

WHEN TO USE WHAT STATISTICAL TEST IN RESEARCH

There are several statistical test types for analyzing Research Data. When to use what is often the challenge. This piece provides a simplification

1 t-test:*

- Use when: You want to compare the means of two groups to determine if there's a significant difference.
- Example: You want to compare the average score of students who received traditional teaching vs. those who received innovative teaching.

2 ANOVA (Analysis of Variance):*

- Use when: You want to compare the means of three or more groups to determine if there are significant differences.
- Example: You want to compare the average score of students from different schools to determine if there are significant differences in their performance.

3 Regression (Simple and Multiple):*

- Use when: You want to examine the relationship between a dependent variable and one or more independent variables.
- Example: You want to examine the relationship between hours studied and exam scores (simple regression), or the relationship between hours studied, exam scores, and student motivation (multiple regression).

4 Chi-squared test:*

- Use when: You want to determine if there's a significant association between two categorical variables.
- Example: You want to determine if there's a significant association between smoking and lung cancer.

55 Wilcoxon rank-sum test (Mann-Whitney U test):*

- Use when: You want to compare the distributions of two independent groups.
- Example: You want to compare the distribution of scores between students who received traditional teaching and those who received innovative teaching.

66 Kruskal-Wallis H test:*

- Use when: You want to compare the distributions of three or more independent groups.
- Example: You want to compare the distribution of scores among students from different schools.

77 Friedman test:*

- Use when: You want to compare the distributions of three or more related groups.
- Example: You want to compare the distribution of scores among students at different time points.

88 Pearson correlation coefficient:*

- Use when: You want to examine the linear relationship between two continuous variables.
- Example: You want to examine the relationship between hours studied and exam scores.

99 Spearman rank correlation coefficient:*

- Use when: You want to examine the relationship between two variables when data is not normally distributed.

- Example: You want to examine the relationship between ranking of favorite foods and ranking of nutritional value.

* Kendall's tau correlation coefficient: *

- Use when: You want to examine the relationship between two variables when data is ordinal or categorical.

- Example: You want to examine the relationship between socioeconomic status and education level.

11 ARIMA models:

- Use when: You want to forecast future values in a time series data.

- Example: You want to predict stock prices based on past trends.

12 Exponential smoothing (ES):

- Use when: You want to forecast future values in a time series data with a simple exponential smoothing method.

- Example: You want to predict sales based on past trends.

13 Seasonal decomposition:

- Use when: You want to decompose time series data into trend, seasonality, and residuals.

- Example: You want to analyze website traffic data to identify seasonal patterns.

14 Kaplan-Meier estimator:

- Use when: You want to estimate the survival function of a population.
- Example: You want to analyze the survival rate of patients with a specific disease.

15 Cox proportional hazards model:

- Use when: You want to examine the relationship between covariates and survival time.
- Example: You want to investigate the effect of treatment on survival time.

16 Log-rank test:

- Use when: You want to compare the survival curves of two or more groups.
- Example: You want to compare the survival rates of patients with different treatments.

17 K-means clustering:

- Use when: You want to group similar observations into clusters based on features.
- Example: You want to segment customers based on buying behavior.

18 Hierarchical clustering:

- Use when: You want to group similar observations into clusters based on features, with a hierarchical structure.
- Example: You want to analyze gene expression data to identify clusters of genes.

19 DBSCAN (density-based spatial clustering of applications with noise):

- Use when: You want to group similar observations into clusters based on features, with noise handling.
- Example: You want to analyze spatial data to identify clusters of high density.

20 Principal component analysis (PCA):

- Use when: You want to reduce the dimensionality of a dataset by identifying principal components.
- Example: You want to analyze stock prices to identify principal components of variation.

21 Discriminant analysis:

- Use when: You want to predict group membership based on multivariate data.
- Example: You want to predict customer churn based on usage patterns.

22 Canonical correlation analysis:

- Use when: You want to examine the relationship between two sets of multivariate data.
- Example: You want to investigate the relationship between personality traits and behavior.

23 Bayesian inference:

- Use when: You want to update probabilities based on new data.
- Example: You want to update the probability of a hypothesis based on new evidence.

24 Bayesian regression:

- Use when: You want to model the relationship between variables using Bayesian methods.
- Example:

25 Bayesian networks:

- Use when: You want to model complex relationships between variables using Bayesian methods.
- Example: You want to model the relationship between genes and diseases.

26 Decision trees:

- Use when: You want to classify observations based on a tree-like model.
- Example: You want to predict customer churn based on usage patterns.

27 Random forests:

- Use when: You want to classify observations based on an ensemble of decision trees.
- Example: You want to predict disease diagnosis based on symptoms.

28 Support vector machines (SVMs):

- Use when: You want to classify observations based on a hyperplane.
- Example: You want to predict customer churn based on usage patterns.

29 Cluster analysis:

- Use when: You want to group similar observations into clusters based on features.
- Example: You want to segment customers based on buying behavior.

30 Factor analysis:

- Use when: You want to reduce the dimensionality of a dataset by identifying underlying factors.
- Example: You want to analyze survey data to identify underlying factors of satisfaction.

31 Survival analysis:

- Use when: You want to analyze the time-to-event data.
- Example: You want to analyze the survival rate of patients with a specific disease.

32 Time-series analysis:

- Use when: You want to analyze data that is ordered in time.
- Example: You want to analyze stock prices to identify patterns and trends.

33 Non-parametric tests:

- Use when: You want to analyze data without assuming a specific distribution.
- Example: You want to compare the median scores of students who received traditional teaching vs. those who received innovative teaching.

34 Machine learning algorithms:

- Use when: You want to predict outcomes or classify observations based on large datasets.
- Example: You want to predict customer churn based on usage patterns.

The specific test or technique used depends on the research question, data type, and study design.

You can learn all these and more at the Research Data Analysis Course

#mosesudoisoh #researchdataanalysis