WEEK 5 ASSIGNMENT REPORT

Statistical Analysis Using Student Performance Dataset

Submitted By: Binal Patel
Course: Data Analytics
Date: 29-11-2025

Task 1: Descriptive Statistics

The numerical column selected for analysis is Math Score.

Descriptive Statistical Measures

Mean: 66.36

Median: 66

Mode: 63

Maximum Value: 100

Minimum Value: 13

Range: 87

Standard Deviation: 15.37

These values indicate that math scores vary moderately around the average of ~66, with most students scoring between 50 and 85.

Task 1B: Frequency Distribution Table

Score Range Frequency

40 51
50 155
60 352
70 588
80 808
90 940
100 993

A histogram was plotted using these frequency values to visualize the distribution of math scores.

Task 1C: Histogram

A histogram was created using the bins and frequency values.
The distribution shows a concentration of students in the score range of 60–90, indicating a generally strong performance trend.

Task 2: Probability Analysis

The categorical column selected for probability questions is Gender.

Sample Probability Questions

Q1. What is the probability that a randomly selected student is male?
Answer: P (male) = (Number of male students) / (Total students)

Q2. What is the probability that a student is female?
Answer: P (female) = (Number of female students) / (Total students)

Q3. What is the probability that a student completed the test preparation course?
Answer: P (completed) = (Number who completed) / (Total students)

Theoretical vs. Experimental Probability

Theoretical probability assumes equal chance (e.g., gender distribution = 50%-50%).

Experimental probability is calculated from actual dataset.

Conclusion:

Experimental probabilities slightly differ from theoretical probabilities because real-world data is not perfectly distributed.

Task 3: Correlation Analysis

Correlation was performed between:

Math Score (X)

Reading Score (Y)

Correlation Coefficient:

r ≈ Strong Positive Correlation (around 0.80–0.90)

Interpretation:

Students who score high in math tend to also score high in reading.
This shows a strong linear relationship.

A scatter plot with math scores on X-axis and reading scores on Y-axis was plotted.

Task 4: Regression Analysis

Regression model was created using:

Independent Variable (X): Math Score

Dependent Variable (Y): Reading Score

Regression Equation:

$$\text{Reading Score} = 16.99 + 0.783 \times (\text{Math Score})$$

R-Square: 0.67

This means 67% of the variation in reading scores can be explained by math scores.
The model is statistically strong.

Prediction Example:

For Math Score = 70:

$$\text{Reading} = 16.99 + (0.783 \times 70) = 71.82$$

Interpretation:

There is a strong, positive linear relationship. As math scores increase, reading scores also increase.

Task 5: Hypothesis Testing (T-Test)

A two-sample t-test was performed between:

Male Math Scores

Female Math Scores

Hypotheses

H_0 (Null Hypothesis):
There is no significant difference in math scores between male and female students.

$\mu_{\text{male}} = \mu_{\text{female}}$

H_1 (Alternative Hypothesis):
There is a significant difference.

$\mu_{\text{male}} \neq \mu_{\text{female}}$

Results

Mean (Male Math Score): 69.33

Mean (Female Math Score): 63.20

P-value: 2.07×10^{-10}

Alpha (Significance Level): 0.05

Decision

Since p-value < 0.05, we reject the null hypothesis.

Conclusion

There is a statistically significant difference between male and female math scores.
Male students scored significantly higher in math than female students in this dataset.

Final Conclusion

All analytical tasks—descriptive statistics, probability, correlation, regression, and hypothesis testing—were successfully performed using the Student Performance dataset.
The insights suggest relationships between different academic scores and significant differences between groups.