21AB202 STATISTICAL METHODS WITH DATA VISUALIZATION

Hours Per Week :

L	T	P	C
3	-	2	4

Total Hours :

L	T	P
45	-	30

COURSE DESCRIPTION AND OBJECTIVES:

This course makes a student familiar with the python interactive environment whichincludes the installation of statistical packages, data structures, statistical functions,different data format conversions and various plots. In addition, it teaches to performstatistical manipulations and programming in a more efficient way when compared withtraditional
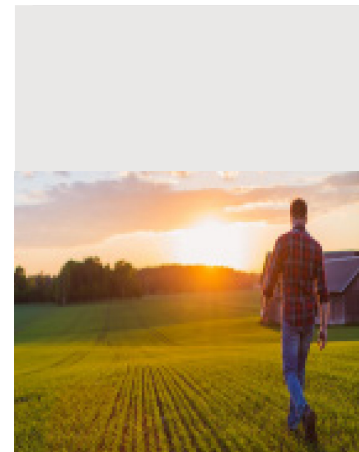
statistical analysis

COURSE OUTCOMES:

Upon completion of the course, the student will be able to achieve the following outcomes:

COs	Course Outcomes
1	Hands on practice in exploratory data analysis
2	Use of several distributions
3	Write functions including generic functions using various methods and loops
4	Select and perform manipulations on values as required for a specific statistical

SKILLS:

✓ *Create a large corpus using lists and data frames*

✓ *Perform Linear algebra operations on data for finding correlation and covariance*

✓ *Develop functions for data manipulation algorithms*

UNIT - 1

Python and Statistics: Why Statistics? Python Packages for Statistics, First Python Programs, Pandas: Data Structures for Statistics, Data Input: Input from Text Files: Visual Inspection, Reading ASCII-Data into Python, Input from MS Excel, Data types: Categorical, Numerical

UNIT - 2

Displaying Statistical Datasets: Univariate Data: Scatter Plots, Histograms, Kernel-Density - Estimation (KDE) Plots, Cumulative Frequencies, Error-Bars, Box Plots, Grouped Bar Charts, PieCharts

Bivariate and Multivariate Plots: Bivariate Scatter Plots, 3D Plots

UNIT - 3

Populations and Samples, Distribution Center: Mean, Median, Mode, Geometric Mean Quantifying Variability: Range, Percentiles, Standard Deviation and Variance. Discrete Distributions - Bernoulli Distribution, Binomial Distribution, Poisson Distribution

UNIT - 4

Normal Distribution - Examples of Normal Distributions, Central Limit Theorem

Continuous Distri butions Derived from the Normal Distribution: t-Distribution, Chi-Square Distribution, F-Distribution

Hypothesis Tests: Typical Analysis Procedure: Data Screening and Outliers, Normality Check, Hypothesis Concept, Errors, p-Value, and Sample Size - Generalization and Applications, The Interpretation of the p-Value, Types of Error, Sensitivity and Specificity

UNIT - 5

Analysis of Variance (ANOVA): One-Way ANOVA, Two-Way ANOVA, One-Way Chi-Square Test, Chi-

Square Contingency Test: Linear Regression Models-Linear Correlation-Correlation Coefficient, Rank Correlation, General Linear Regression Model, Coefficient of Determination, Linear Regression Analysis with Python

LABORATORY EXPERIMENTS

LIST OF EXPERIMENTS

1. Compute statistical measures like Mean, Median, Mode, Standard Deviation, Variance and Co-variance for a given series of values

2. Plot histograms and box plots for a given scenario

3. Determine outliers (Plots, quantiles, box plots, percentiles) froma given sample data

4. Write a python program to compute continuous distributions(exponential distribution, normal distribution)

5. Sample problems on computation of Probability

6. Write a python program to perform Hypothesis testing

7. Write a python program to perform Resampling

8. Write a program to carry out regression analysis to understand the output and diagnostics

9. Evaluate the working process of A/B testing

REFERENCES:

1. An Introduction to Statistics with Python With Applications in the Life Sciences, Thomas Haslwanter - Springer- ISSN 1431-8784 - ISBN 978-3-319-28315-9 Springer International Publishing Switzerland 2016