

Skill Based/Specialized Laboratory (SB) Courses

Skill Based/Specialized Laboratory Course 2a: Data Analysis Using Python

Course Title & Code	Credits	Credit Distribution of the Course			Eligibility Criteria	Prerequisite of the course (if any)
		Lecture (00 Hours)	Tutorial (00 Hours)	Practical (60 Hours)		
SB 2a: Data Analysis Using Python	2	0	0	2	NIL	NIL

Course Objective:

- To introduces Python as a tool for statistical analysis, covering data visualization, probability, statistical inference, regression, sampling, and matrix algebra.
- To develop understanding of fitting probability distributions, conducting hypothesis testing, and applying statistical models to real-world data using key Python libraries.

Course Learning Outcomes:

- Understand Python programming for statistical analysis.
- Develop computational skills for probability, inference, regression, and matrix algebra.
- Perform hypothesis testing, distribution fitting and ANOVA.
- Apply regression techniques and model selection for statistical analysis.

Software & Libraries Required:

1. Python (Jupyter Notebook / Google Colab).
2. Libraries: NumPy, Pandas, SciPy, Statsmodels, Seaborn, Matplotlib, Scikit-learn.

Unit 1 (15 Hours)

Python basics: Variables, loops, functions, list comprehensions, NumPy: One-dimensional & two-dimensional arrays, operations, Pandas: Data Frames, filtering, grouping, handling missing data, Data visualization: Histograms, boxplots, scatterplots, KDE plots, bar charts, pie charts, Probability distributions & simulations: Binomial, Poisson, Normal, Exponential, Monte Carlo methods, Law of Large Numbers & Central Limit Theorem.

Unit 2 (15 Hours)

Estimation: MLE, confidence intervals, bootstrapping, Hypothesis testing: Z-test, t-test, F-test, Wilcoxon signed-rank test, Kruskal-Wallis test, Analysis of variance (ANOVA): One-way, two-way, post-hoc tests (Tukey's HSD), Interpretation & visualization of hypothesis tests and ANOVA results.

Unit 3 (15 Hours)

Simple & multiple linear regression, assumptions, residual analysis and diagnostic plots, Model selection: Adjusted R^2 , AIC, BIC, stepwise selection. Survey sampling: Simple random sampling, stratified sampling, estimation of population parameters. Handling missing data & graphical representation of survey results. Markov Chains: transition probability matrix, steady-state probabilities. Simulation of Markov Chains in Python.

Unit 4 (15 Hours)

Matrix operations: Addition, multiplication, inversion, eigenvalues, eigenvectors Spectral decomposition & solving linear systems. Fitting probability distributions using MLE & SciPy.stats Goodness-of-Fit Tests: Kolmogorov-Smirnov, Anderson-Darling, Chi-Square Visualization of fitted distributions using histograms & QQ-plots.

Essential Readings:

1. Balagurusamy, E. (2016). *Introduction to Computing and Problem-Solving Using Python*, McGraw Hill.
2. Halswanter, T. (2016). *An Introduction to Statistics with Python: With Applications in the Life Sciences*, Springer.

Suggested Readings:

1. McKinney, W. (2012). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc."
2. Perkovic, L. (2015). *Introduction to Computing Using Python: An Application Development Focus*. John Wiley & Sons.
3. Vander Plas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. " O'Reilly Media, Inc."