
Postgraduate Certificate in Multivariate Analysis with R

Advanced Topics in Multivariate Analysis

Advanced Topics in Multivariate Analysis Glossary:

A:

- ANOVA (Analysis of Variance): A statistical method used to analyze the differences among group means in a sample.
- Assumption: A condition that must be met before statistical techniques can be applied.

B:

- Bootstrapping: A resampling technique used to estimate the distribution of a statistic by repeatedly sampling with replacement from the original data.
- Bayesian Analysis: A statistical approach that uses Bayes' theorem to update the probability of a hypothesis as new evidence becomes available.

C:

- Canonical Correlation Analysis: A multivariate technique used to assess the relationship between two sets of variables.
- Cluster Analysis: A method used to group similar objects into clusters based on their characteristics.

D:

- Discriminant Analysis: A technique used to classify objects into predefined categories based on their characteristics.
- Dimensionality Reduction: Techniques used to reduce the number of variables in a dataset while preserving important information.

E:

- Exploratory Factor Analysis: A statistical method used to identify underlying factors that explain patterns of correlations among variables.
- Eigenvalue: A scalar value that represents the amount of variance explained by a principal component in PCA.

F:

- Factor Analysis: A statistical method used to identify underlying factors that explain patterns of correlations among observed variables.
- Factor Loading: The correlation between an observed variable and a factor in factor analysis.

G:

- Generalized Linear Models: A class of models that generalizes linear regression to accommodate non-normal error distributions.
- Goodness of Fit: A measure of how well a model fits the data.

H:

- Hierarchical Clustering: A method of cluster analysis that builds a hierarchy of clusters.
- Hotelling's T-squared: A multivariate statistical test used to compare the means of two groups.

I:

- Independent Component Analysis (ICA): A technique used to separate a multivariate signal into additive, independent components.
- Interpretation: The process of explaining the meaning of statistical results in the context of the research question.

J:

- Jackknife Resampling: A resampling technique used to estimate the bias and variance of a statistic.
- Joint Distribution: The probability distribution of two or more random variables.

K:

- K-means Clustering: A method of cluster analysis that partitions data into k clusters.
- Kurtosis: A measure of the "tailedness" of the probability distribution of a real-valued random variable.

L:

- Latent Variable: A variable that is not directly observed but is inferred from observed variables.
- Linear Discriminant Analysis (LDA): A technique used to find a linear combination of features that best separates classes.

M:

- MANOVA (Multivariate Analysis of Variance): A statistical method used to analyze the differences among group means in multiple dependent variables.
- Missing Data: Data that is not available for some observations in a dataset.

N:

- Nonparametric Methods: Statistical techniques that do not make assumptions about the distribution of the data.
- Normality: A condition where the data follows a normal distribution.

O:

- Outlier: An observation that deviates significantly from the rest of the data.
- Ordination: A method used to visualize the relationships between objects in a dataset.

P:

- PCA (Principal Component Analysis): A technique used to reduce the dimensionality of a dataset by finding orthogonal linear combinations of variables.
- Permutation Test: A nonparametric test that assesses the significance of a statistic by permuting the data.

Q:

- Quantile Regression: A regression technique that estimates the conditional quantiles of a response variable.
- Q-Q Plot: A graphical tool used to assess whether the data comes from a specific distribution.

R:

- Regression Analysis: A statistical technique used to model the relationship between a dependent variable and one or more independent variables.
- Residual Analysis: The examination of the difference between observed and predicted values in a regression model.

S:

- Scree Plot: A graphical tool used to determine the number of components to retain in factor analysis or PCA.

- Standardization: The process of transforming variables to have a mean of 0 and a standard deviation of 1.

T:

- Test Statistics: A value calculated from sample data that is used to make inferences about a population parameter.

- Time Series Analysis: A statistical technique used to model and analyze time-dependent data.

U:

- Unsupervised Learning: Machine learning techniques that do not require labeled data for training.
- Univariate Analysis: Statistical analysis of a single variable at a time.

V:

- Variance-Covariance Matrix: A square matrix that contains the variances of variables on the diagonal and covariances off-diagonal.
- Variable Selection: The process of choosing a subset of variables that best predict the outcome in a model.

W:

- Wilks' Lambda: A multivariate statistical test used to determine the significance of the differences among group means in MANOVA.
- Ward's Method: A hierarchical clustering algorithm that minimizes the within-cluster variance.

X:

- X-means Clustering: An extension of K-means clustering that automatically determines the number of clusters.
- X-bar Chart: A control chart used to monitor the mean of a process.

Y:

- Yates' Correction: A correction factor used in contingency table analysis to adjust for small sample sizes.
- Yield Analysis: A statistical technique used to optimize the yield of a manufacturing process.

Z:

- Z-score: A standardized score that indicates how many standard deviations a data point is from the mean.

- Z-test: A statistical test used to determine whether the mean of a sample differs significantly from a known population mean.