

Introduction to Multivariate Analysis

Introduction to Multivariate Analysis

(A concise, well-organized guide for students and practitioners)

1. What Is Multivariate Analysis?

Aspect	Description
Definition	Statistical techniques that examine more than two variables simultaneously to understand relationships, patterns, and underlying structures.
Goal	• Summarize complex data sets • Test hypotheses about inter-variable relationships • Reduce dimensionality while retaining information • Predict outcomes using multiple predictors
When to Use	• Surveys with many questions • Biological/medical data (e.g., gene expression) • Market research (multiple product attributes) • Any situation where variables are inter-correlated and a single-variable analysis would be misleading.

2. Core Concepts & Terminology

Term	Meaning
Variables	Measured attributes (continuous, ordinal, or categorical).
Observations	Individual cases/subjects (rows in a data matrix).
Covariance Matrix (?)	Square matrix of covariances between every pair of variables; captures joint variability.
Correlation Matrix (R)	Standardized covariances (values between -1 and 1).
Dimensionality	Number of variables (p). High-dimensional data: $p \gg n$ (observations).
Factor / Component	Latent construct derived from observed variables (e.g., principal components).
Eigenvalues / Eigenvectors	Numbers/vectors that describe variance explained and direction of new axes in methods like PCA.

Introduction to Multivariate Analysis

Term	Meaning
Multicollinearity	High correlation among predictors; can inflate variance of coefficient estimates. • Linearity (for many methods)
Assumptions	• Multivariate normality (for inference) • Homogeneity of variance?covariance matrices (MANOVA) • Independence of observations

3. Major Families of Techniques

Family	Typical Purpose	Representative Methods
Exploratory	Discover structure, reduce dimensions	• Principal Component Analysis (PCA) • Factor Analysis (FA) • Multidimensional Scaling (MDS) • Multivariate Analysis of Variance (MANOVA)
Confirmatory / Inferential	Test hypotheses about group differences or relationships	• Canonical Correlation Analysis (CCA) • Discriminant Analysis • Multivariate Linear Regression
Predictive / Regression?Based	Model a set of dependent variables using predictors	• Partial Least Squares (PLS) Regression • Generalized Linear Models (GLM) for multivariate outcomes • K?means / Hierarchical clustering (applied to multivariate space)
Cluster & Classification	Group observations or assign class labels	• Support Vector Machines (SVM) with multiple features • Random Forests, Neural Networks (multivariate inputs)

Introduction to Multivariate Analysis

4. Step-by-Step Workflow

1. Define the Research Question

- Identify dependent and independent variables.
- Clarify whether the aim is description, inference, or prediction.

2. Collect & Organize Data

- Structure data in a matrix $\mathbf{X}$ ($n \times p$) where n = observations, p = variables.
- Include a separate matrix $\mathbf{Y}$ for multivariate responses