Outline of Master Degree Course Curricula
(Following BSMA Committee's Recommendations)
Department of Agricultural Statistics
Faculty of Agriculture
Uttar Banga Krishi Viswavidyalaya

Course Code	Course Title	Credit Hours	Semester
*AST 552	Probability Theory	2+0	I
*AST 553	Statistical Methods	2+1	I
*AST 562	Statistical Inference	2+1	II
*AST 563	Design of Experiments	2+1	II
*AST 564	Sampling Techniques	2+1	II
*AST 565	Statistical Genetics	2+1	II
*AST 571	Multivariate Analysis	2+1	III
*AST 572	Regression Analysis	1+1	III
*AST 573	Statistical Computing	1+1	III

*: Major/Core Courses

Course Contents

I. Course Title : Probability Theory
II. Course Code : AST 552
III. Credit Hours : 2+0

IV. Aim of the course

This is a fundamental course in Statistics. This course lays the foundation of probability theory, random variable, probability distribution, mathematical expectation, etc. which forms the basis of basic statistics. The students are also exposed to law of large numbers and central limit theorem. The students also get introduced to stochastic processes.

V. Theory

Unit I

Basic concepts of probability. Elements of measure theory: class of sets, field, sigma field, minimal sigma field, Borel sigma field in R, measure- probability measure. Axiomatic approach to probability. Properties of probability based on axiomatic definition. Addition and multiplication theorems. Conditional probability and independence of events. Bayes theorem.

Unit II

Random variables: definition of random variable, discrete and continuous, functionsof random variables. Probability mass function and Probability density function, Distribution function and its properties. Notion of bivariate random variables, bivariate distribution function and its properties. Joint, marginal and conditional distributions. Independence of random variables. Transformation of random variables (two dimensional case only). Mathematical expectation: Mathematical expectation of functions of a random variable. Raw and central moments and their relation, covariance, skewness and kurtosis. Addition and multiplication theorems of expectation. Definition of moment generating function, cumulating generating function, probability generating function and statements of their properties.

Unit III

Conditional expectation and conditional variance. Characteristic function and its properties. Inversion and uniqueness theorems. Chebyshev, Markov, Cauchy- Schwartz, Sequence of random variables and modes of convergence (convergence in distribution in probability, almost surely, and quadratic mean) and their interrelations.

Unit IV

Laws of large numbers: WLLN, Bernoulli and Kintchin's WLLN. Kolmogorov inequality, Kolmogorov's SLLNs. Central Limit theorems: Demoviere-Laplace CLT, Lindberg- Levy CLT and simple applications.

VI. Suggested Reading

- Ash RB. 2000. *Probability and Measure Theory*. 2nd Ed. Academic Press. Billingsley P. 1986. *Probability and Measure*. 2nd Ed. John Wiley.
- Capinski M and Zastawniah. 2001. *Probability Through Problems*. Springer. Dudewicz EJ & Mishra SN. 1988. *Modern Mathematical Statistics*. John Wiley.
- Feller W. 1972. *An Introduction to Probability Theory and its Applications*. Vols. I., II. John Wiley.
- Loeve M. 1978. *Probability Theory*. 4th Ed. Springer.
- Marek C, Tomasz JZ. 2003. *Probability* Through Problems (Problem Books in Mathematics) Corrected Ed.
- Marek F. 1963. *Probability Theory and Mathematical Statistics*. John Wiley.
- Rohatgi VK & Saleh AK Md. E. 2005. *An Introduction to Probability and* Statistics. 2nd Ed. John Wiley.

I. Course Title : Statistical Methods
II. Course Code : AST 553
III. Credit Hours : 2+1

IV. Aim of the course

This course lays the foundation of probability distributions and sampling distributions

and their application which forms the basis of Statistical Inference. Together with probability theory, this course is fundamental to the discipline of Statistics. The students are also exposed to correlation and regression, and order statistics and their distributions. Categorical data analysis is also covered in this course.

V. Theory

Unit I

Descriptive statistics: probability distributions: Discrete probability distributions ~ Bernoulli, Binomial, Poisson, Negative-binomial, Geometric and Hyper Geometric, uniform, multinomial ~ Properties of these distributions and real life examples. Continuous probability distributions ~ rectangular, exponential, Cauchy, normal, gamma, beta of two kinds, Weibull, lognormal, logistic, Pareto. Properties of these distributions. Probability distributions of functions of random variables.

Unit II

Concepts of compound, truncated and mixture distributions (definitions and examples). Sampling distributions of sample mean and sample variance from Normal population, central and non–central chi-Square, t and F distributions, their properties and inter relationships.

Unit III

Concepts of random vectors, moments and their distributions. Bivariate Normal distribution - marginal and conditional distributions. Distribution of quadratic forms. Cochran theorem. Correlation, rank correlation, correlation ratio and intra-class correlation. Regression analysis, partial and multiple correlation and regression.

Unit IV

Sampling distribution of correlation coefficient, regression coefficient. Categorical data analysis, Association between attributes. Variance Stabilizing Transformations.

Unit V

Order statistics, distribution of r-th order statistics, joint distribution of several order statistics and their functions, marginal distributions of order statistics.

VI. Practical

- Fitting of discrete distributions and test for goodness of fit;
- Fitting of continuous distributions and test for goodness of fit; Fitting of truncated distribution;
- Computation of simple, multiple and partial correlation coefficient, correlationratio and intra-class correlation;
- Regression coefficients and regression equations;
- Fitting of Pearsonian curves;
- Computation of simple, multiple and partial correlation coefficient, correlationratio and intra-class correlation;
- Regression coefficients and regression equations;
- Fitting of Pearsonian curves;
- Analysis of association between attributes, categorical data and log-linear models.

VII. Suggested Reading

- Agresti, A. 2012. *Categorical Data Analysis* 3rd Ed. John Wiley.
- Arnold BC, Balakrishnan N and Nagaraja HN. 1992. *A First Course in Order Statistics*. JohnWiley.
- David HA and Nagaraja HN. 2003. *Order Statistics*. 3rd Ed. John Wiley.
- Dudewicz EJ and Mishra SN. 1988. *Modern Mathematical Statistics*. John Wiley.
- Huber PJ. 1981. *Robust Statistics*. John Wiley.
- Johnson NL, Kotz S and Balakrishnan N. 2000. *Continuous Univariate Distributions*. JohnWiley.
- Johnson NL, Kotz S and Balakrishnan N. 2000. *Discrete Univariate Distributions*. JohnWiley.
- Marek F.1963. *Probability Theory and Mathematical Statistics*. John Wiley.
- Rao CR. 1965. *Linear Statistical Inference and its Applications*. John Wiley.
- Rohatgi VK and Saleh AK Md. E. 2005. *An Introduction to Probability and Statistics*. 2nd Ed. John Wiley.
- Gupta. S.P 2008. *Statistical Methods*. Sultan Chand & sons Educational Publisher

I. Course Title : Statistical Inference

II. Course Code : AST 562

III. Credit Hours : 2+1

IV. Aim of the course

This course lays the foundation of Statistical Inference. The students would betaught the problems related to point and confidence interval estimation and testing of hypothesis. They would also be given the concepts of nonparametric and sequentialtest procedures and elements of decision theory.

V. Theory

Unit I

Concepts of point estimation: unbiasedness, consistency, efficiency and sufficiency. Statement of Neyman's Factorization theorem with applications. MVUE, Rao- Blackwell theorem, completeness, Lehmann- Scheffe theorem. Fisher information, Cramer-Rao lower bound and its applications.

Unit II

Moments, minimum chi-square, least square and maximum likelihood methods of estimation and their properties. Interval estimation-Confidence level, shortest length CI. CI for the parameters of Normal, Exponential, Binomial and Poisson distributions.

Unit III

Fundamentals of hypothesis testing-statistical hypothesis, statistical test, critical region, types of errors, test function, randomized and non- randomized tests, level of significance, power function, most powerful tests: Neyman-Pearson fundamental lemma, MLR families and UMP tests for one parameter exponential families. Concepts of consistency, unbiasedness and invariance of tests. Likelihood Ratio tests, asymptotic properties of LR tests with applications (including homogeneity of means and variances). Relation between confidence interval estimation and testing of hypothesis.

Unit IV

Sequential Probability ratio test, Properties of SPRT. Termination property of SPRT, SPRT for Binomial, Poisson, Normal and Exponential distributions. Concepts of loss, risk and decision functions, admissible and optimal decision functions, estimation and testing viewed as decision problems, conjugate families, Bayes and Minimax decision functions with applications to estimation with quadratic loss.

Unit V

Non-parametric tests: Sign test, Wilcoxon signed rank test, Runs test for randomness, Kolmogorov – Smirnov test for goodness of fit, Median test and Wilcoxon-Mann-Whitney U-test. Chi-square test for goodness of fit and test for independence of attributes. Spearman's rank correlation and Kendall's Tau tests for independence.

VI. Practical

- Methods of estimation - Maximum Likelihood, Minimum chi-square and Moments;
- Confidence Interval Estimation;
- MP and UMP tests;
- Large Sample tests;
- Non-parametric tests, Sequential Probability Ratio Test;
- Decision functions.

VII. Suggested Reading

- Box G.E.P. and Tiao G.C. 1992. *Bayesian Inference in Statistical Analysis*. John Wiley.
- Casela G and Berger R.L. 2001. *Statistical Inference*. Duxbury Thompson Learning.
- Christensen R. 1990. *Log Linear Models*. Springer.
- Conover W.J. 1980. *Practical Nonparametric Statistics*. John Wiley.
- Dudewicz EJ and Mishra SN. 1988. *Modern Mathematical Statistics*. John Wiley.
- Gibbons J.D. 1985. *Non Parametric Statistical Inference*. 2nd Ed. Marcel Dekker.
- Kiefer J.C. 1987. *Introduction to Statistical Inference*. Springer.
- Lehmann EL. 1986. *Testing Statistical Hypotheses*. John Wiley.
- Lehmann EL. 1986. *Theory of Point Estimation*. John Wiley.
- Randles R.H and Wolfe D.S. 1979. *Introduction to the Theory of Nonparametric Statistics*. John Wiley.
- Rao C.R. 2009. *Linear Statistical Inference and Its Applications*, 3rdEd. John Wiley.
- Rohatgi V.K. and Saleh A.K. Md. E. 2005. *An Introduction to Probability and Statistics*. 2nd Ed.

John Wiley.
- Rohtagi V.K. 1984. *Statistical Inference*. John Wiley
- Sidney S and Castellan N.J. Jr. 1988. *Non Parametric Statistical Methods for Behavioral Sciences*. McGraw Hill.
- Wald A. 2004. *Sequential Analysis*. Dover Publ.
- Michael J. Panik. 2012. *Statistical Inference*. A John Wiley & Sons, INC, publication

I. Course Title : Design of Experiments

II. Course Code : AST 563

III. Credit Hours : 2+1

IV. Aim of the course

Design of Experiments provides the statistical tools to get maximum information from least amount of resources. This course is meant to expose the students to the basic principles of design of experiments. The students would also be provided with mathematical background of various basic designs involving one-way and two-way elimination of heterogeneity and their characterization properties. This course would also prepare the students in deriving the experimental data.

V. Theory

Unit I

Elements of linear estimation, Gauss Markoff Theorem, relationship between BLUEs and linear zero-functions. Aitken's transformation, test of hypothesis, Analysis of Variance, Partitioning of degrees of freedom.

Unit II

Orthogonality, contrasts, mutually orthogonal contrasts, analysis of covariance; Basic principles of design of experiments, uniformity trials, size and shape of plots and blocks, Randomization procedure.

Unit III

Basic designs - completely randomized design, randomized complete block design and Latin square design; Construction of orthogonal Latin squares, mutually orthogonal Latin squares (MOLS), Youden square designs, Graeco Latin squares.

Unit IV

Balanced Incomplete Block (BIB) designs – general properties and analysis withoutand with recovery of intra block information, construction of BIB designs. Partially balanced incomplete block designs with two associate classes - properties, analysis and construction, Lattice designs, alpha designs, cyclic designs, augmented designs.

Unit V

Factorial experiments, confounding in symmetrical factorial experiments ($2^n$and 3^n series), partial and total confounding, asymmetrical factorials.

Unit VI

Cross-over designs. Missing plot technique; Split plot and Strip plot design; Groupsof experiments. Sampling in field experiments.

V. Practical

- Determination of size and shape of plots and blocks from uniformity trials data;
- Analysis of data generated from completely randomized design, randomizedcomplete block design;
- Latin square design, Youden square design; Analysis of data generated from aBIB design, lattice design, PBIB designs;
- 2^n, 3^n factorial experiments without and with confounding;
- Split and strip plot designs, repeated measurement design;
- Missing plot techniques,
- Analysis of covariance;
- Analysis of Groups of experiments,
- Analysis of clinical trial experiments.

VI. Suggested Reading

- Chakrabarti M.C. 1962. *Mathematics of Design and Analysis of Experiments*. Asia Publ. House.
- Cochran W.G. and Cox D.R. 1957. *Experimental Designs*. 2nd Ed. John Wiley.

- Dean A.M. and Voss D. 1999. *Design and Analysis of Experiments*. Springer.
- Dey A and Mukerjee R. 1999. *Fractional Factorial Plans*. John Wiley.
- Dey A 1986. *Theory of Block Designs*. Wiley Eastern. Hall M Jr. 1986. *Combinatorial Theory*.John Wiley.
- John J.A. and Quenouille M.H. 1977. *Experiments: Design and Analysis*. Charles & Griffin.
- Kempthorne, O. 1976. Design and Analysis of EXperiments. John Wiley. Khuri AI & Cornell JA. 1996. *Response Surface Designs and Analysis*. 2nd Ed. Marcel Dekker.
- Kshirsagar A.M. 1983. *A Course in Linear Models*. Marcel Dekker.
- Montgomery D.C. 2013. *Design and Analysis of Experiments*. John Wiley & Sons
- Raghavarao D. 1971. *Construction and Combinatorial Problems in Design of Experiments*. John Wiley.
- Searle S.R. 2006. *Linear Models*. John Wiley.
- Street A.P. and Street D.J. 1987. *Combinatorics of Experimental Designs*. OXford Science Publ.
- Design Resources Server. *Indian Agricultural Statistics Research Institute (ICAR), New Delhi-110 012, India*. Hyperlink "http: //www.iasri.res.in/design" www.drs.icar.gov.in.

I. Course Title : Sampling Techniques
II. Course Code : AST 564
III. Credit Hours : 2+1
IV. Aim of the course

This course is meant to expose the students to the techniques of drawing representative samples from various populations and then preparing them on the mathematical formulations of estimating the population parameters based on the sample data. The students would also be exposed to the real life applications of sampling techniques and estimation of parameters.

V. Theory

Unit I

Sample survey vs complete enumeration, probability sampling, sample space, sampling design, sampling strategy; Determination of sample size; Confidence-interval; Simple random sampling, Estimation of population proportion, Stratified random sampling, Proportional allocation and optimal allocation, Inverse sampling.

Unit II

Ratio, Product and regression methods of estimation, Cluster sampling, Systematic sampling, Multistage sampling with equal probability, Separate and combined ratio estimator, Double sampling, Successive sampling - two occasions. Unbiased ratio type estimators

Unit III

Non-sampling errors – sources and classification, Non-response in surveys, Randomized response techniques, Response errors/Measurement error – interpenetrating sub-sampling.

Unit IV

PPS Sampling with and without replacement, Cumulative method and Lahiri's method of selection, Horvitz-Thompson estimator, Ordered and unordered estimators, Sampling strategies due to Midzuno-Sen and Rao-Hartley-Cochran. Inclusion probability proportional to size sampling.

VI. Practical

- Determination of sample size and selection of sample;
- Simple random sampling, Inverse sampling, Stratified random sampling, Cluster sampling, systematic sampling;
- Ratio and regression methods of estimation;
- Double sampling, multi-stage sampling, Imputation methods;
- Randomized response techniques;
- Sampling with varying probabilities.

VII. Suggested Reading

- Cassel C.M., Sarndal C.E. and Wretman J.H. 1977. *Foundations of Inference in Survey Sampling*. John Wiley.
- Chaudhari A and Stenger H. 2005. *Survey Sampling Theory and Methods*. 2nd Ed. Chapman & Hall.

- Chaudhari A and Voss J.W.E. 1988. *Unified Theory and Strategies of Survey Sampling*. North Holland.
- Cochran W.G. 1977. *Sampling Techniques*. John Wiley.
- Hedayat A.S. and Sinha B.K. 1991. *Design and Inference in Finite Population Sampling*.John Wiley.
- Kish L. 1965. *Survey Sampling*. John Wiley.
- Mukhopadhyay, P. 2008.
- *Theory and Methods of Survey Sampling*, John Wiley & Sons
- Murthy M.N. 1977. *Sampling Theory and Methods*. 2nd Ed. Statistical Publ. Society, Calcutta.
- Sukhatme P.V., Sukhatme B.V., Sukhatme S and Asok C. 1984. *Sampling Theory of Surveys with Applications*. Iowa State University Press and Indian Society of Agricultural Statistics,New Delhi.
- Thompson SK. 2000. *Sampling*. John Wiley.
- Cochran WG. 2007. *Sampling Techniques*. A John Wiley & Sons *Publication*

I. Course Title : Statistical Genetics

II. Course Code : AST 565

III. Credit Hours : 2+1

IV. Aim of the course

This course is meant to prepare the students in applications of statistics in quantitative genetics and breeding. The students would be exposed to the physical basis of inheritance, detection and estimation of linkage, estimation of genetic parameters and development of selection indices.

V. Theory

Unit I

Physical basis of inheritance. Analysis of segregation, detection and estimation of linkage for qualitative characters. Amount of information about linkage, combined estimation, disturbed segregation.

Unit II

Gene and genotypic frequencies, Random mating and Hardy -Weinberg law, Application and extension of the equilibrium law, Fisher's fundamental theorem of natural selection. Disequilibrium due to linkage for two pairs of genes, sex-linked genes, Theory of path coefficients.

Unit III

Concepts of inbreeding, Regular system of inbreeding. Forces affecting gene frequency - selection, mutation and migration, equilibrium between forces in large populations, Random genetic drift, Effect of finite population size.

Unit IV

Polygenic system for quantitative characters, concepts of breeding value and dominance deviation. Genetic variance and its partitioning, Effect of inbreeding on quantitative characters, Multiple allelism in continuous variation, Sex-linked genes, Maternal effects - estimation of their contribution.

Unit V

Correlations between relatives, Heritability, Repeatability and Genetic correlation. Response due to selection, Selection index and its applications in plants and animals' improvement programmes, Correlated response to selection.

Unit VI

Restricted selection index. Variance component approach and linear regression approach for the analysis of GE interactions. Measurement of stability and adaptability for genotypes. Concepts of general and specific combining ability. Diallel and partial diallel crosses - construction and analysis.

VI. Practical

- Test for the single factor segregation ratios, homogeneity of the families with regard to single factor segregation;
- Detection and estimation of linkage parameter by different procedures;
- Estimation of genotypic and gene frequency from a given data.
- Hardy-Weinberg law;

- Estimation of changes in gene frequency due to systematic forces, inbreeding coefficient, genetic components of variation, heritability and repeatability coefficient, genetic correlation coefficient;
- Examination of effect of linkage, epistasis and inbreeding on mean and varianceof metric traits;
- Mating designs;
- Construction of selection index including phenotypic index, restricted selectionindex. Correlated response to selection.

VII. Suggested Reading

- Agarwal BL and Agarwal SP. 2007. *Statistical Analysis of Quantitative Genetics*. New Age International Publisher.
- Bailey NTJ. 1961. *The Mathematical Theory of Genetic Linkage*. Clarendon Press.
- Balding DJ, Bishop M and Cannings C. 2001. *Hand Book of Statistical Genetics*. John Wiley.
- Crow JF and Kimura M. 1970. *An Introduction of Population Genetics Theory*. Harper and Row.
- Dahlberg G. 1948. *Mathematical Methods for Population Genetics*. Inter Science Publ.
- East EM and Jones DF. 1919. *Inbreeding and Outbreeding*.
- Lippincott JB & Co. Ewens WJ. 1979. *Mathematics of Population Genetics*. Springer.
- Falconer DS. 1985. *Introduction to Quantitative Genetics*. ELBL.
- Fisher RA. 1949. *The Theory of Inbreeding*. Oliver & Boyd.
- Fisher RA. 1950. *Statistical Methods for Research Workers*. *Oliver*& Boyd.
- Fisher RA. 1958. *The Genetical Theory of Natural Selection*. *Dover* Publ.
- Kempthorne O. 1957. *An Introduction to Genetic Statistics*. The Iowa State Univ. Press.
- Lerner IM. 1950. *Population Genetics and Animal Improvement*. Cambridge Univ. Press.
- Lerner IM. 1954. *Genetic Homeostasis*. Oliver & Boyd.
- Lerner IM. 1958. *The Genetic Theory of Selection*. John Wiley.
- Li CC. 1982. *Population Genetics*. The University of Chicago Press.
- K & Jinks JL. 1977. *Introduction to Biometrical Genetics*. Chapman & Hall.
- Mather K and Jinks JL. 1982. *Biometrical Genetics*. Chapman & Hall.
- Mather K. 1949. *Biometrical Genetics*. Methuen.
- Mather K. 1951. *The Measurement of Linkage in Heredity*.
- Methuen. N. P. 1990. *Statistical Genetics*. Wiley Eastern.

I. Course Title : Multivariate Analysis

II. Course Code : AST 571

III. Credit Hours : 2+1

IV. Aim of the course

This course lays the foundation of Multivariate data analysis. Most of the data sets in agricultural sciences are multivariate in nature. The exposure provided to multivariate data structure, multinomial and multivariate normal distribution, estimation and testing of parameters, various data reduction methods would help the students in having a better understanding of agricultural research data, its presentation and analysis.

V. Theory

Unit I

Concept of random vector, its expectation and Variance-Covariance matrix. Marginal and joint distributions. Conditional distributions and Independence of random vectors. Multinomial distribution. Multivariate Normal distribution, marginal and conditional distributions. Sample mean vector and its distribution. Maximum likelihood estimates of mean vector and dispersion about mean vector.

Unit II

Wishart distribution and its simple properties. Hotelling's T^2 and Mahalanobis D^2 statistics. Null distribution of Hotelling's T^2. Rao's U statistics and its distribution. Wilks' λ criterion and its properties. Concepts of discriminant analysis, computationof linear discriminant function, classification between k (≥ 2) multivariate normal populations based on LDF and Mahalanobis D^2.

Unit III

Principal Component Analysis, factor analysis. Canonical variables and canonical correlations. Cluster analysis: similarities and dissimilarities of qualitative and

quantitative characteristics, Hierarchical clustering. Single, Complete and Average linkage methods. K-means cluster analysis.

Unit IV

Path analysis and computation of path coefficients, introduction to multidimensional scaling, some theoretical results, similarities, metric and non-metric scaling methods.

VI. Practical

- Maximum likelihood estimates of mean-vector and dispersion matrix;
- Testing of hypothesis on mean vectors of multivariate normal populations;
- Cluster analysis, Discriminant function, Canonical correlation, Principal component analysis, Factor analysis;
- Multivariate analysis of variance and covariance, multidimensional scaling.

VII. Suggested Reading

- Abdelmonem A, Virginia AC and Susanne M. 2004. Computer Aided Multivariate Analysis. Chapman & Hall/CRC.
- Anderson TW. 1984. *An Introduction to Multivariate Statistical Analysis*. 2nd Ed. John Wiley.
- Arnold SF. 1981. *The Theory of Linear Models and Multivariate Analysis*. John Wiley.
- Giri NC. 1977. *Multivariate Statistical Inference*. Academic Press.
- Johnson RA and Wichern DW. 1988. *Applied Multivariate Statistical Analysis*. Prentice Hall.
- Kshirsagar AM. 1972. *Multivariate Analysis*. Marcel Dekker.
- Muirhead RJ. 1982. *Aspects of Multivariate Statistical Theory*. John Wiley. Muirhead, RJ. (2005) *Aspects of Multivariate Statistical Theory*. 2nd Ed. John Wiley.
- Rao CR. 1973. *Linear Statistical Inference and its Applications*. 2nd Ed. John Wiley.
- Rencher AC. 2012. *Methods of Multivariate Analysis*. 3rd Ed. John Wiley.
- Srivastava MS and Khatri CG. 1979. *An Introduction to Multivariate Statistics*. North Holland.

I. Course Title : Regression Analysis
II. Course Code : AST 572
III. Credit Hours : 1+1

IV. Aim of the course

This course is meant to prepare the students in linear and non-linear regression methods useful for statistical data analysis. They would also be provided a mathematical foundation behind these techniques and their applications in agricultural data.

V. Theory

Unit I

Simple and Multiple linear regressions: Least squares fit, Properties and examples. Polynomial regression: Use of orthogonal polynomials.

Unit II

Assumptions of regression; diagnostics and transformations; residual analysis ~ Studentized residuals, applications of residuals in detecting outliers, identification of influential observations. Lack of fit, Pure error. Test of normality, test of linearity, Testing homoscedasticity and normality of errors, Durbin-Watson test. Testof goodness of fit for the model evaluation and validation. Concept of multi-collinearity.

Unit III

Weighted least squares method: Properties, and examples. Box-Cox family of transformations. Use of dummy variables, Over fitting and under fitting of model, Selection of variables: Forward selection, Backward elimination. Stepwise and Stagewise regressions.

Unit IV

Introduction to non-linear models, nonlinear estimation: Least squares for nonlinear models.

VI. Practical

- Multiple regression fitting with three and four independent variables;
- Estimation of residuals, their applications in outlier detection, distribution ofresiduals;
- Test of homoscedasticity, and normality, Box-Cox transformation;
- Restricted estimation of parameters in the model, hypothesis testing, Step wise regression analysis;
- Least median of squares norm, Orthogonal polynomial fitting.

VII. Suggested Reading

- Barnett V and Lewis T. 1984. *Outliers in Statistical Data*. John Wiley.
- Belsley DA, Kuh E and Welsch RE. 2004. *Regression Diagnostics-Identifying Influential Data and Sources of Collinearity*. John Wiley.
- Chatterjee S and Hadi AS. 2013. *Regression Analysis* by *Example*. A John Wiley & sons Publication.
- Draper NR and Smith H. 1998. *Applied Regression Analysis*. 3rd Ed. John Wiley.
- McCullagh P and Nelder JA. 1999. *Generalized Linear Models*. 2nd Ed. *Chapman* & Hall.
- Montgomery DC, Peck EA and Vining GG. 2003. *Introduction to Linear Regression Analysis*.3rd Ed. John Wiley.
- Rao CR. 1973. *Linear Statistical Inference and its Applications*. 2nd*Ed*. John Wiley.

I. Course Title : Statistical Computing

II. Course Code : AST 573

III. Credit Hours : 1+1

IV. Aim of the course

This course is meant for exposing the students in the concepts of computational techniques. Various statistical packages would be used for teaching the concepts of computational techniques.

V. Theory

Unit I

Introduction to statistical packages and computing: data types and structures, Use of Software packages like, SAS, SPSS or "R: The R Project for Statistical Computing". Data analysis principles and practice, Summarization and tabulation of data, Exploratory data analysis; Graphical representation of data. Statistical Distributions: Fitting and testing the goodness of fit of discrete and continuous probability distributions;

Unit II

ANOVA, regression and categorical data methods; model formulation, fitting, diagnostics and validation; Matrix computations in linear models. Analysis of discrete data. Multiple comparisons, Contrast analysis.

Unit III

Numerical linear algebra, numerical optimization, graphical techniques, numerical approximations, Time Series Analysis.

Unit IV

Analysis of mixed models; Estimation of variance components, Analysis of Covariance, Fitting of non-linear model, Discriminant function; Principal component analysis. Techniques in the analysis of survival data and longitudinal studies, Approaches to handling missing data, and meta-analysis.

VI. Practical

- Data management, Graphical representation of data, Descriptive statistics;
- General linear models ~ fitting and analysis of residuals, outlier detection;
- Fitting and testing the goodness of fit of probability distributions;
- Testing the hypothesis for one sample *t*-test, two sample *t*-test, paired *t*-test, testfor large samples - Chi-squares test, F test, One way analysis of variance, contrastand its testing, pairwise comparisons;
- Mixed effect models, estimation of variance components;
- Categorical data analysis, dissimilarity measures, similarity measures;
- Analysis of discrete data, analysis of binary data;
- Numerical algorithms;
- Spatial modeling, cohort studies;
- Clinical trials, analysis of survival data;
- Handling missing data. Analysis of time series data - fitting of ARIMA models.

VII. Suggested Reading

- Agresti A. 2013. *Categorical Data Analysis*. 3rd Ed. John Wiley.
- Everitt BS and Dunn G. 1991. *Advanced Multivariate Data Analysis*. 2nd Ed. Arnold.
- Geisser S. 1993. *Predictive Inference: An Introduction*. Chapman & Hall.
- Gelman A & Hill J. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical*

Models. Cambridge Univ. Press.
- Gentle JE, Härdle W and Mori Y. 2012. *Handbook of Computational Statistics - Concepts and Methods*. 2nd Ed. Springer.
- Han J and Kamber M. 2000. *Data Mining: Concepts and Techniques*. Morgan.
- Hastie T, Tibshirani R and Friedman R. 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.
- Kennedy WJ & Gentle JE. 1980. *Statistical Computing*. Marcel Dekker.
- Miller RG Jr. 1986. *Beyond ANOVA, Basics of Applied Statistics*. John Wiley.
- Rajaraman V. 1993. *Computer Oriented Numerical Methods*. Prentice-Hall.
- Ross S. 2000. *Introduction to Probability Models*. Academic Press.
- Ryan BF and Joiner BL. 1994. *MINITAB Handbook*. 3rd Ed. DuXbury Press.
- Simonoff JS. 1996. *Smoothing Methods in Statistics*. Springer.
- Singh, AK. 2016. *Practical R-Book by Examples for Agricultural Statistics*. Deptt. Of Ag. Statistics, IGKV. Raipur
- Snell EJ. 1987. *Applied Statistics: A Handbook of BMDP Analyses*. Chapman & Hall.
- Thisted RA. 1988. *Elements of Statistical Computing*. Chapman & Hall.
- Venables WN and Ripley BD. 1999. *Modern Applied Statistics With S-Plus*. 3rd Ed. Springer.
- http://www.r-project.org/
- http://www.stat.sc.edu/~grego/courses/stat706/. Design Resources Server: www.drs.icar.gov.in.
