

**SHRIMATHI DEVKUNVAR NANALAL BHATT VAISHNAV
COLLEGE FOR WOMEN (AUTONOMOUS)**

CHENNAI - 600044.

Re accredited with A+ Grade by NAAC

MASTER OF SCIENCE

(Shift – I)

Under the faculty of Science

PG DEPARTMENT OF BIOSTATISTICS



**CHOICE BASED CREDIT SYSTEM (CBCS)
OUTCOME BASED EDUCATION (OBE)**

(Effective from the Academic Year 2020-21)

PROGRAMME PROFILE
(PG Department of Biostatistics – Batch 2020)

TOTAL CREDITS: 91

TOTAL TEACHING HRS: 94

PART	COURSE	TITLE OF THE PAPER	CODE	L	T	H	C
I SEMESTER							
I	Core Theory -I	Probability and Distribution Theory	20PBSCT1001	4	1	5	4
I	Core Theory -II	Statistical Inference – I	20PBSCT1002	4	1	5	4
I	Core Theory -III	Population Studies	20PBSCT1003	4	1	5	4
I	Core Practical - I	Practical - I (Statistical Computing –I)	20PBSCP1001	4		4	4
I	Elective Theory - I	Statistical Genetics	20PBSET1001	4	2	6	3
II	Skill Based Elective	Teaching Skills	18PSSTS1001				3
II SEMESTER							
I	Core Theory -IV	Statistical Inference – II	20PBSCT2004	4	1	5	4
I	Core Theory -V	Applied Regression Analysis	20PBSCT2005	4	1	5	4
I	Core Theory -VI	Research Design and Analysis	20PBSCT2006	4	1	5	4
I	Core Practical – II	Practical - II (Statistical Computing - II)	20PBSCP2002	4		4	4
I	Elective Theory - II	Categorical Data Analysis	20PBSET2002	4	1	5	3
II	Soft Skill	Swayam (MOOC)	18MOOC2002				4
III SEMESTER							
I	Core Theory -VII	Applied Multivariate Analysis	20PBSCT3007	4	1	5	4
I	Core Theory -VIII	Sample Survey Designs	20PBSCT3008	4	1	5	4
I	Core Theory -IX	Survival Analysis	20PBSCT3009	4	1	5	4
I	Core Theory -X	Machine Learning Techniques	20PBSCT3010	4	1	5	4
I	Elective Practical - I	Exploratory Data Analysis using SPSS	20PBSEP3001	4		4	3
	Internship	Internship	20PBSIP3001				2
II	Skill based elective	Research Skills	18PSSRS3003				3
IV SEMESTER							
I	Core Theory -XI	Clinical Trials and Epidemiology	20PBSCT4011	4	1	5	4
I	Core Practical- III	Machine Learning using R	20PBSCP4003	4		4	4
I	Elective Theory - III	SAS Programming	20PBSET4003	4	2	6	3
I	Elective Theory - IV	Forecasting and decision making Techniques	20PBSET4004	4	2	6	3
I	Project I	Project*	20PBSPR4001	6			4
II	Soft Skill	SWAYAM (MOOC)	18MOOC4004				4

L =Lecture Hrs; T =Tutorial Hrs; H = Hrs per week; C =Credits

PROGRAMME PROFILE
(PG Department of Biostatistics – Batch 2021)

TOTAL CREDITS: 91

TOTAL TEACHING HRS: 94

PART	COURSE	TITLE OF THE PAPER	CODE	L	T	H	C
I SEMESTER							
I	Core Theory -I	Probability and Distribution Theory	20PBSCT1001	4	1	5	4
I	Core Theory -II	Statistical Inference – I	20PBSCT1002	4	1	5	4
I	Core Theory -III	Population Studies	20PBSCT1003	4	1	5	4
I	Core Practical - I	Core Practical - I (Statistical Computing –I Using Excel)		4		4	4
I	Elective Theory - I	Statistical Genetics	20PBSET1001	4	2	6	3
II	Skill Based Elective	Teaching Skills	18PSSTS1001				3
II SEMESTER							
I	Core Theory -IV	Statistical Inference – II	20PBSCT2004	4	1	5	4
I	Core Theory -V	Applied Regression Analysis		4	1	5	4
I	Core Theory -VI	Research Design and Analysis		4	1	5	4
I	Core Practical – II	Core Practical - II (Statistical Computing - II Using Python)		4		4	4
I	Elective Theory - II	Categorical Data Analysis	20PBSET2002	4	1	5	3
II	Soft Skill	Swayam (MOOC)	18MOOC2002				4
III SEMESTER							
I	Core Theory -VII	Applied Multivariate Analysis	20PBSCT3007	4	1	5	4
I	Core Theory -VIII	Sample Survey Designs	20PBSCT3008	4	1	5	4
I	Core Theory -IX	Survival Analysis	20PBSCT3009	4	1	5	4
I	Core Theory -X	Machine Learning Techniques	20PBSCT3010	4	1	5	4
I	Elective Practical - I	Exploratory Data Analysis using SPSS	20PBSEP3001	4		4	3
	Internship	Internship	20PBSIP3001				2
II	Skill based elective	Research Skills	18PSSRS3003				3
IV SEMESTER							
I	Core Theory -XI	Clinical Trials and Epidemiology	20PBSCT4011	4	1	5	4
I	Core Practical- III	Machine Learning using R	20PBSCP4003	4		4	4
I	Elective Theory - III	SAS Programming	20PBSET4003	4	2	6	3
I	Elective Theory - IV	Forecasting and decision making Techniques	20PBSET4004	4	2	6	3
I	Project I	Project*	20PBSPR4001	6			4
II	Soft Skill	SWAYAM (MOOC)	18MOOC4004				4

L =Lecture Hrs; T =Tutorial Hrs; H = Hrs per week; C =Credits

DEPARTMENT OF STATISTICS
SDNB VAISHNAV COLLEGE FOR WOMEN
(AUTONOMOUS) CHENNAI-600044.
(PG Department of Biostatistics)

COURSE FRAME WORK FOR THE BATCH 2021

SEMESTER I

SEM	COURSE CODE	COURSE TITLE	TITLE OF THE PAPER	HRS	CREDITS	CA	SE	T
I	20PBSCT1001	Core Paper-I	Probability and Distribution Theory	5	4	40	60	100
	20PBSCT1002	Core Paper-II	Statistical Inference - I	5	4	40	60	100
	20PBSCT1003	Core Paper-III	Population Studies	5	4	40	60	100
	20PBSCP1001	Core Practical – I	Core Practical - I (Statistical Computing - I)	3	4	40	60	100
	20PBSET1001	Elective Theory – I	Statistical Genetics	6	3	40	60	100
	18PSSTS1001	Skill Based Elective	Teaching Skills		3	50	---	100
				TOTAL		22	250	300

SEMESTER I

PROBABILITY AND DISTRIBUTION THEORY

TOTAL HOURS: 75

Credits: 4

SUB CODE: 20PBSCT1001

L-T-P: 4 – 1 -- 5

Course objectives:

1. The course explores the concepts of modern probability theory and its applications for decision-making in economics, business, and other fields of social sciences.
2. To understand the uncertain occurrence situations with logical manner.
3. The course is heavily oriented towards the formulation of mathematical concepts on probability and probability distributions and densities with practical applications.

Course outcome:

On successful completion of the course the students will be able to

CO	CO STATEMENT
CO1	Discuss about the fundamental concepts on probability of an event and explore the knowledge of Conditional probability and understand the concept of Baye's theorem and its applications. Apply problem-solving techniques to solving real-world events.
CO2	Recognize common probability distributions for discrete and continuous random variables. Understand the concept of convergence, central limit theorem, and large approximation, common methods for evaluating an inequalities performance and properties of desirable estimators.
CO3	Recall well known distributions such as Bernoulli, geometric, uniform, exponential, and normal, Cauchy, gamma and beta distributions.
CO4	Compare the Families of discrete distributions (binomial, Poisson, Multinomial) Continuous distribution for Exponential and its applications.
CO5	Calculate the probabilities relevant to multivariate distributions, including marginal and conditional probabilities and the covariance of two random variables.

SYLLABUS

UNIT 1:

Total number of hour:10

Probability: Sample space - Probability axioms - discrete and continuous probability space - conditional probability – Bayes’ theorem and its applications - screening tests, sensitivity, specificity, predictive value positive and negative - independence of events - simple problems.

UNIT 2:

Total number of hour:15

Random variables - Probability distribution of a random variable - discrete and continuous random variables - functions of random variable -multiple random variable-independent random variable, Convolution, covariance, correlation, moments, conditional expectation-Classical Central Limit theorem and its applications.

UNIT 3:

Total number of hour: 16

Univariate distributions: Negative binomial, Hyper-geometric, Normal, Exponential, Lognormal, Cauchy, Beta, Gamma, Laplace, Logarithmic, Pareto and Weibull (mean & variance only)-Compound and mixture of distributions relating to Binomial, Poisson and Normal distributions -Truncated distributions – Order statistics and their distributions.

UNIT 4:

Total number of hour:18

Bivariate discrete distributions: Binomial, Poisson, Multinomial distribution. (Mean, variance, covariance & correlation and applications of the above distributions). Bivariate Continuous distribution: Bivariate Exponential.

UNIT 5:

Total number of hour: 16

Bivariate Normal - Multivariate Normal Distributions – M.g.f - marginal and conditional distributions- independence of variables - linear transformation - distribution of sample mean vector.

TEXT BOOKS:

1. Daniel, W.W (2005): A foundation for analysis in the Health Sciences, John Wiley & Sons.(chapter 3 and 4)
2. Gupta, S.C and Kapoor, V.K. (2002): Fundamentals of Statistics, Sultan Chand & Sons Pvt. Ltd., New Delhi.
3. Hogg, R.V., Mckean, J. W. and Craig, A. T. (2006): Introduction to Mathematical Statistics, Sixth edition, Pearson Education. (for unit 5).
4. Meyer, P.L. (1970): Introductory probability and statistical Applications, 2ed, Oxford & IBH, New Delhi.
5. Rohatgi, V.K. and Saleh, A.K. (2001): An Introduction to Probability and Statistics, John Wiley & Sons. (Chapters 1,2 and 6 only)

BOOK FOR REFERENCES:

1. Casella, G. and Berger, R.L. (2002): Statistical Inference, Thomson Duxbury, Second edition.
2. Lehmann, E.L. and Casella, G. (2006): Theory of Point Estimation, Second edition, Springer Texts in Statistics
3. Rao, C.R. (2002): Linear Statistical Inference and its applications, Second edition, Wiley series in Probability and Statistics
4. Miller, I. and Miller, Marylees. John E. Freund's Mathematical Statistics with Application, 7th ed., New Jersey: Prentice Hall, 2010.
5. Wackerly, D, Mendenhall, W. and Scheaffer, R. Mathematical Statistics with Applications. Duxbury:California, United States, 2008.

E- LEARNING RESOURCES:

- <http://moodle.wolfware.ncsu.edu/>
- <http://www.e-booksdirectory.com/listing.php?category=15>
- <https://youtu.be/6x1pL9Yov1k>
- <https://youtu.be/UrOXRvG9oYE>
- <http://www.analyticsvidhya.com/blog/2016/02/free-read-books-statisticsmathematics-data-science/>

Note: Question Patterns based on theory part only.

MAPPING OF CO WITH PO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	2	2	3	2	2	1
CO2	2	2	2	2	2	1
CO3	2	3	3	3	3	2
CO4	2	2	2	2	2	2
CO5	2	3	3	3	2	2
AVG	2	2	3	2	2	2

PEDAGOGY

- Formal face-to-face lectures
- Chalk and Talk
- Problem solving and allow time for students to resolve problems in understanding of lecture material.

SEMESTER I
STATISTICAL INFERENCE – I

TOTALHOURS:75

SUB CODE: 20PBSCT1002

CREDIT:4

L-T-P: 4-1-5

Course objectives:

1. To learn the advance level basic concepts and statistical inference used in decision making which help them in their higher studies and solve involved decision making problems.
2. Be able to assist researchers for drawing inferences using their experimental outcomes.
3. The estimation approach can serve as an adjuvant toward better inference: it pushes back against over-confident claims from inadequate samples, improves comparisons of results across contexts, normalizes the publication of negligible effects, and provides a straight forward approach for planning informative studies.

Course outcomes: On completion of the course the students will be able to...

CO	CO Statement
CO1	Understand the basic concepts of Point estimation methods, including optimal properties of a point estimator.
CO2	Demonstrates to apply Sufficient Statistic, factorization criterion, exponential family , completeness and invariance.
CO3	Discuss the various concepts of obtaining estimators using estimation methods and its properties based on grouped data, Censored data and to explore the steps based on EM algorithm.
CO4	Understand the basic form of a confidence interval based on small and large samples, deals with confidence interval estimation.
CO5	Discuss Bayesian inference its Prior and posterior distributions and its rules, applications and estimation. Different types of loss functions and their applicability in different situations are discussed.

SYLLABUS

UNIT 1: **(Total number of hour: 12)**

Point Estimation: Optimal properties of a point estimator - consistency and unbiasedness. Minimum variance unbiased estimators - Cramer-Rao lower bound - Chapman-Robbins bound. Examples

UNIT 2 : **(Total number of hour:14)**

Sufficient Statistic - factorization criterion –One parameter exponential family (examples) - completeness - bounded completeness - invariance (definition only) -Rao-Blackwell theorem - Lehmann-Scheffe theorem

UNIT 3 : **(Total number of hour:15)**

Methods of estimation: Method of Maximum likelihood (Statement of properties only) - maximum likelihood estimation based on grouped data, Censored data (type I, II and random censoring - MLE for exponential distribution only) - Method of minimum chi-square and method of moments.

UNIT 4 : **(Total number of hour:15)**

Interval estimation – Based on small and large samples - sufficient statistic and confidence limits- fiducial interval - Confidence Interval for parameters of Normal distribution(s), Confidence Intervals based on large samples – examples.

UNIT 5: **(Total number of hour:12)**

Bayesian inference: Prior and posterior distributions, Baye’s rule, Baye’s estimator, conjugate prior, minimax estimator, admissibility rule, Bayesian Interval estimation

TEXT BOOKS:

1. Daniel, W. W (2005): A foundation for analysis in the Health Sciences, John Wiley & Sons.
2. Hogg, R. V., McKean, J. W. and Craig, A. T. (2006): Introduction to Mathematical Statistics, Sixth edition, Pearson Education.
3. Rohatgi, V.K. and Saleh, A.K.(2001):An Introduction to Probability and Statistics, John Wiley & Sons. (Chapter 8,unit 8.1 to 8.7)
4. Casella, G. and Berger, R.L. (2002): Statistical Inference, Thomson Duxbury, Second edition.
5. Rao, C.R. (2002): Linear Statistical Inference and its applications, Second edition, Wiley series in Probability and Statistics.

BOOKS FOR REFERENCE:

1. Lehmann, E.L. and Casella, G. (2006): Theory of Point Estimation, Second edition, Springer Texts in Statistics
2. Pagano, M. and Gauvreau, K. (2000): Principles of Biostatistics, Second Edition, Duxbury.
3. *Statistical Inference*. Second Edition. *George Casella*. University of Florida. *Roger L Berger*. 2002
4. Statistical Inference: Theory of Estimation – 1 January 2014 by Kumar S M (Author)
5. Statistical Inference – 1 November 2006 by George Casella (Author), Roger L. Berger (Author)

E-LEARNING RESOURCES:

- <https://corporatefinanceinstitute.com/resources/knowledge/other/point-estimators/>
- <http://www2.math.ou.edu/~cremling/teaching/lecturenotes/stat/ln4.pdf>
- http://www.maths.qmul.ac.uk/~bb/MS_NotesWeek10.pdf
- https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_confidence_intervals/BS704_Confidence_Intervals2.html
- <https://towardsdatascience.com/probability-concepts-explained-bayesian-inference-for-parameter-estimation-90e8930e5348>
- <https://statswithr.github.io/book/the-basics-of-bayesian-statistics.html>

Note: Question Patterns based on theory part only.

Mapping of CO with PSO:

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	2	2	3	2	3	2
CO2	2	2	2	2	2	2
CO3	3	3	3	3	3	3
CO4	3	2	3	3	3	3
CO5	2	3	2	3	2	3
Average	2	2	3	3	3	3

PEDAGOGY

- Chalk and Talk
- Power point presentation
- Group Discussions
- Seminars
- Problem solving to understand the concepts of theory

SEMESTER I
POPULATION STUDIES

TOTAL HOURS: 75

SUB CODE: 20PBSCT1003

CREDIT: 4

L-T-P: 4 – 1 -- 5

Course objectives:

1. Students will become familiar with basic concepts and sources of data in Demography and also will be able to comprehend the processes and events in Demography and their interactions.
2. The course also tries to discuss the various factors affecting population growth and its proximate determinants.
3. This course helps to understand all the mathematical procedures that measure population change and its underlying factors and help in visualizing the future prospects of population growth.

Course outcome: On successful completion of the course the students will be able to

CO	CO Statements
CO1	Discuss the various sources of data in Demography and the students would be able to comprehend the basic concepts and definitions.
CO2	Interpret the definitions in terms of fertility, mortality, Migration and construction of life table.
CO3	Analyze the Population Growth and fit the data using various models such as Arithmetic, Geometric, Exponential, Logistic.
CO4	Explain the Rates and Ratios – Person years lived, Crude and Specific Rates, Standardization – Direct and Indirect Methods : Components of Rates
CO5	Relate the components of population change-Fertility, Mortality and Migration, Causes and consequences of change in the population.

SYLLABUS

UNIT 1:

Total number of hour: 10

Introduction to Demography: Sources of Demographic data – completeness of vital registration – Chandrasekar Deming index - Nature, Scope and importance of demography – relationship with other disciplines. Analysis of age distribution - percent distribution – percent change by age – graphical representation of age data – population pyramid – sex ratio – aging of population – measures of aging of population.

UNIT 2:

Total number of hour: 15

Mortality and life tables: Crude and specific rates – infant mortality rate – standardized death rates – direct and indirect method of standardization. Life tables – constructions and uses – abridged life table – construction – Reed Merrell method – Greville’s method – Chiang’s method

UNIT 3:

Total number of hour: 15

Fertility: Crude and specific rates – General fertility rate – Total fertility rate – Age specific fertility rate - Gross reproduction rate – Net reproduction rate – parity progression ratio – child women ratio – fertility differential – determinants of fertility.

UNIT 4:

Total number of hour: 15

Migration: Definition and concepts – measures of migration – national growth rate method – residual method – balancing equation – survival ratio method – selectivity and differential – push and pull factors.

UNIT 5:

Total number of hour: 20

Population growth: Concept of stable and stationary population – measurement of population growth – arithmetic, geometric and exponential - population projection and estimation – different methods of projection – logistic curve – Pearl-Reed method – method of selected points – component method population projection .Urbanism and Urbanization – urban population growth – components of urban population growth.

TEXT BOOKS:

1. Samuel Preston , Patrick Heuveline and Michel Guillot (2000):Demography: Measuring and Modeling Population Processes
2. Goon A.M., Gupta M. K., Dasgupta B (2008): Fundamentals of Statistics, Published by Prentice Hall, 2nd edition.
3. Gupta S.C.& Kapoor V.K, (2000): Fundamentals of & Mathematical Statistics, Sultan Chand Sons 10th edition.
4. Pathak, K.B. and F.Ram, (1998): Techniques of Demographic Analysis, Mumbai, Himalaya Publishing House, Chapter 4
5. Jacob S. Siegel, David A. Swanson (2004): The methods and Materials of Demography, Elsevier Inc.

BOOK FOR REFERENCES:

1. Hinde, Andrew (1998), Demographic Methods, London: Edward Arnold, 1st edition.
2. Cox, P. (1959): Demography, Cambridge University Press, 2nd edition.
3. Keyfitz, (1985): Applied mathematical Demography, Springer-Verlag, New York, 2nd edition.
4. Shrivastava, O.S. (1995): Demography and population Studies, Vikas Publishing house private limited, 2nd edition.

E LEARNING RESOURCES:

- https://shodhganga.inflibnet.ac.in/bitstream/10603/169842/7/07_chapter%202.pdf
- <http://ocw.jhsph.edu/courses/PopulationChange/PDFs/Lecture4.pdf>
- http://www.chsj.org/uploads/1/0/2/1/10215849/booklet_2-_understanding_numbers-_populations_and_demography.pdf
- <https://www.khanacademy.org/science/biology/ecology/population-growth-and-regulation/a/exponential-logistic-growth>

Note: Question Patterns based on theory part only.

Mapping of CO with PSO:

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	2	2	2	2	2	1
CO2	2	3	3	3	1	1
CO3	2	3	3	2	1	1
CO4	2	2	3	2	2	1
CO5	2	1	1	1	1	1
Average	2	2	2	2	1	1

PEDAGOGY

- Formal face-to-face lectures
- Chalk and Talk
- Assignments
- Allow time for students to resolve problems in understanding of lecture material.

SEMESTER I
STATISTICAL COMPUTING –I
Using Excel

TOTAL HOURS: 45

SUB CODE:

CREDIT: 4

L-T-P: 3 - 0 - 3

Course objectives:

- Understand the role that statistical data analysis plays in managerial decision making under risk.
- Recognize the important characteristics, assumptions, terms and concepts of statistical data analysis.
- Understand the principles of inference.

Course outcome: On successful completion of the course the students will be able to

CO	CO Statement
CO1	Demonstrate knowledge of probability and the standard statistical distributions.
CO2	Demonstrate knowledge of fixed-sample and large-sample statistical properties of point and interval estimators.
CO3	Demonstrate knowledge of the properties of parametric and non parametric testing procedures.
CO4	Understand the measures of mortality, fertility and describe the concept of life tables.

SYLLABUS

1. Presentation of data – graphical and numerical
2. Calculation of simple, partial and multiple correlation coefficients (three variables only)
3. Fitting of Binomial, Poisson, Negative Binomial, Hyper geometric and normal distributions.
4. Estimation of parameters by the method of moments, maximum likelihood.
5. Confidence interval for proportion(s), mean(s), variance and ratio of variances.
6. Standard asymptotic and exact tests of significance with regard to proportion(s), mean(s), variance(s), coefficient of correlation(s), regression coefficient, one-way analysis of variance, Independence tests by contingency tables of order (p x q), (p, q <= 5), determination of sample sizes.
7. Rates and ratios, measures of fertility and mortality.
8. Construction of life tables and Abridged life table
9. Population estimation using logistic, Makhem and Gompertz curves.

MAPPING OF CO WITH PO:

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	3	1	1	1	1
CO2	2	2	2	1	3	1
CO3	3	2	2	1	2	1
CO4	3	2	2	2	1	1
Average	3	2	2	1	2	1

PEDAGOGY:

- Formal class room lectures(Chalk and Talk)
- Problem solving and allow time for students to resolve problems in understanding of lectures material.

SEMESTER I
STATISTICAL GENETICS

TOTAL HOURS: 90

SUB CODE: 20PBSET1001

CREDIT: 3

L-T-P: 6- 0 -- 6

Course objectives:

1. To enable students to understand the structure and function of genes.
2. To extend students abilities in understanding the role of genetic technologies applying in Statistics.
3. To present theory and techniques for the qualitative and quantitative analysis of statistical genetics.

Course outcome:

On successful completion of the course the students will be able to

CO	CO Statement
CO1	Equip the students with the information of various fundamentals of genetics.
CO2	Understand basic principles of Mendel's law of inheritance and to apply chi square tests using genetical problems.
CO3	Evaluating the principles to describe the genetics profile of populations as specified by Hardy-Weinberg law and apply the principles of Selection and breeding methods in plants and animals.
CO4	Analyzing the results of multiple alleles and gene inheritance to provide statistical solutions to specific problems in the field.
CO5	Obtain qualitative and quantitative traits by using various selection methods.

SYLLABUS

UNIT 1:

Total number of hour: 15

Elements of genetics – Physical basis of heredity – cell structure, chromosomes and genes –concept of genotype and phenotypes – linkage and crossing - over genetic maps

UNIT 2:

Total number of hour: 15

Mendel's law of inheritance - Laws of segregation and independent assortment – interaction of genes – concept of generation – Application of Chi-square tests

UNIT 3:

Total number of hour: 15

Sex – linked genes – concept of gene frequency, random mapping, Hardy Weinberg equilibrium, matrix theory of random mapping with applications – detection and estimation of linkage from back cross

UNIT 4:

Total number of hour: 15

Multiple allelic system -elementary aspects of the study of the human blood group – Multiple gene inheritance - statistical problems in human genetics- Steps in genetic algorithm

UNIT 5:

Total number of hour: 15

Quantitative Genetics: Qualitative Vs Quantitative traits- Polygenic traits-normal distribution-Types of gene actions-Heritability -Selection methods-mating methods

TEST BOOKS:

1. Arora, M.P and Sandhu, G.S. (1992): Genetics, Himalaya Publishing House, Delhi.
2. Elrod, S.L. (2006): Theory and Problems of Genetics, Fourth Edition, Tata McGraw Hill, New Delhi
3. Meyyan, R. (2002): Genetics and Biostatistics, First Edition, Sara's Publications India.
4. Ziegler, A. and Konig I.R (2006).: A Statistical Approach to Genetic Epidemiology, Wiley-VCH.
5. Nan M. Laird and ChristophLange (2011): The Fundamentals of Modern Statistical Genetics Springer-Verlag, New York.

BOOKS FOR REFERENCE:

1. Kowles, R. (2006): Solving Problems in Genetics, Springer (India), New Delhi.
2. Verma, P.S. and Agarwal, V.K.(1985): Genetics, S.Chand co Ltd.
3. Larose D. T (2005): Wiley 5th edition chapter 6, Data mining Methods and models.
4. Statistical Genetics of Quantitative Traits: Linkage, Map and QTL by Rongling Wu,Chang-Xiang Ma and George Casella, Springer-Verlag, New York.
5. Andreas Ziegler, Inke R. Koenig (2006): A Statistical Approach to Genetic Epidemiology: Concepts and Applications.

E- LEARNING RESOURCES

- <https://www.britannica.com/science/heredity-genetics/During-meiosis>
- https://en.wikipedia.org/wiki/Mendelian_inheritance
- <https://www.jove.com/science-education/10962/what-is-population-genetics>
- <https://courses.lumenlearning.com/wm-biology1/chapter/reading-multiple-alleles/>
- <http://eagri.org/eagri50/GBPR111/lec06.pdf>

Note: Question Patterns based on 75 % theory part and 25 % problem .

Mapping of CO with PSO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	2	2	1	1	1	2
CO2	2	2	2	2	1	2
CO3	3	3	3	3	2	2
CO4	3	3	3	3	3	3
CO5	3	3	3	3	3	3
AVG	3	3	2	2	2	2

PEDAGOGY

- Formal face-to-face lectures.
- Chalk and Talk
- Group Discussion.
- PowerPoint Presentation.

DEPARTMENT OF STATISTICS
SDNB VAISHNAV COLLEGE FOR WOMEN
(AUTONOMOUS) CHENNAI-600044.
(PG Department of Biostatistics)

COURSE FRAME WORK
SEMESTER-II

SEM	COURSE CODE	COURSE TITLE	TITLE OF THE PAPER	HRS	CREDITS	CA	SE	T
II	20PBSCT2004	Core Paper-IV	Statistical Inference - II	5	4	40	60	100
	20PBSCT2005	Core Paper-V	Applied Regression Analysis	5	4	40	60	100
	20PBSCT2006	Core Paper-VI	Research Design and Analysis	5	4	40	60	100
	20PBSCP2002	Core Practical – II	Core Practical - II (Statistical Computing - II)	5	4	40	60	100
	20PBSET2002	Elective Theory – II	Categorical Data Analysis	6	3	40	60	100
	18MOOC2002	Skill Based Elective	Swayam (MOOC)		4	50	----	100
				TOTAL		23	250	300

SEMESTER II
STATISTICAL INFERENCE-II

TOTALHOURS:75
CREDIT: 4

SUB CODE: 20PBSCT2004
L-T-P: 4 – 1 --5

Course objectives:

1. To understand the logical framework of tests of hypotheses.
2. To learn the types of error and non parametric test.
3. To perform test of hypothesis as well as obtain UMP,MP test.

Course outcome: On successful completion of the course the students will be able to

CO	CO STATEMENT
CO1	Formulate the null hypothesis and alternative hypothesis, level of significance, Type I error and Type II error and derive the Neyman Pearson Lemma.
CO2	Explain, Evaluate and interpret the concept of Most Powerful test and Uniformly Most Powerful test for the distributions with Monotonic Likelihood Ratio property, under Neyman Pearson Lemma.
CO3	Demonstrate the concepts of Unbiased test , Uniformly Most Powerful Unbiased Test and Locally most powerful (LMP) tests and able to apply the concept normal distribution.
CO4	Analyse the concept of Likelihood Ratio Test and Sequential Probability Ratio Test
CO5	Utilize the Non Parametric procedures population distribution known test.

SYLLABUS

UNIT 1:

Total no of hour: 18

Statistical hypotheses: formulation of hypotheses testing - Neyman -Pearson fundamental lemma: Most powerful tests - Uniformly most powerful (UMP) tests for distributions with monotone likelihood ratio ($H_0: \theta \leq \theta_0$ versus $H_A: \theta > \theta_0$ and its dual).

UNIT 2:

Total no of hour: 18

Generalized Neyman Pearson Lemma - UMP test of two sided hypotheses ($H_0: \theta \leq \theta_1$ or $\theta \geq \theta_2$ versus $H_A: \theta_1 < \theta < \theta_2$) for one parameter exponential family of distributions - problems relating to normal distribution only. Unbiased tests: definition, UMPU test for one parameter exponential family ($H_0: \theta = \theta_0$ versus $H_A: \theta \neq \theta_0$).

UNIT 3:

Total no of hour: 11

Similar tests: uniformly most powerful similar test for one parameter exponential family ($H_0: \theta_1 \leq \theta \leq \theta_2$ versus $H_A: \theta < \theta_1$ or $\theta > \theta_2$) – Locally most powerful (LMP) tests: definition, statement of the theorem for construction of LMP tests only, problems relating to normal distribution only.

UNIT 4:

Total no of hour: 18

Likelihood ratio tests (LRT): Definition, LRT for Binomial ,LRT for Normal distributions (one and two populations) only.

Sequential probability ratio test (SPRT): Definition, Operating Characteristic function and Average sample number, optimum property of SPRT (statement only).

UNIT 5:

Total no of hour: 10

Non-parametric tests: Measurement scales, Sign test, Wilcoxon Signed –Rank test, Wald – Wolfowitz runs test, Median test, Mann-Whitney test, Kolmogorov-Smirnov Test (one sample and two samples) - Kruskal-Wallis test - Friedman test.

TEXT BOOKS:

1. Conover, W. J. (2006): Practical Non-parametric Methods in Statistics, Second Edition, (Unit 5)
2. Daniel, W.W. (2006): Biostatistics: A foundation for analysis in the Health Sciences, 7th edition John Wiley & Sons. (Unit 5)
3. Rohatgi, V.K. and Saleh, A.K. (2015): An Introduction to Probability and Statistics, John Wiley & Sons. (Chapters 8 - 8.3, 9, 10- 10.1,10.2,10.6, 11 - 11.3)
4. M. Rajagopalan and P. Dhanavanthan (2012): Statistical Inference, Prentice Hall Of India
5. Rao, C.R. (2002): Linear Statistical Inference and its applications, Second edition, Wiley series in Probability and Statistics.

BOOKS FOR REFERENCES:

1. Lehmann, Erich L., Romano, Joseph P.(2005) Testing Statistical Hypotheses third edition, Springer-Verlag New York
2. Casella, G. and Berger, R.L. (2002): Statistical Inference, Thomson Duxbury, Second edition.
3. Statistical Inference: A Short Course Author(s): Michael J. Panik published:18 June 2012 Copyright © 2012 John Wiley & Sons, Inc.

E- LEARNING RESOURCES:

- <https://newonlinecourses.science.psu.edu/statprogram/reviews/statistical-concepts/hypothesis-testing/examples>
- <https://www.statisticssolutions.com/hypothesis-testing/>
- <https://stattrek.com/hypothesis-test/hypothesis-testing.aspx>
- <https://freevideolectures.com/course/3499/statistical-inference>
- <https://nptel.ac.in/courses/103106120/>

Note : Question Patterns based on theory part only .

MAPPING WITH CO & PCO

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	2	1	3	3	2
CO2	3	2	3	2	2	1
CO3	2	2	1	3	2	1
CO4	2	2	2	1	2	1
CO5	2	3	1	1	2	1
Average	2	2	2	2	2	1

PEDAGOGY

- Formal face-to-face lectures
- Chalk and Talk
- Experiential Learning
- Provide systematic feedback and corrections
- Appropriate use of interpersonal and teamwork skill
- Small-group cooperative learning (as one of the ways for teachers to incorporate active learning).

SEMESTER II

APPLIED REGRESSION ANALYSIS

TOTALHOURS: 75

CREDIT: 4

SUB CODE:

L-T-P: 4 – 1 -- 5

Course Objectives:

1. To understand linear and nonlinear relationships between variables and training the students in applications oriented.
2. To teach Linear Regression models, its assumptions and its properties.
3. To perform model adequacy check before using Linear Regression models

Course outcome: On successful completion of the course the students will be able to

CO	CO Statement
CO1	Discuss Simple Linear Regression and Multiple Linear Regression models. Understand Multiple Linear Regression models and how these can be constructed using real time data.
CO2	Obtain estimates of parameters of Linear Regression models and prove their properties. Perform various testing of hypotheses procedures relevant to Linear Regression model and significance of the model.
CO3	Apply model diagnostic checks and variance stabilizing transformations. Discuss the use of generalized and weighted least squares and that of dummy variables and their uses.
CO4	Describe Multi collinearity effect and its effect on model building. Apply methods for detecting Multi collinearity and influential observations.
CO5	Explain the concepts of Nonlinear regression and their linear transformation model

SYLLABUS

UNIT 1: **(Total number of hour:15)**

Simple linear regression-Assumptions, estimation of model parameters, standard error of estimators, testing of hypotheses on slope and intercept (β 's), interval estimation of model parameters, prediction interval of a new observation, coefficient of determination, regression through origin.

UNIT 2: Multiple linear regressions **(Total number of hour:15)**

Standard Gauss Markov setup, least square estimation of model parameters, variance covariance of least squares estimators, estimation of error variance

Tests of hypotheses – significance of regression (ANOVA, R^2 and adjusted R^2), individual regression coefficients, subset of regressor variables, general linear hypotheses- Confidence intervals and regions, prediction intervals, detecting hidden interpolation.

UNIT 3: **(Total number of hour:15)**

Model adequacy checking - residual plots for checking normality homoscedasticity and detection of outliers. Test for Lack of fit of the model. Durbin – Watson test for autocorrelation. Analytical methods for selecting a transformation generalized and weighted least squares- Detection of influential observations – Cooks statistic, DFFITS, DFBETAS.

Variance stabilizing transforms and transforms to linearize the model, analytical methods for selecting a transform, generalized and weighted least squares. Dummy (or indicator variables) – general concepts and their use.

UNIT 4: **(Total number of hour:15)**

Multicollinearity – sources, effects, diagnostics, Methods of dealing with multi collinearity (collection of additional data, model respecification, Ridge regression).

Selection of Variables – forward selection, backward elimination and stepwise regression (algorithms only)

UNIT 5: **(Total number of hour:15)**

Nonlinear regression – transformation to a linear model, their use and limitations, initial estimates (starting values), parameter estimation using iterative procedures – Gauss-Newton, steepest Descent, Marquardt's compromise.

Count data- Poisson Regression – variables selection- Non –parametric regression.

TEXT BOOKS :

1. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003): Introduction to Linear regression analysis, third edition, John Wiley and Sons, Inc.
2. Zar, J.H. (2006): Biostatistical Analysis, fourth edition, Pearson education.
3. Douglas C. Montgomery (2012) Introduction to Linear Regression Analysis.
4. Iain Pardoe (2012): Applied regression Modeling, second edition, Wiley

BOOKS FOR REFERENCE:

1. Draper, N.R. and Smith, H. (2003): Applied Regression Analysis, third edition, John Wiley and Sons, Inc.
2. Johnston, J. (1984): Econometric methods, third edition, McGraw Hill International.
3. A. Sen, M. Srivastava, Regression Analysis — Theory, Methods, and Applications, Springer-Verlag, Berlin, 2011.

E-LEARNING RESOURCES

- <http://home.iitk.ac.in/~shalab/regression/Chapter2-Regression-SimpleLinearRegressionAnalysis.pdf>
- <http://www.mit.edu/~6.s085/notes/lecture3.pdf>
- https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Nonlinear_Regression.pdf
- <https://data.princeton.edu/wws509/notes/c4.pdf>
- <http://home.iitk.ac.in/~shalab/regression/Chapter15-Regression-PoissonRegressionModels.pdf>

Note : Question Patterns based on theory part only .

Mapping of CO with PSO:

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	2	1	3	3	2
CO2	3	2	3	2	2	1
CO3	2	2	1	3	2	1
CO4	2	2	2	1	2	1
CO5	2	3	1	1	2	1
Average	2	2	2	2	2	1

PEDAGOGY

- Formal face-to-face lectures
- Chalk and Talk
- Problem solving and allow time for students to resolve problems in understanding of lecture material.

SEMESTER II

DESIGN OF EXPERIMENTS

TOTALHOURS: 75

CREDIT: 4

SUB CODE:

L-T-P: 4 – 1 – 5

Course objectives:

1. To learn, design and conduct experiments efficiently and effectively, and analyze the resulting data to obtain objective conclusions.
2. To equip the students with various types of designs that are used in practical life and to develop greater skills and understanding of analysis of these designs.
3. To familiarise basic concepts of Factorial experiments its analysis using ANOVA technique.

Course outcomes: On completion of the course the students will be able to...

CO	CO Statement
CO1	Understand the basic concepts of Design of Experiments
CO2	Demonstrates to apply design experiments using Randomized , Latin Square design efficiently and effectively.
CO3	Analyse on measurements to be taken in account (the response), what are the conditions to study and what experimental material to use (the units) to give valid generalization from the given data
CO4	Demonstrates the concepts of factorial experiments, Nested designs and Split- Split plot design experiment .
CO5	Understand the concepts of Response surface methodology

SYLLABUS

UNIT 1

Total number of hour:12

Single factor Analysis of variance-Two-factor ANOVA with unequal and equal replication (with/without interactions)- fixed, random and mixed effects model- Multiple comparison tests- Tukey, Newman-Keul, Scheffe test.

UNIT 2

Total number of hour: 15

Randomized block designs-repeated measures designs with and without blocking, Latin square design, BIBD, ANOCOVA (for CRD and RBD).

UNIT 3

Total number of hour: 18

General factorial design - blocking in a factorial design - 2^k Full factorial design: diagrammatic presentation of main effects and first order interactions, model, analysis of single as well as more than one replicates using ANOVA.

Total confounding of 2^k design in 2^p blocks, $p = 1,2$. Partial confounding. Fractional factorial design-the one-half fraction of the 2^k designs

UNIT 4

Total number of hour:15

Three-level factorial designs - 3^2 designs: contrasts for linear and quadratic effects, statistical analysis of 3^2 designs. Nested designs- two stages, Split-plot, Split- Split plot design experiment .

UNIT 5

Total number of hour:15

Response surface methodology - Experimental Designs for Fitting Response Surfaces- - Analysis of a Second-Order Response Surface for CRD.

TEXT BOOKS:

1. Das, M.N. and Giri N.C. (2006): Design and Analysis of Delhi. (For unit 5)
Experiments, Second Edition New Age International (P) Ltd., New
2. Montgomery D.C (2006): Design and Analysis of Experiments 5th ed., Wiley India Edition. (For Unit 1,2 3,4,5)
3. Zar, J.H. (2007): Biostatistical Analysis, fourth edition, Pearson Education. (for unit 1,2)
4. Designing Experiments and Analyzing Data: A Model Comparison Perspective, Third Edition Kindle (2017) .
5. Design and Analysis of Experiments by Douglas Montgomery ,Heath Rushing , Andrew Karl, James Wisnowski, 2014

BOOKS FOR REFERENCE:

1. Berger, P.D. and Maurer, R.E. (2002): Experimental designs, edition, S. Karger publications
2. Design of experiments statistical principles of research design and analysis Rober O Kuehl (2000)
3. Govindarajulu, Z. (2000): Statistical techniques in Bioassay, 2nd Thomson Duxbury.
4. D.C. Montgomery (2013) "Design and analysis of experiments". Eighth Edition. John Wiley and Sons, Singapore.
5. J. Lawson (2015). "Design and Analysis of Experiments with R", Boca Raton: CRC Press, Taylor & Francis Group.

E-LEARNING RESOURCES:

- <https://online.stat.psu.edu/stat503/lesson/3>
- http://compneurosci.com/wiki/images/9/98/Latin_square_Method.pdf
- <https://www.just.edu.jo/~haalshraideh/Courses/IE710/ch07.pdf>
- <https://home.iitk.ac.in/~shalab/anova/chapter8-anova-factorial-experiment.pdf>
- file:///C:/Users/stats/Downloads/Design_of_experiment_and_Response_surfac-1.pdf
- <https://medcraveonline.com/BBIJ/a-general-overview-of-response-surface-methodology.html>

Note : Question Patterns based on theory part only .

Mapping of CO with PSO:

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	2	2	3	2	3	2
CO2	2	2	2	2	2	2
CO3	3	3	3	3	3	3
CO4	3	2	3	3	3	3
CO5	2	3	2	3	2	3
Average	2	2	3	3	3	3

PEDAGOGY

- Chalk and Talk
- Power point presentation
- Group Discussions
- Seminars
- Problem solving to understand the concepts of theory

SEMESTER II
STATISTICAL COMPUTING –II
Using Python

TOTAL HOURS: 60

SUB CODE:

CREDIT: 4

L-T-P: 4 – 0 -- 4

Course objectives:

1. To provide hands on practical training on application of Design of experiment in real life problems.
2. To provide hands on training to estimate the inference on partial and multiple correlation coefficients and fitting regression models.
3. To teach different models for categorical data.

Course outcomes:

On successful completion of the course the students will be able to

CO	CO Statement
CO1	Apply Regression techniques and discusses on the tests relating to simple and multiple regression models in real life problems.
CO2	Discusses to plan, design and conduct experiments and to analyze the resulting data to obtain objective conclusions.
CO3	Apply appropriate tests of significance to a given problem and draw Inferences.
CO4	Demonstrate knowledge of statistical inference for Contingency tables.
CO5	Evaluate concepts of Interval estimation for odds ratio, log odds and relative risk

SYLLABUS

1. Diagnostic checks for model adequacy
2. Two-way classification with equal / unequal number of observations per cell (model with interaction). Multiple comparison tests.
3. Repeated measures ONE WAY ANOVA.
4. Analysis of covariance in one-way and two-way data.
5. 2ⁿ Factorial Experiments, Analysis of Single Replicate of 2ⁿ, (n=2, 3, 4).
5. 2 x 3 Factorial experiment, 2 x 3 x 3 Factorial experiment with and without blocking.
6. Total and partial confounding
7. Fractional Replication.
8. 3² Factorial Experiment
9. Split plot design
10. Power curve
11. Nonparametric tests
12. Chi-square test for Independence
13. Interval estimation for odds ratio, log odds, relative risk
Cochran Armitage trend test
14. Cochran Mantel Haenzel test
15. Tests relating to simple and multiple regression models

Mapping of CO with PSO:

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	3	1	1	1	1
CO2	2	2	2	1	3	1
CO3	3	2	2	1	2	1
CO4	3	2	2	2	1	1
CO5	3	2	2	1	1	1
Average	3	2	2	1	2	1

PEDAGOGY:

- Formal class room lectures(Chalk and Talk)
- Problem solving and allow time for students to resolve problems in understanding of lecture material.

SEMESTER II

CATEGORICAL DATA ANALYSIS

TOTALHOURS: 75

SUBCODE: 20PBSET2002

CREDIT:

L-T-P: 4 – 1 --5

Course objectives:

1. To develop students abilities in understanding the concepts of categorical data.
2. To present theory and techniques for the analysis of categorical data.
3. To enable students to learn how to choose appropriate techniques, to analyse categorical data, and present results.

Course outcome: On successful completion of the course the students will be able to

CO	CO Statement
CO1	Identify and summaries categorical data into $2 \times 2 / I \times J$ contingency tables.
CO2	Apply and interpret ODDS ratio, RR and perform tests for independent of attributes in contingency tables using various methods.
CO3	Demonstrate skills in using standard analytic methods for single and two-way classification data, including the use of Poisson and multinomial models for data, odds ratios, and Pearson's chi-squared and likelihood ratio statistics.
CO4	Identify the need for, the structure, and the usefulness of, generalised linear models. Fit and interpret the results of fitting, generalised linear models, including log-linear models (for example trend models) and logistic regression models.
CO5	Obtain and fit good models in three-way situations using a Various Matched-Pair Models.

SYLLABUS

UNIT 1: **(Total number of hour: 15)**

Nature of Categorical data - Contingency tables and their distribution: Binomial and Multinomial sampling. Comparing proportions in two-by-two tables: Difference of proportions - Relative risk - Odds Ratio - Properties of Odds Ratio - relationship between Odds Ratio and Relative Risk - Odds ratio in I x J tables -Nominal and Ordinal Measures of Association. Inference for Contingency tables: Interval estimation for difference of proportions, odds ratio, log odds ratio and relative risk. Testing Independence in Two-Way tables: Pearson and Likelihood-ratio chi-square tests - Yate's correction for continuity-Residuals for cells in a contingency table-Partitioning chi-squared.

UNIT 2: **(Total number of hour: 15)**

Testing Independence for Ordinal Data-Trend tests for 2 x J tables-Exact Inference for small samples - Fisher Exact Test for 2 x 2 tables. Association in Three-Way Tables: Partial Tables - Conditional and Marginal Odds Ratios - Homogeneous Association - Cochran-Mantel-Haenszel methods.

UNIT 3: **(Total number of hour:15)**

Components of Generalized Linear Models-Logit models for Binary data-Binomial GLM for 2 x 2 contingency tables
Logistic regression: parameter interpretation - maximum likelihood estimate - test of overall regression and goodness of fit - Wald test, deviance statistic, LR test, score test-Logistic regression diagnostics.

UNIT 4: **(Total number of hour: 15)**

Logit model with Categorical predictors: Linear logit model for I x 2 tables - Cochran-Armitage Trend test. Logit models for multiway contingency tables.
Multinomial Logit models: Logit models for Nominal responses- Base-line category logits-estimating response probability
Logit models for ordinal responses: Cumulative logits - proportional odds model.Probit models – complementary log-log link model (description only).

UNIT 5: **(Total number of hour: 15)**

Poisson log-linear model for count data- ML estimation, goodness of fit test. Loglinear Models for Two-way Tables: Independence Model-Interpretation of parameters- Poisson regression for rates. Models for matched pairs: Comparing dependent proportions-McNemar's test, small sample test for comparing matched proportions.
Measures of agreement-Kappa measure - Weighted Kappa - Bradley-Terry model for paired preferences. Generalized estimating equation methodology (basic idea).

TEXT BOOKS:

1. Agresti, A. (2007). An Introduction to Categorical Data Analysis. 2nd edn. New York: Wiley.
2. Agresti, A. (2010). Analysis of Ordinal Categorical Data. 2nd edn. New York: Wiley.
3. Powers, D.A. and Xie, Y. (2000). Statistical Methods for Categorical Data Analysis. San Diego, CA: Academic Press
4. Stokes, M.E., Davis, C.S. and Koch, G.G. (2000). Categorical Data Analysis Using the SAS System, 2nd edn. Cary, NC: SAS Institute Inc.
5. Hardin, J.W., and Hilbe, J.M. (1994): Generalized Estimating Equation, Chapman and Hall, London, 2nd edition.

BOOK FOR REFERENCES:

1. Agresti, A. (1991): An Introduction to Categorical data analysis, John Wiley & Sons, 2nd edition.
2. Deshpande, J.V., Gore, A.P. and Shanubhogue, A. (1995): Statistical Analysis of Non Normal Data, New Age International Publishers Ltd., New Delhi, 1st edition.
3. Radhakrishna Rao, "Linear Statistical Inference and its Applications" Wiley-Interscience, 2nd | 2001 | ISBN: 0471218758.
4. Leonard, T. (2000). A Course in Categorical Data Analysis. Boca Raton, FL: Chapman and Hall/CRC.
5. McCullagh, P. and Nelder, J.A. (1989). Generalized Linear Models, 2nd edn., London: Chapman and Hall.

E LEARNINGS RESOURCES:

- <http://www.stat.ufl.edu>>categorical data analysis.
- <http://www.stat.yale.edu>>courses>catdat
- <http://www.statgraphics.com>>categorical-data-analysis.
- <http://www.johndcook.com/blug/2018/04/14/>> categorical-data-analysis.
- https://in.sagepub.com/sites/default/files/upm-assets/67534_book_item_67534.pdf

Note : Question Patterns based on theory part only .

Mapping of CO with PSO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	2	1	1	1	3
CO2	2	3	3	2	2	2
CO3	3	3	2	3	3	2
CO4	3	3	3	3	3	2
CO5	2	2	2	3	3	2
AVG	2	2	2	3	3	2

PEDAGOGY

- Formal face-to-face lectures
- Chalk and Talk
- Work with theoretical analysis of the methods applied to a field of study
- PowerPoint Presentation for data analysis.

DEPARTMENT OF STATISTICS
SDNB VAISHNAV COLLEGE FOR WOMEN
(AUTONOMOUS) CHENNAI-600044.
(PG Department of Biostatistics)

COURSE FRAME WORK

SEMESTER III

SEM	COURSE CODE	COURSE TITLE	TITLE OF THE PAPER	HRS	CREDITS	CA	SE	T
III		Core Major-VII	Applied Multivariate Analysis	6	4	25	75	100
		Core Major-VIII	Sample Survey Designs	6	4	25	75	100
		Core Major-IX	Survival Analysis	6	4	25	75	100
		Core Major-X	Machine Learning Techniques	6	4	25	75	100
		Elective Practical	Exploratory Data Analysis using SPSS	4	3	40	60	100
		Internship	Internship		2	50	----	100
		Skill based elective	Research Skills		3			
TOTAL					24	190	360	600

SEMESTER III

APPLIED MULTIVARIATE ANALYSIS

TOTALHOURS: 75

SUB CODE:

CREDIT: 4

L-T-P: 4 – 1 -- 5

Course objectives:

1. To classify the extensions of Univariate to the multivariate statistical techniques used in data analysis.
2. To learn and develop scientific view to deal with multidimensional datasets and its uses in the analysis of research data.
3. To be able to carry out multivariate statistical techniques and methods efficiently and effectively.

Course outcome: On successful completion of the course the students will be able to

CO	CO Statement
CO1	Discussed about multivariate normal distribution and their real life applications. Understand the concept of Wishart distribution, Hotelling T^2 and Mahalanobis D^2 statistic.
CO2	Methods of Multivariate analysis: Discuss about the basic concepts of Principal Component Analysis (PCA) and Factor Analysis (FA) and its applications. Explain the component terms in PCA and prove its properties.
CO3	Explain the concepts of Canonical correlation and how it's differ from PCA and FA. Discuss the concepts of correspondence analysis and its applications. Explain the components of double weighting, weight of the profile, standardization method.
CO4	Understand the fundamental concepts of Discriminant and classification analysis and its applications for solving problem of real-world events.
CO5	Discuss the concept of cluster analysis and explain its various types of measures, types of clusters and how its work on real data field.

SYLLABUS

UNIT 1:

Total number of hour: 15

Hotelling's T2 statistic - applications of T2 statistic - single sample, two samples, paired comparisons - confidence region and simultaneous comparison of component means. One-way and two-way Multivariate Analysis of Variance. Testing of hypothesis for single covariance matrix, testing the equality of covariance matrices.

UNIT 2:

Total number of hour:20

Principal component analysis: Extraction of components - properties and characteristics of components - total variation, relative importance, standardization of variables and components, special covariance structures - interpretation of principal components.

Factor analysis: orthogonal factor model-parameter estimation by principal component and maximum likelihood method-factor rotation - factor scores.

UNIT 3:

Total number of hour: 15

Canonical correlation analysis: Extraction of canonical correlations and their variable - testing the significance of canonical correlation - interpretation of canonical variables.

Correspondence Analysis: Introduction- Attribute independency and contingency table, Graphic representation: file profiles or column profiles, Relationship with main components: double weighting, weight of the profiles, Profiles standardization, Distributional difference between profiles, Concept and interpretation.

UNIT 4:

Total number of hour:20

Discrimination and classification: Classification problem - standards of good classification - procedures of classification into one of two populations with known probability distributions - evaluation of classification function - Fisher's linear discriminant function.

UNIT 5:

Total number of hour:15

Cluster analysis: Distance and similarity measures - hierarchical clustering techniques - agglomerative techniques - single linkage, complete linkage, average linkage methods - nonhierarchical clustering methods-K means method.

TEXT BOOKS:

1. Izenman, Alan Julian (2008), Modern Multivariate Statistical Techniques, Springer Texts in Statistics.
2. Mardia, K. V., Kent, J. T., Bibby, J. M. (1979) Multivariate Analysis. Academic Press.
3. Johnson, R.A. and Wichern, D.W. (2002): Applied Multivariate Statistical Analysis, 5th edition, Pearson Education, Asia.

BOOK FOR REFERENCES:

1. Anderson, T. W. (2003): An Introduction to Multivariate Statistical Analysis, 3rd edition, John Wiley & Sons.
2. Hair, J.F., Anderson, R.E., Tatham. R.L. and Black, W.C. (2006): Multivariate Data Analysis, 5th edition, Pearson Education, Asia.
3. Anderson, T.W. 2009. An Introduction to Multivariate Statistical Analysis, 3rd Edition, John Wiley.
4. Everitt B, Hothorn T, 2011. An Introduction to Applied Multivariate Analysis with R, Springer.
5. Barry J. Babin, Hair, Rolph E Anderson, and William C. Blac, 2013, Multivariate Data Analysis, Pearson New International Edition.

E – LEARNING RESOURCES:

- <https://www.classcentral.com/course/swayam-applied-multivariate-analysis-13898>
- https://www.fun-mooc.fr/courses/cours_v1:agrocampusouest+40001EN+session06/about
- <https://online.stat.psu.edu/stat505/lesson/4>
- <http://www2.math.uu.se/~thulin/mm/L7.pdf>
- <http://staffwww.itn.liu.se/~aidvi/courses/06/dm/lectures/lec8.pdf>

Note : Question Patterns based on theory part only .

MAPPING OF CO WITH PSO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	2	2	2	3	3	2
CO2	2	2	2	3	3	1
CO3	2	2	1	2	2	1
CO4	2	2	1	3	3	1
CO5	2	2	2	3	3	1
AVG	2	2	2	3	3	1

PEDAGOGY

- Formal face-to-face lectures
- Discussions
- Practical applications using R.

SEMESTER III
SAMPLE SURVEY DESIGNS

TOTAL HOURS: 75

SUB CODE: 20PBSCT3008

CREDIT: 4

L-T-P: 4-1-1

Course Objective

1. Acquire knowledge of scientific view to conduct the survey in proper way to collect the data about specific perspective.
2. Assess variety of probability and non probability sampling methods for selecting a sample from a population.
3. Utilize the intellectual facts of the sampling techniques to implement in projects and to motivate the students in carrying out the field projects in scientific manner and statistical skills

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

CO	CO Statement
CO1	Understand concepts and techniques in sampling methods.
CO2	Identify various statistical sampling schemes.
CO3	Able to define and distinguish probability and non probability sampling.
CO4	Implement Cluster sampling, Ratio and Regression estimation in real life problems
CO5	Identify the concept of sampling error and its implications.

SYLLABUS

UNIT 1: **(Total number of hour: 15)**

Basic methods of sample selection: Inclusion probabilities, Simple random sampling with and without replacement, Probability proportional to size sampling with and without replacement – estimation of mean and its sampling variance, Horwitz-Thompson estimator.

UNIT 2: **(Total number of hour: 15)**

Systematic sampling Introduction- Linear systematic sampling, schemes for population with linear trend. Stratified sampling-sample size allocations, comparison with simple random sampling and systematic sampling.

UNIT 3: **(Total number of hour: 15)**

Ratio estimation-unbiased ratio type estimator, almost unbiased ratio estimator, Jackknife ratio estimator, bound for bias. Product estimator, Regression estimator, Difference estimator – bias and mean square error under srswor.

UNIT 4: **(Total number of hour: 15)**

Cluster sampling: Estimate of mean per element and its variance, Optimum cluster size. Multistage sampling: Two stage sampling-units of equal size-variance of estimated mean under SRS in both the stages.

UNIT 5: **(Total number of hour: 15)**

Sources of non-sampling errors, Randomized Response method-Warner's model, Simmons randomized response model-Adaptive sampling and Snowball sampling (procedure only).

TEXT BOOKS:

1. Gupta, A. K. and Kabe D.G, (2011): Theory of Sample Surveys, World Scientific Publishing Co. Pte. Ltd., Singapore
2. Cochran, W.G. (2007). Sampling Techniques. (Third Edition). John Wiley & Sons, New Delhi.
3. Mukhopadyay, P. (2007). Survey Sampling. Narosa Publisher, New Delhi.
4. Ardilly, P. and Yves T. (2006). Sampling Methods: Exercise and Solutions. Springer.
5. Sampth, S. (2005). Sampling Theory and Methods, 2nd Edition, Alpha Science International Ltd.

BOOKS FOR REFERENCE:

1. Thompson (2012). Sampling, Wiley Eastern Private Limited.
2. Daroga Singh & F. S. Chaudhary (2015): Theory and Analysis of Sample Survey Designs, New Age International Publishers
2. Cochran, W.G.(1972): Sam
3. Singh D., and Chowdhary, F. S. (2018). Theory and Analysis of Sample Survey Design, New Age International Private Ltd., New Delhi.
4. Sarjinder Singh (2004): Advanced Sampling – Theory with Applications, Kluwer Publications
5. Arijit Chaudhuri, Horst Stenger(2005): Survey Sampling: Theory and Methods, Second Edition (Statistics: A Series of Textbooks and Monographs Book 181) Kindle Edition

E-LEARNING RESOURCES:

1. <https://nptel.ac.in/courses/111/104/111104073/>
2. <https://nptel.ac.in/content/storage2/courses/111104073/Module14/Lecture42.pdf>
3. <https://www.mooc-list.com/tags/sampling-methods>
4. <http://home.iitk.ac.in/~shalab/sampling/chapter2-sampling-simple-random-sampling.pdf>
5. https://www.tutorialspoint.com/statistics/simple_random_sampling.htm

Note : Question Patterns based on theory part only .

Mapping of CO with PSO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	2	2	2	1	1	1
CO2	2	3	3	2	2	1
CO3	3	3	3	3	2	1
CO4	2	2	2	2	1	1
CO5	2	2	2	1	1	1
AVG	2.2	2.4	2.4	1.8	1.4	1

PEDAGOGY

- Formal face-to-face lectures
- Discussions
- Practical applications using R.

SEMESTER III
SURVIVAL ANALYSIS

TOTAL HOURS: 75

SUB CODE:

CREDIT: 4

L-T-P: 4 – 1- 5

Course objectives:

1. To equip students with the basic concepts and methods employed in survival analysis.
2. To acquire knowledge on Time-to-event data analysis and to estimate the time for an individual or a group to experience an event of interest.
3. To enable students to analyse data from studies in which individuals are followed up until a particular event occurs.
4. To describe the various methods used for modelling and evaluating survival data.

Course outcome:

On successful completion of the course the students will be able to

CO	CO Statement
CO1	Equip the students with the information of concept of survival data and various types of censoring.
CO2	Apply the concepts on parametric life distributions and also will be able to estimate Likelihood ratio test for accelerated failure time model.
CO3	Estimate non-parametric survival function using Kaplan-Meier estimator and to compare average expectation of life based on age specific death using population life table.
CO4	Compare two or more independent groups by interpreting log rank test, Gehan Test, Mantel - Haentzel Test and Tarone - Ware tests.
CO5	Predict Time-to-event data methods to analyze the effect of several risk factor on covariates using Cox-Proportional Hazard model.

SYLLABUS

UNIT 1: **(Total number of hour: 15)**

Concept of Time and event. Order and Random Censoring. Censoring mechanism and truncations. Survival, hazard and density functions. Mean and median residual life and their elementary properties. Ageing classes - IFR, IFRA, NBU, NBUE, HNBUE, DMRL and their duals, Bathtub Failure rate

UNIT 2: **(Total number of hour: 13)**

Life distributions - Exponential Gamma, Weibull, Lognormal, Pareto, Linear Failure rate- Parametric inference: Point estimation, Confidence Intervals, Scores, tests based on LR and MLE, Partial likelihood estimation-log logistic distribution, Accelerated failure time model, Cox-Snell residuals.

UNIT 3: **(Total number of hour: 17)**

Estimation of survival function–reduced sample method, actuarial estimator, Nelson-Aalen estimator, Kaplan-Meier Estimator-life table estimation.

UNIT 4: **(Total number of hour: 15)**

Nonparametric methods: Gehan Test, Log rank test. Peto's test, Mantel - Haentzel Test, Tarone - Ware tests, Efron Tests.

UNIT 5: **(Total number of hour: 15)**

Semi-parametric regression for failure rate - PH assumptions - Cox's proportional hazards model with one and several covariates-goodness of fit. Extended Cox's model.

TEXT BOOKS:

1. Klein, J.P. and Moeschberger, M.L. (2003): Survival Analysis- Techniques for Censored and Truncated data
2. Kleinbaum, DG and Klein M. Survival Analysis (2005).: A Self-Learning Text, 2nd Edition. New York: Springer
3. Lee E T: Statistical Methods for survival Data Analysis, ed. 2. New York, John Wiley & Sons.
4. Armitage P, Berry G (2001): Statistical Methods in Medical Research, ed.4, Wiley Blackwell.
5. Altman D G(2006): Practical Statistics for Medical Research, London:Chapman and Hall,

BOOKS FOR REFERENCE:

1. Hosmer, Jr, DW. and Lemeshow, S. Applied Survival Analysis: Regression Modeling of Time to Event Data, 2nd Edition. New York: John Wiley & Sons (2008).
2. Survival Analysis: Techniques for Censored and Truncated Data / Edition 2 (ISBN:978038793991) by John P. Klein, Melvin L. Moeschberger, J.P. Klein, Alwyn B. Scott, Springer, New York.
3. Modelling Survivaldata in Medical Research, Second Edition (ISBN: 1-58488-325-1) by David Collett, Chapman & Hall/CRC, Boca Raton.
4. Modeling Survival Data, Extending the Cox Model (ISBN: 0-387-98784-3) by Terry M. Therneau and Patricia M. Grambsch, Springer, New York.
5. Barlow, R. E. and Proschan, F. (1975): Statistical Theory of Reliability and Life testing, Holt, Rinehart and Winston, New York.

WEB RESOURCES

- <https://tinyheero.github.io/2016/05/12/survival-analysis.html>
- <https://towardsdatascience.com/introduction-to-survival-analysis-the-kaplan-meier-estimator-94ec5812a97a>
- <https://data.princeton.edu/pop509/ParametricSurvival.pdf>
- <https://www.measureevaluation.org/resources/training/online-courses-and-resources/non-certificate-courses-and-mini-tutorials/population-analysis-for-planners/lesson-7>
- <http://www.sthda.com/english/wiki/cox-proportional-hazards-model>

Note : Question Patterns based on theory part only .

Mapping of CO with PSO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	2	1	1	1	3
CO2	2	3	2	2	2	1
CO3	2	3	3	3	3	2
CO4	1	3	3	3	3	2
CO5	2	2	3	3	3	3
AVG	2	3	2	2	2	2

PEDAGOGY

- Formal face-to-face lectures.
- Chalk and Talk
- Group Discussion.
- PowerPoint Presentation.

**SEMESTER III
MACHINE LEARNING TECHNIQUES**

TOTAL HOURS: 75
Credits: 4

SUB CODE: 20PBSCT3010
L-T-P: 4-2-0

Course Objectives:

1. Acquire theoretical knowledge on setting hypothesis for pattern recognition.
2. Apply suitable machine learning techniques for data handling and to gain knowledge from it.
3. Evaluate the performance of algorithms and to provide solution for various real-world applications.

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

CO	CO Statement
CO1	Recognize the characteristics of machine learning strategies.
CO2	Apply various supervised learning methods to appropriate problems.
CO3	Identify and integrate more than one technique to enhance the performance of learning.
CO4	Create probabilistic and unsupervised learning models for handling unknown pattern.
CO5	Analyze the co-occurrence of data to find interesting frequent patterns.

SYLLABUS

UNIT: 1

(Total number of hours: 13)

Introduction to Machine Learning

Introduction ,Components of Learning , Learning Models , Geometric Models, Probabilistic Models, Logic Models, Grouping and Grading, Designing a Learning System, Types of Learning, Supervised, Unsupervised, Reinforcement, Perspectives and Issues, Version Spaces, PAC Learning, VC Dimension.

UNIT: 2

(Total number of hours: 16)

Supervised and Unsupervised Learning

Neural Networks: Introduction, Perceptron, Multilayer Perceptron, Support vector machines: Linear and Non- Linear, Kernel Functions, K-Nearest Neighbors.

Learning a Class from Examples, Linear, Non-linear, Multi-class and Multi-label classification, Decision Trees: ID3, Classification and Regression Trees (CART).

UNIT: 3

(Total number of hours: 16)

Ensemble and Probabilistic Learning

Ensemble Learning Model Combination Schemes, Voting, Error-Correcting Output Codes, Bagging: Random Forest Trees, Boosting: Adaboost, Stacking.

Bayesian Learning, Bayes Optimal Classifier, Naïve Bayes Classifier, Bayesian Belief Networks, Mining Frequent Patterns.

UNIT: 4

(Total number of hours: 15)

Reinforcement Learning and Evaluating Hypotheses

Introduction, Learning Task, Q Learning, Non deterministic Rewards and actions, temporal-difference learning, Relationship to Dynamic Programming, Active reinforcement learning, Generalization in reinforcement learning. Motivation, Basics of Sampling Theory: Error Estimation and Estimating Binomial Proportions, The Binomial Distribution, Estimators, Bias, and Variance.

UNIT: 5

(Total number of hours: 15)

Association Rule Mining

Association analysis: Association rule-support and confidence-rule mining problem-A priori algorithm: Frequent item set generation. Evaluation of association pattern.

TEXT BOOKS:

1. Ethem Alpaydin, "Introduction to Machine Learning", MIT Press, Prentice Hall of India, 3rd Edition 2014.
2. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar "Foundations of Machine Learning", MIT Press, 2012.
3. Tom Mitchell, "Machine Learning", McGraw Hill, 3rd Edition, 1997.
4. MACHINE LEARNING - An Algorithmic Perspective, Second Edition, Stephen Marsland, 2015.

BOOKS FOR REFERENCE:

1. Charu C. Aggarwal, "Data Classification Algorithms and Applications", CRC Press, 2014.
2. Charu C. Aggarwal, "DATA CLUSTERING Algorithms and Applications", CRC Press, 2014.
3. Kevin P. Murphy "Machine Learning: A Probabilistic Perspective", The MIT Press, 2012
4. Jiawei Han and Micheline Kamber and Jian Pei, "Data Mining Concepts and Techniques", 3rd edition, Morgan Kaufman Publications, 2012.

E-LEARNING RESOURCES:

- <https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-ee6024fdb08>
- <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>
- <https://towardsdatascience.com/probabilistic-machine-learning-series-post-1-c8809652dd60>
- <https://www3.cs.stonybrook.edu/~cse352/16L14association.pdf>

Note : Question Patterns based on theory part only .

Mapping of CO with PSO:

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	2	1	3	3	2
CO2	3	2	3	2	2	1
CO3	2	2	1	3	2	1
CO4	2	2	2	1	2	1
CO5	2	3	1	1	2	1
Average	2	2	2	2	2	1

PEDAGOGY

- Formal face-to-face lectures
- Chalk and Talk
- Problem solving and allow time for students to resolve problems in understanding of lecture material Statistical software's

SEMESTER III
EXPLORATORY DATA ANALYSIS
(USING SPSS)

TOTAL HOURS: 60
CREDIT: 3

SUB CODE:
L-T-P: 4 – 0 -- 4

Course Objective:

1. Acquire the statistical techniques and their implementation using comprehensive SPSS software.
2. Enrich the technical skills in Data analytics through the statistical methods using software orientation towards job roles.
3. Enhance the intuition of the students towards the insights of the theoretical concepts with its application in real time domain through software.

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

CO	CO Statement
CO1	Create and edit the data files, plot graphs using SPSS.
CO2	Compute descriptive statistics and perform inferential statistics using SPSS.
CO3	Able to find the associations, relationships and variations among the Bivariate's and Multivariate's.
CO4	Use the applications of theory in real time problems through software and summarize the results.
CO5	Acquire the statistical software knowledge, which is essential required for projects in all disciplines.

SYLLABUS

1. Descriptive measures.
2. Simple, Partial and Multiple correlation
3. Simple Linear Regression
4. Multiple Regression
5. Regression Diagnostics: Examining Residuals, identifying and
 - a. handling influential observations – Collinearity
6. Parametric and Non parametric tests
7. Power curve
8. ANOVA, MANOVA
9. Generalized linear models
10. Categorical Data Analysis: Tests of Association – Simple Logistic Regression
11. Factor analysis
13. Cluster analysis
14. Discriminant analysis
15. Survival analysis: Actuarial and Kaplan-Meir method,
16. Non parametric tests- Gehan's test, Tarone Ware test, Log rank test

Books for Study and References:

1. Andy Field(2011). Discovering Statistics Using SPSS 3rd edition, Sage Publications Ltd
2. George & Mallery(2011). SPSS for Windows Step by Step, 10th edition, Pearson Education in South Asia

Mapping CO with PSO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	1	1	1	1	1	1
CO2	3	2	3	2	2	3
CO3	3	2	3	2	2	1
CO4	2	2	3	2	2	1
CO5	2	2	2	2	2	1
AVG	2	2	2	2	2	1

DEPARTMENT OF STATISTICS
SDNB VAISHNAV COLLEGE FOR WOMEN
(AUTONOMOUS) CHENNAI-600044.

(PG Department of Biostatistics)

COURSE FRAME WORK

SEMESTER IV

SEM	COURSE CODE	COURSE TITLE	TITLE OF THE PAPER	HRS	CREDITS	CA	SE	T
IV		Core Major-XI	Clinical Trials and Epidemiology	75	4	25	75	100
		Core Major Practical- III	Machine Learning using R		4	40	60	100
		Elective- IV	SAS Programming	75	3	25	75	100
		Elective-V	Forecasting and decision making Techniques	75	3	25	75	100
		Project	Project*		4	40	60	100
		Extra Disciplinary	SWAYAM (MOOC)		4	50	----	100
TOTAL					22	205	345	600

SEMESTER IV

CLINICAL TRIALS AND EPIDEMIOLOGY

TOTAL HOURS: 75

Credits: 4

SUB CODE:

L-T-P: 4-1-5

Course objectives:

1. To prepare students to understand the various concepts and applications in Clinical trials and epidemiology.
2. To acquire knowledge of safety and Intervention of the drug in Clinical research.
3. To demonstrate a basic understanding of epidemiologic methods and study design.
4. To compute theory and project concept for students in clinical research and epidemiology.

Course outcome: On successful completion of the course the students will be able to

CO No	Co Statement
CO1	Equip the students with the information of concept of protocol in clinical trial experiment.
CO2	Demonstrate various terminologies in clinical experiment involving different phases and the ethics, principles and conduct of clinical trial experiments with an overview of Phase I-IV trials. Various clinical trial design commonly employed in practice.
CO3	Evaluating the principles of blinding in various trials and to determine trial size using statistical methods.
CO4	Distinguish the roles and relationships between epidemiology and biostatistics in the prevention of disease and the improvement of health.
CO5	Obtain concepts, methods, and tools of public health data collection, analysis and interpretation by using various epidemiological study designs.

SYLLABUS

UNIT 1: **(Total number of hour: 13)**

Introduction to Clinical Trials: Historical background - The need and ethics of clinical trials. Organization and Planning: Main features of the study protocol - Selection of patients - Treatment schedule - Evaluation of patient response - follow-up studies-GCP/ICH guidelines.

UNIT 2: **(Total number of hour: 15)**

Phases of clinical trials: Phase I, II, III and IV trials.

Basic study designs: Randomized control study, Nonrandomized concurrent control study, historical controls, cross-over design, withdrawal studies, group allocation design, hybrid designs, studies of equivalency (description only). Methods of Randomization: Fixed allocation randomization - Stratified randomization- Adaptive randomization - Unequal Randomization.

UNIT 3: **(Total number of hour: 16)**

Blinding and Placebos: Unblinded, Single blind and Double-blind trials - conduct of double blind trials- Crossover trials- Design, analysis and interpretation.

Statistical methods for determining Trial size: Method for dichotomous response variable, continuous response variables, repeated measures, cluster randomization and equivalency of intervention. Multicenter trials.

UNIT 4: **(Total number of hour: 15)**

Introduction: Definition and Scope of Epidemiology –Epidemiology measures: Prevalence and incidence rates – cumulative incidence rate or risk – case fatality.

Causation and Prevention in Epidemiology: Guidelines for causation –causation pie-Hill's Bradford model. The scope of Prevention – Levels of Prevention- epidemics and endemic investigation, surveillance, screening.

UNIT 5: **(Total number of hour: 16)**

Types of Epidemiological studies: Observational and experimental studies – Descriptive studies – Ecological studies – Cross-sectional studies – Cohort studies - Random control trials –Field trials and Community trials.

Planning an Epidemiological Project (Basic Concepts):Introduction – Choosing a project – Describing the population – Questionnaire Preparation- Analysis and Interpretation of data.

TEXT BOOKS

1. Pocock, S.J. (2009): Clinical Trials – A Practical Approach, John Wiley and Sons.
2. Lawrence M.Friedman.Curt D.Furberg David L.DeMets.David M. Reboussin Christopher B. Granger (2015): Fundamentals of Clinical trials (Fifth Edition)
3. Beaglehole, R., Bonita, R. and Kjellstrom, T. (2005): Basic Epidemiology, World Health Organization. (WHO).
4. Park, K.(2005) : Textbook of Preventive and Social Medicine.20th edn. Jabalpur: Bhanot Publishers.
5. Timmreck, T.C.(2002): An Introduction to Epidemiology, Third Edition,Jones and Bartlett.

BOOKS FOR REFERENCE:

1. Piantadosi, S. (2005): Clinical Trials - A Methodological Perspective, Second Edition, Wiley series in probability and Statistics.
2. Gerstman, B.B.(2003) : Epidemiology Kept Simple: An Introduction to traditional and Modern Epidemiology ,Second Edition, John Wiley and Sons.
3. Gordis, L.(2008): Epidemiology, Fourth Edition, Saunders.
4. Jekel, J.F., Katz, D.L. and Elmore, J.G. (2001): Epidemiology,
5. Biostatistics and preventive Medicine, second edition, Saunders.

E-LEARNING RESOURCES

- <http://www2.cdc.gov/mmwr>
- <http://www.cdc.gov/ophss/csels/dsepd/ss1978/ss1978.pdf>
- http://apps.who.int/iris/bitstream/10665/43541/1/9241547073_eng.pdf
- http://samples.sainsburysebooks.co.uk/9781118031179_sample_388791.pdf
- http://onlinelibrary.wiley.com/doi/10.1002/0470842563.fmatter_indsup/pdf

Note : Question Patterns based on theory part only .

Mapping of CO with PSO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	2	1	1	1	3
CO2	2	3	2	3	3	1
CO3	2	3	3	3	2	2
CO4	3	3	2	2	2	2
CO5	2	2	2	3	3	3
AVG	2	3	2	2	2	2

PEDAGOGY

- Formal face-to-face lectures.
- Chalk and Talk
- Group Discussion.
- PowerPoint Presentation.

SEMESTER IV
MACHINE LEARNING USING R

TOTAL HOURS: 60

SUB CODE:

CREDIT: 4

L-T-P: 4 – 0 -- 4

Course Objectives:

The main objectives of this course are to:

1. To identify the operations and functions of R Programming and Perform Statistical analysis using built-in functions.
2. To impart the analytical skills comprising the Statistical Methods, predictive analytics, data mining, and machine learning using software and its implications in data science
3. To give adequate acquaintance with the technical world to elevate the job competency.

Course Outcomes:

On successful completion of the course the students will be able to

CO1	Understand basics of R environment and perform various operations on data using R
CO2	Perform basic statistical procedures.
CO3	Write and execute the code for Bivariate and Multivariate analyses.
CO4	Create customized program for statistical problems. Access various machine learning methods and applications in R.
CO5	Specialized in Time Series and Survival Analysis.

SYLLABUS

1. Multivariate Techniques
 - MANOVA
 - Factor Analysis
 - Cluster Analysis
 - Discriminant Analysis
2. Classification Techniques
 - Multiple Logistic Regression
 - Association rule mining(Apriori)
 - Support Vector Machine
 - Random Forest
 - k-Nearest Neighbors (k-NN)
3. Survival analysis:
 - Actuarial and Kaplan-Meir method
 - Gehan's test & Tarone Ware test
 - Log rank test
 - Cox Regression
4. Forecasting Techniques
 - Fitting ARIMA Models

TEXT BOOKS:

1. Sudha G. Purohit, Sharad D. Gore and Shilaja R. Desmukh (2009), Statistics Using R Language, Narosa, Chennai.
2. Brian S Everitt, TorstenHothorn (2009), A Handbook of Statistical Analyses Using R, Chapman & Hall/CRC, Second edition, England.
3. Deshpande.J.V and Purohit.G.S (2016), Life time Statistical Models and Methods, 2nd Edition, Worlds Scientific Publishing co Private limited.
4. Alan Agresti (2002): Categorical Data Analysis. John Wiley & Sons.
5. Gardener(2017).: Beginning R- The Statistical Programming Language, 1st edition, Wiley India Pvt ltd

BOOKS FOR REFERENCE:

1. Chambers, J. (2008). Software for Data Analysis: Programming with R, Springer.
2. Crawley, M.J. (2017). The R Book, John Wiley & Sons.
3. Matloff, N. (2011). The Art of R Programming, No Starch Press, Inc.
4. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani(2013). An Introduction to Statistical Learning: with Applications in R, Springer, 1st edition.
5. Garrett Golemund, O'Reilley,(2014) Hands-on programming with R, 1st Edition.
6. Michael J. Crawley, Statistics (2015) : An Introduction Using R, **E-LEARNING RESOURCES: WILEY**, Second Edition

E-LEARNING RESOURCES:

- <https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>
- <https://data-flair.training/blogs/machine-learning-for-r-programming/>
- <https://cran.r-project.org/doc/contrib/usingR.pdf>
- <http://little-book-of-r-for-multivariate-analysis.readthedocs.org/>.
- <https://media.readthedocs.org/pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-for-time-series.pdf>

Mapping of CO's with PSO's :

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	1	1	1	1	1	1
CO2	3	3	2	2	1	1
CO3	3	3	2	2	2	1
CO4	2	3	2	2	2	1
CO5	2	3	1	2	2	1
AVG	2	3	2	2	2	1

PEDAGOGY

- Formal face-to-face lectures.
- Group Discussion.
- PowerPoint Presentation.

SEMESTER IV
SAS Programming

TOTALHOURS: 75

SUB CODE:

CREDIT: 3

L-T-P: 4 – 1 -- 5

Course objectives:

1. To provide a clear explanation that demonstrates your knowledge of SAS programming that will inculcate the students in handling modern statistical software in the analysis of real-time data.
2. To enhance the knowledge of SAS which seeks to provide students with an analytical toolset that enables them to address modern, data-intensive business problems?
3. To understand the application of SAS for bio-statistical and epidemiological analysis.

Course outcomes: On successful completion of the course the students will be able to

CO1	Understand the basic concepts to computer programming and presents a portion of the SAS, SAS Studio programming environment.
CO2	Explain the topics related to computing in statistics such as creating, reading dataset from external file and examining errors. Work and understand the concepts of SAS datasets: sort, subset, merge, with SAS Data set
CO3	Demonstrate the programming environment and major aspects about the SAS basic syntax that is necessary to write SAS code to perform basic statistical inference.
CO4	Acquire specialized knowledge in the advanced areas on manipulating data sets and basic statistical data analysis using SAS
CO5	Gain knowledge in necessary skills on the concepts of Connecting to Oracle and Other Database with SAS programming and also apply the learned techniques in Warehousing Business Intelligence Concepts.

SYLLABUS

UNIT 1

(Total number of hour: 15)

Introduction to SAS programs - Components and Fundamental concepts of SAS Programming - Naming Conventions in SAS – SAS Library – Exploring SAS libraries and files – Creating and Maintaining SAS Libraries – SAS Datasets - DATA STEP – PROC STEP - Diagnosing and Correcting syntax errors — PROC PRINT - Producing list report – Sub-setting – PROC COMPARE - Logical and Special Operators – Enhancing Output -Formatting Data Values

UNIT 2:

(Total number of hour: 15)

Creating Dataset using In stream Data - Reading form External raw data files -Examining Data errors and log errors – Assigning attributes to variables – Integration with Excel spreadsheets – Reading SAS data sets and Creating variables – SAS FORMATS & INFORMATS – Date Formats - Conditional processing – Dropping and keeping variables – Concatenating , Merging and Combining SAS data sets (one-to-one, one-to-many, many-to-many)- Match-Merging SAS datasets .

UNIT 3:

(Total number of hour: 15)

Iterative Data processing: DO Loop and SAS ARRAY processing. Producing Summary reports using PROC FREQ, PROC MEANS, PROC REPORT and PROC TABULATE procedures – Producing Graphical Charts (Bar, Pie etc.) using PROC CHART, PROC GPLOT and PROC GCHART Procedures – Enhancing Charts

UNIT 4 :

(Total number of hour:15)

Descriptive Statistics and ANOVA: Selecting Random samples from SAS datasets - Performing Descriptive statistical analysis, Examining distributions, Estimating Confidence intervals for mean and Hypothesis testing of single Mean - Two-Sample t- Tests using PROC CORR, PROC TTEST and PROC UNIVARIATE procedures –Paired t-Tests. Regression Analysis using PROC REG Procedure-PROC PHREG and PROC LIFE-One-Way ANOVA, Two-Way ANOVA with/ without interactions using GLM procedure –Multiple comparison tests - Non-Parametric ANOVA – Categorical data analysis.

UNIT 5:**(Total number of hour:15)**

Connecting to Oracle and Other Databases and Creating Linked Libraries – PROC ACCESS
Procedure for Mining Data from RDBMSs and Excel Files – PROC DBLOAD for Loading
Data into RDBMSs and EXCEL files – Introduction to Data warehousing Business Intelligence
Concepts using SAS – Introduction to SAS Macros

TEXT BOOKS:

1. Delwiche, L.E and Slaughter, S.J. (2003): The Little SAS book, A primer 3rdEd. by SAS publishing
2. Everitt, B.S and Der, G.(2004): Handbook of Statistical Analyses Using SAS, 2nd edition by SAS publishing
3. SAS Certification Prep Guide: Statistical Business Analysis Using SAS ,JONI N.SHREVE AND DONNA DEA HOLLAND ,2018
4. An Introduction to SAS University Edition ,Ron Cody ,II Edition(2018).
5. Carpenter, Art (1998): Carpenter's Complete Guide to the SAS Macro Language. SAS Institute, Cary, NC.

BOOKS FOR REFERENCE :

1. O'Rourke, N, Hatcher, L and Stepanski, E (2005): A Step-by-Step Approach to Using SAS for Univariate and Multivariate, Statistics,2ndEdition, SAS Press and John Wiley Sons Inc.
2. Michele M. Burlew (2006): SAS Macro Programming Made Easy, 2nd Edition by SAS Publishing [ISBN-10: 1-59047-882-7]
3. Learning SAS by Example: A Programmer's Guide, Second Edition Paperback – 2018
4. SAS Certification Prep Guide: Base Programming for SAS 9, Third Edition, By SAS institute, (2011)
5. Carpenter's Guide to Innovative SAS Techniques , By Art Carpenter,SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414,(2012)

E-LEARNING RESOURCES:

- <https://www.edureka.co/blog/sas-programming/>
- https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.5&docsetId=lrcon&docsetTarget=n08fs0rt7fikeln1uh0t8v5pt25d.htm&locale=en
- <https://online.stat.psu.edu/stat481/book/export/html/667>
- https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.5&docsetId=proc&docsetTarget=n1qnc9bddfvhzqn105kqitnf29cp.htm&locale=en
- https://www.lexjansen.com/wuss/2018/26_Final_Paper_PDF.pdf

Note : Question Patterns based on theory part only .

Mapping of CO with PSO:

CO / PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	1	1	2	2	2	2
CO2	1	1	2	2	2	3
CO3	2	2	2	2	3	3
CO4	2	2	2	3	2	3
CO5	2	2	2	2	3	3
Average	2	2	2	2	2	3

PEDAGOGY

- Chalk and Talk
- Power point presentation
- Group Discussions
- Seminars
- Problem solving to understand the concepts of theory

SEMESTER IV

FORECASTING AND DECISION MAKING TECHNIQUES

TOTALHOURS:75

SUB CODE:

CREDIT:4

L-T-P: 4 – 1 -- 5

Course objectives:

1. To understand the functioning of industries and business strategies.
2. To provide a clear explanation of the fundamental theory of time series analysis and forecasting models, which are significant for business study.
3. To provide students with a practical experience that will help them in handling modern statistical software in the analysis of real-life data
4. The main goal in business forecasting is simply to minimize forecast error and produce better forecast.

Course outcome: On successful completion of the course the students will be able to

CO	CO Statement
CO1	Understand sources of demand variability. Outline the role, methods and response in business forecasting and its applications. Describe Qualitative and Quantative forecasting.
CO2	Identify suitable time series models for a given real-life problem.
CO3	Briefly describe averaging techniques, trend and seasonal techniques, and regression analysis.
CO4	Demonstrate advanced knowledge of the types of data analysis problems that can be appropriately dealt with using forecasting techniques.
CO5	To learn various decision rules theories and its applications of decision making as individuals, in groups, and in organizations. Understand decision problem, loss function, risk function and decision rules.

SYLLABUS

UNIT 1:

Total number of hour: 15

Business forecasting – role of forecasting in Business – steps in forecasting – methods of forecasting, exploring data patterns- measuring forecasting errors

UNIT 2:

Total number of hour: 15

Moving average and smoothing methods: Naïve models-forecasting methods based on averages: simple, moving averages and double moving averages-Exponential smoothing methods: Holt's method, Winters' methods- forecasting based on smoothing, adaptive smoothing.

UNIT 3:

Total number of hour: 15

Exploratory Time Series Analysis: Components of time series- forecasting trend-seasonality-seasonally adjusted data-cyclical and irregular variations-forecasting a seasonal time series.

UNIT 4:

Total number of hour: 15

Stationary processes: a) moving average) (MA), b) Auto Regressive (AR), c) ARMA and (d) AR integrated MA (ARIMA) models, Box-Jenkins methodology-steps only-choosing a forecasting technique.

UNIT 5:

Total number of hour: 15

Introduction, decision making-problem, process and environment, decision under uncertainty-Laplace criterion, Maximin or Minimax criterion, Maximax or Minimin criterion and Horwicz criterion, Decision under risk-Expected Monetary Value (EMV), Expected Opportunity Loss (EOL) and Expected Value of Perfect Information (EVPI), Decision tree analysis

TEXT BOOKS:

1. Hanke, J.E. and Wichern, D.W. (2007): Business Forecasting, eighth edition, Pearson education, Asia.
2. Kanthi Swarup, Gupta, P.K. and Manmohan (2003): Operations Reasearch . Sultan Chand & Sons
3. Taha, H.A. (2006): Operations Research, 7th edition, Collier MacMillan.

BOOKS FOR REFERENCE :

1. Casella, G. and Berger, R.L. (2002): Statistical Inference, Thomson Duxbury, Second edition.
2. Lehmann, E.L. and Casella, G. (2006): Theory of Point Estimation, Second edition, Springer Texts in Statistics
3. Rao, C.R. (2002): Linear Statistical Inference and its applications, Second edition, Wiley series in Probability and Statistics
4. Miller, I. and Miller, Marylees. John E. Freund's Mathematical Statistics with Application, 7th ed.,, New Jersey: Prentice Hall, 2010.

E-LEARNING RESOURCES:

- <http://moodle.wolfware.ncsu.edu/>
- <http://www.e-booksdirectory.com/listing.php?category=15>
- <https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique>
- <http://www.analyticsvidhya.com/blog/2016/02/free-read-books-statisticsmathematics-data-science/>

Note : Question Patterns based on theory part only .

Mapping Of CO With PSO:

CO \ PSO	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6
CO1	3	2	2	1	1	1
CO2	2	2	3	3	1	1
CO3	2	2	2	2	2	1
CO4	2	2	3	3	2	1
CO5	2	2	3	3	1	1
AVG	2	2	3	2	1	1

PEDAGOGY:

- Formal face-to-face lectures
- Chalk and Talk
- Problem solving and allow time for students to resolve problems in understanding of lecture material.

QUESTION PAPER PATTERN:

QUESTION PAPER PATTERN FOR OBE

(2020-21 onwards)

Theory

PG –Question paper Pattern- conventional on- papermode

Bloom's Category Level	Sections	Marks	Word limit	Total	Meaning of K's
K1, K2	Multiple Choice Questions	30	Correct choice	75	K 1 & K2 - Understanding Level K 3 - Apply Level K 4 - Analyze Level K 5 – Evaluate Level K 6 – Create Level
K3, K4	Section B 5 Questions out of 7 questions *5 Marks	25	Short answers {approx. 500 Words)		
K4, K5,k6	Section C 1 out of 3 Questions *10 Marks + Compulsory Question 10 Marks	20	Elaborate answers (approx. 1000 Words)		

*** 75 marks to be converted as 60 marks.**

UG/PG QUESTION PAPER PATTERN FOR OBE ONLINE
ASSESSMENT (2020 - 2021)

Bloom's Category Level	Sections	Marks	Description of answer	Total	Meaning of K's
INTERNAL SETTING					
K1,K2,K3	Section A Multiple Choice Questions 25 Questions *1 Marks (No Choice)	25X1=25	Choose the write option.	50	K 1 & K2 - Understanding Level K 3 - Apply Level K 4 - Analyze Level K 5 – Evaluate Level K 6 – Create Level
EXTERNAL SETTING					
K2,k3,K4, K5,K6	Section B 5 out of 7 Questions *5 Marks	25	Short answers/500 Words		

*** 50 marks to be converted as 60 marks.**

BLOOM'S CATEGORY LEVEL (ANNEXURE chart)

S.no	K component scale	Verbs for question
I.	K 1& K2 Verbs	Verbs to be used for questioning are “choose, find, identify, indicate, match, name, state, what, when, where, which, who, cite, label, reproduce. define, list, quote, revise, explain, show, sketch, illustrate, interpret, describe, substitute, convert, give example, rephrase
2.	K2 &k3	The questions may contain the verbs such as explain, show, sketch, illustrate, interpret, describe, substitute, convert, examFle, rephrase, apply, relate, solve, classify, predict, compute, prepare
3.	K4	The questions may contain verbs - Apply, relate, solve, classify, predict, compute, prepare.
4.	K5	The questions may contain any of the following verbs : Ascertain, diagnose, distinguish, infer, associate, examine, differentiate, reduce, discriminate, dissect, determine, justify, organize, recommend, solve.
5	K6	The questions may contain any of the following verbs: Appraise, conclude, critique, judge, assess, contrast, deduce, weigh. Compare, criticize, evaluate.

Question paper pattern for Continuous Assessment Test (CAT)

(The online assessment pattern)

U.G/P.G PROGRAMME

SHRIMATHI DEVKUNVAR NANALAL BHATT VAISHNAV

COLLEGE FOR WOMEN

B.Sc/M.Sc/B.A/M.A/B.Com/M.com DEGREE EXAMINATION,, 2020.

..... YEAR..... SEMESTER

CAT – I/II/III

Sub Title:

Max. Marks: 50

Sub Code:

Date:

Time: 2hrs.

Question paper Pattern-Two Components: (Max marks=50) - 3hrs

I. Multiple Choice Questions (MCQ) - 20 marks (10x2=20)

II. Google Class Room (GCR) - 30 marks (Structured)

A. Section A: 5 out of 6 – each carries 2 marks (5x2=10)

B. Section B: 4 out of 5 – each carries 5 marks (4x5=20)

- The answers for the questions for QP uploaded in GCR will be as uploads (images of hand written answer sheets converted to .pdf) in Google Class Room.
- The duration for each GCR session (answering and uploading) would be 3 hours (maximum).
- The structured component (30 marks) SHOULD be conducted in GCR as per the CAT schedule.MCQ (10X2=20) CAN be conducted out of schedule also, but should be completed during the CAT examination scheduled.

Note: The GCR question paper and MCQ assessment links to be shared with the COE office for approval and validity on or before the respective allotted dates.

RUBRICS FOR CONTINUOUS ASSESSMENT

Assignment	
Seminar	
Field visit	
Participatory Learning	
Group Discussion	
Flipped/Blended Learning	

Assessment Model (from 2020 – 21 onwards)
Post graduation programme
40% Internal 60% External

S.No	Assessment Component	Marks	Weighted %
A.	Theory		
1	INTERNAL ASSESSMENTS		
	Continuous Assessment Test(best two out of three)	2 x 50 = 100	15
2	Quiz/Group Discussion/Seminar/Assignment/Role Play/ Case Study/ Open Book/ snap Test/ Video Presentation/ Review (any three to be considered)	3 x 10 = 30	15
3	MCQ (one test to be conducted online during the semester)	20	10
4	EXTERNAL ASSESSMENT		
	End semester examinations	75	60
	Grand Total		100
B	Practical		
1	INTERNAL ASSESSMENTS		
	Continuous Assessment Test(best two out of three)	2 x 50 = 100	15
2	Record + Observation	10 +10 = 20	15
3	MCQ (one test to be conducted online during the semester)	20	10
4	EXTERNAL ASSESSMENT		
	End semester Examinations	60	60
	Grand Total		100

Shrimathi DevkunvarNanalal Bhatt Vaishnav College for Women (Autonomous)
Re-accredited with “A+” Grade by NAAC

Amendments in the regulations from 2020 – 2021 onwards

PG

Changes in Part-II

Semester – I

Title	Internal Marks	External Marks	Credits
Skill based elective-Teaching Skills	50	-	3

Semester – II

Title	Internal Marks	External Marks	Credits
Soft Skills – SWAYAM (MOOC)	50	-	4

Semester – III

Title	Internal Marks	External Marks	Credits
Skill based elective -Research Skills	50	-	3

Semester – IV

Title	Internal Marks	External Marks	Credits
Extra Disciplinary– SWAYAM (MOOC)	50	-	4