2,m}\}$ represent the 2 sets or random sample from the population. Using the sample variance formula $s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]$, on each of the 2 rows of the table, we get:

$$s_1^2 = \frac{1}{5}\left[\sum x_1^2 - \frac{60^2}{6}\right] = 0.8$$

$$\sum x_1^2 = 4 + 600 = 604$$

And

$$s_2^2 = \frac{1}{6}\left[\sum x_2^2 - \frac{77^2}{7}\right] = 1$$

$$\sum x_2^2 = 6 + 7 \cdot 11^2 = 6 + 7 \cdot 121 = 853$$

We now calculate the sample variance of the combined sample as follows:

$$s^2 = \frac{1}{12}\left[\sum x_1^2 + \sum x_2^2 - \frac{(60+77)^2}{13}\right] = \frac{604 + 853 - \frac{137^2}{13}}{12} = 1.103$$

Question 9 Emma is a data analyst at a Grab studying how food delivery time Y (in minutes) depends on the distance traveled X (in kilometers).

After analyzing past data, she finds a simple linear relationship: Y=[b]X+15.

Emma has already determined that the variance in driving distances is [a]. What is the standard deviation in food delivery time?

solution The variance of delivery times is

$$Var(Y) = \sigma_Y^2 = b^2 Var(X) = b^2 \cdot a$$

$$\sigma_Y = b\sqrt{a}$$

Question 10 A researcher is analyzing two variables, X and Y, and finds that they are uncorrelated, meaning their correlation coefficient is 0 (i.e., $\rho(X,Y) = 0$) and they both have mean 0.

Now, the researcher defines a new variable: $Z = X + Y$

The variances of X and Y are 1 and 3 respectively.

Given the above:

The standard deviation of Z is equal to [A]

The correlation coefficient $\rho(X,Z)$ is equal to [B]

solution Using variance law, we get:

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 = 1 + 3 = 4$$

Thus,

$$\sigma_Z = 2$$

For the correlation coefficient of X and Z, we have:

$$\rho(X, Z) = \frac{E[(X - \mu_X)(Z - \mu_Z)]}{\sigma_X \sigma_Z}$$

Since, $\mu_x = \mu_Z = 0$, we get that $\mu_Z = \mu_X + \mu_Y = 0$ and thus:

$$\rho(X, Z) = \frac{E[XZ]}{\sigma_X \sigma_Z} = \frac{E[X(X + Y)]}{\sigma_X \sigma_Z} = \frac{E[X^2] + E[XY]}{\sigma_X \sigma_Z} = \frac{E[X^2]}{\sigma_X \sigma_Z} = \frac{Var(X) - (E[X])^2}{\sigma_X \sigma_Z}$$

$$\rho(X, Z) = \frac{\sigma_X^2}{\sigma_X \sigma_Z} = \frac{1}{1 \cdot 2} = 0.5$$

Question 11 Let the universal set be $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Define the following sets:

- $A = \{1, 2, 3, 4, 5\}$
- $B = \{4, 5, 6, 4, 7\}$
- $C = \{2, 4, 6, 8, 10\}$

Which of the following sets is equal to the set: $(A \cap B) \cup C'$?

solution

$$A \cap B = \{4, 5\}$$

$$C' = \{1, 3, 5, 7, 9\}$$

$$(A \cap B) \cup C' = \{1, 3, 4, 5, 7, 9\}$$

Question 12 A security system uses 5-digit access codes, where each digit can be any number from 0 to 9. However, the following restrictions apply:

- The first digit cannot be 0.
- Digits cannot be repeated within the code.

How many different 5-digit codes can be created under these restrictions?

solution The number of possible access codes is:

$$9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 27216$$

as there are 9 choices for the first digit and due to no repetition there are 9 for the second digit and the number drops by 1 for each subsequent digit.

Question 13 A coffee shop offers a custom drink builder where customers can create a 5-ingredient drink by choosing from 8 different ingredients (e.g., milk, chocolate, caramel, vanilla, etc.).

A customer can select the same ingredient multiple times (e.g., choosing chocolate twice).

How many different 5-ingredient drinks can be created?

solution This is an example of unordered sampling with replacement because the order of ingredients does not matter for making the drink (i.e a drink is specified by just the set of 5 ingredients).

Thus, the number of possible drinks is:

$$\binom{8+5-1}{5} = \binom{12}{5} = 792$$

Question 14 A box contains 6 red balls, and 5 green balls. You randomly draw 3 balls from the box without replacement.

What is the probability that the balls you draw are not all from the same color?

solution The probability that the balls are not all from the same color is:

$$P(not-all-same-color) = 1 - P(all-same-color)$$

$$P(not-all-same-color) = 1 - [P(all-red) + P(all-blue)]$$

Now, we have:

$$P(all-red) = \frac{\binom{6}{3}}{\binom{11}{3}} = \frac{20}{165}$$

and similarly

$$P(all-green) = \frac{\binom{5}{3}}{\binom{11}{3}} = \frac{10}{165}$$

Thus,

$$P(not-all-same-color) = 1 - [P(all-red) + P(all-blue)] = \frac{135}{165} = 0.818$$

Question 15 A company is hiring for three positions: Manager, Assistant Manager, and Office Clerk. There are [a] applicants for the job, but with the following restrictions:

- The Manager must have at least 5 years of experience. There are [b] applicants with at least 5 years of experience.
- The Assistant Manager must not have more than 3 years of experience. There are [c] applicants with 3 years of experience or fewer.

- The Office Clerk can be anyone (with no restrictions on years of experience).

How many different ways can the company assign these three positions?

solution For the Manager position there are b possibilities, for the assistant manager, there are c and after having selected these 2 roles, there are $a-2$ possibilities for Office Clerk.

Thus, the total number is: $b \cdot c \cdot (a-2)$.

Question 16 Two events are mutually exclusive if:

solution The occurrence of one prevents the occurrence of the other

Question 17 A deck of 52 playing cards is shuffled, and one card is drawn at random. Let:

- A be the event of drawing a red card (hearts or diamonds).
- B be the event of drawing a face card (jack, queen, or king).

What is $P(A \cup B)$, the probability of drawing a red card or a face card?

solution (This question will not be counted towards the total score, since the statement had incomplete information. That is, not mentioning the number of red cards in a deck.)

We have that:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Since there are 26 red cards, 12 face cards and 6 red face cards in a deck, we get:

$$P(A \cup B) = \frac{26}{52} + \frac{12}{52} - \frac{6}{52} = \frac{32}{52} = 0.615$$

Question 18 A symmetric distribution has mean 20 and its inter-quartile range (IQR) is equal to 8.

The 25th percentile of the distribution is: [A]

The 75th percentile of the distribution is: [B]

solution Since the distribution is symmetric,we have that: mean=median=trimean=μ. Thus,

$$\frac{P_{25} + 2\mu + P_{75}}{4} = \mu$$

$$P_{25} + P_{75} = 2\mu = 40$$

Furthermore,

$$IQR = P_{75} - P_{25} = 8$$

Adding the above, we get:

$$2P_{75} = 48$$

$$P_{75} = 24$$

Thus,

$$P_{25} = P_{75} - 8 = 24 - 8 = 16$$

Question 19 When rolling a fair 6-sided die with possible outcomes in the set 1,2,3,4,5,6, the events

- A: "the outcome is even"
- B: "the outcome is strictly greater than 3"

are independent.

solution We have the following:

$$P(A) = P(B) = \frac{3}{6}$$

$$P(A \cap B) = \frac{2}{6} = \frac{1}{3} \neq P(A) \cdot P(B) = \frac{1}{4}$$

Therefore, events A and B are **not** independent.
Answer:**FALSE**

Question 20 A box contains 6 red balls, 4 green balls, and 3 yellow balls. Two balls are drawn without replacement.

What is the (conditional) probability that the second ball drawn is green, given that the first ball drawn was not yellow?

solution Consider the following events:

- Y_1: The first ball is yellow
- R_1: The first ball is red
- G_1: The first ball is green
- G_2: The second ball is green

If we condition of the first ball not being yellow, then: $P(R_1|Y_1') = \frac{6}{6+4} = 0.6$ and $P(G_1|Y_1') = \frac{4}{6+4} = 0.4$.

We have the following:

$$P(G_2|Y_1') = 0.6 \cdot P(G_2|R_1) + 0.4 \cdot P(G_2|G_1) = 0.6 \cdot \frac{4}{12} + 0.4 \cdot \frac{3}{12} = 0.2 + 0.1 = 0.3$$

since there are 4 out of 12 green balls remaining after a red ball is picked and 3 out of 12 green balls remaining after a green ball is picked in the first draw.