SC2000: Practice Quiz 1b solutions

February 9, 2025

1 Solutions/answers

Question 1 A set of data shows that the mean service time at a certain service counter is much higher than the median service time, what can you say about the shape of the distribution of service times?

solution The shape of the distribution is positively skewed.

Question 2 Consider that we randomly pick 2 numbers successively from the set 0,1,2,3,4,5,6,7,8,9 with replacement. What is the probability that the first number is greater than the second?

solution Let A be the first number and B be the second number we draw. Since all numbers are equally likely, We have that $P(B = A) = 0.1$ regardless of the value of A. Due to the symmetry of the experiment (i.e for any event $A = x, B = y$ for $x > y$, the symmetric event $A = y, B = x$ has the same probability) we conclude that $P(A > B) = P(A < B) = p$. Furthermore,

$$P(A > B) + P(A < B) + P(A = B) = 1$$

due to the events being exhaustive and mutually exclusive. Thus, we have:

$$2p + 0.1 = 1$$

$$p = \frac{0.9}{2} = 0.45$$

Question 3 The following table shows the means and the variances of 3 sets of random samples taken from a population. Determine the variance of the combined data of the 3 sets of samples.

solution Let $X_1 = \{x_{1,1}, \ldots, x_{1,n}\}, X_2 = \{x_{2,1}, \ldots, x_{2,m}, X_3 = \{x_{3,1}, \ldots, x_{3,m}\}$ represent the 3 sets or random sample from the population. From the first row of the table, we get:

$$\frac{1}{9}\left[\sum x_1^2 - \frac{(3.2 \cdot 10)^2}{10}\right] = a$$

$$\sum x_1^2 = 9a + 102.4$$

From the second row of the table, we get:

$$\frac{1}{5}\left[\sum x_2^2 - \frac{(2.9 \cdot 6)^2}{6}\right] = b$$

$$\sum x_1^2 = 5b + 50.46$$

From the third row of the table, we get:

$$\frac{1}{3}\left[\sum x_3^2 - \frac{(2.5 \cdot 4)^2}{4}\right] = c$$

$$\sum x_3^2 = 3c + 25$$

Finally, we use again the formula $s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]$ to compute the sample variance of the combined set of samples as follows:

$$s^2 = \frac{1}{19}\left[\sum x_1^2 + \sum x_2^2 + \sum x_3^2 - \frac{(3.2 \cdot 10 + 2.9 \cdot 6 + 2.5 \cdot 4)^2}{20}\right]$$

$$s^2 = \frac{1}{19}(9a + 5b + 3c + 177.86 - \frac{59.4^2}{20})$$

$$s^2 = \frac{9a + 5b + 3c + 177.86 - 176.418}{19} = \frac{9a + 5b + 3c + 1.442}{19}$$

Question 4 Given the data set {[a], 2, -2, 1, 0.3, [a], -1.3, [a]}, determine the standard deviation.

solution To compute the variance, we will use the formula

$$\sigma^2 = \frac{1}{n}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]$$

$$\sigma^2 = \frac{1}{8}\left[(3a^2 + 4 + 4 + 1 + 0.09 + 1.69) - \frac{(3a)^2}{8}\right] = \frac{1}{8}\left[(3a^2 + 10.78) - \frac{(3a)^2}{8}\right]$$

$$\sigma^2 = \frac{3a^2 + 10.78}{8} - \left(\frac{3a}{8}\right)^2$$

Thus the standard deviation is:

$$\sigma = \sqrt{\frac{3a^2 + 10.78}{8} - \left(\frac{3a}{8}\right)^2}$$

Question 5 Given the data set {[a], 2.0, -2.1, -[a], [a], 1.3, [a]}, determine the mode.

solution The mode is a, as it is most frequently appearing value.

Question 6 Calculate the mean for the data set {-5.6, 5.2, -1.3, -2.4, 1, 1.1, -6.5, -3.1}.

solution To compute the mean we take the sum of the 8 values and divide by 8: $\mu = (-5.6 + 5.2 - 1.3 - 2.4 + 1 + 1.1 - 6.5 - 3.1)/8 = -11.6/8 = -1.45$

Question 7 A tour agency asks all of its customers to fill up a feedback form at the end of their organised tours. With the collected data, it calculates the proportion of these customers who would recommend the tour to their friends and family. This is an example of ...

solution descriptive statistics

Question 8 A set of [N] numbers is transformed by taking the log2 of each number. The sum of the transformed data is [X]. Calculate the geometric mean of the untransformed data.

solution Let $\{a_1, \ldots, a_N\}$ ne the numbers.
The geometric mean is: $GM = (\prod_{i=1}^{N} a_i)^{1/N}$ Thus,

$$\log_2(GM) = \frac{1}{N}\sum_{i=1}^{N} \log_2 a_i = \frac{X}{N}$$

$$GM = 2^{\frac{X}{N}}$$

Question 9 Gary uses the 4-digit passcode to lock his mobile phone. It is noted that there are 2 smudges over 2 digits on the phone touch screen. What is the probability that one could correctly guess his passcode? Smudges are the blurred marks on the screen.

solution From the fact that there are 2 marks, we know that the passcode contains only 2 specific digits. Without loss of generality, we can assume these numbers are 0 and 1. Thus, the number of possible passcodes is the number of possible 4-digit binary numbers except 0000 and 1111 (as there would be only 1 mark in these cases). Since there are $2^4 = 16$ 4-digit binary numbers, there are 16-2=14 possible codewords. So, the probability of a correct guess is $\frac{1}{14} = 0.071$.

Question 10 A committee of 4 persons is to be selected randomly from 3 men and 5 women. What is the probability that the committee consists of 2 men and 2 women? Express your answer in decimal number to 3 decimal places.

solution The are $\binom{3}{2}$ ways to select 2 out of 3 men and $\binom{5}{2}$ ways to select 2 out of 5 women, while there are in total $\binom{8}{4}$ ways to select the 4 committee members out of 8 people.

Thus, the probability of 2 men and 2 women being selected is:

$$\frac{\binom{3}{2}\binom{5}{2}}{\binom{8}{4}} = \frac{3 \cdot 10}{70} = 0.429$$

Question 11 Given that Y = X (X + [b]), where X is a quantitative variable with mean = [a] and variance = 0.5. Determine the mean of Y.

solution

$$E[Y] = E[X(X+b)] = E[X^2] + bE[X] = Var(X) + (E[X])^2 + bE[X]$$

$$E[Y] = 0.5 + a^2 + b \cdot a$$

Question 12 The following table shows the means and the variances of 3 sets of random samples taken from a population. Determine the variance of the combined data of the 3 sets of samples.

solution Let $X_1 = \{x_{1,1}, \ldots, x_{1,n}\}, X_2 = \{x_{2,1}, \ldots, x_{2,m}, X_3 = \{x_{3,1}, \ldots, x_{3,m}\}$ represent the 3 sets or random sample from the population. From the first row of the table, we get:

$$\frac{1}{9}\left[\sum x_1^2 - \frac{(3.2 \cdot 10)^2}{10}\right] = a$$

$$\sum x_1^2 = 9a + 102.4$$

From the second row of the table, we get:

$$\frac{1}{5}\left[\sum x_2^2 - \frac{(2.9 \cdot 6)^2}{6}\right] = b$$

$$\sum x_1^2 = 5b + 50.46$$

From the third row of the table, we get:

$$\frac{1}{3}\left[\sum x_3^2 - \frac{(2.5 \cdot 4)^2}{4}\right] = c$$

$$\sum x_3^2 = 3c + 25$$

Finally, we use again the formula $s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]$ to compute the sample variance of the combined set of samples as follows:

$$s^2 = \frac{1}{19}\left[\sum x_1^2 + \sum x_2^2 + \sum x_3^2 - \frac{(3.2 \cdot 10 + 2.9 \cdot 6 + 2.5 \cdot 4)^2}{20}\right]$$

$$s^2 = \frac{1}{19}(9a + 5b + 3c + 177.86 - \frac{59.4^2}{20})$$

$$s^2 = \frac{9a + 5b + 3c + 177.86 - 176.418}{19} = \frac{9a + 5b + 3c + 1.442}{19}$$

Question 13 Consider a survey of which [k] participants were randomly selected from your class comprises [n] students. What is the probability that you or your friend Dave are among the chosen participants?

solution The probability that neither of us are chosen is: $\frac{\binom{n-2}{k}}{\binom{n}{k}} = \frac{(n-2)! \cdot k! \cdot (n-k)!}{k! \cdot (n-2-k)! \cdot n!} = \frac{(n-k) \cdot (n-k-1)}{n \cdot (n-1)}$.

Thus, the probability at least one of us is selected is:

$$1 - \frac{(n-k) \cdot (n-k-1)}{n \cdot (n-1)}$$

Question 14 A box contains 5 balls. Some are blue and some are yellow. When 2 balls are randomly picked (without replacement), the probability that both are yellow is 0.6. Determine the number of yellow balls in the box.

solution Let y be the number of yellow balls in the box. The probability that both balls selected are yellow is:

$$\frac{\binom{y}{2}}{\binom{5}{2}} = \frac{y(y-1)}{5 \cdot 4} = \frac{y(y-1)}{20} = 0.6$$

$$y^2 - y - 12 = 0$$

$$(y-4)(y+3) = 0$$

Since y is a positive integer, we have:

$$y = 4$$

Question 15 Fifty percents of the observations in a data set are greater than or equal to the

solution median

Question 16 An airport limousine can carry a maximum of 6 passengers in a trip. Given the following information:

The probability of having at least 3 passengers on a trip is equal to the probability of having 2 passengers. The probability of having 2 passengers is the twice the probability of having one passenger. The minimum number of passengers per trip is 1. Calculate the probability of having 1 passenger on a trip.

solution Let X be the number of passengers on the trip.

We have the following:

- $P(X \geq 3) = P(X = 2)$
- $P(X = 2) = 2 \cdot P(X = 1)$
- $P(X = 0) = 0$

We know that:

$$P(X = 0)+P(X = 1)+P(X = 2)+P(X = 3)+P(X = 4)+P(X = 5)+P(X = 6) = 1$$

since they are mutually exclusive and exhaustive events.

Thus, we get:

$$P(X = 1) + 2 \cdot P(X = 1) + 2 \cdot P(X = 1) = 1$$

$$P(X = 1) = 0.2$$

Question 17 Given that $Y = a \cdot X^2 + bX + c$, where X is a random variable with mean = 1 and variance = 0.5. Determine the mean of Y.

solution We have:

$$E[Y] = E[a \cdot X^2 + bX + c] = aE[X^2] + bE[X] + c = a(Var(X) + (E[X])^2) + bE[X] + c$$

$$E[Y] = 1.5a + b + c$$

Question 18 In how many ways can 8 people sit around a circular table having dinner?

solution There are 8! ways to occupy 8 seats. Howwever, since the table is circular for each arrangement, if everyone moves to the left (or right) by 1 seat, the relative arrangement does not change (everyone still has the same neighbors). Therefore, there are 8!/8 equivalence classes each containing 8 arrangements. So, the number of possible ways would be $8!/8 = 7! = 5040$.

Question 19 Which of the following presentation types are suitable for showing the percentage of students majoring in different fields, e.g. psychology, biology, or computer science?

solution Pie chart and Bar chart

Question 20 The figure below shows a typical boxplot for a data set. Given the data set {9, 9, 8, 7, 12, 7, 4}, determine the upper and the lower adjacent values A and E respectively.

How many outliers (not the values) are there in the data set?

solution Sorted data: {3,7,7,8,9,9,13} To calculate the 25th percentile, we have:
Lower hinge $= R_{25} = \frac{25}{100}(N-1)+1 = \frac{25}{100}(7-1)+1 = 2.5$
$P_{25} = 7 + (7-7) \cdot 0.5 = 7$

To calculate the 75th percentile, we have: $R_{75} = \frac{75}{100}(N-1)+1 = \frac{75}{100}(7-1)+1 = 5.5$
Upper hinge= $P_{75} = 9 + (9-9) \cdot 0.5 = 9$
$Step = 1.5 \cdot IQR = 1.5 \cdot (9-7) = 3$

Upper inner fence $= 9 + 3 = 12$
Lower inner fence $= 7 - 3 = 4$

Upper adjacent value=9
Lower adjacent value=7
# of Outliers=2

Question 21 Two datasets X and Y have mean 0 and variance 4 and 1 respectively. If the Pearson correlation coefficient is $\rho(X, Y) = -1$, calculate the following:

$\rho(X, X+Y)$ is [A]
$\rho(Y, X+Y)$ is [B]

solution Let $\sigma_X^2, \sigma_Y^2, \sigma_Z^2$ be the variances of X, Y and Z=X+Y respectively. From variance law II, we get:

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y = 4 + 1 - 2 \cdot \sqrt{4}\sqrt{1} = 1$$

From the definition of Pearson's correlation coefficient, we get:

$$\rho(X, X+Y) = \frac{Cov(X, X+Y)}{\sigma_X \sigma_Z} = \frac{E[(X-\mu_X)(X+Y-\mu_X-\mu_Y)]}{\sigma_X \sigma_Z} = \frac{E[X(X+Y)]}{\sigma_X \sigma_Z}$$

where the last equality is due to the means μ_X, μ_Y being equal to 0. So, we get:

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

We also have that:

$$\rho(X, Y) = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} = \frac{E[XY]}{\sigma_X \sigma_Y}$$

$$E[XY] = (-1) \cdot \sigma_X \sigma_Y = -2$$

Substituting above, we get:

$$\rho(X, X+Y) = \frac{\sigma_X^2 + E[XY]}{\sigma_X \sigma_Z} = \frac{4-2}{2} = 1$$

and similarly

$$\rho(Y, X+Y) = \frac{\sigma_Y^2 + E[XY]}{\sigma_Y \sigma_Z} = \frac{1-2}{1} = -1$$

The above can be verified by the vector representation of the standard deviations of X,Y,Z=X+Y seen below: