Generic Elective (GE) Course 3b: Applied Multivariate Statistics

Structure 1: PG Curricular Structure with only Course Work
Structure 2: PG Curricular Structure with Course Work + Research

Course Title & Code	Credits	Credit Distribution of the Course			Eligibility Criteria	Prerequisite of the course (if any)
		Lecture (45 Hours)	Tutorial (00 Hours)	Practical (30 hours)		
GE 3b: Applied Multivariate Statistics	**4**	**3**	**0**	**1**	**NIL**	**Basic knowledge of Probability Theory and Linear Algebra**

Course Objectives:

- Introduces students to multivariate statistical techniques used in applied research.
- Emphasis on understanding concepts, interpreting results, and applying methods to real data using statistical software and tools.

Course Learning Outcomes: After successful completion of this course, the students will be able to:

- Understand multivariate data structures and matrix notation.
- Apply techniques for data reduction and visualization.
- Analyze relationships among multiple variables simultaneously.
- Use multivariate methods for classification, prediction, and inference.
- Interpret multivariate output from software in practical contexts.

Unit I (9 hours)

Concept of bivariate and trivariate data, covariance and correlation (Pearson's r, correlation matrix), scatterplots, correlation heatmaps, pairwise plots, idea of random vectors and covariance matrix (without derivations), concept of multivariate normal distribution: definition, properties, and testing for normality, idea of linear combinations and marginal distributions (concept only), outlier detection using Mahalanobis distance, checking multivariate normality (graphical and numerical methods).

Unit II (12 hours)

Multiple and partial correlation; multiple linear regression – concept, interpretation of coefficients, R^2, adjusted R^2, and residuals; idea of multicollinearity; model checking (linearity, normality, and homoscedasticity) and visualization using diagnostic plots; introduction to canonical correlation – concept and interpretation; practical interpretation of correlation and regression output.

Unit III (10 hours)

Principal Component Analysis (PCA): eigenvalues and eigenvectors, total variance explained, scree plot, component loadings, interpretation, Factor Analysis: common factor model, estimation of loadings, rotation (varimax), communalities, Comparison of PCA and Factor Analysis, use of software to perform PCA/FA and interpret outputs.

Unit IV (14 hours)

Concept of grouping similar observations; hierarchical clustering (concept, dendrogram interpretation) and k-means clustering (steps, centroids, and output interpretation), Discriminant Analysis: concept of classifying observations into known groups using predictor variables; linear discriminant function, idea of classification accuracy, confusion matrix, and misclassification rate; Comparison between clustering (unsupervised) and discriminant (supervised) methods, overview of real-world applications and interpretation.

Essential Readings:

1. Everitt, B.S. and Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*, Springer.
2. Hair, J.F., Babin, B.J., Anderson, R.E. and Black, W.C. (2022). *Multivariate Data Analysis*, Cengage Learning.
3. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021). *An introduction to statistical learning: with applications in R*, Springer.
4. Johnson, R.A. and Wichern, D.W. (2019). *Applied Multivariate Statistical Analysis*, Pearson.

Suggested Readings:

1. Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons.
2. Bakker, J.D. (2024). *Applied Multivariate Statistics in R.* University of Washington.
3. Chatfield, C. and Collins, A.J. (1980). *Introduction to Multivariate Analysis*, Chapman & Hall.

4. Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*, Sage Publications.
5. Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E (2010). *Multivariate Data Analysis*, Pearson.
6. Izenman, A.J. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer.
7. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R*, Springer.
8. Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, Wiley.
9. Rencher, A.C. and Christensen, W.F. (2012). *Methods of Multivariate Analysis*, John Wiley & Sons.

List of Practicals:

1. Data Import and Exploration
2. Visualization and Correlation Matrix
3. Multiple and Partial Correlation
4. Multiple Regression and Diagnostics
5. Principal Component Analysis (PCA)
6. Factor Analysis
7. Cluster Analysis (Hierarchical & K-Means)
8. Discriminant Analysis and Classification