

Suggested Teaching Guidelines for
Advanced Analytics using Statistics
PG-DBDA September 2022

Objective: To perform advanced analytics using Python & R skills and important mathematical concepts.

Prerequisites: Good Knowledge of Basic Mathematics

List of Books / Other training materials

Text Book:

1. Statics Using R by Sudha Purohit, Pub: Narosa
2. Practical Statistics For Data Scientists 2/Ed 50+ Essential Concepts Using R and Python by Peter Bruce (Shroff/O'Reilly Publisher)

Reference:

1. Beginning R – The Statistical Programming Language by Dr. Mark Gardener PUB: WILEY
2. Art of Programming in R, by Norman Matloff
3. Statistics for Management by Levin
4. Business Analytics: Methods, Models, and Decisions by James R Evans
5. Introductory Statistics with R (Statistics and Computing) by Peter Dalgaard
6. R in a Nutshell by Joseph Adler (O'REILLY)
7. R Cookbook by Paul Teetor (O'REILLY)
8. The R Book, Second Edition
9. Statistics Using R, Shailaja Deshmukh, Sudha Purohit, Sharad Gore, Pub: Narosa
10. Statistical Inference via Data Science: A Modern Dive into R and the Tidyverse by Chester Ismay (Chapman & Hall Publisher)
11. Statistics for Machine Learning: Implement Statistical methods used in Machine Learning using Python by Himanshu Singh (BPB Publications)
12. Statistics: Statistics for Beginners in Data Science: Theory and Applications of Essential Statistics Concepts using Python by AI Publishing

Note:

- Each session mentioned is for theory and of 2 hours' duration. Lab assignments are indicatives; faculty needs to assign more assignments for better practice.
- Trainer has to teach the statistical and probability concepts involved here in detail

Session 1, 2 and 3:

- Introduction to Analytics
- Data analytics Life Cycle:
- Discovery,
- Types of Data
- Pulling the data from CSVs
- Data preparation
- Model planning
- Model building implementation
- Quality assurance
- Documentation

Suggested Teaching Guidelines for
Advanced Analytics using Statistics
PG-DBDA September 2022

- Management approval
- Installation
- Acceptance and operation
- Intelligent data analysis
- Calculating frequency counts of univariate and bivariate crosstabs and interpreting the normalize option of pandas.crosstab

Session 4 & 5:

- Random Variable
- Concepts of Correlation
- Covariance
- Outliers
- Producing graphs like bar chart, pie chart, histogram, boxplot, density plot, scatter plot with different options in pandas, matplotlib and seaborn libraries
- Detecting Outliers using Boxplot

Session 6 & 7:

- Sample Spaces and Events
- Concept of Probability: Addition, Multiplicative, Complement Rules
- Joint, Conditional and Marginal Probability
- Bayes' Theorem
- Usage of sklearn.BernoulliNB function to predict probabilities (predict_proba () method)

Session 8 & 9:

- Probability Distribution
 - Discrete distribution – (Binomial, Poisson) Probability Mass Functions, Distribution Functions
 - Continuous distribution – (Normal) Probability Density Function, Distribution Function, Inverse of Distribution Function

Session 10:

- Descriptive Statistical Measures
- Summary Statistics - Central Tendency & Dispersion (Mean, Median, Mode, Quartiles, Percentiles, Range, Interquartile Range, Standard Deviation, Variance, and Coefficient of Variation)

Suggested Teaching Guidelines for
Advanced Analytics using Statistics
PG-DBDA September 2022

Session 11:

- Sample & population, Uni-variate and bi-variate sampling, re-sampling
- Sampling and Estimation: Sampling Distribution
- Concept of Confidence Interval
- Central Limit Theorem

Session 12 & 13:

- Statistical Inference Terminology (types of errors, tails of test, confidence intervals etc.)
- Hypothesis Testing
- Parametric Tests: One sample t-test, paired t-test, 2 independent samples t-test, 1-Way ANOVA
- Non-parametric Tests- chi-Square, U-Test

Session 14:

- Predictive Modelling (From Correlation to Supervised Segmentation):
 - Identifying Informative Attributes,
 - Segmenting Data by Progressive Attributive,
 - Models,
 - Induction and Prediction,
 - Supervised Segmentation,
 - Visualizing Segmentations,
 - Trees as Set of Rules,
 - Probability Estimation;

Session 15 & 16:

- Simulation and Risk Analysis
- Monte Carlo Simulation Method
- Optimization, Linear
- Formulating any Linear Programming Problem (LPP) and solving it using Excel and Python options

Session 17:

- Decision Analytics:
 - Evaluating Classifiers,
 - Analytical Framework,
 - Evaluation,
 - Baseline,
 - Performance and Implications for Investments in Data;

Suggested Teaching Guidelines for
Advanced Analytics using Statistics
PG-DBDA September 2022

Session 18:

- Evidence and Probabilities:
 - Explicit Evidence Combination with Bayes Rule,
 - Probabilistic Reasoning;

Session 19:

- Business Strategy:
 - Achieving Competitive Advantages,
 - Sustaining Competitive Advantages

Session 20:

- Factor Analysis,
- Directional Data Analytics,

Session 21 & 22:

- Interactivity with ipwidgets in Jupyter Notebook
- Creating simple interactive graphics with ipwidgets
- Creating a simple What-if tool for predicting using ipwidgets

Session 23:

- Generating Simple Interactive Applications in Shiny App with R
 - Interactive Boxplots, Histograms
 - Interactive scatter plots