Approval : 19th Senate Meeting

IIT Mandi
Proposal for a New Course

Course Name	: Statistical Methods
Course Number	: HS550
Credits	: 3-0-2-4
Prerequisites	: A prior course in probability, statistics and random processes; or, consent of the instructor
Intended for	: Ph.D. and Masters
Distribution	: Discipline Elective for M.A. Development Studies, Free elective for others
Semester	: Even/Odd

1. 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, probability and statistics) that are mainly targeted to varied applications in engineering and sciences (such as MA524). 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) with significant practical applications that are targeted for understanding the research problems in social sciences. 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. In the lab sessions, the course intends to provide a hands-on training to the students on practical applications of several statistical methods that are covered. The course will use statistical packages, which are available for teaching on campus.

2. Modules with lecture hours:

Module 1: Representation of Data and Descriptive Statistics (4 hours+2 lab 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.

Lab: Exercise on various ways of representing quantitative data; Measuring Central tendency, dispersion, skewness, and kurtosis of a given dataset; deriving quartile and percentile; Deriving Gini coefficient and Lorenz curve to understand inequality present in a dataset; calculation of correlation coefficients of various forms.

Module 2: Probability and Random Variables (4 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 (8 hours+6 lab 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.

Lab: Drawing random samples from the population –simple random sampling (with and without replacement); Fitting distribution curves to a given dataset; Statistical estimation – parametric point estimation and interval estimation; Maximum Likelihood Estimator, Hypothesis testing and calculation of effect size.

Module 4: Non-parametric Statistical Inference (8 hours + 6 lab 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.

Lab: Carrying out non-parametric tests, estimation of location and dispersion, tolerance interval and tests of association.

Module 5: Designs of Experiment (8 hours + 6 lab hours)
Experimental design strategies; Blocking and Randomization; Factorial design of experiments.

Lab: Designing a suitable experiment to test a given hypothesis, testing the hypotheses by changing the variables within the experiment. Carrying out experiments based on Blocking and Randomization; Factorial design of experiments.

Module 6: Regression Analysis and Analysis of Variance (10 hours + 8 lab 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; diagnostic tests; binary

explanatory variables; multiple regression analysis; two-way independent ANOVA and two-way Mixed ANOVA.

Lab: Estimate multiple linear regressions to carry out the diagnostic tests; finding out the key determinants; interpretation of estimates, testing the significance and carrying out an Analysis of Variance (ANOVA).

3. **Textbooks:**

- Field, A. P., Miles, J., and Field, Z. (2012). *Discovering statistics using R*. London: Sage.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South-Western, Cengage Learning. Ohio, USA. ISBN-13: 978-1-111-53104-1.

4. **References:**

- STATA Manual: https://www.stata.com/manuals13/u.pdf
- Heiman, G. W. (2011). Basic Statistics for the Behavioral Sciences. Sixth Edition. Wadsworth.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics*. London: Sage.
- Agresti, A., and Finlay, B. Statistical Methods for the Social Sciences. (1997). Dellen, San Francisco.
- Arnold, J. C., and Milton, J. S. (2003). Introduction to Probability and Statistics. McGraw-Hill.
- Gibbons, J. D. and Chakraborti, S. (2003). Nonparametric Statistical Inference, Fifth Edition. Marcel Dekker, Inc.
- Johnston, J. and DiNardo, J. (2006). Econometric Methods, 4th Edition. McGraw-Hill
- Montgomery, D. C. and Runger, G.C. (2011) Applied Statistics and Probability for Engineers. 5th ed. New Delhi: Wiley-India.
- Montgomery, D. C. (2012). Design and Analysis of Experiments, 8th Edition. John Wiley & Sons, Inc.
- Ross, S. M. (2014). Introduction to probability and statistics for engineers and scientists. Academic Press.
- Rohatgi, V. K. and Saleh, A. K. E. (2015). An Introduction to Probability and Statistics.

4. **Similarity Content Declaration with Existing Courses:** Note that existing HS550 (Statistical Methods- 3-0-0-3) and HS550P (Statistical Methods: Practical 0-0-2-1) have been merged and proposed as HS550: Statistical Methods (3-0-2-4). Existing HS550 (Statistical Methods- 3-0-0-3) and HS550P (Statistical Methods: Practical 0-0-2-1) will not be offered separately once HS550: Statistical Methods (3-0-2-4) is approved.

Course	Content	Overlap
MA 524	Probability and random variable; Sampling distributions; point and interval estimation, Testing of	~25%

	hypothesis, Goodness of fit, linear regression, ANOVA.	

5. **Justification for new course proposal if cumulative similarity content is > 30%:N/A**