Statistical Techniques for Data Science
Basics

Duration – 1.5 Days / 12 Hours

Program Description

This course introduces foundational concepts in statistics and linear algebra. Learners will explore methods for data collection, categorization, and presentation, while also understanding key measures of central tendency (mean, median, mode) and dispersion (range, variance, standard deviation). The course also covers elementary linear algebra, focusing on scalars,

Learning Goals

- Understand key statistical terms and concepts like mean, median, mode, variance, and standard deviation.
- Perform basic operations with scalars, vectors, and matrices.
- Grasp fundamental concepts of probability and its real-world applications.
- Implement basic data collection, categorization, and presentation methods.

Course Topics

- Introduction to Statistics and Data Collection: Data categorization and representation methods, measures of central tendency, measures of dispersion.
- Introduction to Linear Algebra: Scalars, vectors, matrices, and operations on them (addition, multiplication, transpose, inverse).
- Basic Probability Concepts: Introduction to probability, sample spaces, and simple events

Statistical Techniques for Data Science
Intermediate

Duration – 2 Days / 16 Hours

Program Description

This course builds on the foundational concepts, diving deeper into probability theory, probability distributions, and hypothesis testing. Learners will explore various probability distributions such as Normal, Binomial, and Poisson, and understand how to apply them to real-world scenarios. This phase also covers sampling techniques and estimation methods, which are essential for making inferences about larger populations. Hypothesis testing is introduced, with a focus on parametric tests like t-tests and ANOVA, enabling students to analyze differences between groups and datasets.

Learning Goals

- Apply advanced probability distributions such as Normal, Binomial, and Poisson distributions.
- Understand and implement sampling techniques and point/interval estimates.
- Apply the Central Limit Theorem and develop confidence intervals for population parameters.
- Perform parametric hypothesis testing using t-tests, z-tests, and ANOVA.

Course Topics

- Probability Distributions: Understanding Normal, Binomial, and Poisson distributions and their applications.
- Sampling and Estimation: Random sampling, estimation of sample size, and calculating standard error.
- Hypothesis Testing: Introduction to parametric tests (t-test, z-test, ANOVA) and confidence intervals

Statistical Techniques for Data Science
Advanced

Duration – 1.5 Days / 12 Hours

Program Description

This course equips learners with cutting-edge statistical tools and techniques for handling more complex data science challenges. The course introduces non-parametric tests such as chi-square, Wilcoxon, and Kruskal-Wallis, which are crucial when data does not meet parametric test assumptions. Students will also explore correlation analysis using Pearson, Kendall, and Spearman methods to measure relationships between variables. An in-depth focus on handling outliers and understanding causation vs. correlation is provided. Finally, learners will delve into advanced sampling techniques like SMOTE to address class imbalance in datasets.

Learning Goals

- Apply non-parametric tests such as chi-square, Wilcoxon, and Kruskal-Wallis tests.
- Understand and apply correlation analysis using Pearson, Spearman, and Kendall methods.
- Regression (Statistical Perspective)
- Interpret and handle correlation cautions, especially in the presence of outliers and causality concerns.
- Use advanced sampling techniques like SMOTE for handling class imbalance in datasets.

Course Topics

- Non-Parametric Testing: Application of chi-square, Wilcoxon, Kolmogorov-Smirnov, and Kruskal-Wallis tests.
- Data Relationship Analysis: Correlation techniques (Pearson, Kendall, Spearman) and understanding causation.
- Advanced Sampling Techniques: SMOTE for addressing class imbalance, over-sampling, and under-sampling.