Rayat Shikshan Sanstha's

Sadguru Gadage Maharaj College, Karad.

(An Autonomous, Affiliated to Shivaji University, Kolhapur)

Accredited By NAAC With 'A+' Grade

CHOICE BASED CREDIT SYSTEM

Syllabus For

Bachelor of Science (Part III) Statistics

Syllabus to be implemented from June 2021 onwards

Sadguru Gadge Maharaj College, Karad
(An Autonomous)
B.Sc.-III Statistics
Structure of the course
To be implemented from June 2021
Theory
Course Structure

Semester V									
Theory					Practical				
Paper No.	Paper Code	Title	Lecture Per Week	Credits	Paper No.	Paper Code	Title	Lect. Per Week	Credits
IX	BST 501	Probability Distribution-I	3	2	VII	BSP 506	Probability Distribution-I	5	2
X	BST 502	Statistical Inference – I	3	2	VIII	BSP 507	Statistical Inference – I	5	2
XI	BST 503	Operations Research	3	2	IX	BSP 508	Operations Research	5	2
XII	BST 504	Elective – I Designs of Experiments	3	2	X	BSP 509	Designs of Experiments	5	2
		Elective – II Demography and Vital Statistics					Demography and Vital Statistics		
		Elective – III Data Mining					Data Mining		

Semester VI									
Theory					Practical				
Paper No.	Paper Code	Title	Lect. Per Week	Credits	Paper No.	Paper Code	Title	Lect. Per Week	Credis
XIV	BST 601	Probability Distribution-II	3	2	XII	BSP 606	Probability Distribution-II	5	2
XV	BST602	Statistical Inference – II	3	2	XIII	BSP 607	Statistical Inference II	5	2
XVI	BST603	Industrial Statistics	3	2	XIV	BSP 608	Industrial Statistics	5	2
XVII	BST604	Elective – I Sampling Theory	3	2	XV	BSP 609	Sampling Theory	5	2
		Elective – II Survey Sampling and Official Statistics					Survey Sampling and Official Statistics		
		Elective – III Reliability Theory					Reliability Theory		

Evaluation Pattern For B.Sc.-III

Semester-V

	SEE	Internal Exam		Total		Practical		Submission			Total
		CCE-I	CCE-II			Exam (Prac.+Oral +Journal)	Student Performance	Seminar/Educational tour	Project/Literature Survey		
Paper-XIV	40	5	5	50	Practical-XII	40	5	5	-		50
Paper-XV	40	5	5	50	Practical-XIII						
Paper-XVI	40	5	5	50	Practical-XIV	40	-	-	10		50
Paper-XVII	40	5	5	50	Practical-XV						
AECC-English	40	5	5	50	-	-	-		-		-
Total	200	25	25	200+50	Total	80	5	5	10		100

Semester-VI

	SEE	Internal Exam		Total		Practical		Submission			Total
		CCE-I	CCE-II			Exam (Prac.+Oral +Journal)	Student Performance	Seminar/Educational tour	Project/Literature Survey		
Paper-XIV	40	5	5	50	Practical-VII	40	5	5	-		50
Paper-XV	40	5	5	50							
Paper-XVI	40	5	5	50	Practical-VIII	40	-	-	10		50
Paper-XVII	40	5	5	50							
AECC-English	40	5	5	50	-	-	-		-		-
Total	200	25	25	200+50	Total	80	5	5	10		100

4. Instructions

(i) While attempting questions based on R-software students have to write the commands of R-software on their Answer-book. Final result should be shown to the examiner online or the printout may be attached.

(ii) Duration of each practical paper should be of four hours.

(iii) Student can use MS-Excel or electronic calculators for other practical.

5. Requirements

(i) There should be two subject experts at the time of practical examination.

(ii) Laboratory should be well equipped with 20 scientific calculators, 20 computers, 2 printers with sufficient backup facility (UPS/Inverter /Generator)

EQUIVALENCE FOR THEORY PAPERS

(From June 2021)

Old Syllabus		New Syllabus	
Paper No.	Title of the paper	Paper No.	Title of the paper
IX	Probability Distributions	BST 501	Probability Distributions- I
X	Statistical Inference-I	BST 502	Statistical Inference-I
XI	Design of Experiments	BST 503	Design of Experiments
XII	Operation Research	BST 504	Operation Research
XIII	Probability Theory	BST 601	Probability Distributions II
XIV	Statistical Inference-II	BST 602	Statistical Inference-II
XV	Sampling Theory	BST 603	Sampling Theory
XVI	Quality Management and Data Mining	BST 604	Industrial Statistics

PAER- IX

BST 501 :PROBABILITY DISTRIBUTION-I

Theory: 36 hours Marks-50 (Credit 02)

Course Outcomes: The students will acquire

a) knowledge of important univariate distributions such as Laplace, Cauchy, Lognormal, Weibull, Logistic, Pareto, Power Series Distribution.

b) knowledge of Multinomial and Bivariate Normal Distribution.

c) knowledge of Truncated Distributions.

d) information of various measures of these probability distributions.

e) acumen to apply standard continuous probability distributions to different situations.

Unit-1: Univariate Continuous Probability Distributions (15 hours)

1.1 Laplace (Double Exponential) Distribution: P. d. f. with parameters (μ , λ), Nature of the probability curve, Distribution function, quartiles, m. g. f., mean, variance, moments, $\beta 1$, $\beta 2$, $\gamma 1$ and $\gamma 2$, Laplace distribution as the distribution of the difference of two i. i. d. exponential variates with parameter θ, examples and problems

1.2 Lognormal Distribution: P.d.f. with parameters(μ, $\sigma 2$), Nature of the probability curve, mean, variance, median, mode, moments, $\beta 1$, $\beta 2$, $\gamma 1$ and $\gamma 2$ coefficients, Relation with N(μ, $\sigma 2$),examples and problems.

1.3 Cauchy Distribution: P. d. f. with parameters (μ, λ), nature of the probability curve, distribution function, quartiles, non-existence of moments, additive property for two independent Cauchy variates (statement only), statement of distribution of the sample mean, relationship with uniform and Student's 't' distribution, distribution of X/Y where X and Y are i. i. d. N(0, 1), examples and

problems.

1.4 Weibull Distribution: p.d. f. with parameters (α, β), distribution function, quartiles, mean and variance, coefficient of variation, relation with gamma and exponential distribution, examples and problems.

Unit-2: Univariate and Multinomial Probability Distributions (12 hours)

2.1: Logistic distribution: P.d.f. with parameters (μ, σ), c.d.f., mean, mode, variance, skewness using mode, applications.

2.2 : Pareto distribution: P.d.f. with parameters (α,β),mean, variance, mode, skewness using mode, applications.

2.3 :Power series distribution: P.m.f. mean, mode, variance, Binomial, Poisson, Geometric and negative binomial distribution as particular cases of power series distribution.

2.4 : Multinomial distribution: P.m.f, m.g.f.,marginal distribution, mean, variance, covariance, variance and covariance matrix, correlation coefficient, additive property, Trinomial distribution as particular case of multinomial distribution.

Unit-3: Truncated Distributions (8 hours)

3.1 : Truncated distribution as conditional distribution, truncation to the right, left and on both sides.

3.2 : Binomial distribution B(n, p) left truncated at X = 0(value zero not observable), its p.m.f, mean, variance.

3.3 : Poisson distribution P(m), left truncated at X = 0 (value zero not observable), its p.m.f., mean and variance.

3.4 : Normal distribution N(μ, σ2) truncated

(i) to the left below a

(ii) to the right above b

(iii)to the left below a and to the right above b, its p.d.f. and mean. **3.5:**Exponential distribution with parameter θ left truncated below a, its p.d.f., mean and variance.

3.6: Examples and problems.

Unit4:GeneratingFunctions (10 hours)

Definitions of generating function and probability generating function. Expression for mean and variance in terms of generating functions. Definition of a convolution of two or more sequences Generating function of a convolution. Generating

function of standard discrete distributions. Relation between: i) Bernoulli and Binomial distributions ii) Geometric and Negative Binomial distributions in terms of convolutions.

Books Recommended

1. Cramer H.: Mathematical Methods of Statistics, Asia Publishing House, Mumbai.
2. Mood, A. M., Graybill K, Bose. D. C. : Introduction to Theory of Statistics. (Third edition) Mc-GrawHiil Series.
3. Lindgren B. W.: Statistical Theory (Third Edition), Collier Macmillan International Edition, Macmillan Publishing Co. Inc. New York.
4. Hogg, R. V. and Craig A. T. :Introduction to Mathematical Statistics (Third Edition),Macmillan Publishing Company, Inc. 866, 34d Avenue, New York, 10022.
5. Sanjay Arora and BansiLal : New Mathematical Statistics (First Edition),Satya Prakashan, 16/17698, New Market, New Delhi, 5 (1989).
6. Gupta S. C and Kapoor V. K. : Fundamentals of Mathematical Statistics, Sultan Chand and Sons, 88, Daryaganj, New Delhi 2.
7. Rohatgi V. K.: An Introduction to Probability Theory and MathematicalStatistics, Wiley Eastern Ltd., New Delhi.
8. Feller. W. : An Introduction of Probability Theory and its Applications, Wiley Eastern Ltd.. Mumbai.
9. Jhonson and Kotz: Continuous Univariate Distributions I and II
10. Bhat B. R.: Modern Probability Theory. New Age International.
11. Morgan.B.J.T.(1984) Elements of simulation.
12.Robert. C.P and Casella.G. Monte Carlo statistics method,(Sprinnger)
13. Ross.S.M.(2006) Simulation, (Academic press).
14. Elements of Simulation: Byron.J.T. Morgan

Paper X : BST 502 :Statistical Inference – I

Theory: 36 hours Marks-50 (Credit 02)

Course Outcomes: The students will acquire

a) knowledge about important inferential aspect of point estimation.

b) concept of random sample from a distribution, sampling distribution of a statistic, standard error of important estimates such as mean and proportions.

c) knowledge of various important properties of estimator

d) knowledge about inference of parameters of standard discrete and continuous distributions.

e) concept of Fisher information and CR inequality.

f) knowledge of different methods of estimation.

Unit - 1: Point Estimation (15 hours)

1.1 : Notion of a parameter, parameter space, general problem of estimation, estimating an unknown parameter by point and interval estimation.

1.2 : Point estimation: Definition of an estimator (statistic) & its S.E., distinction between estimator and estimate, illustrative examples.

1.3 : Properties of estimator: Unbiased estimator, biased estimator, positive and negative bias, examples of unbiased and biased estimators. Proofs of the following results regarding the unbiased estimators:

(a) Two distinct unbiased estimators of $\phi(\theta)$ give rise to infinitely many unbiased estimators of $\phi(\theta)$

(b) If T is unbiased estimator of θ then ϕ (T) is an unbiased estimator of ϕ (θ)provided ϕ(.) is a linear function. Sample variance is a biased estimator of the population variance. Illustration of unbiased estimator for the parameter and parametric function.**1.4:**Relative efficiency of T1 with respect to T2, where T1 and T2 are unbiased estimators. Use of mean square error to modify the above definition for biased estimator. Minimum Variance Unbiased Estimator (MVUE) and Uniformly Minimum Variance Unbiased Estimator (UMVUE), uniqueness of UMVUE whenever it exists.Illustrative examples.

1.5 : Consistency : Definition, proof of the following :

(a) Sufficient condition for consistency,

(b) If T is consistent for θ and ϕ (.) is a continuous function then ϕ (T) is consistent for$\phi(\theta)$Illustrative examples.

Unit - 2: Likelihood and Sufficiency (12 hours)

2.1 : Definition of likelihood function as a function of the parameter θ for a

random sample from discrete and continuous distributions. Illustrative examples.

2.2: Sufficiency: Concept of sufficiency, definition of sufficient statistic through (i) conditional distribution (ii) Neyman factorization criterion. Pitman Koopman form and sufficient statistic. Proof of the following properties of sufficient statistic:

(a) If T is sufficient for θ then ϕ (T) is also sufficient for θ provided ϕ (.) isa bijective function.

(b) If T is sufficient for θ then T is sufficient for ϕ (θ).

2.3: Fisher information function: Definition of information function, amount of information contained in a sample. Statement regarding equality of the information in (x1, x2,. .....,xn) and in a sufficient statistic T, concept of minimal sufficient statistic. With illustrations to exponential family.

2.4: Illustrative examples.

Unit - 3: Cramer-Rao Inequality **(7)**

Statement and proof of Cramer Rao inequality. Definition of Minimum Variance Bound Unbiased Estimator (MVBUE) of ϕ (θ). Proof of the following results:

(i) If MVBUE exists for θ then MVBUE exists for ϕ (θ), if ϕ(.) is a linear function.

(ii) If T is MVBUE for θ then T is sufficient for θ. Examples and problems.

Unit - 4: Methods of Estimation **(11)**

4.1: Method of maximum likelihood, derivation of maximum likelihood estimators for parameters of standard distributions. Use of iterative procedure to derive MLE of location parameter μ of Cauchy distribution, invariance property of MLE, relation between MLE and sufficient statistic. Illustrative examples.

4.2: Method of moments: Derivation of moment estimators for standard distributions. Illustrations of situations where MLE and moment estimators are distinct and their comparison using mean square error (for uniform distribution). Illustrative examples. **4.3:** Method of minimum chi-square: Definition, derivation of minimum chi-square estimator for the parameter. Illustrative examples.

Books Recommended

1. Kale, B. K.: A first Course on Parametric Inference
2. Rohatgi, V. K.: Statistical Inference
3. Rohatgi, V. K.: An introduction to Probability Theory and Mathematical Statistics
4. Saxena H. C. and Surenderan : Statistical Inference
5. Kendall M. G. and Stuart A.: An advanced Theory of Statistics
6. Lindgren, B. W.: Statistical Theory

7. Lehmann, E. L.: Theory of Point Estimation
8. Rao, C. R.: Linear Statistical Inference
9. Dudewicz C. J. and Mishra S. N. : Modern Mathematical Statistics
10. Fergusson, T. S.: Mathematical statistics.
11. Zacks, S.: Theory of Statistical Inference.
12. Cramer, H.: Mathematical Methods of Statistics.
13. Cassela G. and Berger R. L.: Statistical Inference

Paper XI : BST 503 : Operations Research

Theory: 36 hours Marks-50 (Credit 02)

Course Outcomes: The students will acquire

a) Concept of Linear programming problem.

b) Knowledge of solving LPP by graphical and Simplex method.

c) Knowledge of Transportation, Assignment and Sequencing problems.

d) Concept of queuing theory.

e) Knowledge of simulation technique and Monte Carlo technique of simulation.

Unit-1: Linear programming (15)

1.1: Basic concepts

Statement of the Linear Programming Problem (LPP), formulation of problem as.L.P. problem. Definition of (i) a slack variable, (ii) a surplus variable.L.P. problem in (i) canonical form, (ii) standard form. Definition of (i) a solution, (ii) a feasible solution,

(iii) basic variable and non-basic variable, (iv) a basic feasible solution, (v) a degenerate and a non-generate solution, (vi) an optimal solution.

1.2 : Solution of L.P.P.

i. Graphical Method: Solution space, obtaining an optimal solution, unique and nonunique optimal solutions.

ii. Simplex Method:

(a) Initial basic feasible solution (IBFS) is readily available: obtaining an IBFS, criteria for deciding whether obtained solution is optimal, criteria for unbounded solution, , more than one optimal solutions.

(b) IBFS not readily available: introduction of artificial variable, Big-M method, modified objective function, modifications and applications of simplex method to L.P.P., criterion for no solution.

1.3 : Duality Theory:

Writing dual of a primal problem, solution of L.P.P. with artificial variable.

1.4 : Examples and problems.

Unit-2: Transportation and Assignment Problems (12)

2.1: Transportation problem

i. Transportation problem (T. P.), statement of T. P., balanced and unbalance T. P.

ii. Methods of obtaining initial basic feasible solution of T.P. (a) North West corner rule (b)Method of matrix minima (least cost method), (c) Vogel's approximation(VAM).

iii. MODI method of obtaining Optimal solution of T. P, uniqueness and

non- uniqueness of optimal solutions, degenerate solution.
iv. Examples and problems.

2.2 : Assignment Problem

i. Statement of an assignment problem, balanced and unbalanced assignment problem,relation with T.P, optimal solution of an assignment problem using Hungarian method.
ii. Examples and problems.

2.3 : Sequencing Problem

i. Introduction. Statement of problem.
ii. Procedure of processing n jobs on two machines.
iii.Procedure of processing n jobs on three machines and m machines. Computations of elapsed time and idle times.
iv.Examples and problems.

Unit-3: Decision Theory **(8)**

i. Introduction, steps in decision theory approach.
ii. Type of decision making environments.
iii. Decision making under uncertainty: Criteria of optimism, criteria of pessimism, equally likely decision criterion, criterion of regret.
iv. Decision making under risk: Expected monetary value, expected opportunity loss, expected value of perfect information.
v. Examples **a**nd problems.

Unit-4: Network Analysis: **(10)**

4.1 Scope and definition of network model, minimal spanning tree algorithm, shortest route problem.
4.2 CPM, PERT: Network Representation, critical path computation, construction of time schedule, linear programming formulation of CPM. PERT calculations.

Book Recommended:

1.Taha H. A.: Operation research – An Introduction, Fifth Edition, Prentice Hall of India, New Delhi.
2.Shrinath, L. S. : Linear Programming, Affiliated East-West Press Pvt. Ltd., New Delhi. 3.Sharma, J. K. : Mathematical Models in Operations Research, Tau McGraw Hill Publishing Company Ltd., New Delhi.
4.Kapoor, V. K. ; Operations Research, Sultan Chand and Sons, New Delhi.
5.Gupta, P. K. and Hira, D. S. : Operations Research, S. Chand and Company Ltd., New Delhi.

Elective :I : Paper XII : BST 504 : Designs of Experiments

Theory: 36 hours Marks-50 (Credit 02)

Course Outcomes:

The students will acquire: a) knowledge of basic terms used in design of experiments.
b) concept of one-way and two-way analysis of variance.
c) knowledge of various designs of experiments such as CRD, RBD, LSD and factorial experiments.
d) knowledge of using an appropriate experimental design to analyze the experimental data.

Unit – 1: Simple Designs of Experiments (10)

1.1: Basic Concepts:

i. Basic terms in design of experiments: Experimental unit, treatment, layout of an experiment.
ii. Basic principles of design of experiments: Replication, randomization and local control.
iii. Choice of size and shape of a plot for uniformity trials, the empirical formula for the variance per unit area of plots.
iv Concept and definition of efficiency of a design.

1.2: Completely Randomized Design (CRD)

i. Application of the principles of design of experiments in CRD, layout, model, assumptions and interpretations:
ii. Estimation of parameters, expected values of mean sum of squares, components of variance.
iii. Breakup of total sum of squares in to components.
iv. Techniqueof one way analysis of variance (ANOVA) and its applications to CRD. v.Testing for equality for treatment effects and its interpretation. F-test for testing H0, test for equality of two specified treatment effects.

Unit- 2: Simple Design of Experiments II : (15)

2.1: Randomized Block Design(RBD):

i.Application of the principles of design of experiments in RBD, layout, model , assumptions and interpretations:
ii. Estimation of parameters, expected values of mean sum of squares, components of variance.
iii. Breakup of total sum of squares into components.
iv. Technique of two way analysis of variance (ANOVA) and its applications to RBD. v.Tests and their interpretations , test for equality of

two specified treatment effects, comparison of treatment effects using critical difference (C.D.).

vi. Idea of missing plot technique.

vii. Situations where missing plot technique is applicable.

viii. Analysis of RBD with single missing observation.

ix. Efficiency of RBD over CRD.

2.2 Latin Square Design (LSD)

i. Application of the principles of design of experiments in LSD, layout, model, assumptions and interpretations:

ii. Breakup of total sum squares into components.

iii. Estimation of parameters, expected values of mean sum of squares,components of variance. preparation of analysis of variance (ANOVA) table.

iv. Tests and their interpretations, test for equality of two specified treatment effects, comparison of treatment effects using critical difference (C.D.).

v. Analysis of LSD with single missing observation.

vi. Identification of real life situations where CRD, RBD AND LSD are used. vii.Efficiency of LSD over CRD and LSD over RBD.

Unit – 3 SP Design and ANOCOVA (10)

3.1. Split Plot Design:

Application of the principles of design of experiments in layout, model , assumptions and interpretations:

ii. Estimation of parameters, expected values of mean sum of squares, components of variance.

3.2 Strip Plot Design:

i. Application of the principles of design of experiments, layout, model, assumptions and interpretations:

ii. Estimation of parameters, expected values of mean sum of squares, components of variance.

3.2 Analysis of Covariance (ANOCOVA) with one concomitant variable

i. Purpose of analysis of covariance.

ii. Practical situations where analysis of covariance is applicable.

iii. Model for analysis of covariance in CRD and RBD. Estimation of parameters (derivations are not expected).

iv. Preparation of analysis of covariance (ANOCOVA) table, test for $\beta = 0$, test for equality of treatment effects (computational technique only).

Note :- For given data, irrespective of the outcome of the test of regression coefficient (β), ANOCOVA should be carried out.

Unit – 4 : Factorial Experiments **(10)**

i. General description of factorial experiments, 2^2 and 2^3 factorial experiments arranged in RBD.

ii. Definitions of main effects and interaction effects in 2^2 and 2^3factorialexperiments.

iii. Model , assumptions and its interpretation.

iv. Preparation of ANOVA table by Yate's procedure, test for main effects and interaction effects.

v. General idea and purpose of confounding in factorial experiments.

vi. Total confounding (Confounding only one interaction) : ANOVA table, testing main effects and interaction effects.

vii. Partial Confounding (Confounding only one interaction per replicate): ANOVAtable,

testing main effects and interaction effects.

viii. Construction of layout in total confounding and partial confounding in 23factorial experiment.

Books Recommended

1. Federer, W.T. : Experimental Design, Oxford and IBH publishing Company, New Delhi.
2. Cochran, W.G. and Cox, G.M. : Experimental Design, John Wiley and Sons, Inc., New York.
3. Montgomery, D.C.: Design and Analysis of Experiments, Wiley Eastern Ltd., New Delhi. 12
4. Das, M.N. and Giri, N.C. : Design and Analysis of Experiments, Wiley Eastern Ltd., New Delhi.
5. Goulden, G.H. : Methods of Statistical Analysis, Asia Publishing House, Mumbai.
6. Kempthrone, O. : Design and Analysis of Experiments, Wiley Eastern Ltd., New Delhi.
7. Snedecor, G.W. and Cochran, W.G. : Statistical Methods, Affiliated East-West Press, New Delhi.
8. Goon, Gupta, Dasgupta : Fundamental of Statistics, Vol. I and II, The World Press Pvt. Ltd. Kolkata.
9. Gupta, S.C. and Kapoor, V.K. : Fundamentals of Applied Statistics, S. Chand & Sons, New Delhi.
10. C.F. Jeff Wu, Michael Hamada : Experiments, Planning Analysis and Parameter Design Optimization.

Elective :II : Paper XII : BST 504 : Demography and Vital Statistics

Theory: 36 hours Marks-50 (Credit 02)

Unit1 **(10)**

Population Theories: Coverage and content errors in demographic data, use of balancing equations and
Chandrasekaran-Deming formula to check completeness of registration data.
Adjustment of age data, use of Myer and
UN indices, Population composition, dependency ratio.

Unit 2 **(15)**

Introduction: Sources of data on vital statistics, errors in census and registration data.
Measurement of population,
Rates and ratios of vital events.
Measurements of Mortality: Crude Death Rate (CDR), Specific Death Rate (SDR), Standardized Death Rate, Cause of
Death Rate, Case Fatality Rate, Infant Mortality Rate (IMR), Maternal
Mortality Rate (MMR), Neonatal and Perinatal
Mortality Rates.

Unit 3 **(10)**

Life (Mortality) Tables: Assumption, descriptions of Complete and Abridged Life Tables, Cohort vs. Current Life
Tables, Stationary and Stable population, Construction of Complete Life
Table from population and death statistics,
Central Mortality Rates and Force of Mortality, Uses of Life Tables.
Measurements of Morbidity: Morbidity Incidence and Morbidity
Prevalence Rates.
Measurements of Fertility: Crude Birth Rate (CBR), General Fertility Rate (GFR), Specific Fertility Rate (SFR) and
Total Fertility Rate (TFR).

Unit 4 **(10)**

Measurement of Population Growth: Crude rates of natural increase, Pearl's Vital Index, Gross Reproduction Rate (GRR)
and Net Reproduction Rate (NRR).
Population Estimation, Projection and Forecasting: Use of A.P. and G.P. methods for population estimates, Use of
component method for population projection, Fitting of Logistic curve for population forecasting using Rhode's method.

Elective :III : BST 504 : Paper XII : Data Mining

Theory: 36 hours Marks-50 (Credit 02)

Unit-1:(10)

Data understanding and data cleaning, concept of supervised and unsupervised learning. Problem of classification, classification techniques: k-nearest neighbor, decision tree, Naïve Bayesian, classification based on logistic regression, Bayesian belief Network., CART(classification and regression trees)

Unit-2: (15)

Model evaluation and selection: Metrics for Evaluating Classifier Performance, Holdout Method and Random Subsampling, Cross-Validation, Bootstrap, Model Selection Using Statistical Tests of Significance, Comparing Classifiers Based on Cost–Benefit and ROC Curves.

Unit 3: (8)

Techniques to Improve Classification Accuracy: Introduction to Ensemble Methods, Bagging, Boosting and AdaBoost, Random Forests, Improving Classification Accuracy of Class-Imbalanced Data.

Unit-4: **(12)**

Unsupervised learning: Clustering: k-mediods, CLARA, DENCLUE, DBSCAN, Probabilistic model based clustering. Market Basket Analysis: Association rules and prediction, Apriori Algorithm, data attributes, applications to electronic commerce

Semester V : Practicals

Practical Paper- VII: BSP-506: Probability Distributions-I

1. Model sampling from Laplace and Cauchy distributions
2. Model sampling from Pareto distribution.
3. Model sampling from truncated binomial and poison distributions.
4. Model sampling from truncated normal and exponential distributions.
5. Fitting of Weibull distribution.
6. Fitting of truncated Binomial distribution.
7. Fitting of truncated Poisson distribution.
8. Applications of multinomial distribution.
9. Applications of Weibull, Laplace and Pareto distribution.

Practical Paper –VIII : BSP-507: Statistical Inference-I

1. Point estimation by method of moments for discrete distributions.
2. Point estimation by method of moment for continuous distributions.
3. Point estimation by method of maximum likelihood (one parameter).
4. Point estimation by method of maximum likelihood (two parameters).
5. Point estimation by method of minimum chi-square.
6. Interval estimation of location and scale parameters of normal distribution (single sample).
7. Interval estimation of difference of location and ratio of scale parameters of normal distribution (two samples).
8. Interval estimation for population proportion and difference between two population proportions.
9. Interval estimation for population median using order statistics.

PRACTICAL PAPER – IX : BSP-508: Operations Research

1. L.P.P. by simplex method I (Slack variable)
2. L.P.P. by simplex method II (Big M method)
3. Transformation problem-I.
4. Transformation problem-II. (Degeneracy)
5. Assignment problem.
6. Sequencing Problem.
7. Decision Theory.
8. Construction of CPM and PERT.
9. Minimal Path and cut.

PRACTICAL PAPER X : BSP-509: Design of Experiments

1. Analysis of CRD and RBD.
2. Analysis of Latin Square Design (LSD).
3. Missing Plot Technique for RBD and LSD with one missing observation.
4. Efficiency of i) RBD over CRD and ii) LSD over CRD and RBD.
5. Analysis of Split plot and strip plot Design.
6 Analysis of Covariance in CRD

7 Analysis of Covariance in RBD

8 Analysis of 22 and 23 Factorial Experiment

9. Total and partial Confounding

B. Sc. III Statistics: Semester VI

Paper XIV : BST 601 : Probability Distribution – II

Theory: 36 hours Marks-50 (Credit 02)

Course Outcomes: The students will acquire

a) knowledge about order statistics and associated distributions
b) concept of convergence and Chebychev'sinequality and its uses
c) concept of law large numbers and central limit theorem and its uses.
d) knowledge of terms involved in reliability theory as well as concepts and measures.

Unit-1: Convergence and Central Limit Theorem (10)

1.1: Convergence

Definition of convergence of sequence of random variables (a) in probability, (b) in distribution, (c) in quadratic mean.

If $X_n \xrightarrow{P}$ then $g(X)$

$g(X_n) \xrightarrow{P}$ where g is continuous function without proof.

i. Examples and problems.

1.2: Weak Law of Large Numbers and Central Limit Theorem

i. Weak law of large numbers (WLLN) statement and proof for i. i. d. random variables with finite variance.

ii. Central limit theorem: Statement and proof for i. i. d. random variables with finite variance, proof based on m. g. f..

iii. Simple examples based on Bernoulli, binomial, Poisson and chi-square distribution.

Unit-2: MARKOV CHAIN (12)

a) Definition of Stochastic process, state space ,parameter space, types of stochastic processes ,first order Markov property,Markov chains (MC), finite MC, time homogeneous M.C. onestep transition probabilities, and transition probability matrix (t.p.m.),stochastic matrix, Chapman Kolmogorov equation,nsteptransition probabilitiy matrix , n- step tpm of two stateM.C. and some typical t. p. m.initial distribution, finitedimensional distribution functions , partial sum (and functions)of independent and identically distributed random variables asMarkov chain, illustrations such as random walk,Gambler'sruin problem,

Unit-4: Queuing Theory **(11)**

i. Introduction, essential features of queuing system, input source, queue configuration, queue discipline, service mechanism.

16

ii. Operating characteristics of queuing system, transient- state and steady state, queue length, general relationship among system characteristics.

iii. Probability distribution in queuing system: Distribution of arrival, distribution of inter arrival time, distribution of departure and distribution of service time (Derivations are not expected).

iv.Types of queuing models:

v.Solution of queuing Model: M/M/1, using FCFS queue discipline. vi.Problems and examples.

Books Recommended

1. Cramer H.: Mathematical Methods of Statistics, Asia Publishing House, Mumbai.
3. Lindgren B. W.: Statistical Theory (Third Edition), Collier Macmillan International Edition, Macmillan Publishing Co. Inc. New York. .. . .
4. Hogg, R. V. and Craig A. T. : Introduction to Mathematical Statistics (Third Edition), Macmillan Publishing Company, Inc. 866, 34d Avenue, New York, 10022.
5. Sanjay Arora and BansiLal : New Mathematical Statistics (First Edition), Satya Prakashan, 16/17698, New Market, New Delhi, 5 (1989).
6. Gupta S. C and Kapoor V. K. : Fundamentals of Mathematical Statistics, Sultan Chand and Sons, 88, Daryaganj, New Delhi 2.
7. Rohatgi V. K.: An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd., New Delhi.
8. MedhiJ : Stochastic Processes. Wiley Eastern Ltd. New Delhi.
9. Hoel, Port and Stone: Introduction to Stochastic Processes, Houghton Mifflin.
10. Feller. W. : An Introduction of Probability Theory and its Applications.Wiley Eastern Ltd.. Mumbai.
12. Bhat B. R.: Modern Probability Theory.
13. Karlin and Taylor: Stochastic Process. 14.Ross S: Probability Theory.

15.Bhat B. R.: Stochastic Models : Analysis and Applications. New Age International. 16.Zacks S. : Introduction to Reliability Analysis, Probability Models and Statistical Methods, Springer Verlag.

17. Taha H. A.: Operation research – An Introduction, Fifth edition, Prentice Hall of India, New Delhi
18. Barlow R. E. and ProschanFrank : Statistical Theory of Reliability and Life Testing. Holt Rinebart and Winston Inc., New Yark.
19. Sinha S. K. : Reliability and Life Testing, Second Edition, Wiley Eastern Publishers, New Delhi.
20. Trivedi R. S. : Probability and Statistics with Reliability and Computer Science Application, Prentice – Hall of India Pvt. Ltd., New Delhi.

21. ParimalMukhopadhyaya : An Introduction to the Theory of Probability. World Scientific Publishing.

Paper XV : BST 602 : Statistical Inference – II

Theory: 36 hours Marks-50 (Credit 02)

Course Outcomes:

The students will acquire

a) concept of interval estimation.

b) knowledge of interval estimation of mean, variance and population proportion.

c) knowledge of important aspect of test of hypothesis and associated concept.

d) concept about parametric and non-parametric methods.

e) Knowledge of some important parametric as well as non–parametric tests.

Unit - 1: Interval Estimation (11)

1.1 : Notion of interval estimation, definition of confidence interval, length of confidence interval, confidence bounds. Definition of Pivotal quantity and its us in obtaining confidence intervals and bounds.

1.2 : Interval estimation for the following cases:

(i) Mean μ of normal distribution ($\sigma 2$ known and $\sigma 2$ unknown).

(ii) Variance $\sigma 2$ of normal distribution (μknown and μ unknown).

(iii) Difference between two means $\mu 1 - \mu 2$,

(a) for a sample from bivariate normal population,

(b) for samples from two independent normal populations.

(iv) Ratio of variances for samples from two independent normal populations.

(v) Mean of exponenential distribution.

(vi) Population proportion and difference of two population proportions of two independent large samples.

(vii) Population median using order statistics. Illustrative examples.

Unit - 2: Parametric Tests (13)

2.1 : Statistical hypothesis, problems of testing of hypothesis, definitions and illustrations of

(i) simple hypothesis (ii) composite hypothesis, critical region, type I and type II error, probabilities of type I & type II errors. Power of a test, p-value, size of a test, level of significance, problem of controlling probabilities of type I & type II errors.

2.2 : Definition of Most Powerful (MP) test. Statement and proof (sufficient part) of Neyman

- Pearson (NP) lemma for simple null hypothesis against simple alternative hypothesis for construction of MP test. Examples of construction of MP test of level α.

2.3 : Power function of a test, power curve, definition of uniformly most powerful (UMP) level α test. Use of NP lemma for constructing UMP level

Ehrenfest chain

b) Communicating states , first return probability
,probability of ever return Classification of states ,
as persistent and transient states , irreducible MC.

Unit 3: STOCHASTIC PROCESSES **(12)**

a) Definition of stochastic process.
b) Postulates and difference differential equations for :
i) Pure birth process
ii) Poisson process with initially 'a' members,
for a =0 and a >0
iii) Yule Furry process
iv) Pure death process
v) Death process with $\mu n=\mu$
vi) Death process with $\mu n=n\mu$
vii) Birth and death process

α test for one-sided alternative.Illustrative examples.

2.4 : Likelihood Ratio Test: Procedure of likelihood ratio test, statement of its properties, Likelihood Ratio test involving mean and variance of normal population.

Unit - 3: Sequential Tests **(9)**

General theory of sequential analysis and its comparison with fixed sample procedure. Wald's SPRT of strength (α, β), for simple null hypothesis against simple alternative hypothesis. Illustrations for standard distributions like binomial, Poisson, exponential and normal. Graphical and tabular procedure for carrying out the test. Illustrative examples.

Unit - 4: Non- parametric Test **(12)**

Notion of non-parametric statistical inference (test) and its comparison with parametric statistical inference.Concept of distribution free statistic. Test procedure of:

(i) Run test for one sample (i.e. test for randomness) and run test for two independent sample problems.

(ii) Sign test for one sample and two sample paired observations

(iii) Wilcoxon's signed rank test for one sample and two sample paired observations.

(iv) Mann-Whitney U - test (two independent samples)

(v) Median test

(vi) Kolmogorov Smirnov test for one and for two independent samples.

Books Recommended

1. Kale, B. K.: A first Course on Parametric Inference
2. Rohatgi, V. K.: Statistical Inference
3. Rohatgi, V. K.: An introduction to Probability Theory andMathematical Statistics
4. Saxena H. C. and Surenderan : Statistical Inference
5. Kendall M. G. and Stuart A.: An advanced Theory of Statistics
6. Lindgren, B. W.: Statistical Theory
7. Cassela G. and Berger R. L.: Statistical Inference
8. Lehmann, E. L: Testing of Statistical Hypothesis
9. Rao, C. R.: Linear Statistical Inference
10. Dudewicz C. J. and Mishra S. N. : Modern Mathematical Statistics
11. Fergusson, T. S.: Mathematical statistics.
12. Zacks, S.: Theory of Statistical Inference.
13. Cramer, H.: Mathematical Methods of Statistics.
14. Gibbons, J. D.: Non-parametric Statistical Inference.
15. Doniel: Applied Non-parametric Statistics
16. Siegel, S.: Non-parametric Methods for the behavioral sciences.
17. Kunte, S. ;Purophit, S. G. and Wanjale, S.K.: Lecture notes on Non-parametric Tests.

Course Outcomes:

The students shall get a) basic knowledge of complete enumeration and sample, sampling frame sampling distribution, sampling and non-sampling errors, principle steps in sample surveys, sample size determination, limitations of sampling etc.
b) concept of various sampling methods such as simple random sampling, stratified random sampling, systematic sampling and cluster sampling.
c) an idea of conducting sample surveys and selecting appropriate sampling techniques.
d) knowledge of comparing various sampling techniques. e) knowledge of ratio and regression estimators.

Unit – 1: Basic Terminology and Simple Random Sampling (15)

1.1: Basic Terminology

Concept of distinguishable elementary units, sampling units, sampling frame, random sampling and non-random sampling. Advantages of sampling method over census method, objectives of a sample survey, Designing a questionnaire, Characteristics of a good questionnaire, Concept of sampling and non-sampling errors. Handling of nonresponse cases.

1.2:Simple random sampling

i. Simple random sampling from finite population of size N with replacement (SRSWR) and without replacement (SRSWOR): Definitions, population mean and population total as parameters, inclusion probabilities.

ii. Sample mean y as an estimator of population mean, derivation of its expectation, standard error and estimator of standard error.

iii. $N\,y$ as an estimator of population total, derivation of its expectation, standard error and estimator of standard error.

iv. Sampling for dichotomous attributes. Estimation of population proportion Sample proportion (p) as an estimator of population proportion (P), derivation of its expectation, standard error and estimator of standard error using SRSWOR..Np as an estimator of total number of units in the population possessing the attribute of interest, derivation of its expectation, standard error and estimator of standard error. **1.3:Determination of the sample size.**

Determination of the sample size (n) for the given:

i.Margin of error and confidence coefficient.

ii.Coefficient of variation of the estimator and confidence coefficient.

Unit – 2: Stratified Sampling **(15)**

i. Real life situations where stratification can be used.

ii. Description of stratified sampling method where sample is drawn from individual stratum using SRSWOR method.

iii. (a) *st y* as an estimator of population mean $\bar{Y}$, derivation of its expectation, standard error and estimator of standard error.

(b) *st N y* as an estimator of population total, derivation of its expectation, standard error and estimator of standard error.

iv. Problem of allocation: Proportional allocation, Neyman's allocation and optimum allocation, derivation of the expressions for the standard errors of the above estimators when these allocations are used.

v. Comparison amongst SRSWOR, stratification with proportional allocation and stratification with optimum allocation.

vi. Cost and variance analysis in stratified random sampling, minimization of variance for fixed cost, minimization of cost for fixed variance, optimum allocation as a particular case of optimization in cost and variance analysis.

Unit-3 Other Sampling Methods **(10)**

3.1 : Systematic Sampling

i. Real life situations where systematic sampling is appropriate. Technique of drawing a sample using systematic sampling.

ii. Estimation of population mean and population total, standard error of these estimators.

iii. Comparison of systematic sampling with SRSWOR.

iv. Comparison of systematic sampling with SRSWOR and stratified sampling in the presence of linear trend.

v. Idea of Circular Systematic Sampling.

3.2 : Cluster Sampling

i. Real life situations where cluster sampling is appropriate. Technique of drawing a sample using cluster sampling.

ii. Estimation of population mean and population total (with equal sizeclusters), standard error of these estimators

iii. Systematic sampling as a particular case of cluster sampling.

3.3 Two Stage and Multi Stage Sampling

Idea of two-stage and multistage sampling.

Unit - 4: Sampling Methods using Auxiliary variables(5)

4.1: Ratio Method

i. Concept of auxiliary variable and its use in estimation

ii. Situations where Ratio method is appropriate.

iii. Ratio estimators of the population mean and population total and their standard errors (without derivations), estimators of these standard errors.

iv. Relative efficiency of ratio estimators with that of SRSWOR

4.2: Regression Method

i. Situations where Regression method is appropriate.

ii. Regression estimators of the population mean and population total and their standard errors (without derivations), estimators of these standard errors.

iii. Comments regarding bias in estimation

iv. Relative efficiency of regression estimators with that of

a) SRSWOR b) Ratio estimator.

Books Recommended

1. Cochran, W.G: Sampling Techniques, Wiley Eastern Ltd., New Delhi.
2. Sukhatme, P.V. and Sukhatme, B.V. : Sampling Theory of Surveys with Applications, Indian Society of Agricultural Statistics, New Delhi.
3. Des Raj : Sampling Theory.
4. Daroga Singh and Choudhary F.S.; Theory and Analysis of Sample Survey Designs, Wiley Eastern Ltd., New Delhi.
5. Murthy, M.N: Sampling Methods, Indian Statistical Institute, Kolkata.
6. Mukhopadhay, Parimal: Theory and Methods of Survey Sampling, Prentice Hall.

Paper XVI : BST 603 :Industrial Statistics

Theory: 36 hours Marks-50 (Credit 02)

Unit.1: Quality Tools (10)

Meaning and dimensions of quality, quality philosophy,Magnificent tools of quality: Histogram, Check sheet, Pareto diagram, cause and effect diagram, scatter diagram, control chart, flow chart. Deming's PDCA cycle for continuous improvements and its applications.

Unit 2: Process Control (12)

CUSUM chart, tabular form, use of these charts for monitoring process mean. Moving average and exponentially weighted moving average charts. Introduction to six-sigma methodology, DMAIC cycle and case studies.

Unit 3: Product Control (13)

Sampling Inspection plans for attribute inspection: Concept of AQL, LTPD, Consumer's risk, producer's risk, AOQ, AOQL, OC, ASN and ATI. Descriptionof Single and double sampling plans with determination of above constants.

Unit 4: Lean and Six Sigma (10)

Overview of Lean and Six Sigma with principles. Methodologies – Introduction to SCORE, DMAIC, Six Sigma Roles and Responsibilities. Tools used in Define Phase. Tools used in Measure Phase. Spaghetti diagram. Tools used in Analyze Phase. Various Statistical

Techniques used in analyze Phase (Revision), Tools used in Improve/Design Phase. Tools used in Control/Verify Phase.

Books Recommended

1. Introduction to quality Control – Montgomery D. C.
2. Quality Control and Industrial statistics Duncan A J
3. Statistical Quality Control by E L Grant

Elective I: Paper XVII :BST 604 :Sampling Theory

Theory: 36 hours Marks-50 (Credit 02)

Elective II: Paper XVII :BST 604 : Survey Sampling and Official Statistics

Theory: 36 hours Marks-50 (Credit 02)

Course Outcomes:

The students shall get a) basic knowledge of complete enumeration and sample, sampling frame sampling distribution, sampling and non-sampling errors, principle steps in sample surveys, sample size determination, limitations of sampling etc.
b) concept of various sampling methods such as simple random sampling, stratified random sampling, systematic sampling and cluster sampling.
c) an idea of conducting sample surveys and selecting appropriate sampling techniques.
d) knowledge of comparing various sampling techniques. e) knowledge of ratio and regression estimators.

Unit I: (10)

Concept of population and sample, complete enumeration verses sampling, sampling and non-sampling error, Types of sampling: Non-Probability and Probability sampling, basic principle of sample survey, Simple Random Sampling With and Without Replacement, definition and procedure of selecting a sample, estimates of: population mean, total and proportion, variances of these estimates, estimates of their variances and sample size determination.

Unit II: (10)

Stratified Random Sampling: Technique, estimates of population mean and total, variances of these estimates, proportional and optimum allocation and their comparison with SRS. Practical difficulties in allocation, estimation of gain in precision, post stratification and its performance. Systematic Sampling: Technique, estimates of population mean and total, variances of these estimates (N=n k). Comparison of systematic sampling with SRS and Stratified Sampling in the presence of linear trend and corrections

Unit III: (10)

Introduction to Ratio and Regression method of estimation, first approximation to the population mean and total(for SRS of large size), MSE of these estimates and estimates of these variances, MSE in term of correlation coefficient for regression method of estimation and their comparison with SRS. Cluster sampling (Equal clusters only) estimation of population mean and variance, comparison(with and

without randomly formed clusters). Concept of sub sampling, Two-stage sampling. Estimation of Population mean and variance of the estimate

Unit IV: **(15)**

An outline of present official statistical system in India, Methods of collection of official statistics, their reliability and limitation. Role of Ministry of Statistics and Problem Implementation (MoSPI), Central Statistical Office (CSO), National Sample Survey Office (NSSO), Registered General Office and National Statistical Commission, Government of India's Principal publications containing data on topics such as Agriculture, price, population, industry, finance and employment Consumer price index, Wholesale price Index number and Index of industrial production National Income: Basic idea and a brief description of income, expenditure and production approaches

Elective III:Paper XVII :BST 604 :Reliability Theory

Theory: 36 hours Marks-50 (Credit 02)

Unit-1: **(12)**

Structure function, dual of a structure, cuts and paths, components & systems, coherent systems, redundancy, Pivotal decomposition, Associated random variables and their properties. Birnbaum's measure of structural importance. Reliability concepts and measures, reliability of coherent systems, bounds on system reliability, Modular decomposition.

Unit-2: **(10)**

Life time distributions, survival functions, failure rate function, cumulative hazard function, residual life time, survival function of residual life time, mean residual life time, Computation of these functions for Common life time distributions: exponential, Weibull, Gamma, Makeham, Pareto, Rayleigh.

Unit-3: **(13)**

Notion of ageing: IFR, DFR, IFRA, DFRA, DMRL, NBU, NWU, NBUE, NWUE classes, ageing properties of common life time distributions, closure properties under formation of coherent structures, convolutions and mixtures of these classes. Damage model, cumulative damage model, univariate shock models and life distributions arising from shock models, bivariate exponential distribution.

Unit-4: **(10)**

Stochastic ordering: usual stochastic ordering, hazarad rate ordering,reverse hazard rate ordering, dispersive ordering, mean residual lifeordering and their implications. Availability, interval reliability, availability of a system with a single spare and a repair facility

B.Sc. III (Statistics) : Semester VI : Practicals

PRACTICAL PAPER XV :Sampling Methods

1. Simple Random Sampling for Variables.
2. Simple Random Sampling for Attributes.
3. Determination of Sample Size in SRS for Variables and Attributes.
4. Stratified Random Sampling – I
5. Stratified Random Sampling – II
6. Ratio Method of Estimation.
7. Regression Method of Estimation.
8. Systematic Sampling.
9. Cluster Sampling.

Practical XII: Probability Distributions and R – Software

1. Model sampling from Log-normal and Weibull distributions using R-Software.
2. Model sampling from Logistic distribution using R-Software.
3. Fitting of Binomial and Poisson distributionsusing R-Software.
4. Fitting of Normal distributionusing R-Software.
5. Fitting of Log-normal distribution using R-Software.
6. Analysis of Completely Randomized Design (CRD) using R.
7. Analysis of Randomized Block Design (RBD) using R.
8. Classification of TPM, States and computation of higher transition probabilities.
9. Applications of Queueing Systems.

Reference Books:

1. PurohitSudha : Lecture notes on R.
2. Verzani : Using R for introductory Statistics.

Practical –XIII: Statistical Inference-II

1. Construction of MP test.
2. Construction of UMP test.
3. Construction of SPRT for binomial, Poisson distributions, graphical representation of procedure.
4. Construction of SPRT for exponential and normal distribution, graphical representation of procedure.
5. NP test- -Run test (for one and two independent samples).
6. NP test –Sign test and Wilcoxon's signed rank test (for one and twosamples paired observation).

7 NP test-- Mann-whitney U- test (for two independent samples).

8. NP test –Median test (for two large independent samples)
9. NP test—Kolmogorov - smirnov test (for one and two independent samples).

PRACTICAL PAPER – XIV : Industrial Statistics

1. EWMA-Chart.
2. CUSUM chart.
3. Six sigma limits for mean.
4. Single sampling plan-I (Small sample).
5. Single sampling plan-II (Large sample).
6. Double sampling plan-I (Small sample.
7. Double sampling plan-II (Large sample).
8. k-nearest neighbor technique for classification. 9.k-means technique for clustering.