Case Studies: Statistics, Central Tendency, Dispersion, Covariance, and Correlation

Case Study 1 : Analyzing Monthly Expenses

Scenario:
You are given a dataset that represents the monthly expenses (in dollars) of 6 individuals in a household over the past year. The goal is to analyze the expenses and understand their general spending behavior.

Dataset (Monthly Expenses in $): [1200, 1500, 900, 1100, 1000, 950]

Questions:

1. Calculate the mean, median, and mode of the monthly expenses.
2. What is the range of the expenses?
3. Calculate the variance and standard deviation to measure the consistency of the expenses.
4. How would the analysis change if an additional person with an expense of $2500 is added to the dataset? Recalculate the mean, median, and standard deviation with this outlier included. Discuss the impact of the outlier on the results.

Case Study 2 : Employee Performance Scores

Scenario:
The performance scores of 8 employees in a company, based on a recent evaluation, are given. The goal is to determine the overall performance pattern and examine the consistency of the scores.

Dataset (Employee Performance Scores): [75, 80, 85, 60, 90, 95, 70, 85]

Questions:

1. Find the mean, median, and mode of the employee performance scores.

2. What is the variance and standard deviation of the performance scores?
3. Is the data skewed or symmetric? Use the mean and median to interpret.
4. If an additional employee has a score of 40 (significantly lower), calculate the new mean and standard deviation. How does this change the analysis, and what can you infer about the effect of a low outlier?

Case Study 3 : Relationship Between Hours Studied and Test Scores

Scenario:
A teacher wants to understand if there is a relationship between the number of hours students study and their test scores. You are given the data on hours studied and corresponding test scores of 7 students.

Dataset:

Hours Studied (X): [2, 4, 6, 8, 10, 12, 14]

Test Scores (Y): [55, 60, 65, 70, 75, 85, 90]

Questions:

1. Calculate the mean, variance, and standard deviation of both the hours studied and test scores.
2. Find the covariance between the hours studied and test scores.
3. Calculate the correlation coefficient between the two variables. What does the value tell you about the relationship between hours studied and test scores?
4. Based on the correlation, can you conclude that studying more hours leads to better test performance? Justify your answer.

Case Study 4 (Hard): Income vs. Spending Behavior

Scenario:
A researcher is studying the relationship between annual income and monthly spending habits. The data for 10 individuals includes their annual income (in $1000) and their average monthly spending (in $). The goal is to analyze the relationship and infer spending patterns.

Dataset:

Annual Income (X in $1000): [30, 40, 45, 50, 60, 65, 75, 80, 85, 95]

Monthly Spending (Y in $): [1500, 2000, 2100, 2400, 3000, 3200, 3500, 3600, 3700, 4000]

Questions:

1. Compute the mean, median, and standard deviation for both income and spending.
2. Calculate the covariance and correlation coefficient between income and spending. Interpret the result and discuss the strength of the relationship.
3. Identify any outliers and discuss how they affect the correlation. Recalculate the correlation after removing any outliers.
4. Based on the correlation and data patterns, what general conclusion can you draw about the relationship between income and spending behavior?
5. Propose how this analysis could be used to predict future spending for a person earning $100,000 annually.