CS2204	Data Analytics	3L: 0T: 0P	3 Credits

Pre-requisites	Probability and Statistics, Python

Course objectives

1. To understand the fundamentals of statistics
2. To calculate the measures of central tendency, asymmetry and variability
3. To distinguish and work with different types of distributions
4. To estimate confidence intervals for population parameters
5. To understand z-test, t-test, chi-square test, ANOVA
6. To understand mechanics of regression analysis
7. To calculate correlation and covariance

Detailed contents

Unit-I

Introduction to Analytics: why analytics, types of analytics, framework for data driven decision making. Descriptive Analytics: types of data measurement scales, population, sample, measures of central tendency, measures of variation, shape, data visualization techniques. Probability distributions: Random variables, Probability Density Function, Cumulative Distribution Function, PMF, Discrete distributions- binomial, Poisson, geometric distributions, Continuous distributions- uniform, exponential, normal distribution and other distributions(chi-square, t, F)

Unit-II

Hypothesis testing-I: sampling & estimation- sampling, types of sampling, sampling distribution, central limit theorem, sample size estimation, estimation of population parameters, confidence intervals- Confidence intervals for population mean, population proportion. Introduction to hypothesis testing- basics, z-test, t-test, one tailed & two-tailed tests, type-I error, type-II error.

Unit-III

Hypothesis Testing-II: Comparing two populations –two sample z-test & t-test, hypothesis test for difference in population proportion, hypothesis test for equality of population variances, chi-square test, goodness of fit test, F-test, analysis of variance(ANOVA) –multiple t-tests for comparing several means, one way ANOVA.

Unit-IV

Correlation & Regression: Introduction to correlation, correlation coefficient, correlation Vs Causation, coefficient of determination. Simple linear regression –model building, estimation of parameters using ordinary least squares, validation of regression model, multiple linear regression- ordinary least square estimation of MLR, validation, logistic regression.

Unit-V

Case studies with python or R: problems related to data visualization, hypothesis testing, ANOVA, classification and regression.

Course outcomes:

Students will be able to

1. Plot different types of data
2. Perform hypothesis testing
3. Carry out regression analysis
4. Make data driven decisions

Suggested textbooks:

5. U Dinesh Kumar, Business Analytics-The science of Data driven decision making, Wiley publications