

Statistical Procedures

Technique	Use	Assumptions	Example
Regression Analysis	Predict a continuous dependent variable based on one or more independent variables	Linearity Homoscedasticity Independence of errors	Predicting house prices based on square footage.
Pearson Correlation	Measure linear association between 2 continuous variables	Linearity Homoscedasticity	Examine the relationship between age and income
Tests	Use	Assumptions	Example
t- test	Compare means of two groups	Normal distribution Equal variances	Compare test scores of two teaching methods
ANOVA	Compare means of more than two groups	Normal distribution Equal variances	Compare test scores among three teaching groups
Chi-Squared Tests	Test independence between categorical variables	Random sample Large sample size	Test if there is an association between gender and tea/coffee preference
Mann-Whitney U Test	Compare distributions of two independent groups	Independence of observations	Compare test scores between two different schools
Kruskal-Wallis Test	Compare distributions of more than two independent groups	Independence of observations Similar distributions	Compare test scores of different teaching methods across multiple schools
Wilcoxon Signed Rank Test	Compare distributions of two related groups	Dependent variable is continuous or ordinal	Compare pre-test and post-test scores within a group
McNemar's Test	Compare paired proportions or frequencies in a 2x2 contingency table	Binary data Dependent groups	Compare the effectiveness of two treatments on a binary outcome
Fisher's Exact Test	Compare proportions in a 2x2 contingency table	Low sample size Independence of observations	Assess the association between gender and smoking status in a small sample

Definitions

Contingency Table

A matrix format that displays the multivariate frequency distribution of the variables.

Data

Data can be qualitative (categorical) or quantitative

Quantitative data can be discrete or continuous

Qualitative data can be nominal or ordinal

Nominal Data is used to label variables without any order or quantitative value.

The colour of hair can be considered nominal data, as one colour can't be compared with another colour.

Ordinal

Ordinal data have natural ordering where a number is present in some kind of order by their position on the scale. These data are used for observation like customer satisfaction, happiness, etc., but we can't do any arithmetical tasks on them. Compared to nominal data, ordinal data have some kind of order that is not present in nominal data.

Homoscedasticity

Also termed homogeneity of variance, it means that the variance of the error terms (residuals) in a regression model is constant across all levels of the independent variable(s).

Linearity

The assumption that the relationship between two variables is linear

Measures of Association

Any of various factors or coefficients used to indicate a relationship between two or more variables.

Multivariate

Involving two or more variable quantities

Statistical Procedures

Technique	Use	Assumptions	Example
Regression Analysis	Predict a continuous dependent variable based on one or more independent variables	Linearity Homoscedasticity Independence of errors	Predicting house prices based on square footage.
Pearson Correlation	Measure linear association between 2 continuous variables	Linearity Homoscedasticity	Examine the relationship between age and income
Tests	Use	Assumptions	Example
t- test	Compare means of two groups	Normal distribution Equal variances	Compare test scores of two teaching methods
ANOVA	Compare means of more than two groups	Normal distribution Equal variances	Compare test scores among three teaching groups
Chi-Squared Tests	Test independence between categorical variables	Random sample Large sample size	Test if there is an association between gender and tea/coffee preference
Mann-Whitney U Test	Compare distributions of two independent groups	Independence of observations	Compare test scores between two different schools
Kruskal-Wallis Test	Compare distributions of more than two independent groups	Independence of observations Similar distributions	Compare test scores of different teaching methods across multiple schools
Wilcoxon Signed Rank Test	Compare distributions of two related groups	Dependent variable is continuous or ordinal	Compare pre-test and post-test scores within a group
McNemar's Test	Compare paired proportions or frequencies in a 2x2 contingency table	Binary data Dependent groups	Compare the effectiveness of two treatments on a binary outcome
Fisher's Exact Test	Compare proportions in a 2x2 contingency table	Low sample size Independence of observations	Assess the association between gender and smoking status in a small sample

Definitions

Contingency Table

A matrix format that displays the multivariate frequency distribution of the variables.

Data

Data can be qualitative (categorical) or quantitative

Quantitative data can be discrete or continuous

Qualitative data can be nominal or ordinal

Nominal Data is used to label variables without any order or quantitative value.

The colour of hair can be considered nominal data, as one colour can't be compared with another colour.

Ordinal

Ordinal data have natural ordering where a number is present in some kind of order by their position on the scale. These data are used for observation like customer satisfaction, happiness, etc., but we can't do any arithmetical tasks on them. Compared to nominal data, ordinal data have some kind of order that is not present in nominal data.

Homoscedasticity

Also termed homogeneity of variance, it means that the variance of the error terms (residuals) in a regression model is constant across all levels of the independent variable(s).

Linearity

The assumption that the relationship between two variables is linear

Measures of Association

Any of various factors or coefficients used to indicate a relationship between two or more variables.

Multivariate

Involving two or more variable quantities