

Multivariate Analysis

Multivariate Analysis – Quick Reference Notes

1. What Is Multivariate Analysis?

- **Definition:** Statistical techniques that examine **more than two variables** simultaneously to understand relationships, patterns, and underlying structure in a dataset.
 - **Why Use It?**
 - Captures the **interdependence** among variables.
 - Reduces dimensionality while preserving information.
 - Improves prediction and classification accuracy.
 - Helps in **exploratory data analysis** (EDA) and hypothesis testing.
-

2. Core Concepts & Terminology

Term	Meaning
Variable	Measured attribute (continuous, categorical, ordinal).
Dimension	Number of variables (p).
Observation / Case	One row of data (n).
Covariance Matrix (?)	$p \times p$ matrix summarizing pairwise covariances; central to many methods.
Correlation Matrix (R)	Standardized covariances (values between -1 and 1).
Eigenvalues / Eigenvectors	Decompose Σ or R; eigenvalues indicate variance explained, eigenvectors give direction of new axes.
Factor Loadings	Correlations between observed variables and latent factors.
Score	Projected value of an observation on a derived component/factor.
Residual	Difference between observed and model-predicted values.

3. Major Families of Techniques

Family	Goal	Typical Methods	Typical Output
--------	------	-----------------	----------------

Multivariate Analysis

Dimension? reduction	Summarize many variables with few latent constructs	Principal Component Analysis (PCA), Factor Analysis (FA),	Component/factor scores, loadings, scree plot
		Multidimensional Scaling (MDS)	
Classification / Discrimination	Assign observations to predefined groups	Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (multinomial)	Discriminant functions, classification tables
Clustering	Group similar observations without prior labels	K?means, Hierarchical clustering, Gaussian Mixture Models	Cluster assignments, dendrograms
Multivariate Regression	Model several dependent variables simultaneously	Multivariate Linear Regression (MLR), Partial Least Squares (PLS), Canonical Correlation Analysis (CCA)	Coefficient matrices, canonical variates
MANOVA / MANCOVA	Test mean differences across groups for multiple outcomes	MANOVA, MANCOVA (covariate adjustment)	Wilks' ?, Pillai's trace, Hotelling's T ²
Time?Series / Panel Multivariate	Analyze multiple series over time	Vector Autoregression (VAR), Dynamic Factor Models	Impulse?response functions, Granger causality tests
Survival / Event? History (multivariate)	Model time?to?event with several predictors	Cox proportional hazards with multiple covariates, Competing?risk models	Hazard ratios, baseline hazard

4. Step?by?Step Workflow (Typical for PCA/FA)

1. Data Preparation

- Clean missing values (impute or remove).
- Standardize variables (z?scores) if units differ.

2. Assess Suitability

Multivariate Analysis

- **Kaiser-Meyer-Olkin (KMO)** test ? >0.6 good.
- **Bartlett's test of sphericity** ? $p < 0.05$ indicates sufficient correlations.

3.