



B.Com Honours

Semester III

Calicut University

DATA ANALYTICS WITH STATISTICAL SOFTWARE

Course Code: COM3MN210 • Module 4 Notes

Module IV: Predictive Modeling & Variable Reduction

Predictive modeling utilizes mathematical algorithms to forecast future outcomes, while variable reduction techniques compress high-dimensional survey datasets into underlying latent factors.

Unit 14: Predictive Modeling – Meaning, Definition, & Importance

Process of building statistical models using historical business data to estimate the probability of future events (e.g., customer churn prediction, loan default forecasting).

Unit 15: Classification of Business Problems

Regression Problems: Predicting continuous numeric targets (e.g., quarterly revenue amount).

Classification Problems: Predicting discrete categorical outcomes (e.g., Default vs. Non-Default).

Unit 16: Mapping Analytical Techniques to Business Problems

Matching algorithms to business needs: Linear Regression for revenue estimation, Logistic Regression for binary classification, Decision Trees for credit risk rules.

Unit 17: Phases of Predictive Modeling

Phases: Problem Definition → Data Collection & Prep → Exploratory Data Analysis (EDA) → Model Training & Validation (Train/Test Split) → Deployment.

Unit 18: Data Preparation – Missing Value Imputation

Treating incomplete survey records via Listwise Deletion, Pairwise Deletion, or Mean/Median Imputation techniques.

Unit 19: Outlier Detection & Treatment Methods

Identifying extreme anomalies using Z-Score ($|Z| > 3$) or Interquartile Range ($IQR = Q3 - Q1$) Boxplot fences, followed by Winsorization or removal.

Unit 20: Variable Reduction & Linear Regression

Linear Regression Model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

Goodness-of-Fit (R^2): Measures percentage of target variance explained by predictors.

Unit 21: Exploratory Factor Analysis (EFA) – KMO & Bartlett's Test

- 1. Purpose:** Reduces 20+ survey items into 3-4 underlying latent factors (e.g., brand trust, service quality).
- 2. Diagnostics:** Kaiser-Meyer-Olkin (KMO > 0.6) & Bartlett's Test of Sphericity ((p < 0.05)) confirm data factorability. Eigenvalues > 1 criterion determines factor extraction.

Practical Case Framework (Factor Analysis Diagnostics & Regression Output Matrix): Diagnostic criteria for evaluating Linear Regression and Exploratory Factor Analysis in SPSS.

Predictive & Reduction Diagnostic Benchmark Table:			
Analytical Technique	SPSS Diagnostic Metric	Acceptable Threshold	Statistical Interpretation
Multiple Regression	R-Squared ((R^2))	(R^2 > 0.60) (60%+)	Proportion of sales variance explained by marketing spend
Multiple Regression	VIF (Multicollinearity)	(VIF < 5.0)	Ensures independent predictors are not highly correlated
Factor Analysis (EFA)	KMO Measure of Sampling	(KMO > 0.60)	Confirms sample size adequacy for factor extraction
Factor Analysis (EFA)	Bartlett's Test Sig.	(p < 0.05)	Rejects identity matrix; item correlation matrix is factorable
Factor Extraction	Kaiser's Eigenvalue Rule	Eigenvalue > 1.0	Retains factors explaining more variance than a single variable

Acing your exams is just a click away!

Visit www.degreeelive.in to download the next module for free.

DegreeLive