Shivansh

DA-Session-11 Correlation & Causation Assignment

Question 1: Define correlation. Why is correlation important in data analytics?

Answer:
Correlation is a statistical measure that describes the strength and direction of the relationship between two variables.
Example 1: As study hours increase, exam scores often increase.
Example 2: As mobile usage increases, exam scores may decrease.

Importance:
→ Helps identify relationships between variables
→ Assists in prediction and trend analysis
→ Guides business and academic decision-making

Question 2: Explain the different types of correlation based on the direction of relationship with examples.

Answer:
1. **Positive Correlation:** Both variables increase or decrease together.
   Example: Study Hours and Exam Score.
2. **Negative Correlation:** One variable increases while the other decreases.
   Example: Mobile Usage and Exam Score.
3. **Zero Correlation:** No observable relationship.
   Example: Shoe size and exam score.

Question 3: Differentiate between linear and non-linear correlation with suitable examples.

Answer:
Linear Correlation:
→Relationship forms a straight line.
Example: Study Hours vs Exam Score.

Non-linear Correlation:
→Relationship forms a curve.

Example: Stress level vs performance (moderate stress improves performance, high stress reduces it).

Question 4: What is Pearson's correlation coefficient? Explain how its values are interpreted.

Answer:
Pearson's correlation coefficient measures the strength and direction of a linear relationship between two continuous variables.

Value Interpretation:
→ +1 : Perfect positive correlation
→ -1 : Perfect negative correlation
→ 0 : No correlation
Example:
A value of 0.85 indicates a strong positive relationship.

Question 5: Why is it risky to assume causation from correlation?

Answer:
Correlation does not imply causation because:
→ A third variable may influence both variables
→ Relationship may be coincidental
Example:
Ice cream sales and drowning incidents both increase in summer, but ice cream does not cause drowning.

LINK FOR EXCEL SHEET - EXCEL ANALYSIS SHEET

Question 6: Calculate Pearson's correlation between Study Hours and Exam Score using Excel. Interpret the result.

Solution:
Using Excel's CORREL() function, the Pearson correlation is calculated.
Result: 0.99
Interpretation:
There is a strong positive correlation, meaning higher study hours are associated with higher exam scores.

Question 7: Identify whether Mobile Usage and Exam Score have a positive or negative correlation. Justify your answer.

Solution:
Using Excel CORREL(), the value obtained is:
-0.99
Interpretation:
This indicates a negative correlation. Increased mobile usage is associated with lower exam scores.

Question 8: Does high attendance cause better exam scores? Justify using data.

Answer:
Data shows a positive correlation between attendance and exam score.

However, this does not prove causation.
Other factors like study habits and sleep also influence scores.

Question 9: Create a scatter plot for Attendance vs Exam Score and comment on the nature of correlation.

Solution:
A scatter plot was created in Excel.
Observation:
Points trend upward, indicating a strong positive linear correlation.

Question 10: Identify a possible confounding variable in the relationship between Study Hours and Exam Score. Explain.

Answer:
Sleep Hours is a confounding variable.

Students who sleep well can study more effectively and perform better, influencing both variables.