

Data Science Notes by Sarowar Ahmed

III Chapter: inferential statistics

Topic: Correlation Analysis

What is Correlation Analysis?

• Correlation analysis is a statistical method used to measure the strength and direction of the relationship between two quantitative variables. It helps us understand how changes in one variable are associated with changes in another, without implying causation.

+ Formula:

The Pearson correlation coefficient (r) is commonly used to measure the strength and direction of the linear relationship between two variables. The formula for r is:

$$r=\sum(xi-x^{-})(yi-y^{-})/\sqrt{\sum(xi-x^{-})^{2}}\cdot\sum(yi-y^{-})^{2}$$

Where:

- xi and yi are individual data points
- x^- and y^- are the means of x and y, respectively

Example:

- Suppose we want to investigate the relationship between study hours and exam scores. We collect data from 20 students and calculate a correlation coefficient (r) of 0.75. This positive value indicates a strong positive correlation between study hours and exam scores, suggesting that as study hours increase, exam scores tend to increase as well.
- Why is Correlation Analysis Useful?
- Correlation analysis allows us to:
 - 1. Identify associations: Determine if there's a relationship between variables.

- 2. Make predictions: Understand how changes in one variable may predict changes in another.
- 3. Guide decision-making: Use data-driven insights to inform decisions in various fields, from education to healthcare to finance.

Key Takeaway:

 Understanding correlation analysis empowers us to uncover meaningful patterns and relationships in data, leading to better insights and informed decision-making.

Got any questions on the Correlation Analysis? Feel free to ask me via Linkedin! Let's keep learning together.

My LinkedIn Date: 12/05/2024

