**STATISTICS USING PYTHON (16CS353)**

**Course Description and Objectives:**

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

The students will be able to:

* Hands on practice in exploratory data analysis.
* Use of several distributions.
* Write functions including generic functions using various methods and loops.
* Select and perform manipulations on values as required for a specific statistical

**SKILLS ACQUIRED:**

* 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.

**ACTIVITIES:**

* Taking the large amounts of data and finding vector length, min, mode, max and average of that data.
* Applying list operations on big data for data interpretation.
* Converting list data into data frames and performing various operations on the data frames.
* Reading the data from various formats and converting them into required formats.

**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, Pie Charts.

***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 Distributions 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.

**Text Book:**

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.