**Statistic**

**Q-1.** A university wants to understand the relationship between the SAT scores of its applicants and their college GPA. They collect data on 500 students, including their SAT scores (out of 1600) and their college GPA (on a 4.0 scale). They find that the correlation coefficient between SAT scores and college GPA is 0.7. What does this correlation coefficient indicate about the relationship between SAT scores and college GPA?

**Ans-1**

A correlation coefficient of 0.7 indicates a strong positive relationship between SAT scores and college GPA. The correlation coefficient measures the strength and direction of the linear relationship between two variables. In this case, a correlation coefficient of 0.7 suggests that there is a strong positive association between SAT scores and college GPA.

The positive sign indicates that as SAT scores increase, college GPA tends to increase as well. It suggests that students who achieve higher SAT scores are more likely to have higher college GPAs, on average.

**Q-2.** Consider a dataset containing the heights (in centimeters) of 1000 individuals. The mean height is 170 cm with a standard deviation of 10 cm. The dataset is approximately normally distributed, and its skewness is approximately zero. Based on this information, answer the following questions:

a. What percentage of individuals in the dataset have heights between 160 cm and 180 cm?

b. If we randomly select 100 individuals from the dataset, what is the probability that their average height is greater than 175 cm?

c. Assuming the dataset follows a normal distribution, what is the z-score corresponding to a height of 185 cm?

d. We know that 5% of the dataset has heights below a certain value. What is the approximate height corresponding to this threshold?

e. Calculate the coefficient of variation (CV) for the dataset.

f. Calculate the skewness of the dataset and interpret the result.

**Answer 2:**

Based on this information, answer the following questions:

1. To find the percentage of individuals with heights between 160 cm and 180 cm.calculate the z-scores for these values and find the area under the normal distribution curve between these z-scores. The z-score formula is given by:

z = (x - μ) / σ

Where x is the height, μ is the mean, and σ is the standard deviation.The z-scores for 160 cm and 180 cm:

For 160 cm:

z1 = (160 - 170) / 10 = -1

For 180 cm:

z2 = (180 - 170) / 10 = 1

Now, use a standard normal distribution table find the area between z-scores. The area between -1 and 1 represents the percentage of individuals with heights between 160 cm and 180 cm.

Using a standard normal distribution table, the area between -1 and 1 is approximately 0.6826 or 68.26%.

Therefore, approximately 68.26% of individuals in the dataset have heights between 160 cm and 180 cm.

1. When randomly selecting 100 individuals from a normally distributed dataset, the average height will also follow a normal distribution.

The mean of the average height will be the same as the mean of the original dataset (170 cm), but the standard deviation of the average height will be the standard deviation of the original dataset divided by the square root of the sample size

(√100 = 10/√100 = 1).

find the probability that avg. height is greater than 175 cm, we have to calculate the z-score for 175 cm and find the area to the right of this z-score.

z = (x - μ) / (σ/√n)

= (175 - 170) / (10/√100)

= 5/1 = 5

Using a standard normal distribution table, find the area to the right of z = 5. It is very small value, close to 0. Therefore, the probability that the average height of the randomly selected 100 individuals is greater than 175 cm is approximately 0.

1. Find z-score corresponding to a height of 185 cm, use the z-score formula:

z = (x - μ) / σ

Put in the values:

z = (185 - 170) / 10 = 15/10 = 1.5

Therefore, the z-score corresponding to a height of 185 cm = 1.5.

1. To find height corresponding to a threshold where 5% of the dataset has heights below that value, here its need to find the z-score corresponding to the 5th percentile of the standard normal distribution. This z-score represents the number of standard deviations below the mean where 5% of the data lies.

Using a standard normal distribution table, to find that the z-score for a 5th percentile is approximately -1.645.

Now calculate the height using the z-score formula:

z = (x - μ) / σ

-1.645 = (x - 170) / 10

Solving for x:

-16.45 = x - 170

x ≈ 170 - 16.45

x ≈ 153.55

Therefore, the approximate height corresponding to the threshold where 5% of the dataset has heights below that value is approximately 153.55 cm.

e. The coefficient of variation (CV) is a measure of relative variability and is calculated by dividing the standard deviation (σ) by the mean (μ) and multiplying by 100 to express it as a percentage:

CV = (σ / μ) \* 100

In this case, the standard deviation is 10 cm and the mean is 170 cm.

CV = (10 / 170) \* 100

≈ 0.0588 \* 100

≈ 5.88

Therefore, the coefficient of variation for the dataset is approximately 5.88%.

Q-3. Consider the ‘Blood Pressure Before’ and ‘Blood Pressure After’ columns from the

data and calculate the following :

<https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k->nFxt8Y/view?usp=share\_

a. Measure the dispersion in both and interpret the results.

b. Calculate mean and 5% confidence interval and plot it in a graph

c. Calculate the Mean absolute deviation and Standard deviation and interpret the results.

d. Calculate the correlation coefficient and check the significance of it at 1% level of significance.

Q-4. A group of 20 friends decide to play a game in which they each write a number between 1 and 20 on a slip of paper and put it into a hat. They then draw one slip of paper at random. What is the probability that the number on the slip of paper is a perfect square (i.e., 1, 4, 9, or 16)?

Ans 4 : To calculate the probability that the number drawn from the slip of paper is a perfect square the number of favorable outcomes (numbers that are perfect squares) and the total number of possible outcomes (numbers from 1 to 20).

The perfect squares between 1 and 20 are 1, 4, 9, and 16. These are the favorable outcomes.

The total number of possible outcomes is 20 since there are 20 slips of paper in the hat, each with a unique number from 1 to 20.

Therefore, the probability that the number on the slip of paper is a perfect square is:

P(Perfect Square) = (Number of Favorable Outcomes) / (Total Number of Possible Outcomes)

P(Perfect Square) = 4 / 20

Simplifying, we get:

P(Perfect Square) = 1 / 5

So, probability that the number drawn from the hat is a perfect square is 1/5 or 0.2, which is equivalent to 20%.

Q-5. A certain city has two taxi companies: Company A has 80% of the taxis and Company B has 20% of the taxis. Company A's taxis have a 95% success rate for picking up passengers on time, while Company B's taxis have a 90% success rate. If a randomly selected taxi is late, what is the probability that it belongs to Company A?

Ans -5

To calculate the probability that a randomly selected late taxi belongs to Company A, we can use Bayes' theorem.

Let's define the following events:

A: The taxi belongs to Company A.

B: The taxi is late.

We want to find P(A|B), the probability that the taxi belongs to Company A given that it is late.

According to Bayes' theorem:

P(A|B) = (P(B|A) \* P(A)) / P(B)

P(B|A) = probability that the taxi is late given that it belongs to Company A, which is 1 - the success rate of Company A, or 1 - 0.95 = 0.05.

P(A) = probability that a randomly selected taxi belongs to Company A, which is 0.8 since Company A has 80% of the taxis.

P(B) = probability that the taxi is late. To calculate this, consider the probabilities of being late for both companies.

P(B) = P(B|A) \* P(A) + P(B|not A) \* P(not A)

P(B|not A) = probability that the taxi is late given that it does not belong to Company A. Since Company B has a 10% (1 - 0.9) failure rate, this probability is 0.1.

P(not A= probability that a randomly selected taxi does not belong to Company A, which is 0.2 since Company B has 20% of the taxis.

Therefore:

P(B) = (0.05 \* 0.8) + (0.1 \* 0.2) = 0.04 + 0.02 = 0.06

Now we can calculate P(A|B) using Bayes' theorem:

P(A|B) = (0.05 \* 0.8) / 0.06

P(A|B) ≈ 0.6667 or 66.67%

Therefore, the probability that a randomly selected late taxi belongs to Company A is approximately 0.6667 or 66.67%.

**Q-6**. A pharmaceutical company is developing a drug that is supposed to reduce blood

pressure. They conduct a clinical trial with 100 patients and record their blood pressure before and after taking the drug. The company wants to know if the change in blood pressure follows a normal distribution.

<https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share_>

**Q-7**. The equations of two lines of regression, obtained in a correlation analysis between variables X and Y are as follows:

and . 2𝑋 + 3 − 8 = 0 2𝑌 + 𝑋 − 5 = 0 The variance of 𝑋 = 4

Find the

a. Variance of Y

b. Coefficient of determination of C and Y

c. Standard error of estimate of X on Y and of Y on X.

Ans 7 : To find the variance of Y, the slope and intercept of the regression line of Y on X.

From the given equation 2Y + X - 5 = 0:

2Y = -X + 5

Y = (-1/2)X + 5/2

Comparing this equation with the general form of a line, Y = mx + c, we find that the slope (m) is -1/2 and the intercept (c) is 5/2.

Now, the variance of Y can be calculated using the formula:

Var(Y) = Var(Y|X) + Var(e)

Since Var(Y|X) represents the variance of Y given X and Var(e) represents the variance of the error term. Therefore, Var(Y|X).

Var(Y|X) = Var(α + βX + e) = Var(α + βX) = Var(-1/2X + 5/2) = Var(-1/2X) = (1/4)Var(X)

Given that the variance of X is 4, we can substitute this value into the equation:

Var(Y) = (1/4)Var(X) = (1/4)(4) = 1

a. The variance of Y is 1.