

Statistical Foundations of Data Science (SFDS)

27 December 2020

SFDS-01: Basic Statistics with R

8-hr online course via Google Meet delivered over two successive weekends

Dates will be announced on social media

All exercises will be conducted using R, thus SFDS-1 is very fundamental

➤ **Part-1: Introduction to R**

- ❖ History of R
- ❖ Basic Operations in R – Objects, Functions, Plots

➤ **Part-2: Probability and Probability Distributions, 2 hr**

- ❖ Probability: Addition and multiplication rules, Bayes' Theorem
- ❖ Probability Distributions: Binomial, Poisson and Normal

➤ **Part-3: Sampling Distributions and Hypothesis Testing**

- ❖ T-Distribution
- ❖ Confidence Intervals
- ❖ Type-I and Type-II errors
- ❖ T-tests and ANOVA

➤ **Part-4: Hypothesis Testing (cont'd)**

- ❖ F-Test and Homogeneity of Variance
- ❖ Chi-Squared
- ❖ Non Parametric Tests – can be elaborated separately if required

SFDS-02: Data Visualization and Time Series

9-hr online course via Google Meet delivered over two successive weekends
Dates will be announced on social media

➤ **Part-1: Three Distributions and Basic Data Analysis**

- ❖ Normal Distribution and Descriptive Statistics
- ❖ Binomial and Poisson Distributions, and their descriptive statistics

➤ **Part-2: The Visual Presentation of Data**

- ❖ Histogram, Stem and Leaf Plot, Probability Plotting, Bar Chart, Pareto, Pie Chart, Dot Plot,
- ❖ Box Plot, Scatter Plot, Matrix Plots, Chernoff-Flury Faces

➤ **Part-3: Elements of Time Series-A**

- ❖ Basics of Forecasting
- ❖ Time Series Graphics

➤ **Part-4: Elements of Time Series-B**

- ❖ Time Series Decomposition
- ❖ The Forecaster's Toolbox
- ❖ Time Series Models: ETS, Holt, Holt-Winters and ARIMA
- ❖ Predictions using Models

SFDS-03: Statistical Learning – Regression

8-hr online course via Google Meet delivered over two successive weekends
Dates will be announced on social media

➤ **Part-1: Statistical Learning and Linear Regression, 2 hr**

- ❖ Introduction to Statistical Learning – Supervised and Unsupervised
- ❖ Regression vs Classification
- ❖ Method of Least Squares
- ❖ Linear Regression

➤ **Part-2: Linear and Multiple Regression, 2 hr**

- ❖ Model Adequacy for Linear Regression
- ❖ Multiple Regression
- ❖ Handling Qualitative Predictors

➤ **Part-3: Classification Techniques, 2 hr**

- ❖ Logistic Regression
- ❖ Linear Discriminant Analysis
- ❖ K-Nearest Neighbors

➤ **Part-4: Resampling Methods**

- ❖ Cross Validation
- ❖ Bootstrap

SFDS-4: Statistical Learning – Advanced Topics

8-hr online course via Google Meet delivered over two successive weekends
Dates will be announced on social media

➤ **Part-1: Model Selection and Regularization**

- ❖ Subset Selection
- ❖ Shrinkage Methods: Ridge Regression and Lasso
- ❖ Dimension Reduction Methods: Principal Components and Partial Least Squares

➤ **Part-2: Tree Based Methods, 2 hr**

- ❖ Classification and Regression Trees
- ❖ Bagging and Random Forests

➤ **Part-3: Support Vector Machines, 2 hr**

- ❖ Maximum Marginal Classifier
- ❖ Support Vector Classifier

➤ **Part-4: Unsupervised Learning**

- ❖ Principal Components Analysis (PCA)
- ❖ K-Means Clustering