Approval: 10th Senate Meeting

Course Name	: Statistical Methods
Course Number	: HS-650
Credits	: 3-0-0-3
Prerequisites	: A prior course in probability, statistics and random processes; or, discretion of the instructors
Intended for	:Postgraduate and B.Tech.
Distribution	: HSS Course – Elective
Semester	: Even/Odd

Course preamble:

Statistical methods form the basic ingredients for research in engineering, basic sciences, and humanities and social sciences. Currently, there exist Institute courses that cover basics of technical writing, research methodology, and certain advanced statistical methods (e.g., time-series analyses). These courses are important as they expose students to relevant research topics. However, there exists no single course that holistically covers topics in basic data analyses(descriptive statistics), probability and random processes, parametric and non-parametric inferential statistics, experimental design, and advanced statistical methods (e.g., analysis of variance and regression). This course fills this gap by providing a holistic coverage of these statistical topics over one-semester of study. This course is an elective course primarily designed for graduate students to help them in their research. However, undergraduate students are may also take this course as it is likely to help them in their project courses like Design Practicum, Interactive Socio-Technical Practicum, and Major Technical Project.

Course outline:

The course is distributed in six modules. It starts with the fundamental methods of *descriptive statistics* (Module 1) that includes the representation of data and calculation of various measures such as averages, variation, and correlation etc. Then, following an introduction to the *theory of probability* (Module 2), it enters into the area of *inferential statistics*– both *parametric* (Module 3) and *non-parametric* (Module 4). These modules will discuss the methods of drawing inference regarding population parameters based on sample statistics. The next part deals with the *designs of experiments* and discusses two frequently used processes – randomization and factorial design (Module 5). Finally, the course provides an exposure to certain frequently-used advanced statistical techniques – *analysis of variance* and *regression* (Module 6). The course will use statistical packages like R, Matlab, or SPSS, which are available for teaching on campus.

Modules:

Module 1: Representation of Data and Descriptive Statistics (6 Hours)

Raw data and frequency data- tabular and diagrammatic representation; concept of moments; measures of central tendency, dispersion, skewness, and Kurtosis; quartile and percentile – their use in the measurement of inequality, Gini Coefficient and Lorenz curve; Bivariate frequency distribution, correlation coefficients- Pearson and Spearman coefficients.

Module 2: Probability and Random Variables (11 Hours)

Basic concepts in set theory as applied in probability; concept of probability- classical, frequency based, axiomatic approach, Bayesian probability; conditional probability, Bayes theorem, statistical independence of events; random variables – discrete and continuous, probability distribution functions, cumulative distribution functions, Expectation and Variance of a random variable, joint distribution of two random variables and their correlation, law of large number.

Module 3: Random Sampling and Parametric Statistical Inference (9 Hours)

Concepts of population and sample, parameter and statistic, random sampling and sampling distribution, Central Limit Theorem; Expectation and Standard Error of sample mean and sample proportion; concepts of theoretical distribution: Normal distributions and four fundamental distributions derived from Normal distribution – Standard Normal, Chi-square, t and F distribution; estimation and testing of hypothesis – point estimation and interval estimation of parameters, Maximum Likelihood Estimator, hypothesis testing, and calculation of effect size.

Module 4: Non-parametric Statistical Inference (6 Hours)

Need for non-parametric tests, estimation of location and dispersion, tolerance interval; one sample and two sample non-parametric tests for location and dispersion (involving independent and related samples); non-parametric measures and tests of association.

Module 5: Designs of Experiment (6 Hours)

Experimental design strategies; Blocking and Randomization; Factorial design of experiments.

Module 6: Regression Analysis and Analysis of Variance (8 Hours)

Gauss Markov theorem and Ordinary Linear Least Square regression; interpreting regression coefficients, concepts of residual, fitted value and goodness of fit, test of significance; multiple regression analysis; two-way independent ANOVA and two-way Mixed ANOVA

Textbooks:

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage.
Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics*. London: Sage.

References:

Agresti, A., & Finlay, B. Statistical Methods for the Social Sciences, 1997. Dellen, San Francisco.
Arnold, J. C., & Milton, J. S. (2003). Introduction to probability and statistics.
Gibbons, J. D., Chakraborti, S., Nonparametric Statistical Inference, Fifth Edition. Marcel Dekker, Inc., 2003.
Johnston, J., DiNardo, J., Econometric Methods, 4th Edition. McGraw-Hill, 1996
Montgomery, D. C., G.C. Runger, Applied Statistics and Probability for Engineers. 5th ed.New Delhi: Wiley-India, 2011.
Montgomery, D. C., Design and Analysis of Experiments, 8th Edition. John Wiley & Sons, Inc., 2012.
Ross, S. M. (2014). Introduction to probability and statistics for engineers and scientists. Academic Press.
Rohatgi, V. K. &Saleh, A. K. E. (2015). An Introduction to Probability and Statistics.