Ph. D. Programme

Course No.	Course Title	Credit Hrs.(T+P)
1st Semester		
STAT- 701	Advanced Statistical Computing	3: (2+1)
STAT - 702	Simulation Techniques	2: (1+1)
STAT - 703	Advanced Statistical Methods	2: (2+0)
STAT - 704	Advanced Statistical Inference	3: (3+0)
STAT - 705	Advanced Statistical Genetics	2: (2+0)
STAT- 706	Stochastic Processes	2: (2+0)
STAT - 707	Survival Analysis	2: (2+0)
STAT - 708	Theory of Reliability and Life Testing	2: (2+0)
STAT - 709	Spatial Analysis	3: (2+1)
STAT - 710	Regression Diagnostics	3: (2+1)
2nd Semester		
STAT - 751	Advanced Design of Experiments	2: (2+0)
STAT - 752	Advanced Sampling Techniques	2: (2+0)
STAT - 753	Statistical Modeling	2: (1+1)
STAT - 754	Advanced Time Series Analysis	2: (2+0)
STAT - 755	Advanced Bioinformatics	2: (2+0)
STAT – 756	Advanced Econometrics	2: (2+0)
STAT- 799	Doctoral seminar I	1: (1+0)
3rd Semester		
STAT - 801	Recent Advances in the Field of Specialization	1: (1+0)
4th Semester		
STAT - 899	Doctoral seminar II	1: (1+0)
5th Semester		
nil		
6th Semester		
STAT - 999	Doctoral seminar III	1: (1+0)
STAT - 1000	Doctoral Research	45: (0+45)

Ph. D. Programme:

STAT- 701: ADVANCED STATISTICAL COMPUTING 3(2+1)

Theory

Measures of association. Structural models for discrete data in two or more Dimensions. Estimation in complete tables. Goodness of fit, choice of a model. Generalized Linear Model for discrete data, Poisson and Logistic regression models. Log-linear models.

Elements of inference for cross-classification tables. Models for nominal

and ordinal response.

Computational problems and techniques for robust linear regression, nonlinear and generalized linear regression problem, tree-structured regression and classification, cluster analysis, smoothing and function estimation, robust multivariate analysis.

Analysis of incomplete data: EM algorithm, single and multiple imputations. Markov Chain, Monte Carlo and annealing techniques, Neural Networks, Association Rules and learning algorithms.

Linear mixed effects models, generalized linear models for correlated data (including generalized estimating equations), computational issues and methods for fitting models, and dropout or other missing data.

Multivariate tests of linear hypotheses, multiple comparisons, confidence regions, prediction intervals, statistical power, transformations and diagnostics, growth curve models, dose-response models.

STAT- 702: SIMULATION TECHNIQUES 2(1+1)

Theory

Review of simulation methods; Implementation of simulation methods – for various probability models, and resampling methods: theory and application of the jackknife and the bootstrap.

Randomization tests, analysis using computer software packages. Simulating multivariate distributions, MCMC methods and Gibbs sampler.

Correlograms, periodograms, fast Fourier transforms, power spectra, crossspectra,

coherences, ARMA and transfer-function models, spectral-domain regression. Simulated data sets to be analyzed using popular computer software packages

Stochastic simulation: Markov Chain, Monte Carlo, Gibbs' sampling, Hastings- etropolis algorithms, critical slowing-down and remedies, auxiliary variables, simulated tempering, reversible- jump MCMC and multi-grid methods.

STAT 703: ADVANCED STATISTICAL METHODS 2(2+0)

Theory

Ridge regression: Basic form, Use as a selection procedure. Robust regression: Least absolute deviations regression, M-estimators, Least median of squares regression. Nonparametric regression.

Introduction to the theory and applications of generalized linear models,fixed effects, random effects and mixed effects models, estimation of variance components from unbalanced data. Unified theory of least - squares, MINQUE, MIVQUE, REML.

Quasi-likelihoods, and generalized estimating equations – logistic regression, over-dispersion, Poisson regression, log-linear models, conditional likelihoods, generalized mixed models, and regression diagnostics. Theory of statistical methods for analyzing categorical data by means of linear models; multifactor and multi-response situations; interpretation of interactions.

Fitting of a generalized linear model, mixed model and variance components estimation, MINQUE, MIVQUE, REML.

Fitting of Logistic regression, Poisson regression, ridge regression, robust regression, non-parametric regression.

STAT- 704: ADVANCED STATISTICAL INFERENCE 3(3+0)

Theory

Robust estimation and robust tests, Robustness, M-estimates. L-estimates, asymptotic techniques, Bayesian inference. Detection and handling of outliers in statistical data.

Loglinear models, saturated models, hierarchical models, Analysis of multi - dimensional contingency tables. Non-parametric maximum likelihood estimation.

Density Estimation: Density Estimation in the Exploration and Presentation of Data. Survey of existing methods. The Kernel method for Univariate Data: Rosenblatts naïve estimator, its bias and variance. Consistency of general Kernel estimators, MSE and IMSE. Asymptotic normality of Kernel estimates of density. Estimation of distribution by method of kernels.

Consistency and asymptotic normality (CAN) of real and vector parameters. Invariance of consistency under continuous transformation. Invariance of CAN estimators under differentiable transformations, generation of CAN estimators using central limit theorem. Exponential class of densities and multinomial distribution, Cramer-Huzurbazar

theorem, method of scoring.

Efficiency: asymptotic relative efficiency and Pitman's theorem. Concepts and examples of Bahadur efficiency and Hodges-Lehmanns efficiency with examples. The concepts of Rao's second order efficiency and Hodges- Lehmann's Deficiency with examples. Rank tests, permutation tests, asymptotic theory of rank tests under null and alternative (contiguous) hypotheses.

Inference on Markov Chains: Maximum likelihood estimation and testing of Transition Probability Matrix of a Markov Chain, testing for order of a Markov chain, estimation of functions of transition probabilities.

Concept of loss, risk and decision functions, admissible and optimal decision functions, a-priori and posteriori distributions, conjugate families. Bayes and Minimax decision rules and some basic results on them. Estimation and testing viewed as cases of decision problems. Bayes and Minimax decision functions with applications to estimation with quadratic loss function. Concept of Bayesian sequential analysis. Bayes sequential decision rule. The SPRT as a Bayes procedure. Minimax sequential procedure.

U-Statistics: definitions of estimable parametric function, kernel, symmetric kernel and U-statistics. Variance and covariance of U-statistics. Hoeffding's decomposition of U-statistics –examples. U-statistics based on sampling from finite populations and weighted U-statistics with examples. Some convergence results on U-statistics. Asymptotic normality of Ustatistics with examples.

Resampling Plans : Estimation of standard and biased deviation of point estimator by the Jackknife, the Bootstrap, the Infinitesimal Jackknife, the Delta and the Influence function methods. Estimation of excess error in regression by cross validation, the Jackknife and Bootstrap methods. Nonparametric confidence interval for the median by the Percentile

method.

STAT 705: ADVANCED STATISTICAL GENETICS 2(2+0)

Theory

Hardy-Weinberg law with multiple allelic systems, auto-tetraploids and self-sterility alleles. Complex cases of selection with two or more loci.

Different approaches to study inbreeding process, methods of path coefficient, probability and generation matrix. Fisher's approach to inbreeding. Stochastic process of gene frequency change, transition matrix approach using finite Markov chains, diffusion approximation, Steady decay and distribution of gene frequency, Probability of fixation of a gene, Conditional process - Markov chains and diffusion

approaches, Distribution of time until fixation, random fluctuations in selection intensity, stationary distribution of gene frequency. Effective population size.

Prediction and estimation of genetic merit. Best linear unbiased prediction, Use of mixed model methodology in analysis of animal and plant breeding experiments. Newer reproductive technology and its effect in genetic evaluation of individual merit. Estimation of genetic parameters – problems relating to computational aspects of genetic variance components, parameter estimation in variance component models for binary response data.

Identification of genes with large effects, Use of molecular markers (RFLP, PCR-AFLP, RAPD and SSR), Gene mapping and Quantitative trait loci. Molecular manipulation for genetic variability.

Survival analysis and concept of censored observation in animal breeding. Phylogeny and analysis of molecular variance.

STAT- 706: STOCHASTIC PROCESSES 2(2+0)

Theory

Introduction to stochastic process - classification according to state space and time domain. Finite and countable state Markov chains; timehomogeneity; Chapman-Kolmogorov equations, marginal distribution and finite dimensional distributions. Classification of Markov chain. Canonical form of transition probability matrix of a Markov chain. Fundamental matrix; probabilities of absorption from transient states into recurrent classes in a finite Markov chain, mean time for absorption. Ergodic state and Ergodic chain. Stationary distribution of a Markov chain, existence and evaluation of stationary distribution. Random walk and gamblers ruin problem.

Discrete state continuous time Markov process: Kolmogorov difference – differential equations. Birth and death process, pure birth process (Yule- Fury process). Immigration-Emigration process. Linear growth process, pure death process.

Renewal process: renewal process when time is discrete and continuous. Renewal function and renewal density. Statements of Elementary renewal theorem and Key renewal theorem.

Stochastic process in biological sciences: Markov models in population genetics, compartmental analysis. Simple deterministic and stochastic epidemic model. General epidemic models-Karmack and McKendrick's threshold theorem. Recurrent epidemics.

Elements of queueing process; the queuing model M/M/1: steady state behaviors. Birth and death process in queuing theory- Multi channel models. Net work of Markovian queuing system.

Branching process: Galton-Watson branching process. Mean and variance of size of nth generation, probability of ultimate extinction of a branching process. Fundamental theorem of branching process and applications.

Wiener process- Wiener process as a limit of random walk. First passage time for Wiener process. Kolmogorov backward and forward diffusion equations and their applications.

STAT- 707: SURVIVAL ANALYSIS 2(2+0)

Theory

Measures of Mortality and Morbidity: Ratios and proportions, rates of continuous process, rates of repetitive events ,crude birth rate, Mortality measures used in vital statistics relationships between crude and age specific rates, standardized mortality ratios ,evaluation of person-year of exposed to risk in long term studies, prevalence and incidence of a disease,

relative risk and odds ratio. Survival Distribution: Survival functions, hazard rate, hazard function, review of survival distributions: exponential, Weibull, Gamma, Rayleigh, Pareto, Lognormal~ IFR and TFRA, Gompertz and Makeham. Gompertz and logistic distributions. Parametric (m.l.e) estimation. Types of Censoring: Type I, Type II, random and other types of censoring, right and left truncated distributions. Expectation and variance of future life time, series and parallel system of failures. Life Tables: Fundamental and construction.

Complete Mortality data, Estimation of Survival Function : Empirical survival function , estimation of survival function from grouped mortality data, joint distribution of the number of deaths, distribution of the estimation Pi covariance of estimate, estimation of curves of deaths and central death rate and force of mortality rate . Incomplete Mortality data (non-parametric models): Actuarial method, m.1.e method, moment

and reduced sample method of estimation and their comparison. Product limit (Kaplan-Meier) method and cumulative hazard function (CHF) of estimation of survival function.

Fitting Parametric Survival Distribution : Special form of survival function cumulative hazard function (CHF) plots, Nelson's method of ungrouped data, construction of the likelihood function for survival data, least squares fitting, fitting a Gompertz distribution to grouped data. Some tests of Goodness of fit: Graphical, Kolmogorov-Smirnov statistics for complete, censored and truncated data, Chi-Square test and Anderson- Darling A2-statistics.

Comparison of Mortality Experiences: Comparison of two life tables, some distribution- free methods (two samples) for ungrouped data, Two samples Kolmogorov-Smirnov test, Wilcoxon test for complete data and modified Wilcoxon test for incomplete data .Gilbert and Gehan's test, mean and variance of Wilcoxon statistics, generalization of Gehan's test. Testing for Consistent Differences in Mortality : Mantel-Haenszel and log rank test. Generalized Mantel-Haenszel test (k-sample).

Concomitant Variables: General parametric model for hazard function with observed concomitant variables. Additive and multiplicative models of hazard rate functions. Estimating multiplicative models, selection of concomitant variables. Logistic linear model, Concomitant Variable regarded as random variable. Age of onset distributions: Models of onset distributions and their estimation. Gompertz distribution, parallel system and Weibull distribution, Fatal short models of failure. Two component series system.

STAT 708: Theory of Reliability and Life Testing 2(2+0)

Life Distribution. Failure distribution, Definitions and Properties. Failure rate function. Failure rate average. Conditional mean remaining life. Loss of memory property. Failure potential. Weibull, Gamma Normal, Log normal. System reliability: series and parallel and stand by system. Reliability estimation with complete and censored sample using MLE and UMVUE.

STAT 709: Spatial Analysis 2(2+1)

What is spatial Analysis? Definition and concepts. Role of spatial analysis. Different models used to accommodate spatial correlation. Stationary and non- stationary spatial models. Application of spatial models in experimental designs. Comparative studies of different spatial models.

STAT 710: Regression Diagnostics 2(2+1)

What is regression diagnostics? Influential variables, leverage, outlier- detection of outlier. Winsorised regression Jack- Knife estimation- its application in different fields of statistical research. Non- normality- reasons and remedy. Heteroscedasticity- reasons and remedy. Multicollinearity- reasons , detection and detection measures. Problem of specification, detection and remedial measures.

2nd Semester:

STAT 751: ADVANCED DESIGN OF EXPERIMENTS 2(2+0)

Theory

General properties and analysis of block designs. Balancing criteria. *m* associate

PBIB designs, and their association schemes including lattice designs - properties and construction, Designs for test treatment – control(s) comparisons; Nested block designs, Mating designs.

General properties and analysis of two-way heterogeneity designs, Youden type designs, generalized Youden designs, Pseudo Youden designs. Structurally Incomplete block designs, Designs for two sets of treatments.

Balanced factorial experiments - characterization and analysis (symmetrical and asymmetrical factorials). Factorial experiments with extra treatment(s). Orthogonal arrays, Mixed orthogonal arrays, balanced arrays, Fractional replication, Regular and irregular fractions.

Response surface designs - Symmetrical and asymmetrical factorials, Response optimization and slope estimation, Blocking. Canonical analysis and ridge analysis. Experiments with mixtures: design and analysis. Experiments with qualitative cum quantitative factors.

Optimality criteria and optimality of designs, robustness of designs against loss of data, outliers, etc. Diagnostics in design of experiments.

STAT 752: ADVANCED SAMPLING TECHNIQUES 2(2+0)

Theory

Controlled selection. Two way stratification, collapsed strata. Systematic

sampling in two dimensions. Use of combinatorics in controlled selection. Integration of surveys - Lahiri and Keyfitz's procedures.

Variance estimation in complex surveys. Taylor's series linearisation, balanced repeated replication, Jackknife and bootstrap methods.

Unified theory of sampling from finite populations. UMV - Non-existence theorem and existence theorem under restricted conditions. Concept of sufficiency and likelihood in survey sampling. Admissibility and hyperadmissibility.

Inference under super population models - concept of designs and model unbiasedness, prediction approach. Regression analysis and categorical data analysis with data from complex surveys. Domain estimation. Small area estimation.

Stochastic parameter models, Bayes' linear predictor, Bayesian models with multi-stage sampling. Measurement error and small area estimation, Time series approach in survey sampling. Dynamic Bayesian prediction, Kalman filter, Empirical and Hierarchical Bayes predictors, Robust linear prediction, Bayesian robustness.

STAT 753: STATISTICAL MODELING 2(1+1)

Theory

Empirical and mechanistic models. Nonlinear growth models like monomolecular, logistic, Gompertz, Richards. Applications in agriculture and fisheries.

Nonlinear estimation: Least squares for nonlinear models, Methods for estimation of parameters like Linearization, Steepest, and Levenberg- Marquardt's Reparameterization.

Two-species systems. Lotka-Volterra, Leslie-Gower and Holling-Tanner non-linear prey-predator models. Volterra's principle and its applications. Gause competition model.

Compartmental modelling - First and second order input-output systems, Dynamics of a multivariable system.

STAT 754: ADVANCED TIME SERIES ANALYSIS 2(2+0)

Theory

Multivariate time series: modelling the mean, stationary VAR models: properties, estimation, analysis and forecasting, VAR models with elements of nonlinearity, Non-stationary multivariate time series: spurious regression, co-integration, common trends.

Volatility: Modelling the variance, The class of ARCH models: properties, estimation, analysis and forecasting, stochastic volatility, realized volatility, Extensions: IGARCH, ARCH-t, ARCD, Multivariate GARCH, Timevarying risk and ARCH-in-mean.

Structural time-series modelling: State space models, Kalman filter. Local level model, Local linear trend model, Seasonal models, Cyclical models. Nonlinear time-series models: Parametric and nonparametric approaches. Autoregressive conditional heteroscedastic model and its extensions. Threshold and Functional coefficient autoregressive models.

Non-linear programming, Kuhn-Tucker sufficient conditions, Elements of multiple objective programming, Dynamic Programming, Optimal control theory - Pontryagin's maximum principle, Time-optimal control problems.

STAT- 755: ADVANCED BIOINFORMATICS 2(2+0)

Theory

Genomic databases and analysis of high-throughput data sets, sequence annotation, ESTs, SNPs. BLAST and related sequence comparison methods. EM algorithm and other statistical methods to discover common motifs in biosequences. Multiple alignment and database search using motif models, ClustalW and others. Concepts in phylogeny. Gene prediction based on codons, Decision trees, Classificatory analysis, Neural Networks, Genetic algorithms, Pattern recognition, Hidden Markov models.

Computational analysis of protein sequence, structure and function. Expression profiling by microarray/gene chip, proteomics etc., Multiple alignment of protein sequences, Modelling and prediction of structure of proteins, Designer proteins, Drug designing.

Analysis of one DNA sequence (Modeling signals in DNA; Analysis of patterns; Overlaps and Generalizations), Analysis of multiple DNA or protein sequences (Alignment algorithms – Gapped global comparisons and Dynamic programming; use of linear gap models; protein sequences and substitution matrices – BLOSUM, PAM; Multiple sequences), BLAST (Comparison of two aligned sequences – Parameter calculation; Choice of a score; Bounds for P-value; Normalized and Bit scores, Karlin – Altschul sum statistic; comparison of two unaligned sequences; Minimum significance Lengths).

Markov chains (MC with no absorbing states; Higher order Markov dependence; patterns in sequences; Markov chain Monte Carlo – Hastings- Metropolis algorithm, Gibbs sampling, Simulated Annealing; MC with absorbing States, Continuous-Time Markov chains) Hidden Markov Models (Forward and Backward algorithm; Viterbi algorithms; Estimation algorithm;

Modeling protein families; Multiple sequence alignments; Pfam; Gene finding), Computationally intensive methods (Classical estimation methods; Bootstrap estimation and Confidence Intervals; Hypothesis testing; Multiple Hypothesis testing), Evolutionary models (Models of Nucleotide substitution; Discrete time models – The Jukes-Cantor Model, The Kimura Model, The Felsenstein Model; Continuous-time models).

Phylogenetic tree estimation (Distances; Tree reconstruction – Ultrametric and Neighbor-Joining cases; Surrogate distances; Tree reconstruction; Parsimony and Maximum Likelihood; Modeling, Estimation and Hypothesis Testing;) Neural Networks (Universal Approximation Properties; Priors and Likelihoods, Learning Algorithms – Backpropagation; Sequence encoding and output interpretation; Prediction of Protein Secondary Structure; Prediction of Signal Peptides and their cleavage sites; Application for DNA and RNA Nucleotide Sequences), Analysis of SNPs and Haplotypes.

STAT- 756 : ADVANCED ECONOMETRICS 2(2+0)

Theory

Quantile regression, binary quantile regression, extreme values, copula, loss functions, Point and interval forecasting, unconditional and conditional forecasting, forecasting with serially correlated errors, bootstrap: asymptotic expansion, bootstrap consistency, asymptotic refinement, recent developments for dependent time series

Multivariate time series: modelling the mean, stationary VAR models: properties, estimation, analysis and forecasting, VAR models with elements of nonlinearity, Non-stationary multivariate time series: spurious regression, co-integration, common trends; Volatility: Modelling the variance, The class of ARCH models: properties, estimation, analysis and forecasting, stochastic volatility, realized volatility.

Basic Concepts of Bayesian Inference, Probability and Inference, Posterior Distributions and Inference, Prior Distributions. The Bayesian linear model and autoregressive (AR) processes; Model selection with marginal likelihoods and fractional priors, Comparison of Bayesian Methods with Classical approaches, Bayes risk and their applications, and Sample Selection Monte Carlo integration, importance sampling and Gibbs

sampling, The Regression Model with General Error Covariance Matrix, Qualitative Choice Models, Bayesian information criterion (BIC), Markov Chain Monte Carlo (MCMC) Model Composition and stochastic search variable selection, BUGS [Bayesian Inference Using Gibbs Sampling] , BUCC [Bayesian Analysis, Computation and Communication].

3rd Semester:

STAT 801 : RECENT ADVANCES IN THE FIELD OF SPECIALIZATION 1(1+0)

Theory

Recent advances in the field of specialization - sample surveys / design of experiments /statistical genetics / statistical modeling / econometrics / statistical inference, etc. will be covered by various speakers from the University / Institute as well as from outside the University / Institute in the form of seminar talks.