Statistics

Answer 1 :

The correlation coefficient of 0.7 indicates a strong positive relationship between SAT scores and college GPA. A correlation coefficient ranges from -1 to +1, where -1 represents a perfect negative correlation, +1 represents a perfect positive correlation, and 0 represents no correlation.

In this case, a correlation coefficient of 0.7 suggests a strong positive relationship, meaning that as SAT scores increase, college GPAs tend to increase as well. This indicates that there is a tendency for students with higher SAT scores to have higher college GPAs.

However, it's important to note that correlation does not imply causation. While there is a strong relationship between SAT scores and college GPAs, it doesn't necessarily mean that higher SAT scores directly cause higher GPAs. Other factors such as study habits, motivation, and personal circumstances can also influence a student's college performance.

Answer 2:

a. To calculate the percentage of individuals with heights between 160 cm and 180 cm, we need to find the area under the normal distribution curve between these two values.

First, we convert the heights to z-scores using the formula:

z = (x - mean) / standard deviation

For 160 cm: z1 = (160 - 170) / 10 = -1

For 180 cm: z2 = (180 - 170) / 10 = 1

Next, we find the cumulative probability associated with these z-scores using a standard normal distribution table or a statistical software.

The percentage of individuals with heights between 160 cm and 180 cm is the difference between these cumulative probabilities:

Percentage = P(z1 < Z < z2) = P(Z < z2) - P(Z < z1)

b. To calculate the probability that the average height of a randomly selected sample of 100 individuals is greater than 175 cm, we need to consider the sampling distribution of the sample means.

The mean of the sampling distribution is equal to the population mean, which is 170 cm. The standard deviation of the sampling distribution, also known as the standard error, is calculated by dividing the population standard deviation by the square root of the sample size:

standard error = standard deviation / sqrt(sample size) = 10 / sqrt(100) = 10 / 10 = 1

Next, we convert the average height of 175 cm to a z-score using the formula:

z = (x - mean) / standard error

z = (175 - 170) / 1 = 5

Finally, we find the cumulative probability associated with this z-score:

Probability = P(Z > z)

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

d. To find the approximate height corresponding to the threshold where 5% of the dataset has heights below that value, we need to find the z-score associated with this percentile.

Using a standard normal distribution table or statistical software, we can find the z-score corresponding to the 5th percentile (0.05). Let's denote this z-score as z_threshold.

Using the z-score formula:

z_threshold = -1.645 (approximately)

Now, we can convert this z-score back to the original height scale:

height_threshold = mean + (z_threshold * standard deviation)

height_threshold = 170 + (-1.645 * 10) = 170 - 16.45 = 153.55 cm

e. The coefficient of variation (CV) is a measure of relative variability, expressed as a percentage. It is calculated by dividing the standard deviation by the mean and multiplying by 100.

CV = (standard deviation / mean) * 100

CV = (10 / 170) * 100 = 5.88%

f. Since the skewness is approximately zero, it indicates that the dataset is symmetrically distributed. In a symmetric distribution, the mean, median, and mode coincide, and the tails of the distribution are balanced. This means that there are an equal number of individuals with heights above and below the mean, resulting in a bell-shaped curve.

Answer3:

Saved as a Python file

Answer4:

To calculate the probability of drawing a perfect square from the hat, we need to determine the number of favorable outcomes (slips with perfect squares) and the total number of possible outcomes (all slips).

Number of favorable outcomes: There are four perfect squares between 1 and 20: 1, 4, 9, and 16. Therefore, there are four favorable outcomes.

Total number of possible outcomes: Since each person writes a number between 1 and 20 on a slip of paper, there are 20 possible outcomes.

Now, we can calculate the probability: Probability = (Number of favorable outcomes) / (Total number of possible outcomes) = 4 / 20 = 1 / 5 = 0.2

Therefore, the probability of drawing a slip with a perfect square number is 0.2 or 20%.

Answer5:

To solve this problem, we can use Bayes' theorem. Let's define the following events:

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) = 0.1 (Company B's taxis have a 90% success rate, so the probability of being late is 1 - 0.9 = 0.1)

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

By applying Bayes' theorem, we have:

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

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

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

We know that P(not A) = 1 - P(A), so we can substitute that in:

P(B) = P(B|A) * P(A) + P(B|not A) * (1 - P(A))

Now we can substitute these values into the formula for P(A|B):

P(A|B) = (P(B|A) * P(A)) / (P(B|A) * P(A) + P(B|not A) * (1 - P(A)))

Let's calculate the result:

$P(A|B) = (0.05 * 0.8) / (0.05 * 0.8 + 0.1 * (1 - 0.8)) = 0.04 / (0.04 + 0.1 * 0.2) = 0.04 / (0.04 + 0.02) = 0.04 / 0.06 = 0.6667$

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

Answer6:
Saved as a python file

Answer7:

To calculate the requested values, let's go through each step:

Given equations of the regression lines:

1. $2X + 3Y - 8 = 0$
2. $2Y + X - 5 = 0$

a. Variance of Y: The variance of Y can be calculated using the formula:

Variance of Y = Variance of X * (Slope of Y on X)^2

From the given equations, we can see that the slope of Y on X is 2. Therefore, we need to find the variance of X to calculate the variance of Y.

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

Variance of Y = 4 * (2^2) = 16

Therefore, the variance of Y is 16.

b. Coefficient of determination of X and Y: The coefficient of determination (R^2) represents the proportion of the variance in the dependent variable (Y) that can be explained by the independent variable (X). It can be calculated using the formula:

R^2 = (Regression sum of squares) / (Total sum of squares)

To find the coefficient of determination, we need the regression sum of squares and the total sum of squares. However, these values are not provided in the given information. Please provide the relevant values so that I can calculate the coefficient of determination accurately.

c. Standard error of estimate of X on Y and of Y on X: The standard error of estimate represents the average distance between the observed values and the predicted values of the dependent variable.

To calculate the standard error of estimate of X on Y, we use the formula:

Standard error of estimate (X on Y) = $\sqrt{(\sum(Y - \hat{Y})^2) / (n - 2)}$

Where Y represents the observed values, Ŷ represents the predicted values of X on Y, and n represents the number of data points.

Similarly, to calculate the standard error of estimate of Y on X, we use the formula:

Standard error of estimate (Y on X) = √((∑(X - X̂)^2) / (n - 2))

Where X represents the observed values, X̂ represents the predicted values of Y on X, and n represents the number of data points.

However, we don't have the observed values or the predicted values from the given information. Please provide the necessary data to calculate these standard errors accurately.

```
import math

variance_x = 4

# a. Variance of Y
variance_y = 4 * (2**2)
print("Variance of Y:", variance_y)

# b. Coefficient of determination of X and Y
r_squared = (variance_x - variance_y) / variance_x
print("Coefficient of determination (X and Y):", r_squared)

# c. Standard error of estimate of X on Y and Y on X
std_error_x_on_y = math.sqrt((1 - r_squared) * variance_x)
std_error_y_on_x = math.sqrt((1 - r_squared) * variance_y)
print("Standard error of estimate (X on Y):", std_error_x_on_y)
print("Standard error of estimate (Y on X):", std_error_y_on_x)
```

Answer8:

Saved as a python file

Answer9:
Saved as a python file

Answer10:
Saved as a python file

Answer11:
Saved as a python file

Answer12:

To conduct a t-test and compare the mean improvement scores between Group A and Group B, we can use the independent samples t-test. Here's how you can perform the analysis:

a. State the null and alternative hypotheses:

- Null hypothesis (H_0): There is no significant difference in the mean improvement scores between Group A and Group B.
- Alternative hypothesis (H_1): There is a significant difference in the mean improvement scores between Group A and Group B.

b. Choose the appropriate t-test: Since we have two independent groups (Group A and Group B) and want to compare their means, we'll use a two-sample independent t-test.

c. Calculate the t-statistic: The formula for the independent samples t-test is: t = (mean1 - mean2) / sqrt((s1²/n1) + (s2²/n2)) where mean1 and mean2 are the sample means, s1 and s2 are the sample standard deviations, and n1 and n2 are the sample sizes.

Given: For Group A: mean1 = 2.5 s1 = 0.8 n1 = 30

For Group B: mean2 = 2.2 s2 = 0.6 n2 = 30

Substituting these values into the formula, we get: t = (2.5 - 2.2) / sqrt((0.8²/30) + (0.6²/30))

Calculating this expression will give us the t-value.

d. Determine the degrees of freedom (df): The degrees of freedom for an independent samples t-test is given by the formula: df = (n1 + n2) - 2

Substituting the values: df = (30 + 30) - 2

Calculating this expression will give us the degrees of freedom.

e. Determine the critical value and p-value: With the t-value and degrees of freedom, you can look up the critical value in the t-distribution table or use statistical software to find the p-value associated with the t-value.

f. Compare the p-value to the significance level: If the p-value is less than the significance level (0.05 in this case), we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

Based on the results of the t-test, you can make the following conclusion:

b. State whether the null hypothesis should be rejected or not and provide a conclusion in the context of the study: If the p-value is less than 0.05, we reject the null hypothesis and conclude that there is a significant difference in the mean improvement scores between Group A and Group B. If the p-value is greater than or equal to 0.05, we fail to reject the null hypothesis, indicating that there is not enough evidence to suggest a significant difference in the mean improvement scores between the two groups.