INTRODUCTION TO STATISTICS

Semester No 2	Code STAT-103	Credit Hours 3-0

COURSE OBJECTIVES:

1. Describe and apply the basic concepts of probability and statistics used for data representation and sampling
2. Use probability theory to analyze data for decision-making and for solving problems.
3. Estimate confidence intervals, hypothesis, quality control and probability to make decisions for industrial/engineering data

COURSE LEARNING OUTCOMES:

After completion of the course students will be able to:

1. Perform graphical profiling of data and analyzing patterns in univariate and multivariate data.
2. Perform quantitative profiling of univariate and multivariate data.
3. Use Induction to generalize from sample to population

PRESCRIBED TEXTS:

1. Advanced Engineering Mathematics, Erwin Kreyszig, 10th edition, Wiley Publishers, 2011

2. Douglas, A, L, Mason, Robert, D. M & Marchal, William, G "Basic Statistics for Business and Economics". McGraw-Hill 6th Edition, 2007 .

REFERENCE MATERIAL:

a. Paul Newbold, William L. Carlson & Betty Thorne. Statistics for Business & Economics, 8th Edition (Pearson-Prentice Hall), 2013.
b. Wonnacott, Thomas H., and Ronald J. Wonnacott, 1990. *Introductory Statistics for Business and Economics*, 4th ed., Wiley.
c. Haider, M. "Getting Started with Data Science: Making Sense of Data", IBM Press, Pearson, 1st Edition, 2016.

Prerequisites:

NIL

Course Description:

A first introduction to probability, statistics and machine learning. This course will provide background to understand and produce rigorous statistical analysis including estimation, confidence intervals, hypothesis testing, regression, logistic regression and a brief introduction to machine learning. Applicability and limitations of these methods will be illustrated using a variety of modern real world data sets and manipulation of the statistical software R. Precepts are based on real data analysis using R.

ASSESSMENT SYSTEM:

Quizzes	10-15%
Assignments	5-10%
Mid Term	20- 25
ESE	40-50%

Weekly breakdown of course contents is as follows:

Week	Topic	Quizzes	Assignments
1	**Descriptive statistics** • Statistics vs. probability, sample vs population; • Summary statistics: Mean, SD, Median, IQR; • Graphical Summary: Pie Charts, Histograms, Box-plots Lecture Notes 1(Link downloads document), Homework 1		01

2	**Probability** • Sample space, event, probability • Conditional Probability, Bayes's Theorem • Independence • Monte Carlo Simulations Lecture Notes 2(Link downloads document), Homework 2		01
3	**Random variables and probability distributions** • Random variables and probability distribution • Expected values and standard deviation • Probability density functions	01	
4-5	**Commonly used distributions** • Binomial distribution • Hypergeometric, negative bionomial • Poisson distributions • Normal distributions • Normal approximations to data histograms • Exponential and Gammas distributions • Quantile-Quantile plot Joint Distributions and Random Samples • Discrete joint distribution • Joint densities • Covariance and correlation • Multivariate random variables • Square root law • Central limit theorem		

6	**Concepts and Methods of Estimation** • Point Estimation • Methods of Estimation • Standard error • Bootstrap	01	
7	**Estimation and Confidence Intervals** • Point Estimates and Confidence Intervals for known standard deviation' or a Large Sample • Unknown Population Standard Deviation and a Small Sample • A Confidence Interval for a Proportion, Finite-Population Correction Factor, Choosing an Appropriate Sample Size,		01
8	**Introduction of Hypothesis, Null and Alternate Hypothesis, Hypothesis Testing** • One-Sample Tests of Hypothesis • Testing a Hypothesis One-Tailed and Two-Tailed Tests of Significance, testing for a Population Mean with a Known Population Standard Deviation • A Two-Tailed Test • One-Tailed Test • p-Value in Hypothesis Testing • Testing for Population Mean: Large Sample, Population Standard Deviation Unknown • Tests Concerning Proportions • Testing for a Population Mean: Small Sample, Population Standard Deviation		

9	**MID-TERM**		

10	**Introduction: Two-Sample Tests of Hypothesis: Independent Samples, Two-Sample Tests** • Two-Sample Test of Hypothesis • Proportions, Comparing Population Means with Small Samples		
11	**Simple linear regression** • Models and summary statistics • Estimation of model parameters • Regression effect and goodness of fit • Inference of model parameters • Prediction • Inference of Correlation		
12	**Multiple and Nonlinear Regression** • Parameter estimation • Variable Selection • Statistical inference and ANOVA • Model diagnostics • Training and Testing • Cross-validation and Prediction errors • Polynomial and nonlinear regression • Model building using dummies	01	
14	**Expected Frequencies** • Goodness-of-Fit Test: Equal Expected Frequencies • Goodness-of Fit Test: Unequal	01	

15	**Introduction to Machine Learning** • Logistic Regression • Supervised learning and Bayesian classifiers • Fisher and nearest neighborhood classification • Support vector machine		01
	• Unsupervised learning		
16	Presentations		
17	Revision		
18	**END SEMESTER EXAMINATION**		