Progressive Education Society's

Modern College Of Arts, Science and Commerce, Ganeshkhind, Pune – 411 016

(NEP)

Syllabus for

M. Sc. Statistics Part II

Choice Based Credit System Syllabus

To be implemented from Academic Year 2024-25

Introduction:

Statistics helps us in measuring the uncertainty and planning the events by collecting a sample of information. Statistical tools are used to find patterns and insights from that data.

In fact, the statistical methods dominate scientific research as they include planning, designing, collecting data, analyzing, drawing meaningful interpretation and reporting of research findings.

Statistics can help you learn to get comfortable with interpreting numbers and sets of data.

We live in a world that depends on technology and data, so it's important that you know the theories behind statistics so that you can interpret data. Statistics is used in lots of different industries.

The scope of Statistics is very immense, the application of statistics goes into diverse fields such

as solving social problems, industrial and scientific problems. Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.

With the advancement and popularity of India in the business world, the demand for statisticians

is increasing day by day. As per the Bureau of Labor Statistics, a career as a statistician is considered to be one of the fastest-growing careers since 2018 and is expected to see a rise of nearly 33% by 2026 along with a rise in the jobs in this sector by 7.4%.

Statisticians practice the science of using a huge amount of data to make decisions, policies and

report a conclusion. They conduct research, design experiments, collect data, analyze and interpret data, and report conclusions. Statisticians use statistical analysis and apply it to a vast number of

fields or issues, like the environment, public safety, health care and sports.

In order to make a good career in statistics, the aspiring candidates have to pursue the undergraduate and postgraduate Statistics Courses.

The M.Sc. Statistics program is of two years which consists of core courses, which may be compulsory or electives. In addition, there are lab courses (practical) and a project course.

Some courses are termed Open Courses (O). The open courses are those which can be offered to other departments.

For every course, there will be Continuous Internal Assessment (CIA) conducted by the department of college and End of Semester Examination (ESE) conducted by the college at the end of semester.

Programme Outcomes:

PO1: Understand the concepts of mathematical statistics.

PO2: Acknowledge the use of statistics in a variety of domains

PO3: Create codes using different statistical software packages for statistical computing.

PO4: Recognize the significance of Statistics in Administration

PO5: Employ statistical concepts while forecasting in real time

Programme Specific Outcomes:

PSO1: Students learn different techniques used in Industries and research used for carrying analysis.

PSO2: Students will be well acquainted with various fields in statistical knowledge.

PSO3: Students learn the team work while completing the project work.

PSO4: Students get knowledge and training of technical subjects and get more employability in different industries

Structure of the Syllabus: **Following is the structure of two year M.Sc. Statistics program.**

T: Theory P: Practical O:Open C: Compulsory E: Elective

M.Sc.(Statistics)-Part II

Course Code	T/P	O/C/ E	Title	Credits	ESE Duration	Marks Assigned
Semester III						
STA63101	T	C	Probability theory	4	4 Hours	100
STA63102	T	C	Probability Distributions	4	4 Hours	100
STA63103	T	C	Statistical Inference	2	2 Hours	50
STA63104	P	C	Practical based on Probability theory, Probability Distributions, Statistical Inference	4	8 Hours	100
STA63205	T	E	Regression Analysis	2	2 Hours	50
STA63206	P	E	Introduction to SAS	2	4 Hours	50
STA63207	T	E	Statistical Quality Control	2	2 Hours	50

STA63208	P	E	Introduction to advanced R	2	2 Hours	50
STA63309	P	C	Research Project	**4**	4 Hours	**100**
Total				**22**		**550**
SEMESTER IV						
STA64101	T	C	Analysis of Clinical Trials	4	4 Hours	100
STA64102:	T	C	Survival Analysis	2	2 Hours	50
STA64103	T	C	Design of Experiment	2	2 Hours	50
STA64104	P	C	Project +practical based on Clinical trials, Survival Analysis, Design of Experiment	4	8 Hours	100
STA64205	T	E	Asymptotic Inference	2	2 Hours	50
STA64206	P	E	Introduction to SQL	2	4 Hours	50
STA64207	T	E	Advanced Operation Research	2	2 Hours	50
STA64208	P	E	Practical based on Advanced Operation Research and Computer intensive statistical methods	2	4 Hours	50
STA64309	P	C	Research Project	**6**	6Hours	**150**
Total				**22**		**550**

Semester III

STA63101: Probability theory **No. of Credits: 04**

Course Outcomes:-On completion of the course, the students will be able to:
CO1) Differentiate between various distributions with respect to their probability function and probability curve
CO2) Compute moments and hence shape of the distribution
CO3) Interrelations between the distributions
CO4) Real life situations where they can use the probability distributions as models

Unit1: Review [15 Hours]

- Algebra of sets, sequence of sets, limsup, liminf and limit of asequence of sets, Classes of sets,
- field, sigma field, minimal sigma field, Borel fields,
- measurable space, monotone classes, Measurable function,

- Real and Vector valued random variables, simple random variable, random variable as a limit of sequence of simple random variables,
- Probability measure on a measurable space, probability space, properties of probability measure: continuity, mixture of probability measures, Lebesgue and Lebesgue-Steltjes measures.

Unit2: CDF and Expectation [15 Hours]

- Distribution function,
- decomposition of a distribution function, discrete and continuous type random variable,
- Correspondence theorem,
- Expectation of simple random variable, non-negative random variable, arbitrary random variable, properties of expectation,
- moments, moment inequalities.

Unit3: Convergence [15 Hours]

- Convergence of a sequence of random variables,
- convergence in probability,
- convergence in distribution,
- convergence in rth mean, almost sure convergence,
- their inter-relations,
- Slutkey's Theorem, convergence theorem for expectations,

Unit4: Independence, Limit theorem [15 Hours]

- Independence of events, class of independent events, independence of classes, independence of random variables,
- expectation of the product of independent random variables, equivalent definitions of independence,
- Kolmogorov 0-1 Law, Borel 0-1 criterion, Khintchin's WLLN, Strong Law of Large Numbers (SLLN) (Statement only),
- Central Limit Theorem (CLT), Levy continuity theorem, CLT for i.i.d. random variables, Liaponove's form, Lindeberg Feller form and their applications.

Books Recommended:

Sr. No.	**Name of the Book**	**Author**	**Publisher**
1	Probability Theory	Athreya, K. B. and Lahiri S. (2006)	Hindustan BookAgency
2	Modern Probability Theory: An Introductory Text Book	Bhat, B. R. (2007)	New Age International
3	Probability and Measure, 3rdEdition,	Billingsley, P. (1995)	John Wiley, NewYork

4	Chung, K. L.	A Course in Probability Theory, Third Edition, (2001)	Academic Press, London
5	Probability: A Graduate Course	Gut, Allan (2005)	Springer, New

STA63102: Probability Distributions **No. of Credits: 04**

Course Outcomes: After completing this course, the students will be able to;

CO1) understand the basic concepts used to describe any probability distribution.

CO2) describe the expectation, moment generating function, probability generating function and characteristic function for some probability distributions.

CO3) find joint, marginal and conditional probability distributions for given random variables.

CO4) understand different multivariate probability distributions and non-central probability distributions. .

CO5) recall the concepts of order statistics and non-parametric tests, understand quadratic forms.

Unit1: **[15 Hours]**

- Random Variable: Brief introduction to sigma field, probability space, set function, measure, finite measure, probability measure, axioms of probability, Borel field, Borel measurable function, random variable as a measurable function on probability space.
- Distribution function: CDF of random variable, continuity theorem of limit of probability (without proof), characteristic properties of CDF (with proof). Decomposition of CDF, mixture of distributions. Identification of a given function whether it is CDF. Results: (i) product of distribution function is CDF (ii) If F(.) is a CDF then F^n (.) , $1-(1-F(.))^n$ are CDF for n positive integers. (iii) Convex combination of CDFs is CDF.
- Bivariate random variable. CDF of a bivariate random variable and its characteristic properties. Identification of a given function whether it is CDF.
- Symmetry: Symmetric probability distribution around a, Concept, mean median mode need not coincide. Results: (i) Odd ordered central moments of symmetric distribution are zero. (ii) Sum and difference of random variables with symmetric distribution is symmetric.
- Transformations:Transformation of random variables of the type (i) one to one on to (ii) monotonic (iii) non monotonic.

Unit2: **[15 Hours]**

- Expectation of random variables, necessity of existence of absolute moments, non uniqueness of moments, existence of r thorder moments if s thorder moment exists ($r \leq s$).
- MGF: Existence of MGF, properties, MGF of symmetric distributions.
- PGF.: Properties, moments using PGF, probability distributions of X+Y ,X-Y when X and Y are not identically distributed random variables. Compound distribution and its PGF. Wald's identity
- Results (i) If $P_1(s)$ and $P_2(s)$ are PGF of independent rv s then $P_1(s)$. $P_2(s)$ is a PGF (ii) $P_1(P_2(s))$ is PGF(iii)[$P_1(s)$]n is a PGF n positive integer.(iv) $P_x(s)= P_{-x}(s)$ if X is symmetric around zero.

- Convolutions of random variables .Distributions of X+Y, X-Y incase of U(0, 1), Normal, exponential etc.
- Characteristic function:characteristic function and properties, conjugate pairs of distributions, Parseval relation, uniqueness theorem
- Random vector:Joint and marginal distributions, mixed moments, covariance matrix, Conditional mean and variance, Results (i)E[E(X|Y)]=E(X) .ii)Var(X)=EVar(X|Y)+VarE(X|Y). (iii)If E(Y|x)=a +bx then b is the regression coefficient of y on x . MGF of random vectors.
- Regression: Conditional expectation as regression function

Unit3: [15 Hours]

- Bivariate Poisson Distribution: Definition, derivation of pmf, mgf of bivariate Poisson distribution, marginal pmf of variables, Cgf, correlation coefficient, conditional distribution of X1|X2=x2, problems.
- Bivariate exponential distribution:
  Types of bivariate exponential distribution:
  1) Marshall –Olkin model: joint distribution, marginal distribution, lack of memory property, Properties of Marshall –Olkin model.
  2) Fruend's bivariate exponential model.
  Theorem: Decomposition of bivariate exponential distribution into absolutely continuous and singular parts.
- Dirichlet Distribution: Definition, derivation of joint pdf ,particular case of dirichlet distribution, Marginal distribution of X and Y (for particular case k=3) .Conditional distribution of X|Y and Y|X.

Unit4: [15 Hours]

- Review of order Statistics:
  Definition, Marginal pdf of $X_{(r)}$,Joint pdf of ($X_{(r)}$, $X_{(s)}$),pdf of sample range, corr($X_{(r)}$, $X_{(s)}$) when random sample from U(0,1), problems based on exponential. Distributions of spacings and standardized spacings.
- Probability integral transformation. Quantile function.
- Distribution free statistics: Sign test, Wilcoxon sign rank test, Kolmogorov –Smirnov test, Run test.
- Quadratic forms: Quadratic forms, Classification of quadratic forms Sampling distribution of quadratic forms and linear forms for random sample from normal distribution, distribution of quadratic forms.All results required for Fisher Cochran theorem, Fisher Cochran theorem.
- Non-central distribution:Non-central chi square distribution, derivation of pdf , mgf , mean variance, applications.Non-central t-distribution distribution, derivation of pdf , mgf , mean variance, applications. Non-central F- distribution, derivation of pdf, mgf , mean variance, applications

Books Recommended:

Sr. No.	Name of the Book	Author	Publisher

1	Statistical Inference,	Berger, R. and Casella G. (2002), Second Edition	Duxbury Resource Center,
2	Fundamentals of Probability: A First Course	Dasgupta, A. (2010)	Springer, NewYork
3	Introduction to Mathematical Statistics,	Hogg, R. V. , McKean, J. W. and Craig, T. T. (2005), Sixth Edition	Pearson Prentice Hall, New Jersey.
4	Linear Statistical Inference and Its Applications	Rao, C. R. (2002)	Wiley
5	Introduction to Probability and Statistics	Rohatgi, V. K. & A. K. M. E Saleh (2001)	Wiley, New York

STA63103: Statistical inference **No. of Credits: 02**

Course Outcomes:-On completion of the course, the students will be able to:
CO1)understand the different families of distributions
CO2)understand the different families of distributions
CO3) estimates with optimal properties from a given sample with appropriate distributional assumptions
CO4) obtain tests and confidence intervals with some with optimal property.
CO5) understand the properties of MLE

Unit1: [09 Hours]

- Concept of Sufficiency
  Likelihood equivalence
  Minimal sufficiency, construction of minimal sufficient statistics
- Special classes of distributions
  - ★ One parameter exponential family
  - ★ Multiparameter exponential family
  - ★ Pitman family
  - ★ Minimal sufficient statistic for special classes of distributions.

Unit2: [09 Hours]

- Completeness, bounded completeness, complete sufficient statistics,
- Special classes of distributions admitting complete sufficient statistics,ancillary statistic
- Basu's theorem and its applications
- Estimability of parametric function
- Cramer-Rao inequality,minimum variance unbiased estimators (MVUE), necessary and sufficient conditions for existence of MVUE,Minimum variance bound unbiased estimators (MVBUE)

- Chapman-Robin Bounds (without proof)
- Bhattacharya Bounds (without proof)
- Rao- Blackwell theorem
- Lehmann- Scheffe theorem

Unit3: [08 Hours]

- Problem of testing of hypothesis
- Simple and composite hypotheses.
- Randomized and non-randomized tests
- Most powerful test
- Neyman-Pearson Lemma (with proof)
- Power function of a test
- Existence of UMP tests for
  - ★ one-sided alternatives in one parameter exponential family and Pitman family
  - ★ two sided alternatives, their existence and non-existence
  - ★ Unbiased test,UMP Unbiased test
- Monotone likelihood ratio property and its applications

Unit4: [4 Hours]

- Concept of confidence intervals
  - ★ uniformly most accurate confidence bounds.
- Introduction to Bayesian estimation
  - ★ prior and posterior distributions
  - ★ loss functions, principle of minimum expected posterior loss, quadratic and other common loss functions
  - ★ conjugate family of prior distributions and its examples.

Books Recommended:

Sr. No.	Name of the Book	Author	Publisher
1	Statistical Inference	Casella, G. and Berger, R. L. (2002)	Duxbury AdvancedSeries, Second Edition
2	Computer Age Statistical Inference: Algorithms,Evidence and Data Science.	Efron, B. and Hastie, T. (2016).	Cambridge University Press
3	Parametric Inference: An Introduction,	Kale, B.K. & Muralidharan, K. (2015)	Alpha Science International Ltd.
4	Theory of Point Estimation.	Lehmann, E.L. and Casella, G. (1998).	Springer, NewYork

5	Testing Statistical Hypotheses,	Lehmann, E. L. and Romano, J. (2005)	Springer
6	Linear Statistical Inference and its Applications,	Rao, C. R. (1995).	Wiley
7	. Introduction to Probability andStatistics	Rohatgi, V. K. and Saleh, A.K. Md. E. (2001)	John Wiley & Sons, New York
8	Mathematical Statistics, Springer-Verlag,	Shao, J. (2003).	New, New York

STA63104: Practical based on Probability theory, Probability distributions and Statistical inference **No. of Credits:04**

Sr No	Title	No of Practicals
1	Computation of Probabilities of events related to bivariate probability distribution for bivariate Normal Distribution.	
2	Computation of Probabilities of events related to bivariate probability distribution for bivariate Poisson Distribution.	
3	Computation of Probabilities of events related to bivariate probability distribution for bivariate Exponential Distribution.	1
4	Model sampling from Gamma ,Chi-Square ,Weibull and Lognormal Distribution.	2
5	Model Sampling from discrete ,continuous and mixture of distribution.	1
6	Model Sampling from bivariate probability Distributions.	1
7	Model Sampling from non - central chisquare ,t ,F- Distributions.	1
8	H-T Estimator and PPS ,π PS designs ,Confidence Interval of estimator	2
9	Minimal sufficient statistic for special classes of distributions	1
10	Construction of minimal sufficient statistic	1
11	Checking the validity of MVUE and MVBUE for some standard distributions on the basis of model sampling.	1

12	Randomized test and Non-randomized test	1
13	One parameter and multiparameter exponential family	1
14	UMP tests for one-sided alternatives in one parameter exponential family ,two sided alternatives, their existence and non-existence	2
15	Testing of hypotheses and construction of power function	2
16	Construction of confidence intervals uniformly most accurate confidence bounds.	1
17	Decomposition of a distribution function	2
18	Application of Central Limit theorem.and Weak law of large numbers.	2
19	Convergence of sequence of random variables, convergence in probability ,convergence in distribution.	1
20	Project (equivalent to 5 Practicals)	
	Total	**30**

STA63205: Regression Analysis No. of Credits:02Course Outcomes**:-**
On completion of the course, the students will be able to:
CO1) Recall the concept of fitting of simple regression models.
CO2) Compare residual diagnostics and apply corrective measures.
CO3)Analyze the multiple linear regression and logistic regression models.
CO4) Determine tests of hypothesis of model parameters, AIC and BIC criteria.

Unit 1 **[7 Hours]**

Confidence intervals and regions, testing of hypothesis for one and more than one linear parametric functions, testing of hypotheses about parallelism (slopes), equality of intercepts, congruence of two simple regression models, lack of fit test

(ii) Polynomial regression model (one and two regressors), orthogonal polynomial regression, cubic spline regression model

Unit 2 **[8 Hours]**

(i) Diagnostic checks and correction: graphical techniques, tests for normality (Shapiro test, Anderson-

Darling test), uncorrelatedness, homoscedasticity; Criteria for model adequacy: R2, adjusted R2,

Mallow's Cp etc.

(ii) Durbin Watson test, estimation of parameters in autocorrelation

(iii) Outlier, leverage points, influential points, PRESS statistic, Cook's D statistic

(iv)Multicollinearity : consequences, tools for detection and remedies,Ridge Regression.

Unit 3 [15 Hours]

(i) Non-linear regression: linearization transforms, their uses and limitations .Box and Cox transformations

(ii) Generalized linear model:introduction to link functions such as binomial, inverse binomial, inverse Gaussian and Gamma.

(iii) Logistic regression: Logit transform, ML estimation, tests of hypothesis, Wald test, LR test, score test, test for overall regression

(iv) Poisson regression: log link transform, ML estimation, tests of hypothesis, Wald test, LR test, score test, test for overall regression.

Books Recommended:

Sr. No.	Name of the Book	Author	Publisher
1	Applied Regression analysis	Draper, N. R. and Smith, H. (1998)	(John Wiley) Third Edition.
2	Introduction to Linear Regression	Montgomery, D. C., Peck, E. A. and Vining, G.G. (2003).	Wiley
3	Applied Linear Statistical Models	Neter, J., W., Kutner, M. H. ;Nachtsheim, C. J. and Wasserman, W. (1996),fourth edition.	Irwin USA
4	Regression Analysis by Example	Chatterjee S. and Hadi A.S.(2012),5thEdition	Wiley.
5	Logistic Regression	Kleinbaum G. and Klein M. (2011),Third Edition	A Self learning text, Springer.

STA63206: Introduction to SAS (Practical) No. of Credits:02

Unit1: [02 Hours]

- Installation of the software
- Introduction to SAS, Early history of SAS, Version history

Unit2: Terminologies: [04 Hours]

- SAS programs, tables, rows, columns / fields, SPSS, graphical interfaces, application programming interfaces, statements and procedures.
- 3 major parts of SAS: Data step, Procedure steps, and macro language.
- File structure, database oriented fourth generation programming languages (SQL focus), operating system, program loop.
- Data set, statistical analysis, macro code, imperative and procedural programming

SAS/IML component.

- Preprocessing runtime, general – purpose programming languages, information technology.
- PComponents:
  - ★ one-sided alternatives in one parameter exponential family and Pitman family
  - ★ two sided alternatives, their existence and non-existence
  - ★ SAS add in for Microsoft Office
  - ★ Base SAS
  - ★ SAS/GRAPH
  - ★ SAS/IML
  - ★ SAS/STAT
- Introduction to SAS Programs
- SAS Data types and Libraries
- Data and Proc Steps
- Format and In Format
- Creating Output Proc Print, Proc Contents
- Output Delivery System (ODS)

Unit3: BASE SAS : [05 Hours]

- Introduction:
  - ★ one-sided alternatives in one parameter exponential family and Pitman family
  - ★ two sided alternatives, their existence and non-existence
  - ★ SAS add in for Microsoft Office
  - ★ Base SAS
  - ★ SAS/GRAPH
  - ★ DATA step
  - ★ ODS
  - ★ SASr9
  - ★ Universal Printing
  - ★ XML engine
  - ★ Preproduction

Unit4: BASE SAS : [04 Hours]

- Introduction:
- Opening SAS enterprise miner
- Creating a new project in SAS enterprise.
- Miner 5.2
- The SAS Enterprise Miner Window.
- Creating a SAS Data Source
- Creating a process flow diagram

Unit5: Regression Models [04 Hours]

- Introduction:
- What types of models can be developed using
- The regression node
- An overview of some properties of the

- Regression node
- Business applications

Unit6: Comparison of Different models **[06 Hours]**

- Introduction:
- Models for Binary targets : an example of
- Predicting attrition models for ordinal targets: an example of predicting accident risk.
- Comparison of all three accident risk models
- The regression node
- An overview of some properties of the
- Regression node
- Business applications

Unit7: Customer Profitability **[05 Hours]**

- Introduction
- Acquisition cost
- Cost of default
- Revenue
- Profit
- The optimum cut –off point
- Alternative Scenarios of response and risk result
- Customer lifetime value
- Suggestions for extending result

Books Recommended:

Sr. No.	Name of the Book	Author	Publisher
1	Little SAS Book	A Primer, Sixth Edition	
2	Learning SAS by Example	A Programmer's Guide, Second Edition	
3	SAS Graphics for Clinical Trials by Example	Getting Started with SAS Programming: Using SAS Studio .	

STA63207: Statistical Quality Control **No. of Credits:02**

Course Outcome: On completion of this course, the students will be able to:

CO1)To use appropriate control charts, sampling plans. To draw inference about process capability.

CO2)Statistical methodology to get rid of defects and improve operational efficiency

CO3) Describe the DMAIC process (define, measure, analyze, improve, and control).

CO4) Perform analysis of process capability and measurement system capability.

Unit1: TQM Total quality Management [10 Hours]

- Meaning and dimensions of quality,
- Quality improvement, Quality Philosophy, Introduction to TQM, six sigma, DMAIC, and other extension of TQM, quality systems, The ISO 9000 and other Quality systems.
- (Control Chart : Revision of theory of control charts, Concepts of stable industrial processes, Systematic variation, random variation, variation within and between subgroups, estimation of process parameters,.
- Equivalence between control charts and testing of hypothesis problems. Choice of control limits Operating characteristic (O C curve) of control chart.
- Probability of false alarm, probability of catching shift in parameter. Concept of Run length , probability distribution of run length ,average run length (ARL).
- Comparison of control chart using ARL,OC curve, criteria of detecting lack of controls (sensitizing rules),patterns on control charts with justification and its effect on Probability of false alarm.

Unit2: Charts [10 Hours]

- X-bar MR chart
- CUSUM Chart :, Chart statistic (Ci^+ ,Ci^-)and chart parameters(k,h),
- construction and working of tabular CUSUM chart for mean and variance ,
- Statement of hypotheses. Estimation of shift in mean of process ,
- Fast initial response or head start feature,
- Siegmund's approximation for ARL and determination of chart parameters.
- CUSUM chart for subgroup size n>1,
- comparison between Shewhart chart and CUSUM chart V mask procedure.
- EWMA chart: Chart statistic, its expectation and variance. Choice of chart parameters (λ,L). Construction and working of EWMA chart for mean and variance.EWMA chart for subgroup size n>1,Comparison of Shewhart control charts with CUSUM charts. Simulation of ARL(δ).
- Process capability: Different Process capability and performance indicesCp,Cpk. Cpm.Propertis and relation between capability indices.Connection between proportion of defectives(DPPM) and Cp.Interval estimation of mean given Cpm $\geq$1.Estimation and confidence intervals of estimators of CpandCpk Testing of hypothesis about Cp .

Unit3: Other control charts and sampling plans [10 Hours]

- Synthetic control chart: Confirming run length (CRL) chart for attributes, Synthetic control chart , computations of chart parameters for given ARL(0),Steady state model Computations of ARL(δ), ATS(δ)., Comparison of with Shewhart control chart and CUSUM charts.
- Hotelling T2 Chart: Testing multivariate normality, Hotelling T^2multivariate control chart for mean vector when (i) dispersion matrix is (i) known (ii) unknown ARL(0), ARL(δ). T^2 control chart when subgroup size n=1
- Control chart for dispersion matrix when the mean vector is (i) known (ii) unknown.
- Multiple and sequential sampling plans for attributes sampling plan, Operating characteristic functions. AOQL, ATI, ASN.

Books Recommended:

Sr.No	Name of the book	Author	Publication
1	Introduction to Quality Control	Montgomery,D.C. (1985)	Wiley
2	Managing Total Quality	Logothetis,N.(1992)	Prentice Hall of India
3	Total Quality Management	Besterfield, D.H.Besterfield -Michana, c, Besterfield,G. H. Besterfield-Sacre,M(2001)	Pearson Education (Singapur)Pte.Ltd.India 2nd Edition
4	Total Quality Management	Oakland J.S.(1989)	Butterworth-Heinemann

STA63208 : Introduction to Advanced R **No of credits : 02**

Sr No	Title	No of Practicals
1	Introduction to R shiny:Setting up R shiny environment and creating projects	1
2	Building a basic UI: Adding UI elements such as buttons, sliders and text input	1
3	Creating server -side logic : Implementing reactive expressions and functions	2
4	Connecting to data source :Retrieving data from CSV,JSON or SQL databases	2
5	Designing interactive plots	2
6	Customization of UI	2
7	Developing own Web application from scratch	2
8	Project	3
	Total	**15**

Semester IV

STA64101: Analysis of Clinical Trials **No. of Credits:04**

Course Outcomes:-On completion of the course, the students will be able to:

CO1) Differentiate between various distributions with respect to their probability function and probability curve

CO2) Compute moments and hence shape of the distribution

CO3) Interrelations between the distributions

CO4) Real life situations where they can use the probability distributions as models

Unit1: Introduction **[12 Hours]**

- Introduction to clinical trials: need and ethics of clinical trials,
- bias and random error in clinical studies,
- conduct of clinical trials, overview of Phase I-IV trials,
- multicenter trials. Data management: data definitions,
- case report forms, database design, data collection systems for good clinical practice.
- Bioavailability, pharmacokinetics and pharmacodynamics, two-compartment model.

Unit2: Design of clinical trials **[12 Hours]**

- Design of clinical trials: parallel vs. cross-over designs, hybrid design, cross-sectional vs. longitudinal designs,
- response surface experiments and group allocation design,
- objectives and endpoints of clinical trials,
- design of Phase I trials, design of single-stage and multi-stage Phase II trials.
- Design and monitoring of Phase III trials with sequential stopping,
- design of bioequivalence trials.
- Inference for2x2 crossover design:
- Classical methods of interval hypothesis testing for bioequivalence,
- Bayesian methods, nonparametric methods.

Unit3: Analysis **[12 Hours]**

- Power and sample size determination, multiplicative (or log-transformed) model,
- ML method of estimation, assessment of inter and intra subject variabilities, detection of outlying subjects.
- Optimal crossover designs: Balaam's design, two-sequence dual design,
- optimal four-period designs,
- assessment of bioequivalence for more than two drugs, Williams design.

Unit4: Designs based on clinical endpoints: **[12 Hours]**

- Weighted least squares method, log-linear models, generalized estimating equations.
- Drug interaction study,
- Dose proportionality study and steady-state analysis, Interim analysis and group sequential tests, alpha spending functions,
- Analysis of categorical data

Books Recommended:

Sr. No.	Name of the Book	Author	Publisher
1	Design and Analysis of Bioavailability and bioequivalence.	Chow S.C. and Liu J.P.(2009), 3rd Edition	CRC Press.
2	Design and Analysis of Clinical Trials.	Chow S.C. and Liu J.P. (2004), 2nd Edition	Marcel Dekkar
3	The Design and Analysis of Clinical Experiments.	Fleiss J. L.(1989).	Wiley.
4	Fundamentals of Clinical Trials,	Friedman L. M.Furburg C. Demets D. L.(1998).	Springer.
5	Group Sequential Methods with Applications to Clinical Trails,	Jennison .C. and Turnbull B. W. (1999).	CRC Press.
6	Analyzing Survival Data from Clinical Trials and Observational Studies,	Marubeni .E. and Valsecchi M. G. (1994).	Wiley.

STA64102: Survival Analysis **No. of Credits:02**

Course Outcomes: On completion of this course, the students will be able to:

CO1) understand different types of censoring and to study the different models from Survival Analysis.

CO2) estimate and interpret survival characteristics .

CO3) compute parametric and non-parametric estimators of survival functions and probability density functions based on incomplete data.

CO4) estimate survival functions and probability density functions based on right- censored, truncated and interval censored data will be considered.

Unit1: Revision **[10 Hours]**

- Concepts of time, order and random censoring (left and right),
- survival function, density function, hazard function (rate),

- cumulative hazard rate, mean residual life function,
- Equilibrium distribution function
- Exponential distribution and its properties: No ageing, Lack of memory property, constant failure rate, Cauchy-function equation, constant mean residual life function, TTT transform, identity function as a TTT transform
- Revision of Ageing classes - IFR, IFRA, NBU, NBUE, DMRL, HNBUE and their duals, and inter relationship among these classes. Bathtub Failure rate.
- Life distributions - Exponential Gamma, Weibull, Lognormal, Pareto, linear Failure rate, Makeham family, Lehman families (proportional hazard rate family), spacing, normalized spacing and results of an exponential distribution based on normalized spacing.
- Revision of Parametric inference for complete data:
  - ★ Exponential distribution:
  - ★ Weibull: Obtaining MLE of scale and shape parameter of Weibull distribution and sample information matrix.
  - ★ Gamma: Obtaining MLE of scale and shape parameter of Gamma distribution and sample information matrix.
- Graphical method for checking exponentiality of data.

Unit2: Parametric inference for Censored data [10 Hours]

- Type I censoring: exponential distribution
- Type II censoring: exponential, gamma,lognormal
- Random censoring: exponential, Lehman family, Weibull distribution, Non-Parametric estimation of survival Function

  a) For complete data: Non parametric estimator of distribution function and survival function, distribution of empirical survival function, confidence band for survival function (by Using Kolmogorov - Smirnov statistics)

  b) For censored data: Actuarial estimator of survival Function, Estimator of variance of actuarial estimator (Greenwood's formula), product limit estimator and its variance, redistribution to right algorithm.

Unit3: Test for Exponentiality: [10 Hours]

- Estimable function of degree r, Kernel,
- symmetric Kernel,U- statistic, variance of U- Statistic, one sample
- U-Statistic theorem,
- Hollander and Proschan Test, Test for exponentiality against positive ageing based n sample spacing,
- Analytical test for exponentiality against NBUE, Deshpande's Test,
- Two sample U- statistic theorem,WilcoxonandMann- Whitneytest, Gehan's Test,Mantel-Haenszel Test, Log rank test, concept of covariates Semi-parametric regression for failure rate-Cox's proportional hazards model with one and several covariates. Baseline model, link function, likelihood function proportional Hazard Rate model
- Nelson-Aalen estimators, introduction to frailty models.

Books Recommended:

Sr. No.	Name of the Book	Author	Publisher
1	,LifeTime Data :Statistical Models and Methods	Deshpande, J.V, Purohit, S.G.,(2005)	
2	Survival Analysis: Techniques for censored and truncated data..	Klein J. P. and Moeschberger M.L. (1997)	Springer, New York
3	Modelling Survival Data in Medical research	Collett D (2003), 2nd Edition	Chapman and Hall/CRC
4	Analysis of Survival Data, Chapman and Hall,	Cox, D.R. and Oakes, D. (1984)	New York.
5	Survival models and Data Analysis,	Elandt-Johnson, R.E., Johnson N.L. (1980)	John Wiley andSons
6	Survival Distributions: Reliability Applications in the Biomedical Sciences,	Gross A.J. and Clark, V. A. (1975)	John Wiley andSons.
7	Survival Analysis,J	Miller, R.G. (1981)	John Wiley andSons
8	Modeling Survival data extending the Cox model.	Therneau T M and Grambsch P M (2000)	Springer, New York.
9	The Frailty model.	Duchateau L Johnson P (2008)	Springer, New York
10	Modeling Survival Data using frailty models	Hanagal D D (2011)	CRC press

STA64103:Design and Analysis of Experiments **No. of Credits:02**

Course Outcomes:

After completion of the course students will able to:

CO1) Understand the concept of BIBD, connectedness, balancedness and orthogonality of design.

CO2) Understand the difference between fixed and random effect models.

CO3) Compare the pairs of treatment means using different methods.

Construct Fractional factorial experiments and apply confounding in real life problems.

CO4) To use appropriate design for solving real life examples.
CO5) To learn the applications of different designs in agricultural experiments

Unit 1: **[06 Hours]**

One way classification with equal and unequal number of observations per cell, Lenege's test, Bartlet's test, Newman Keuls Test, Duncans Multiple Range Test (DMRT), Dunnet test, Non Parametric One way ANOVA (Kruskal Wallis Test), Friedman test (Non- parametric alternative to the one-way ANOVA with repeated measures), two way classification with equal number of observations per cell (with and without interaction), Missing plot techniques.BIBD intra block analysis, incidence matrix, symmetric BIBD, resolvable BIBD, (Results related to all types of BIBD)

Unit 2: **[08 Hours]**

Connectedness, balancedness and orthogonality of design, random effect models for one factor, estimation of variance components and confidence interval for intra class correlation coefficient, random effect model for the two factor, estimation of variance components. 2^k full factorial experiments: diagrammatic presentation of main effects, and first and second order interactions, model analysis using ANOVA, total confounding of 2^k design in 2^p blocks $p \geq 2$, partial confounding in 2^p blocks; p=2,3.

Unit 3: **[08 Hours]**

Fractional factorial experiments, resolution of a design (III, IV & V),aberration of a design, Plackett- Burman designs. 3^2 designs: contrasts for linear and quadratic effects, statistical analysis of 3^2 design, 3^3 designs: contrasts for linear and quadratic effects, statistical analysis of 3^3 design, blocking of 3^2 in three blocks, blocking of 3^3 in 9 blocks, fractional factorial experiment in 3 designs in = 2, 3.

Unit 4: **[06 Hours]**

Response surface methodology (RSM): linear and quadratic model, stationary point, canonical analysis, central composite designs(CCD), ridge systems, multiple responses, concept of rotatable designs, Box- Behnken design for 2 and 3 variables, blocking in Response surface design. Mixture experiments, Simplex lattice design and Simplex centroid design, Taguchi methods: concept of loss function, S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays.

Books Recommended

Sr. No.	Name of the Book	Author	Publisher
1	Experimental Design	Cochran W.G. and Cox, C.M. (1992)	John Wiley and Sons, Inc., New York.
2	Design and Analysis of Experiments	Dass, M.N. and Giri,N.C. (1987) 2nd Edition	Wiley Eastern Ltd., New Delhi
3	Experimental Design	Federer W.T. (1967)	Oxford and IBH Publishing Co., New Delhi
4	Fundamentals of Statistics, Vol.I	Goon, A.M., Gupta,M.K. and Dasgupta, B. (1998)	The World Press Pvt. Ltd. Kolkata
5	Design and Analysis of Experiments	Montgomery, D.C. (2012)	Wiley and Sons Inc., New Delhi.
6	Fundamentals of Applied Statistics	Gupta S.C. and Kapoor V.K.(2006).	S.Chand Sons, New Delhi
7	Statistical Methods	Snedecor, G.W. and Cochran, W.G. (1994), 8th Edition	Affiliated East – West Press, New Delhi

STA64104: Practical based on Analysis of Clinical Trials, Survival Analysis, and Design of experiment+Project No. of Credits:04

Sr No	Title	No of Practicals
1	Parametric analysis of complete data.and censored data	2
2	Computation of Actuarial estimator of survival function and PL – estimator and their variances.	1
3	Testing of hypotheses for various types of clinical trials, Power of the test and sample size determination of clinical trials.	2
4	Estimate survival function, cumulative survival function using Kaplan and Meier estimator and comparison of survival function using log rank test.	1
5	Hypothesis testing and estimation of confidences interval for bioequivalence study, nonparametric methods.	1
6	Estimation of the Pharmacokinetic parameters clinical trial	1

7	CRD	1
8	RBD	2
9	BIBD	2
10	Factorial experiment	1
11	Response surface methodology	1
12	Major Project (2 Credits)	15
Total		**30**

Project Guidelines

1. Projects may be done by individual students or in a group of students not more than 3.
2. As far as possible, students should use LaTeX for dissertation document preparation and paper presentation.
3. All the students should give their details in writing to the Project Coordinator/HOD just before the commencement of the semester or during the first week of semester. This should contain Names of the group members. Title of the project and name of the guide.
4. The project coordinator/ guide shall declare the dates of two rounds of internal presentations at the beginning of semester itself.
5. The project guide should meet his/ her group(s) at least once in a week and keep record of the meeting, attendance and the weekly progress of the project. Submit monthly progress reports to the HOD/ Project co-ordinator.
6. Students should try to use real data sets for their project problems. To the maximum possible extent, text book dada sets should be avoided. If possible, students should conduct an actual experiment and generate data (not by simulation) or contact some of the research organizations/ industry to get real data sets or real statistical problems they are trying to solve and be part of that.
7. Internal Evaluation: Guide will work as an internal examiner. There will be two presentation rounds for the Continuous Internal Assessment (CIA). These presentations will be graded out of 20 marks each (by Guide). Students are expected to submit the presentation to Guide at least one day before the presentation. It is advisable that both the presentations shall be attended by all the faculty members and students from other groups.
8. In the first presentation, students are expected to describe their project problem, the data they are going to analyze and the objectives of their project. In addition to this, they should also mention their methodology (without much detail). Students are to read at least TWO research papers which address similar kinds of problems and they should include the main contents of the papers in their first presentation.
9. In the second presentation, students should discuss the results of their analysis, finding and new methodology they have introduced (if any). Students should make sure that they have something innovative in their project work.
10. Assessment will also be done on the following aspects. Timely submission of the draft of the project report in the proper format which includes (title, abstract, key words, methodology,

conclusion, references, limitations and source of data etc.) is essential.
11. The completed project report in two copies (one for the candidate and other for examiner) should be submitted to the Project coordinator / HOD on or before the last day of teaching of the semester. HOD in consultation with the guide should ensure that whether the project is free from plagiarism and the project is worth presenting finally for end semester examination.
Aspects of Assessment and marks assigned (The following are the guidelines, some modification can be done as and when required by the examiners)
a. Dissertation in the proper pro forma which includes (Title, abstract, Key words, Methodology, conclusion, references, limitations and source of data etc.) (5 mks)
b. Appropriateness of tools used for analysis, testing the assumptions needed for analysis. methodology , program coding(if any) and numerical computations. (10 mks)
c. General understanding about the problem in the project (5 mks)
d. Presentation (5 mks)
e. Validity of conclusions. (5 mks)

Total: 30 marks
13. After the presentation within two days , students should incorporate all suggestions/ corrections suggested by the external examiner as well as the guide and submit the final copy to the department (within two days after the final presentation), failing which he /she will fail the course.

STA64205:Asymptotic Inference **No. of Credits:02**

Course Outcomes:
After completion of the course students will able to:

CO1) Understand the concept of consistency and asymptotic normality.

CO2) Understand method of moments and percentiles, maximum likelihood to Find consistent estimator and Cramer Huzurbazar theorem.

CO3) Apply likelihood ratio tests, Wald, Score and Bartlett's test in real life situations.

CO4) Compare various tests through relative asymptotic efficiency.

Unit 1: **[10 Hours]**

Consistency: real and vector parameters, Invariance under continuous transformation; Methods of obtaining consistent estimators: method of moments, method of percentiles, mean squared error criterion; Asymptotic relative efficiency, Comparison of consistent estimators, minimum sample size required by the estimator to attain certain level of accuracy, Asymptotic Normality; Consistent Asymptotic Normal (CAN)estimators: real and vector parameters; invariance of CAN property under non-vanishing differentiable transformation.Delta method, Methods of obtaining CAN estimators: method of moments and method of percentiles.

Unit-2 **[10 Hours]**

Maximum likelihood estimation, restricted parameter space, Inconsistent MLEs, MLEs in irregular cases. Asymptotic distribution of MLE in special class of distributions: Cramer regularity conditions, Cramer- Huzurbazar theorem, Extension to vector-valued parameters, Super-efficient estimators, BAN estimators, CAN and BAN estimation for multi-parameter exponential family and applications, Solution of likelihood equations, Method of scoring, Newton- Raphson and other iterative procedures.

Unit-3 **[10 Hours]**

Asymptotic theory of tests of hypotheses: Tests based on MLEs, Likelihood Ratio Test (LRT), asymptotic distribution of LRT statistic, Wald Test, Rao's score test, Pearson Chi-square test for goodness of fit, Bartlett's test for Homogeneity of variances, locally most powerful tests.Variance stabilizing transformations (VST): their existence, their applications in obtaining large sample tests and estimators Asymptotic Confidence Intervals: based on CAN estimators, based on VST. Asymptotic Confidence regions in multi-parameter families.

Books Recommended:

Sr. No.	Name of the Book	Author	Publisher
1	Asymptotic Theory of Statistics and Probability	Gupta Anirban Das (2008)	Springer, New York.
2	Parametric Inference: An Introduction	Kale, B.K. and K. Muralidharan (2015	Alpha Science Intl Ltd.
3	Linear Statistical Inference and its Applications	Rao, C.R. (1973)	Wiley, New York
4	Bayesian Statistics: An Introduction	Lee P.M. (2004)	Hodder Arnold, New York.
5	A course on Large Sample Theory	Ferguson, T.S. (1996)	Chapman and Hall, London.
6	Theory of Point Estimation	Lehmann, E.L. and Casella G. (1999)	Springer, New York.

7	Statistical Inference: Theory of Estimation	Manoj Kumar Srivastava, Abdul Hamid Khan and Namita Srivastava (2014)	PHI Learning Pvt. Ltd, Delhi.
8	Introduction to Bayesian Statistics 2nd Ed	Bolstad W. M. (2007)	Wiley, New York

STA64206: Introduction to SQL (Practical) **No. of Credits:02**

Course Outcomes:-On completion of the course, the students will be able to:

CO1)Understand the basic concepts and the applications of database systems.

CO2) Master the basics of SQL and construct queries using SQL.

CO3) Understand the relational database design principles.By understanding SQL, students able to quickly get up to speed with virtually any relational database system used today

CO4) Specific and measurable statements that define the knowledge, skills, and attitudes learners will demonstrate by the completion of a course.

Unit1: Introduction **[04 Hours]**

- Introduction to Basic Database Concepts :
- What is Data, Field, Record and database? ,Limitations of File Management System, Basic Concepts of Advantages of DBMS , Exploring Relational DBMS Understanding Client and Server Module .
- E-R Modeling and Diagram -Analyzing the Requirement ,Identify Entities and their Relationships ,Drawing E-R Diagram ,Conversion of E.R. Diagrams into Tables
- Normalization - First Normal Form , Second Normal Form ,Third Normal Form

Unit2:Introduction to SQL Server **[02 Hours]**

- What is SQL Server Version history and different editions
- Basic Starting and Stopping SQL
- Server Instances / Services Introduction to Management Studio Types of System Databases in SQL

Unit3: Working with Constraints **[03 Hours]**

- Create a constraint,types of constraints
- difference between unique, not null and primary key constraints ,
- Code on dataset with Unique , Not NULL ,Primary Key ,Check ,Default ,foreign

Unit4: DDL and DML Statements and apply on data sets **[03 hours]**

- Why write statements in Frontends?
- DDL Create, Alter and Drop ,Truncate ,Rename Table.
- Insert,Update and Delete Statement Truncate Statement.

Unit5: Working with Queries DQL, DCL ,TCL [09 Hours]

- DQL- Understanding Select Statement ,Use of Top, Distinct, Null etc...keywords
- Using String and Arithmetic Expressions ,Exploring Where Clause with Operators
- Using Advanced Operators Sorting data using Order By clauses
- Working with basic of Sub Queries
- DCL-Creating Users & Roles ,Granting & Revoking of Roles & privileges, Managing using Management Studio
- TCL- Introduction Transactions process ,Types of transactions (Implicit, explicit),Working with Locks, Types of locks .

Books Recommended:

Sr. No.	Name of the Book	Author	Publisher
1	SQL Programming	Brayn Jonson	Tyler MacDonald (25 December 2019)
2	SQL For Beginners: SQL Made Easy; A Step-By-Step Guide to SQL Programming for the Beginner, Intermediate and Advanced User (Including Projects and Exercises)	Craig Berg	Antony Mwau (26 September 2019)

STA64207: Advanced Operation Research No. of Credits:02

Course Outcome:

CO1 Solve linear programming problems using appropriate techniques and optimization solvers, interpret the results obtained.

CO2)Determine optimal strategy for Minimization of Cost of shipping of products from source to Destination/ Maximization of profits of shipping products using various methods, Finding initial basic feasible and optimal solution of the Transportation problems

CO3) Formulate Network models for service and manufacturing systems, and apply operations research techniques and algorithms to solve these Network problems

Unit1:Revision of LPP and Transportation problem : [5 Hours]

- Basic theorems (with Proofs) related to LPP

- Two Phase method, Duality in Linear programming, The Dual simplex method, Revised simplex method, Sensitivity Analysis.
- Ellipsoid method , polynomial time algorithm, Karmakar's polynomial time algorithm.
- Convergence and complexity.
- Revision of Transportation problem: Post optimality analysis in TP.
- Dual of the transportation problem, The trans-shipment problem, Assignment model ,The traveling salesman problem.

Unit2: [10 Hours]

- Integer linear programming problem:Gomory fractional All integer, Mixed IPP, Branch and Bound Method, cutting planes.
- Dynamic Programming problem: Formulation of Dynamic Programming problem, Optimal subdivision problem, Solution of LPP by Dynamic programming problem, Application of Dynamic programming problem.sensitivity analysis. Bellman's optimality principle , Stochastic dynamic programming.

- Nonlinear Programming Problem: Introduction , Local and Global optimum, Concave and Convex function, Kuhn – Tucker Condition, quadratic programming problem, Wolef's Method, Beales Method.

Unit3: [10 Hours]

- Queuing Theory: Introduction, Applications, Waiting time and idle time, classification of queuing models, Kendall's notation for queues various queue disciplines.
- Model: M/M/1:FCFS/ ∞/∞ Single channel Poisson arrival with exponential Service time, infinite population.Derivation of probability of queue length, system length, waiting time in queue and that in system.Little's formule (relation)
- Model: M/M/1:FCFS/ ∞/∞ Single channel Poisson arrival with exponential Service time, infinite population.Derivation of probability of queue length, system length, waiting time in queue and that in system.Little's formule (relation)

  M/M/1:FCFS/a/∞ihood equivalence

  M/M/1:FCFS/ a/b (a, b are finite).

  Arrival rate is function of queue length(n) :$\lambda n=\lambda/(n+1)$

  Service rate is proportional to queue length(n) :$\mu n=n\mu$
- Model: M/M/c: FCFS/ ∞/∞ Multi channel Poisson arrival with exponential Service time,Derivation of probability of queue length, system length, waiting time in queue and that in system
- Model: M/M/∞: FCFS/ ∞/∞ , expected system length and expected waiting time in system.Equivalence with Birth and Death process.

Unit4: [05 Hours]

- Simulation Technique: Introduction, definition , Advantages and disadvantages, Application of Simulation, Monte Carlo simulation, Generation of random number. Acceptance -Rejection method. Simulation from bivariate distributions
- Network analysis: Review of CPM, PERT , Network flow, maximal flow, Transportation problem as network, Transhipment Problem as Network, linear programming formulation.

Books Recommended

Sr.No.	Name of the book	Author	Publication
1	Operation Research	Taha. H.A.(1992) 5th edition	Macmillan Preamkumar Gupta and D.S.Hira
2	Nonlinear Programming	Bertsekas. D.(1999) 2nd edition	Athena Scientific
3	Linear Programming	Hadley .G(1987)	Addison –Wesley
4	Theory and applications	J.K.Sharma, 5th Edition	

STA64208: Practical based on Advanced Operation Research and Computer Intensive Statistical Method **No. of Credits:02**

Sr. No.	Title	No of Practicals
1	Integer programming ,Nonlinear programming	2
2	Dynamic programming	2
3	.CPM	1
4	PERT	1
5	Simulation	1
6	Simulation of M/M/1 queue.	1
7	Metropolis-Hastings and Gibbs Sampling algorithms. Particle Filtering, Rejection algorithms for Approximate Bayes Computation (ABC-Rejection)	2
8	Application of EM algorithm for incomplete data	2
9	EM algorithm for missing value	2
10	EM algorithm for mixture models, stochastic EM algorithm	1
11	Application of Bootstrap methods and Jackknife method	1
Total		15

****