

Correlation, Causality, and Simpson's Paradox in Statistics

Explore the intricate relationships in data through statistics. This presentation delves into correlation, causality, and Simpson's Paradox. Understand how these concepts shape our interpretation of data across various fields.

峰 by 峰 喬

Introduction to Statistics

Definition

Statistics is the science of collecting, analyzing, interpreting, presenting, and organizing data.

Importance

Understanding relationships in data helps make informed decisions. It's essential in economics, healthcare, and social sciences.



Correlation

1

Definition

Correlation expresses the extent to which two variables are linearly related.

2

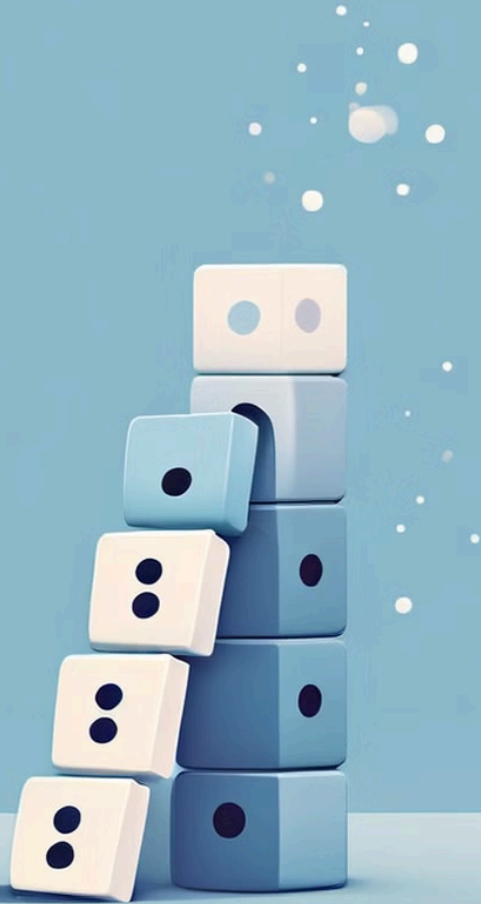
Types

Positive, negative, and no correlation exist between variables.

3

Measurement

Pearson Correlation Coefficient (r) ranges from -1 to +1.



Causality

Definition

Causality indicates that a change in one variable directly results in a change in another.

Key Distinction

Correlation does not imply causation.

Establishing Causality

Requires temporal precedence, covariation, and elimination of alternative explanations.

Examples of Correlation vs. Causation




Ice Cream and Drowning

Both increase in summer, but neither causes the other.
Temperature influences both.



Education and Income

Higher education often correlates with higher income. Other factors also play a role.



Introduction to Simpson's Paradox

1 Definition

A trend appears in different groups but disappears or reverses when combined.

2 Importance

Highlights the need to consider confounding variables and data context.



Example of Simpson's Paradox

Group	Treatment A	Treatment B
Men	90%	80%
Women	70%	60%
Combined	80%	85%

Implications of Simpson's Paradox



Data Analysis

Consider subgroup analysis. Be cautious of conclusions from aggregated data.



Applications

Relevant in healthcare, social sciences, economics, and policy-making.



Insight

Reveals the importance of contextual data interpretation.

Conclusion

1

Distinction

Correlation and causation are distinct concepts.

2

Context

Simpson's Paradox reminds us to analyze data contextually.

3

Importance

Understanding these concepts is crucial for accurate data interpretation.



References

1 Key Texts

"Statistics" by Freedman, Pisani, and Purves. "The Art of Statistics" by Spiegelhalter.

2 Additional Resources

Relevant academic papers and articles on correlation, causality, and Simpson's Paradox.

