Python -

Ans. 1 <https://github.com/saundarya-hash/python/blob/main/python%20question%201>

Ans 2 <https://github.com/saundarya-hash/python/blob/main/python%20question%202>

Ans 3 <https://github.com/saundarya-hash/python/blob/main/python%20question%203>

Ans 4<https://github.com/saundarya-hash/python/blob/main/python%20question%204>

Ans 5 <https://github.com/saundarya-hash/python/blob/main/python%20question%205>

Ans 6

Statistics -

Ans 1 - With a correlation coefficient of 0.7, the relationship between SAT scores and college GPA is strongly positive.

Ans 2 - a) We must first determine the z-scores for people with heights between 160 cm and 180 cm in order to utilise the usual normal distribution table to determine the proportion of people in this height range.

Standard deviation divided by (x - mean) yields the z-score. If you are 160 cm tall, then z1 = (160 - 170) / 10 = -1

If you are 180 cm tall, then z2 = (180 - 170) / 10 = 1

The percentage of people with heights between 160 cm and 180 cm can be calculated using the standard normal distribution table by finding the region between -1 and 1. For this range, the table value is roughly 0.6826, or 68.26%. Therefore, between 160 cm and 180 cm is the range in height for about 68.26% of the people in the dataset.

b)mean of the sample means is still 170 cm

standard deviation of the sample means is 10 cm / √100 = 1 cm.

z-score is calculated as (x - mean) / (standard deviation / √n), where n is the sample size. In this case: z = (175 - 170) / (10 / √100) = 5 / 1 = 5

Using the standard normal distribution table, we can find the area to the right of z = 5, which represents the probability that the average height is greater than 175 cm. This area is very close to 0, so the probability is approximately 0%.

c. To find the z-score corresponding to a height of 185 cm, we use the formula: z = (x - mean) / standard deviation

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

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

d) z-score that corresponds to the 5th percentile is approximately -1.645. We can calculate the height using the formula: x = mean + (z \* standard deviation)

x = 170 + (-1.645 \* 10) = 170 - 16.45 = 153.55

Therefore, the approximate height corresponding to the threshold of 5% is 153.55 cm.

e.) CV = (standard deviation / mean) \* 100% = (10 / 170) \* 100% ≈ 5.88%

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

Ans 4. Probability = Number of favorable outcomes / Total number of possible outcomes

Probability = 4 / 20

Probability = 1 / 5

Therefore, the probability that the number drawn from the hat is a perfect square is 1/5 or 0.2, which can also be expressed as 20%.

Ans 5. Let’s assume A: The taxi belongs to Company A. B: The taxi is late.

We are given the following information: P(A) = 0.8 (Company A has 80% of the taxis) P(B|A) = 0.05 (Company A's taxis have a 95% success rate, so the probability of being late is 1 - 0.95 = 0.05) P(B|not A) = 0.10 (Company B's taxis have a 90% success rate, so the probability of being late is 1 - 0.90 = 0.10)

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

Using Bayes' theorem, we have:

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

To calculate P(B), the probability of a taxi being late, we can use the law of total probability:

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

Since there are only two companies, P(not A) = 1 - P(A).

Substituting the values into the equation:

P(B) = (0.05 \* 0.8) + (0.10 \* (1 - 0.8)) = 0.04 + 0.02 = 0.06

Now we can calculate P(A|B):

P(A|B) = (0.05 \* 0.8) / 0.06 = 0.04 / 0.06 = 2/3 ≈ 0.67

Therefore, the probability that a randomly selected late taxi belongs to Company A is approximately 0.67, or 67%.

Ans 7. The equation of the regression line for X on Y is given by:

X̂ = aX + b, where a is the slope and b is the intercept.

From the given equation 2X + 3 - 8 = 0, we can rewrite it as 2X - 5 = 0. Comparing this to the regression line equation, we have:

aX = 2X, so a = 2, and b = -5.

Similarly, the equation of the regression line for Y on X is given by:

Ŷ = cY + d, where c is the slope and d is the intercept.

From the given equation 2Y + X - 5 = 0, we can rewrite it as X + 2Y - 5 = 0. Comparing this to the regression line equation, we have:

cY = 2Y, so c = 2, and d = -5.

a.) Variance of Y:

The variance of Y can be calculated using the formula:

Var(Y) = Var(Ŷ) + Var(e)

Since Var(e) is the error variance, we can assume Var(e) is equal to the variance of X.

Var(Ŷ) = c^2 \* Var(X) = 2^2 \* 4 = 16

Var(Y) = 16 + 4 = 20

So, the variance of Y is 20.

b.) Coefficient of determination of X and Y:

The coefficient of determination (R^2) is defined as the proportion of the variance in the dependent variable (Y) that can be explained by the independent variable (X).

R^2(X, Y) = Var(X̂) / Var(X)

R^2(X, Y) = Var(aX + b) / Var(X)

R^2(X, Y) = Var(2X - 5) / Var(X)

R^2(X, Y) = Var(2X) / Var(X)

R^2(X, Y) = (2^2 \* Var(X)) / Var(X)

R^2(X, Y) = 4

So, the coefficient of determination between X and Y is 4.

c) Standard error of estimate of X on Y and Y on X:

The standard error of estimate measures the average distance between the actual values and the predicted values.

Standard error of estimate of X on Y (SEest(X|Y)) is given by:

SEest(X|Y) = √[Var(e)]

Since Var(e) is equal to the variance of X, we can substitute it as follows:

SEest(X|Y) = √[Var(X)] = √4 = 2

Standard error of estimate of Y on X (SEest(Y|X)) is given by:

SEest(Y|X) = √[Var(e)]

Since Var(e) is equal to the variance of X, we can substitute it as follows:

SEest(Y|X) = √[Var(X)] = √4 = 2

So, the standard error of estimate of X on Y and Y on X is both 2.

Ans 10 . n = 500 (total number of bulbs in the batch)

p = 0.05 (probability of a bulb being defective)

q = 1 - p = 1 - 0.05 = 0.95 (probability of a bulb not being defective)

a.) Probability of exactly 20 bulbs being defwective:

P(X = 20) = C(n, x) \* p^x \* q^(n-x)

where C(n, x) is the binomial coefficient.

P(X = 20) = C(500, 20) \* (0.05)^20 \* (0.95)^(500-20)

Using a calculator or statistical software to calculate binomial coefficients, we have:

C(500, 20) = 7,400,675

Plugging in the values:

P(X = 20) = 7,400,675 \* (0.05)^20 \* (0.95)^480

Calculating this expression will give you the probability that exactly 20 bulbs are defective.

b.) Probability of at least 10 bulbs being defective:

P(X ≥ 10) = P(X = 10) + P(X = 11) + ... + P(X = 500)

Calculating this probability requires summing up the individual probabilities for all the values of X from 10 to 500. This can be quite time-consuming, but using a calculator or statistical software with the binomial distribution function, you can get the result easily.

c) . Probability that at most 15 bulbs are defective:

P(X ≤ 15) = P(X = 0) + P(X = 1) + ... + P(X = 15)

Similar to part b, you need to sum up the individual probabilities for X values from 0 to 15.

d.) Average number of defective bulbs in a batch of 500:

The average (or expected) number of defective bulbs can be calculated using the formula:

E(X) = n \* p

E(X) = 500 \* 0.05

So, on average, you would expect 25 defective bulbs in a batch of 500.

Ans 12 , a) <https://github.com/saundarya-hash/python/blob/main/statistics%20question%2012.A>

b) In the study, if the null hypothesis is rejected, it shows that there is a significant difference in the mean improvement scores between Group A (new drug) and Group B (standard drug) and if the null hypothesis is not rejected it suggests that there is no difference in the mean improvement scores between the two groups

Machine learning -

Ans 2 - <https://github.com/saundarya-hash/python/blob/main/Machine%20learning%20question%202>