Notes

Association is a fundamental concept in statistics that helps us understand how to or more variables relate to each other. It is widely used in data analysis, machine learning, and scientific research.

Association refers to the relationship between two or more variables. If two variables are associated, knowing the value of one variable provides information about the value of the other.

However, association does not imply causation- just because two variables are related does not mean one causes the other.

Example:

* Stock Prices & interest rates: When interest rates increase, stock prices often decrease.
* Advertising & sales: More advertising is often associated with higher sales.
* Exercise & weight loss: More exercise is usually associated with lower weight.

Types of Association:

* Positive Association
  + Definition – When one variable increases the other variable also increases.
  + Example:
    - Higher education levels are associated with higher salaries.
    - Higher marketing spend is associated with higher sales.
  + Negative Association
    - Definition – When one variable increases, the other variable decreases.
    - Example:
      * Higher fuel prices are associated with lower car sales.
      * More exercise is associated with lower weight.
  + No Association
    - Definition – When two variables do not show any pattern of relationship.
    - Example:
      * Shoe size and intelligence level.
      * Number of pets a person owns and their monthly electricity bill.

Associative vs. Causation

* Example of False Causation (Spurious Correlation)
  + Ice cream sales & drowning deaths – Higher sales of ice cream are associated with more drownings.
    - Reality – Both happen more in summer (confounding variable = hot weather).
  + Number of pirates & Global Warming – As the number of pirates decreased, global temperatures increased.
    - Reality – These events are unrelated.
  + To establish causation, we need controlled experiments or statistical techniques like:
    - Regression analysis
    - Randomized control trials

Python Example

* Analyzed the relationship between Microsoft (MSFT) daily stock returns and NASDAQ-100 index returns over one year (252 trading days).
* Findings:
  + Pearson Correlation Coefficient – 0.72 (Strong Positive correlation)
  + P-value – 4.34e-42 (Extremely low, meaning statistically significant)
  + Conclusion – Microsoft and NASDAQ-100 returns are strongly associated, meaning when the NASDAQ-100 index moves up or down, Microsoft tends to move in the same direction.
  + P-value interpretation:
    - A p-value < 0.05 mean we can reject the null hypotheses (H0) and confirm a significant relationship exists.
    - If the p-value was high, we would not have enough evidence to conclude a meaningful association.
* Investors POV:
  + Microsoft is Highly Tied to NASDAQ-100.
    - When the NASDAQ-100 index rises, Microsoft’s stocks tend to rise.
    - When the NASDAQ-100 index falls, Microsoft’s stock tends to fall.
    - Investors holding Microsoft stock should monitor NASDAQ-100 trends.
  + Portfolio Diversification Insights.
    - If an investor already holds NASDAQ-100 ETF, adding Microsoft may not significantly increase diversification.
    - Instead, they might look for stocks with lower correlation to NASDAQ-100.
  + Market Sentiment & Risk Management
    - If NASDAQ-100 is expected to be volatile, Microsoft’s stock may also see similar price fluctuations.
    - Risk-averse investors might hedge against Microsoft by investing in less correlated assets.