HSNC UNIVERSITY, MUMBAI

(2023-2024)

Ordinances and Regulations With

Respect to

Choice Based Credit System (CBCS)

For the Programmes Under

The Faculty of Science and Technology

Framed According to the National Education Policy (NEP 2020)

To be implemented from Academic Year: 2023-2024

For the Course

Statistics

Curriculum – First Year Postgraduate

Semester-I and Semester -II

2023-2024

HSNC UNIVERSITY, MUMBAI

Board of Faculty of Science & Technology

Board of Studies in the Subjects of Statistics and Data Science & Business Analytics

1) **Name of Chairperson/Co-Chairperson/Coordinator:-**

   a) **Dr Asha Jindal,** Professor and (UG: Head &PG: Coordinator), Department of Statistics, K. C. college, HSNC University Churchgate, Mumbai –400 020. Email ID- asha.jindal@kccollege.edu.in Mobile no- 9821235627

2) **Two to five teachers each having minimum five years teaching experience amongst the full time teachers of the Departments, in the relevant subject.**

   a) **Dr. S. B. Muley,** Assistant Professor, Department of Statistics, K. C. college, HSNC University Churchgate, Mumbai – 400 020. Email ID sakharam.muley@kccollege.edu.in , Mobile No- 9323817918

   b) **Mrs. Pratiksha Kadam,** Assistant Professor, Department of Statistics, K. C. college, HSNC University Churchgate, Mumbai – 400 020. Email ID pratiksha.kadam@kccollege.edu.in , Mobile No- 7507162816

   c) **Ms. Shailaja Rane,** Assistant Professor, Department of Statistics, K. C. college, HSNC University Churchgate, Mumbai – 400 020. Email ID shailaja.rane@kccollege.edu.in, Mobile No- 7506986359

3) **One Professor / Associate Professor from other Universities or professor / Associate Professor from colleges managed by Parent Body;**

   a) **Dr Anjum Ara Ahmed;** Professor and I/C Principal, Rizvi College, Mumbai. Email ID **anjumahmed8@gmail.com, Mobile No-** 8451046220

4) **Four external experts from Industry / Research / eminent scholar in the field relevant to the subject nominated by the Parent Body;**

    a. **Prof. Suresh Kumar Sharma,** Senior Professor, Department of Statistics, Panjab University, Chandigarh.

       Email ID ssharma643@yahoo.co.in, **Mobile No-**9815911381

    b. **Mr Mukesh Jain,** Vice President and Chief Technological Officer, Capgemini. Email ID mdjain@hotmail.com, **Mobile No-**7972637347.

    c. **Dr Santosh Gite,** Professor, Dept. of Statistics, University of Mumbai, Mumbai. Email ID santgite@yahoo.com, **Mobile No-** 9167157717.

    d. **Mr Prashant Kumar Nair,** Director, Geo Spatial Analytics Global Lead, Intelligent Analytics, Nielsen Connect, Email ID prashantkumar.nair@nielsen.com , **Mobile No-**9833747057.

5. Top rankers of the Final Year Graduate and Final Year Post Graduate examination of previous year of the concerned subject as invitee members for discussions on framing or revision of syllabus of that subject or group of subjects for one year.

    a) **Mr. Chinmay Mokal** (Postgraduate student 22-23) Email ID- chinmay30mokal@gmail.com ; Mobile no- 9372323901

    b) **Ms. Aanchal Goyal** (undergraduate student 22-23) Email ID- aanchalgoyal2703@gmail.com ; Mobile no- 7738886488

Statistics
Part 1- Preamble

M. Sc. Statistics program is of minimum 80 credits cover four semesters. Statistics is the language of the uncertainties riddled modern information age. Statistics facilitates the choice making process by quantifying the element of chance or uncertainties. The program emphasizes both theory and modern applications of statistics and is structured to provide knowledge and skills in depth necessary for the employability of students in industry, in academics, government and non-government organizations/agencies. The NEP program has some unique features like independent projects, number of elective courses and extensive computer training of statistical computations including standard software packages like IBM SPSS, SPSS-AMOS, MS-Excel, Meta Essential, MINITAB, R and PYTHON etc. Due to State Public Cluster University, the Statistics department of K. C. College got the academic autonomy and it's been utilized to add the new and need based elective courses. Internship at the end of first year is one among the important components of this program. The syllabus of the first year (two semesters) covers most of the core courses. In the second year of the syllabus, there are six core/major courses, one minor course and 12 credits two projects. The independent project work in both semesters of second year make students ready to supplement and apply classroom and laboratory transactions. It follows the approach of Learning by Doing and Learning by Living. The syllabus has been framed to possess a decent balance of theory, methods and applications of statistics. It is possible for the students to study basic courses from other disciplines like economics, life sciences, computer science and Information Technology in place of optional/electives. The thrust of the course is to prepare students to enter into a promising career after post-graduation, as also provide to them a platform for pursuing higher studies resulting in doctorate degrees.

1. **Program Outcomes:**

   1. **Advanced Disciplinary and Interdisciplinary Knowledge**: Demonstrate a comprehensive understanding of the core and advanced concepts in the specific field of science (e.g., Physics, Chemistry, Mathematical sciences, Life sciences, , Computers and Information Technology etc). Apply the knowledge from related disciplines to address complex scientific challenges.

   2. **Research Proficiency**: Develop the ability to design and conduct experiments, analyze data, and interpret results using modern scientific methods and tools.

   3. **Critical Thinking, Problem Solving, Data Analysis and Interpretation**: Employ logical reasoning and innovative approaches to solve complex scientific problems. Use quantitative and qualitative methods, including statistical and computational tools, to draw valid conclusions.

   4. **Use of Technology, Laboratory and Technical Skills**: Gain proficiency in handling advanced laboratory equipment and techniques specific to the field of study. Effectively use modern software, tools, and technologies relevant to the discipline.

   5. **Communication and Teamwork**: Develop the ability to communicate scientific concepts, methodologies, and findings clearly and effectively in both written and oral formats. Work collaboratively in diverse teams and demonstrate leadership skills in scientific and research projects.

   6. **Ethics and Professionalism**: Uphold ethical principles in research and professional practices, ensuring honesty, integrity, and accountability.

   7. **Environmental and Social Responsibility**: Demonstrate awareness of the environmental, societal, and global impacts of scientific endeavors.

   8. **Independent Learning ability and Preparation for Higher Studies**: Cultivate an attitude of continuous learning to adapt to advancements in science and technology. Be prepared to pursue doctoral programs or other advanced studies in the chosen field.

   9. **Carrier development**: This program brings together the graduates who wish to enhance their skills and gives them an opportunity to develop their careers in a particular direction. The programme provides in-depth knowledge of particular subject and arouses interest of the students towards research in that particular field.

10. **Employability and entrepreneurship**: The masters of Science program provides the candidate with understanding, general proficiency, and methodical abilities on an advanced level required in industry, consultancy, education, entrepreneurship or public administration etc.

2. Course Objectives are

a. **To elucidate** an advanced foundation in statistical theories and methodologies, along with in-depth applications, extensions, and the interdisciplinary relevance of statistics in diverse fields such as machine learning, economics, public health, and scientific research.

b. **To ensure comprehensive and holistic coverage** of the subject, integrating rigorous theoretical frameworks with practical, real-world data analysis, advanced statistical modeling, and exposure to cutting-edge statistical tools and techniques.

c. **To equip graduates** with the advanced statistical skill sets required to meet the high demands of industry and academia, fostering expertise in data analytics, statistical research, and innovation in problem-solving across sectors like finance, technology, and public policy.

d. **To offer stimulating, rewarding, and high-impact career pathways** to postgraduate students, preparing them for leadership roles in research, data science, analytics, academia, or specialized statistical positions in diverse industries.

3. Process adopted for curriculum designing.

The department has conducted multiple meetings with academician, industry experts. After discussion with them, the changes in the syllabus were introduced with the view that students need to learn the core concepts in detail.

4. Salient features, how it has been made more relevant:

The revised syllabus has been made more relevant by:

1. Aligning with Industry Requirements: The syllabus has been revised to align with the requirements of the industries, providing students with skills that are in demand.

2. Incorporating Emerging Trends: The syllabus includes emerging trends in statistics, such as data science and AI branches like machine learning and GenAI, to provide students with cutting-edge knowledge.

3. Providing Practical Experience: The syllabus provides students with practical experience in

statistical analysis and data interpretation through internships, field visits, and software training.

4. Fostering Critical Thinking and Problem-Solving: The syllabus includes seminars, discussions, and project work to foster critical thinking and problem-solving skills in students.

5. **Program-Specific Outcomes (PSOs)**

1. **Advanced Statistical Knowledge**: Acquire a comprehensive understanding of core statistical theories and methodologies, including advanced theories and techniques on Nonparametric Inference, Survival Analysis, Machine learning Techniques, Structural Equation Modeling and Research Methodology (Developing Scales & Analysis, Meta-Analysis and Communication).

2. **Statistical Software Proficiency**: Develop expertise in modern statistical software tools such as SPSS, R, Python, Excel, Jamovi, AMOS, and Meta Essential for data analysis, visualization, and presentation.

3. **Research and Analytical Expertise**: Build skills to design experiments, analyze data using advanced statistical methods, and conduct independent research projects to address complex problems in academia and industry.

4. **Application of Statistical Methods across Disciplines**: Utilize statistical techniques to solve real-world problems in fields like Social Sciences, healthcare, engineering, economics, and data science, integrating interdisciplinary knowledge.

5. **Data-Driven Decision Making**: Learn to apply statistical methods and/or AI for Marketing, forecasting, quality control, Machine Learning, predictive modeling, Text Analytics and Natural Language Processing facilitating informed decision-making in industrial and academic contexts.

6. **Ethical, Responsible Data Usage and Communication**: Understand and uphold ethical principles in statistical practice, ensuring the responsible collection, analysis, and interpretation of data with societal and environmental considerations followed by communication and Publication in well recognised Journals.

7. **Career and Lifelong Learning Skills**: Prepare for successful careers in academia, industry, Government and/or entrepreneurship by acquiring the necessary professional skills and fostering a mind-set for continuous learning to stay updated with evolving statistical methodologies.

6. **Learning Outcomes:** Learning Outcomes of various courses are

Semester I: Course Outcomes (COs)

1. STA503B: Probability and Distribution Theory

- **CO1**: Identify and describe core concepts in probability theory, such as sample spaces, inequalities and events. Apply theorems such as Bayes' theorem.
- **CO2**: Understand random variables to calculate and interpret probabilities in complex scenarios.
- **CO3**: Evaluate and utilize appropriate probability distributions (discrete and continuous) for statistical modeling in various real-world situations.

2. STA504B: Theory of Estimation and Testing of Hypotheses

- **CO1**: Understand and explain key principles of estimation, including point estimation and inequalities. Apply various methods for constructing estimator such as the method of moments and maximum likelihood estimation, and assess their efficiency
- **CO2**: Formulate and test hypotheses using statistical methodologies and interpret the results within the context of statistical inference.
- **CO3**: Understand MP, UMP Likelihood Ratio Test, SPRT and interval estimation and application.

3. STA505B: Linear Models

- **CO1**: Apply linear algebra techniques to formulate and analyze linear models for statistical data.
- **CO2**: Perform analysis of variance (ANOVA) and evaluate models with fixed effect and mixed effects.
- **CO3**: Use Analysis of Variance with mixed &random effects and ANCOVA (regression models) to analyze and predict outcomes, particularly focusing on linear and non-linear models.

4. STA506B: Computational Statistics using R

- **CO1**: Demonstrate proficiency in basic R programming for statistical computing and data analysis.
- **CO2**: Utilize R for data manipulation and advanced statistical techniques, including visualization and model-building.
- **CO3**: Perform statistical computing using R with an emphasis on descriptive statistics, exploratory data analysis. Apply R for text and sentiment analysis.

5. STA507B: Computational Statistics using Python

- **CO1**: Demonstrate proficiency in Python programming and its application to statistical data analysis.
- **CO2**: Use Python libraries such as NumPy and pandas for data manipulation and advanced statistical techniques, including visualization and model-building.
- **CO3**: Perform statistical computing using Python with an emphasis on descriptive statistics, exploratory data analysis. Apply Python for text and sentiment analysis.

6. STA501A: Research Methodology

- **CO1**: Understand research methodologies and design experiments or surveys with appropriate statistical methods.
- **CO2**: Conduct systematic literature reviews and Learn basic of Meta-Analysis.
- **CO3**: Apply meta-analysis, and critically assess research data.
- **CO4**: Develop Scale, Check Reliability and Validity. Apply Survey Techniques using Statistical Softwares and Present research findings effectively with a strong emphasis on statistical accuracy and scientific communication

<div align="center">Semester II: Course Outcomes (COs)</div>

1. STA508B: Multivariate Analysis

- **CO1**: Understand and apply the principles of multivariate distributions, including multivariate normal distribution.
- **CO2**: Implement multivariate analysis techniques such as MANOVA and Hotelling's T^2 to analyze multivariate data.
- **CO3**: Use discriminant analysis and principal component analysis for data classification and dimensionality reduction.

2. STA509B: Sample Surveys

- **CO1**: Apply basic and advanced sampling techniques, such as SRS, stratified sampling, and systematic sampling
- **CO2**: Design and conduct surveys, applying ratio and regression estimators, cluster sampling, Two Stage Sampling to survey data.
- **CO3**: Use probability sampling techniques to handle real-life survey challenges, including unequal probability sampling and double sampling.

3. STA5010B: Regression Analysis

- **CO1**: Build and apply multiple regression models to analyze the relationship between variables in a dataset.
- **CO2**: Perform diagnostics on regression models to identify issues like multicollinearity and heteroscedasticity, and apply appropriate corrections.
- **CO3**: Use generalized linear models including logistic and Poisson regression to model different types of data.

4. STA5011B: Machine Learning Techniques

- **CO1**: Understand key machine learning techniques and their application to statistical data analysis.
- **CO2**: Implement classification techniques including decision trees, Factor Analysis, Cluster Analysis k-NN, and SVM, for predictive modeling.
- **CO3**: Use association rules, market basket analysis and Natural Language Processing to discover patterns in data and make predictions.

5. STA5012B: Demography

- **CO1**: Understand the key principles and theories in demography, including population composition and growth.
- **CO2**: Apply statistical methods for measuring fertility, mortality, and population dynamics.
- **CO3**: Use migration data and demographic projections to analyze and predict population trends.

5. STA502A: Internship

1. **CO1: Application of Statistical Concepts**
   Apply theoretical statistical knowledge to analyze real-world problems in a professional setting.
2. **CO2: Proficiency in Statistical Tools**
   Use statistical software like R, Python, SPSS, or Tableau for data analysis and reporting in practical scenarios.
3. **CO3: Communication and Collaboration**
   Develop clear communication skills to present statistical insights and collaborate effectively within a professional team.
4. **CO4: Ethical and Professional Responsibility**
   Adhere to ethical standards in data analysis, ensuring integrity, confidentiality, and accountability.
5. **CO5: Career Readiness and Networking**
   Gain exposure to industry practices, enhance job readiness, and build professional networks to explore career opportunities.

Part 2- The Scheme of Teaching and Examination is as under: Semester – I Summary

Sr. No.	Choice Based Credit System			Subject Code	Remarks
1	Core Course (**Statistics**)			STA503B,STA504B, STA505B,STA505D,STA506D,STA507D	
2	Elective Course	Discipline Specific Elective (DSE) Course		STA506B/ STA507B and STA508D STA509D	
		2.1	Interdisciplinary Specific Elective (IDSE) Course		
		2.2	Dissertation/Project		
		2.3	Generic Elective (GE) Course		
3	Ability Enhancement Courses (AEC)				
	Skill Enhancement Courses (SEC)				
4	Minor (Research Methodology)			STA501A	

First Year Semester I Internal and External Detailed Evaluation Scheme

Sr. No.	Subject Code	Subject Title	Periods Per Week					Seasonal Evaluation Scheme			Total Marks
			Un its	S. L.	L	T	P	Cred it	S. L. E	CT	SE	
1	STA503B	Probability and Distribution Theory	3	20%*	3	0	0	3	15		60	100
	STA505D	**Practical based on STA503B**					2	1			25	
2	STA504B	Theory of Estimation and Testing of Hypotheses	3	20%*	3	0	0	3	15		60	100
	STA506D	**Practical based on STA504B**					2	1			25	
3	STA505B	Linear Models	3	20%*	3	0	0	3	15		60	100
	STA507D	**Practical based on STA505B**					2	1			25	
4	STA506B	Computational Statistics using R **or**	3	20%*	3	0	0	3	15		60	100
	STA507B	Computational Statistics using Python	3	20%*	3	0	0	3	15		60	

S.No	Subject Code	Subject											
	STA508D or STA509D	**Practical based on STA506B / STA507B**					2	1				25	
5	STA501A		4	20%*	4	0	0	4	15		25	60	100
	Total Hours / Credit												500

*One to two lectures to be taken for CONTINUOUS self –learning Evaluation.

First Year Semester – I Units – Topics – Teaching Hours

S. No	Subject Code		Subject Unit Title	Hours / Lectures	Total No. of hours /lectures	Credit	Tot al Mark s
1	STA503B	I	Basics of Probability.	15	45H	3	
		II	Random Variables.	15			
		III	Distribution Theory.	15			
	STA505D	IV	Practical based on STA503B	30	30H	1	
2	STA504B	I	Introduction to Estimation and Inequalities.	15	45 H	3	100 (60+40)
		II	Methods of Estimation.	15			
		III	Testing of Hypothesis and Interval Estimation	15			
	STA506D	IV	Practical based on STA504B	30	30H	1	
3	STA505B	I	Introduction to Linear Algebra and Gauss-Markoff Model	15	45 H	3	100 (60+40)
		II	Analysis of Variance: One way, Two Way and Fixed Effect Models	15			
		III	Analysis of Variance: Random, Mixed Effect Models and ANCOVA	15			
	STA507D	IV	Practical based on STA505B	30	30H	1	
4	STA506B	I	Introduction to R	15	45 H	3	100 (60+40)
		II	Data Handling using R	15			
		III	Statistical Computing and Text Analytics using R	15			
	STA507B	I	Introduction to Python	15	45 H	3	
		II	Data Handling using R	15			
		III	Statistical Computing and Text Analytics using R	15			
	STA508D / STA509D	IV	Practical based on STA506B A or Practical based on STA506B B	30	30H	1	
5	STA501A	I	Introduction to Research Methodology	15	60H	4	100
		II	Literature Review and Systematic Literature Review	15			

		III	Meta Analysis	15			
		IV	Understanding Steps of Research	15			
			TOTAL			20	500

- **Lecture Duration = 1 hours**
- **One Credit = 15 hours (For Theory)**
- **One Credit = 30 hours (For Practical)**

L: Lecture: Tutorials P: Practical Ct-Core Theory, Cp-Core Practical, SLE- Self learning evaluation
CT-Commutative Test, SEE- Semester End Examination , PA-Project Assessment, AT- Attendance

Detailed Scheme Theory

Curriculum Topics along with Self-Learning topics - to be covered, through self-learning mode along with the respective Unit. Evaluation of self-learning topics to be undertaken before the concluding lecture instructions of the respective UNIT

Course Code: STA503B
Title of the paper: **Probability and Distribution Theory**

Unit	Content	No. of Hours
I	**Basics of Probability:** 1.1 Sample Space and relevance of Probability theory 1.2 Various definitions, Properties of Probability, Basic formulas 1.3 Combination of events, the realization of m among n events 1.4 Conditional Probability, Independent events (Stochastic independence) Bayes theorem. 1.5 Occupancy Problems on runs and recurrent events 1.6 Bonferroni's inequality, Booles' inequality,Markov, Chebyshev and Jensen inequalities, continuity theorem.	**15**
II	**Random Variables:** 3.1. Random variables (discrete and continuous), Distribution Function and its properties, expectation (its properties) and variance. Moments, Bivariate random variable, joint, marginal and conditional pmfs and pdfs, correlation coefficient, conditional expectation. 3.2. Functions of random variables and their distributions using Jacobian of transformation and other tools. Probability Integral transformation..	**15**
III	**Distribution Theory:** 4.1. Moment generating, Characteristic and probability generating functions 4.2. Discrete Distributions: Bernoulli, Binomial, Poisson, Hyper-geometric, geometric, negative binomial and uniform. 4.3. Continuous Distributions: Uniform, normal, exponential, gamma, Beta, Cauchy, Weibull, Pareto, Chi-square, Laplace and Lognormal. Bivariate normal and multinomial distributions. 4.4. Convergence in distribution. De-Moivre-Laplace and Lindeberg-Levy forms of Central Limit Theorem.	**15**

Self-Learning topics (Unit wise)

Unit	Topics
2	2.1 Sets, classes of sets, algebra of sets, limits of sequence of sets 2.2 field, sigma-field.
2	2.4 Bonferroni's inequality, Booles' inequality
3	3.1 Random experiments, sample spaces (finite and infinite), events, algebra of events, three basic approaches to probability.
	3.2 conditional probability, Bayes' formula.
	3.3 Random variables (discrete and continuous), Distribution Function and its properties, expectation (its properties) and variance. Moments, Bivariate random variable, joint, marginal and conditional pmfs and pdfs, correlation coefficient
	3.4 Functions of random variables and their distributions using Jacobian of transformation and other tools. Markov, Chebyshev.

Online Resources

1. 'Measure Theory' by Prof. E. K. Narayanan from IISc Bangalore available on the Swayam portal
   https://nptel.ac.in/courses/111/108/111108135/

2. 'Measure theory' by Prof. Inder Kumar Rana from IIT Bombay available on the Swayam portal
   https://nptel.ac.in/courses/111/101/111101100/

3. 'Probability and Statistics' by Prof. Somesh Kumar from IIT Kharagpur available on the Swayam portal
   https://nptel.ac.in/courses/111/105/111105090/

   for unit II and unit III

4. 'Introduction to Probability in Computing' by Prof.John Augustine from IIT Madras available on the Swayam portal
   https://nptel.ac.in/courses/106/106/106106176/

5. 'Advanced Topics in Probability and Random Processes ' by Prof. P. K. Bora from IIT Guwahati available on the Swayam portal
   https://nptel.ac.in/courses/108/103/108103112/

Course Code: STA504B
Title of the paper: Theory of Estimation and Testing of Hypotheses

Unit	Content	No. of Hours
1	**Introduction to Estimation and Inequalities:** 1.1. **Estimation:** Introduction to the problem of estimation. Unbiased estimator, efficient estimator, sufficiency, consistent estimator, completeness. 1.2. Unbiased estimation: Minimum and uniformly minimum variance unbiased estimation, Rao-Blackwell and Lehmann-Scheffe theorems. 1.3. Ancillary statistic, Basu's theorem and its applications. Fisher information measure. 1.4. Cramer- Rao inequality and MVBUE, Chapman-Robin inequality.	15
2	**Methods of Estimation and Introduction to Tests of Hypotheses** 2.1. Method of moments, maximum likelihood estimation, minimum chi-square method. 2.2. Basic ideas of Bayes and Minimax estimators. 2.3. **Tests of Hypotheses**: Concepts of critical regions, test functions, two kinds of errors, size function, power function, level of significance.	15
3	**Testing of Hypothesis and Interval estimation** **Testing of Hypothesis** 3.1. MP and UMP tests in a class of size α tests. Neyman - Pearson Lemma, MP test for simple null against simple alternative hypothesis. 3.2. UMP tests for simple null hypothesis against one-sided alternatives and for one-sided null against one-sided alternatives in one parameter exponential family. Extension of these results to Pitman family when only upper or lower end depends on the parameter and to distributions with MLR property. Non-existence of UMP test for simple null against two-sided alternatives in one parameter exponential family. 3.3. Likelihood Ratio Tests. Wald's SPRT with prescribed errors of two types. **Interval estimation:** 3.4. Concepts of Confidence Interval, Confidence Coefficient and confidence level, construction of confidence intervals using pivots, Confidence Interval for Proportions using Large Samples, ,confidence interval for correlation coefficient.. Confidence interval for Median and Quartiles using Order Statistics.	15

Self-Learning topics (Unit wise)

Unit	Topics
1	1.1. Introduction to the problem of estimation. Unbiased estimator, efficient estimator, completeness. 1.2. Unbiased estimation: Minimum and uniformly minimum variance unbiased estimation

2	2.1 method of moments, minimum chi-square method. .
2	**2.3 Tests of Hypotheses**: Concepts of critical regions, test functions, two kinds of errors, size function, power function, level of significance

Online Resources

1. 'Probability and Statistics' by Prof. Somesh Kumar from IIT Kharagpur available on the Swayam portal
   https://nptel.ac.in/courses/111/105/111105090/
   for unit I and Unit II
2. 'Statistical Inference' by Prof. Somesh Kumar from IIT Kharagpur
   https://nptel.ac.in/courses/111105043/
3. 'Introduction to Statistical Hypothesis Testing' by Prof. Arun Tangirala from IIT Madras available on the Swayam portal
   https://nptel.ac.in/courses/103/106/103106120/

Course Code: STA505B
Title of the paper: **Linear Models**

Unit	Content	No. of Hours
1	**Introduction to Linear Algebra** 1.1. Basic operations: Vector Spaces, Linear dependence and independence, Determinants, Rank of Matrices, generalized inverse, and Solving linear equations. 1.2. Characteristic roots and characteristic vectors, properties of characteristics roots , Idempotent matrix, Quadratic forms, positive and Positive semi definite matrix **The Gauss-Markoff Model:** 1.3. Linear parametric function and its estimability, Gauss markoff theorem. 1.4. Interval estimates and test of hypothesis, fundamental theorems on conditional error ss, Test of $\Lambda\beta=d$, generalized least squares	15
2	**Analysis of Variance: One way, Two Way and Fixed Effect Models** 2.1. One –way classification model 2.2. Checking assumptions of ANOVA Model 2.3. Simultaneous Confidence Intervals: Scheffe's, Bonferroni and Turkey's interval. 2.4. Two – way classification model with and without interaction effect, one observation per cell and r observations per cell. Tukey's test for non additivity. 2.5. Two – way classification model with and without interaction effect with unequal number of observations per cell.	15
3	**Analysis of Variance: Random, Mixed Effect Models and ANCOVA** **3.1** Analysis of variance with random and Mixed effect models: Estimation and testing of variance components in one-way, two-way and multiway classification models. ANOVA method. **3.2** Analysis of Covariance: Model, BLUE, ANCOVA table, testing of hypothesis, use of ANCOVA for missing observation.	15

Self-Learning topics (Unit wise)

Unit	Topics
1	1.1. Basic operations: Vector Spaces, Linear dependence and independence, Determinants, Rank of Matrices, generalized inverse, and Solving linear equations. 1.2. Characteristic roots and characteristic vectors, properties of characteristics roots , Idempotent matrix, Quadratic forms, positive and Positive semi definite matrix

Online Resources

1. 'Linear Algebra' by Prof. Pranav Haridas from Kerala School of Mathematics available on the Swayam portal
   https://nptel.ac.in/courses/111/106/111106135/
2. 'Basic Linear Algebra' by Prof. I. K .Rana from IIT Bombay available on the Swayam portal
   https://nptel.ac.in/courses/111/101/111101115/

Course Code: STA506B
Title of the paper: Computational Statistics using R

Unit	Content	No. of Hours
1	**Introduction to R:** 1.1. Downloading and installation of R; http://www.r-project.org. Setting of working directory, Entering and manipulating data in R, 1.2. Basic classes of objects (character, numeric, integer, complex, logical), Vectors and their attributes (names, length, type), Arrays and Data types (matrices, frames, list), Combining data (cbind, rbind). 1.3. Matrix operations (addition, subtraction, multiplication, determinant, diagonal, trace, rank), Generating sequences, function repeats, component extraction (for vectors, matrices, list, frames). 1.4. Creating factors, installing packages and library, importing data from other sources (Excel and SPSS). Data input/output functions. Introduction to dplyr and data.table packages	15
2	**Data Handling using R :** 3.1. Data Manipulation: Selecting random N rows, removing duplicate row(s), dropping a variable(s), Renaming variable(s), sub-setting data, creating a new variable(s), selecting of random fraction of row(s), appending of row(s) and column(s), simulation of variables. 3.2. Data Processing: Data import and export, setting working directory, checking structure of Data, Changing type of variable	15

	3.3. Data Visualisation using GGplot: Introduction to various Components of GGplot, Basic Components, Aesthetics, Line, Annotation, Titles and Themes, bar chart,, column chart, clustered column chart, histogram, frequency polygon, Box plot for one and more variables, , violin plot, scatter plot, Quartile plot, Jitter plot, Line plot using R (ggplot)	
4	**Statistical Computing using R** 3.1 Descriptive Statistics: Mean, Median, Mode, Standard Deviation, Variance, Coefficient of Variation, Quartiles. Positional Averages, Skewness, Kurtosis, Correlation, Curve Fitting and Regression. **Text and Sentiment Analytics:** 3.2 Importing Text from a Single File, Multiple Files, Folders, and the Web (Fetching data from Twitter with and /or without an authenticated account.) 3.3 Text Collection and Transformation, Text Cleaning Sequence and Operations, 3.4 Finding Frequent Terms, Word Associations, Frequency Table 3.5 Visualization - Word Cloud and Word Cloud 2 3.6 Comparison Cloud and Common Cloud 3.7 Sentiment Analysis on Individual File 3.8 Sentiment Analysis on Corpus 3.9 Text mining and Modeling. 3.10 Story Telling on ➢ Extract data for your favorite celebrity and perform sentiment analysis ➢ Extract data from Twitter of top 4 metro cities location and compare the polarity and subjectivity analysis ➢ Extract data from your favorite motivational book and performs sentiment analysis and create world cloud	15

Self-Learning topics (Unit wise)

Unit	Topics
4	3.1 Descriptive Statistics: mean, median, mode, standard deviation, variance, coefficient of variation, skewness, kurtosis, quartiles. Positional Averages, Skewness, Kurtosis, Correlation, Curve Fitting and Regression.

Online Resources

1. 'Introduction to R Software' by Prof. Shalabh from IIT Kanpur available on the Swayam portal
   https://nptel.ac.in/courses/111/104/111104100/ for unit 4
2. 'Descriptive Statistics with R Software' by Prof. Shalabh from IIT Kanpur available on the Swayam portal
   https://nptel.ac.in/courses/111/104/111104120/
3. 'Introduction to R' by Santu Ghosh from Rajiv Gandhi University of Health Sciences available on the Swayam portal
   http://ugcmoocs.inflibnet.ac.in/ugcmoocs/view_module_pg.php/1851

Course Code: STA507B
Title of the paper: Computational Statistics using Python

Unit	Content	No. of Hours
1	**Introduction to Python:** 1.1. Python Setup, Python Arithmetic, Basic Data Types, Variables, Lists, Tuples and Strings, Dictionaries and sets. 1.2. Numpy arrays: Creating arrays crating n-dimensional arrays using np.array and array operations(indexing and slicing, transpose, mathematical operations) 1.3. Pandas data frames: Creating series and data frames and Operations on series and data frames 1.4. Reading and writing data: From and to Excel and CSV files 1.5. Control statements: if, if-else, if-else-if, while loop, for loop 1.6. Defining functions: def statement 1.7. Text data operations: len, upper, lower, slice, replace 1.8. Web scrapping and API to connect with online sites	15
2	**Data Handling and Visualisation:** 2.1. **Data Manipulation**: Selecting random N rows, removing duplicate row(s), dropping a variable(s), Renaming variable(s), sub-setting data, creating a new variable(s), selecting of random fraction of row(s), appending of row(s) and column(s), simulation of variables. 2.2. **Data Processing**: Data import and export, setting working directory, checking structure of Data, Changing type of variable, Data split into training and Test 2.3. **Data Visualisation**: Simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram, Box plot for one and more variables, histogram, frequency polygon, scatter plot, correlation plot Time series, Relationship maps, Heat maps, Geo Maps, 3-D Plots, Higher-Dimensional Plots using Matplotlib, Plotly, Seaborn 2.4. **Creating Dashboard, Comparison of Data Sets and Storytelling with** using Panda Profiling, Sweetviz, Autoviz. 2.5. **Story telling** on Data Sets **Iris /Pigeons/Car Design Data Set** for 2.1 to 2.4	15
3	**Statistical Computing using Python** 3.1 Descriptive Statistics: Mean, Median, Mode, Standard Deviation, Variance, Coefficient of Variation, Quartiles. Positional Averages, Skewness, Kurtosis, Correlation, Curve Fitting and Regression. **Text and Sentiment Analytics:** 3.2 Importing Text from a Single File, Multiple Files, Folders, and the Web (Fetching data from Twitter with and /or without an authenticated account.) 3.3 Text Collection and Transformation, Text Cleaning Sequence and Operations, 3.4 Finding Frequent Terms, Word Associations, Frequency Table 3.5 Visualization - Word Cloud and Word Cloud 2	15

| | 3.6 Comparison Cloud and Common Cloud |
| 3.7 Sentiment Analysis on Individual File |
| 3.8 Sentiment Analysis on Corpus |
| 3.9 Text mining and Modeling. |
| 3.10   Story Telling on |

3.6 Comparison Cloud and Common Cloud
3.7 Sentiment Analysis on Individual File
3.8 Sentiment Analysis on Corpus
3.9 Text mining and Modeling.
3.10   Story Telling on
➢ Extract data for your favorite celebrity and perform sentiment analysis
➢ Extract data from Twitter of top 4 metro cities location and compare the polarity and subjectivity analysis
➢ Extract data from your favorite motivational book and performs sentiment analysis and create world cloud

Self-Learning topics (Unit wise)

Unit	Topics
2	1.1. **Data Visualisation**: Simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram, Box plot for one and more variables, histogram, frequency polygon, scatter plot, Higher-Dimensional Plots using Matplotlib, Plotly, Seaborn
4	3.1 Descriptive Statistics: mean, median, mode, standard deviation, variance, coefficient of variation, skewness, kurtosis, quartiles. Positional Averages, Skewness, Kurtosis, Correlation, Curve Fitting and Regression.

Online Resources

1.   'Data Analytics with Python' by Prof. A. Ramesh from IIT Roorkee available on the Swayam portal
   https://nptel.ac.in/courses/106/107/106107220/   for unit 1 and unit 3

Part -4 Detailed Scheme Practical

Computer Application and Practical of Semester I   Total Credit: 04

Paper Code	Title	No. of Hours
STA505D	1. Probability 2. Advance Probability 3. Inequalities 4. Discrete Random Variables 5. Continuous Random Variables 6. Bivariate frequency Distribution 7. Moment Generating and Characteristic Functions 8. Probability Generating Function 9. Inequalities 10. Discrete probability distributions 11. Continuous probability distributions 12. Central Limit Theorem	02 Hours per Practical per Batch*

STA506D	1. Application of Characteristics of Good Estimator 2. Methods of Moments 3. Methods of Maximum likelihood Estimation 4. Minimum Chi-square Method 5. Uniform Minimum Variance Unbiased Estimation 6. Lower bounds for variance 7. Bayes' Estimation 8. Testing of Hypothesis 9. MP test and UMP test 10. Likelihood Ratio Tests 11. Wald's SPRT 12. Interval Estimation	
STA507D	1. Matrix Theory-I (Determinant, Rank of Matrix, Inverse of matrix) 2. Matrix Theory-II (Generalized Inverse, Simultaneous Linear Equations, Characteristics roots & Characteristics Vectors) 3. Linear Model-I 4. Linear Model-II 5. Techniques for Checking Assumptions of ANOVA 6. One way classification model 7. Two way Classification Model -I 8. Two way Classification Model-II 9. Random Effect Models 10. Analysis of Covariance	
STA508D	1. Introduction to R 2. Elementary calculation using R 3. Data processing and Manipulation using R 4. Matrix operations using R 5. Visualization using R –I 6. Visualization using R–II 7. Measures of Central Tendency, Dispersion, Skewness and Kurtosis 8. Visualization - Word Cloud and Word Cloud 2 9. Comparison Cloud and Common Cloud 10. Sentiment Analysis on Individual File 11. Sentiment Analysis on Corpus 12. Text mining and Modeling	
STA509D	1. Introduction to Python 2. Elementary calculation using Python 3. Data processing and Manipulation using Python 4. Matrix operations using Python 5. Visualization using Python –I 6. Visualization using Python–II 7. Measures of Central Tendency, Dispersion , Skewness and	

Kurtosis
8. Visualization - Word Cloud and Word Cloud 2
9. Comparison Cloud and Common Cloud
10. Sentiment Analysis on Individual File
11. Sentiment Analysis on Corpus
12. Text mining and Modeling

Course Code: STA501A
Title of the paper: Research Methodology

Unit	Content	No. of Hours
1	**Introduction to Research Methodology** 1.1 Introduction 1.2 Meaning and Definition of Research 1.3 Characteristics of Research 1.4 Role and Significance of Research 1.5 Objectives of Research 1.6 Types of Research 1.7 Research Process 1.8 Significance of Research 1.9 Problems Encountered by Researchers	15
2	**Literature Review and Systematic Literature Review** **Literature Review** 2.1 Reviewing the literature 2.2 Selecting and evaluating literature 2.3 Writing the literature review 2.4 Using referencing and citation **Systematic Literature Review** 2.5 Theory of planned behaviour, Systematic Literature Review: planning stage, conducting stage, reporting stage, hypothesis development, bibliometric analysis, meta analysis: converting all statistics to Effect Size	15
3	**Meta Analysis** 3.1 Meta Analysis and Its Applications: 3.2 What is Meta Analysis? Why Meta Analysis? When to go for Meta Analysis? 3.3 Concept in Meta Analysis: Assumptions, Summary Effect, Fixed Effect / Random Effect / Mixed Effect, Heterogeneity-Identification and Quantification, Confidence Interval vs Prediction Intervals, Subgroup Analysis, Meta-Regression. Additional Concepts in Meta-Analysis: Vote Counting, Power Analysis, Reporting /Publication Bias, Cumulative Meta Analysis, Reporting Meta- Analysis, Project / Case studies on Meta Analysis	15

4	**Understanding Steps of Research** 4.1 Research Design 4.2 Design of Sample Survey 4.3 Scale: Concepts, Measurement, Construction and Analysis 4.4 Objective wise analysis 4.5 Manuscript writing	15

Self-Learning topics (Unit wise)

Unit	Topics
2	Literature Review
3	Project/ Case studies on Meta Analysis
4	Manuscript writing

Online Resources

Kothari, C.R. & Garg Gaurav, Research Methodology : Methods and Techniques, New Age International Pvt Ltd Publishers
Not Available
'Research Methodology ' available on the NPTEL-NOC IITM portal https://www.youtube.com/@nptel-nociitm9240/search?query=research%20methodology **and** https://www.youtube.com/watch?v=b4GnUdkEAWo

References:
Course Code: STA503B:
1. Apostol, T. M. (1974): Mathematical Analysis. 2nd edition, Narosa Publishing house
2. Bartle G. and Sherbet, D. R. (2000): Introduction to Real Analysis. 3rd edition. Wiley
3. Kumar, A and Kumaresan S. (2015): A Basic course in Real analysis. CRC Press.
4. Malik, S. C. and Arora, S. (2017): Mathematical Analysis. 5th edition. New age International Publishers.
5. Rudin, W. (1976): Principles of Mathematical Analysis. 3rd edition. McGraw-Hill.
6. Bhat B.R. (1999): Modern Probability Theory: An Introductory test book. 3rd edition. New Age International.
7. Bhat B.R. : Modern Probability Theory, New Age International.
8. Rohatgi V.K. & Saleh A.K. Md. Ehasanes (2001) - An Introduction to Probability and Statistics. Wiley
9. Chandra, T. and Gangopadhyay, S. (2017): Fundamentals of Probability Theory. Narosa Publishing House.
10. Ross, S. M. (2014): Introduction to Probability Models. 11th edition. Elsevier.
11. Johnson, N. L., Kotz S. and Balakrishnan, N (2005): Univariate Discrete Distributions. Wiley.
12. Johnson, N. L., Kotz S. and Balakrishnan, N (2004): Continuous Univariate Distributions. Volume-I. Wiley.

13. Johnson, N. L., Kotz S. and Balakrishnan, N (2004): Continuous Univariate Distributions. Volume-II. Wiley.
14. Probability-A Graduate Course, Gut, A., Springer

Course Code: STA504B:
1. Rohatgi, V. K. and Saleh, A.K. M.E.: An Introduction to Probability Theory and Mathematical Statistics,
2. Rao, C.R.: Linear Statistical Inference and its Applications.
3. Lehmann, E.L.: Theory of Point Estimation (Student Edition).
4. Goon, A.M., Gupta, M.K.: An Outline of Statistical Theory
5. Kale, B.K.: A first Course on Parametric Inference
6. Dudewicz, E.J. and Mishra, S.N.: Modern Mathematical Statistics
7. Lehmann, E.L.: Testing Statistical hypotheses (Student Edition).
8. Ferguson, T.S.: Mathematical Statistics.
9. Zacks, S.: Theory of Statistical Inference

Course Code: STA505B:
1. Hohn Franz E : Elementary Matrix Algebra
2. Searle S.R. : Matrix Algebra useful for Statistics
3. Kshirsagar A.M. : A course in Linear Models
4. Wang S. GUI and Chow S.C. : Advanced Linear Models.
5. Healy M. J. R. : Matrices for Statistics
6. Shantinarayan : Textbook of Matrices
7. Finney D, J:- Statistical methods in biological assays.
8. Graybill F.A.:- An introduction to linear statistical models Vol. I.
9. Rao C.R.:- Linear statistical inference and its applications.
10. Searle S.R.:- Linear models.
11. Scheffe H.:- Analysis of variance.

Course Code: STA506B:
1. **Zuur**, Alain, **Ieno**, Elena N., **Meesters**, Erik: A Beginner's Guide to R, Springer
2. Torsten Hothorn, Brian S. Everitt: A Handbook of Statistical Analyses using R, Chapman and Hall/CRC Press, 3rd Edition
3. Pierre-Andre Cornillon, Arnaud Guyader, Francois Husson, Nicolas Jegou, Julie Josse, Maela Kloareg, Eric Matzner-Lober, Laurent Rouvière: R for Statistics, Chapman and Hall/CRC .
4. Asha Jindal(Ed): Analyzing and Visualizing Data with R Software, Shailja Prakashan and K.C.College, 2018

Course Code: STA507B:
1. Mark Lutz: Programming Python, O'Reilly Media, 4th Edition
2. Wes McKinney: Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly Media, 2nd Edition
3. Kenneth A. Lambert: The Fundamentals of Python: First Programs, 2011, Cengage Learning

4. Asha Jindal(Ed): Analyzing and Visualizing Data using Free Open Source Software: Python Programming with Case Studies, Shailja Prakashan and K. C. College, 2020.

Course Code: STA501A:

1. Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.
2. Kumar, R. (2019). Research methodology: A step-by-step guide for beginners. Sage publications.
3. Sekaran, U., & Bougie, R. (2016). Research methods for business: A skill-building approach. John Wiley & Sons.
4. Polit, D. F., & Beck, C. T. (2017). Nursing research: Generating and assessing evidence for nursing practice. Wolters Kluwer.
5. Kothari, C.R. & Garg Gaurav, Research Methodology : Methods and Techniques, New Age International Pvt Ltd Publishers
6. Andy P. Field, ADVENTURE IN STATISTICS: The Reality Enigma, SAGE Publications Ltd; Second edition
7. Andy P. Field, Discovering Statistics Using IBM SPSS Statistics, SAGE Publications Ltd; Fifth edition
8. S.P.Robbins(1998), Organizational Behaviour: Concepts, Contraversies and applications, Seventh Edition,Prentice Hall of India
9. Wikipedia forvK20 formula on Kunder-RichardsonFormula 20
10. Naresh K. Malhotra(2021), Marketing Research: An Applied Orientation, Pearson, 7th edition.
11. Cochrane Handbook for Systematic Reviews of Interventions, Cochrane Book Series Edited by Julian PT Higgins and Sally Gree, Wiley~ Blackwell
12. Michael Borenstein, Larry V. Hedges, Julian P. T.   Higgins, Hannah R. Rothstein, Introduction to Meta Analysis, Wiley
13. John E. Hunter and Frank L. Schmidt, Methods of Meta- Analysis (Correcting Error and Bias in Research Findings), Wiley, Second Edition
14. Mike W.L.Cheung, Meta Analysis: A structural equation modeling Approach,Wiley

Part 5- The Scheme of Teaching and Examination is as under:
First Year Semester – II Summary

Sr. No.	Choice Based Credit System		Subject Code	Remarks
1	Core Course (**Statistics**)		STA508B,STA509B, STA5010B,STA510D,STA511D,STA512D	
2	Elective Course	Discipline Specific Elective (DSE) Course	STA5011B/ STA5012B and STA515D STA516D	
		2.1 Interdisciplinary Specific Elective (IDSE) Course		
		2.2 Dissertation/Project		
		2.3 Generic Elective (GE) Course		
3	Ability Enhancement Courses (AEC)			
	Skill Enhancement Courses (SEC)			
4	Internship		STA502A	

Detail Scheme

Sr. No.	Subject Code	Subject Title	Periods Per Week						Seasonal Evaluation Scheme			Total Marks
			Units	S. L.	L	T	P	Credit	S. L. E	CT	SE	
1	STA508B	Multivariate Analysis	3	20%*	3	0	0	3	15		60	100
	STA510D	**Practical based on STA508B**					2	1			25	
2	STA509B	Sample Surveys	3	20%*	3	0	0	3	15		60	100
	STA511D	**Practical based on STA509B**					2	1			25	
3	STA5010B	Regression Analysis	3	20%*	3	0	0	3	15		60	100
	STA512D	**Practical based on STA5010B**					2	1			25	
4	STA511B	Machine Learning Techniques	3	20%*	3	0	0	3	15		60	100
	STA512B	Demography	3	20%*	3	0	0	3	15		60	
	STA515D or STA516D	**Practical based on STA511B / STA512B**					2	1			25	

| 5 | STA502A | **Internship** | 4 | | | | | 4 | | | | | 100 |
| | Total Hours / Credit | | 20 | | | | | | | | | | 500 |

*One to two lectures to be taken for CONTINUOUS self -learning evaluation.

First Year Semester – II Units – Topics – Teaching Hours

S.N	Subject Code		Subject Unit Title	Hours/Lectures	Total No. of hours/ lectures	Credit	Total Marks
1	STA508B	I	Introduction to Multivariate Distribution	15	45 H	3	100 (60+40)
		II	Hotelling's T2, Regression and MANOVA	15			
		III	Discriminant Analysis and its Application	15			
	STA508D	IV	**Practical based on** STA508B	30	30 H	1	
2	STA509B	I	Basic Sample Designs	15	45 H	3	100 (60+40)
		II	Advance Sample Designs-1	15			
		III	Advance Sample Designs-2	15			
	STA509D	IV	**Practical based on** STA509B	30		1	
3	STA510B	I	Multiple Linear Regression	15	45 H	3	100 (60+40)
		II	Regression Diagnostics	15			
		III	Advanced Regression Models and Generalized Linear Models	15			
	STA510D	IV	**Practical based on** STA510B	30		1	
4	STA511B	I	Introduction to Machine Learning techniques	15	45 H	3	100 (60+40)
		II	Classification Techniques	15			
		III	Naïve Baye's Classification and Market Basket Analysis Using Association Rules	15			
	STA512B	I	Introduction to Demography	15	45 H	3	
		II	Measures of Fertility, Mortality and Construction of Life Tables	15			
		III	Measures of Migration, Population Estimation and Projections	15			
	STA511D / STA512D		**Practical based on** STA511B or STA512B	30		1	
6	STA502A		**Internship**	60	60H	4	100
			TOTAL			20	500

- **Lecture Duration – 1 hour**
- **One Credit =15 Classroom hours**

L: Lecture: Tutorials P: Practical Ct-Core Theory, Cp-Core Practical, SLE- Self learning evaluation
CT-Commutative Test, SEE- Semester End Examination , PA- Project Assessment, AT- Attendance

Curriculum Topics along with Self-Learning topics - to be covered, through self-learning mode along with the respective Unit. Evaluation of self-learning topics to be undertaken before the concluding lecture instructions of the respective UNIT

Course Code: STA508B
Title of Paper: **Multivariate Analysis**

Unit	Content	No. of Hours
1	**Introduction to Multivariate distributions** 1.1 Multivariate normal distribution: definition, conditional & marginal distributions, characteristic function. 1.2 Random sample from multivariate normal distribution. Maximum likelihood estimators of parameters. 1.3 Distributions of sample mean vector and variance-covariance matrix and their independence. 1.4 Null Distribution of partial and multiple correlation coefficients and Interval Estimation and Application in Testing.	15
2	**Hotelling's T^2, Regression and MANOVA** 2.1 Hotelling's T^2 distribution and its applications. 2.2 Wishart distribution and its properties. 2.3 Multivariate Analysis of Variance (MANOVA) for one way and two way.	15
3	**Discriminant analysis,** Principal Component analysis **and its Application** 3.1 Discriminant Analysis and Classification of a discriminant procedure for discriminating between two multivariate normal populations. 3.2 Sample discriminant function and tests associated with discriminant functions, probabilities of misclassification and their estimation. 3.3 Principal Component analysis: Estimation of Linear Functions and their extraction using Orthogonal Rotation.	15

Self-Learning topics (Unit wise)

Unit	Topics
1	1.1. Multivariate normal distribution: definition 1.2. Random sample from multivariate normal distribution.
2	2.4 Multivariate Analysis of Variance (MANOVA) for one way.
4	4.1 Principal Component analysis. 4.2 Factor Analysis

Online Resources

1. 'Applied Multivariate Statistical Modeling' by Prof. J. Maiti from IIT Kharagpur available on the Swayam portal
   https://nptel.ac.in/courses/111/105/111105091/ for unit I, II, III.

2. 'Applied Multivariate Analysis' by Prof. Amit Mitra and Prof. Sharmishtha Mitra from IIT Kanpur
   https://nptel.ac.in/courses/111104024/

Course Code: STA509B
Title of paper: Sample Surveys

Unit	Content	No. of Hours
1	**Basic Sample Designs** 1.1 Sample Surveys: Introduction to usual notations used in sampling. 1.2 Basic finite population sampling techniques: SRSWR, SRSWOR, stratified, systematic and related results on estimation of population mean/ total. 1.3 Relative precision of different sampling techniques. Allocation problem in stratified sampling. 1.4 Determination of Sample size (SRSWR & SRSWOR).	15
2	**Advance Sample Designs-1** 2.1 Ratio and regression estimators based on SRSWOR method of sampling. 2.2 Cluster sampling - equal and unequal size clusters. 2.3 Two-stage sampling with equal size of first stage units.	15
3	**Advance Sample Designs-2** 3.1 Double sampling for ratio and regression methods of estimation. 3.2 Varying Probability Sampling: Basic definitions and concepts, Probability proportional to size and sample drawing methods, PPS in sampling with replacement and related topics, PPS in sampling without replacement and ordered estimators, Unordered estimators, Murthy's Estimator and Horwitz Thompson Estimator and related topics, Horwitz Thompson Estimator and Midzuno System of Sampling	15

” **Swayam Prabha Course, Channel 16, MOE, GOI by Prof. Shalabh, IIT KanpurSelf-Learning topics (Unit wise)**

Unit	Topics
3.2	Varying Probability Sampling: Basic definitions and concepts, Probability proportional to size and sample drawing methods, PPS in sampling with replacement and related topics, PPS in sampling without replacement and ordered estimators, Unordered estimators, Murthy's Estimator and Horwitz Thompson Estimator and related topics, Horwitz Thompson Estimator and Midzuno System of Sampling

Online Resources

“Sampling Theor Source: https://www.youtube.com/playlist?list=PLqMl6r3x6BUTP4XPysDab-RrLAt4_PP6E

Course Code: STA510B
Title of paper: Regression Analysis

Unit	Content	No. of Hours
1	**Multiple Linear Regression** **1.1** Partial and multiple correlations, standard Gauss Markov setup, least squares (LS) estimation, variance-covariance of LS estimators, estimation of error variance, regression analysis with correlated observations, LS estimation with restriction on parameters.; properties of least squares residuals; forward, backward and stepwise regression; different methods for subset selection. **1.2** Dummy Variables: Concept and it's application	15
2	**Regression Diagnostics and Polynomial Regression** 2.1 Lack of fit (linearity): diagnostics and test, Model building. Heteroscedasticity: consequences, diagnostics, tests (including Breusch-Pagan test and White's test) and efficient estimation. 2.2 Autocorrelation: consequences, diagnostics, tests (including Durbin-Watson test, Breusch-Godfrey LM test) and efficient estimation. 2.3 Collinearity: consequences, diagnostics and strategies (including ridge & shrinkage regression). 2.4 Discordant outlier and influential observations: diagnostics and strategies. **2.5** Polynomial Regression Model and Variable selection and model building	15
3	Generalized Linear regression models: 3.1 Introduction 3.2 Logistic regression: Example, model, MLE of parameters, Iterative procedure to solve likelihood equations, multiple regressors. Multinomial and Ordinal Logistic Regression. 3.3 Poisson Regression 3.4 Ridge regression: Ill conditioned matrix, need of ridge regression, biased estimator, Mean square error. Bias and MSE of ridge estimator, ridge trace method.	15

Self-Learning topics (Unit wise)

Unit	Topics
3	3.1 Dummy Variables, Polynomial Regression Model 3.2 Introduction to Generalized Linear Models (GLMs) 3.3 Poisson Regression

Online Resources

1. 'Linear Regression Analysis and Forecasting' by Prof. Shalabh from IIT Kanpur available on the Swayam portal
   https://nptel.ac.in/courses/111/104/111104098/

2. 'Regression Analysis' by Prof. Soumen Maity from IISER Pune available on the Swayam portal
   https://nptel.ac.in/courses/111/105/111105042/
   for unit III and IV.
3. 'Linear Regression Analysis' by Prof. Shalabh from IIT Kanpur
   https://nptel.ac.in/courses/111104074/

Course Code: STA511B
Title of Paper: **Machine Learning Techniques**

	Learning Objective: The purpose of this course is to make use of machine techniques to discover patterns in the data and then make predictions based on often complex findings to answer business questions, detect and analyse trends and help solve problems. Machine learning is effectively a method of data analysis that works by automating the process of building data models.	
Unit	**Content**	**No. of Hours**

1	Introduction to Machine Learning techniques:
	    1.1 Introduction, Data storage, evolution of learning, Logical difference between predictive modelling & ML, Types of data, various algorithm, input data into algorithm, data frame / data manipulation, Prepare mode data set for respective ML technique, Data Preprocessing
	
	Classification Techniques I
	    1.2 Decision Tree
	    Understanding the decision trees, Divide and conquer process, The decision tree algorithm, Choosing best split, Pre-pruning & post pruning technique
	    1.3 Regression trees and model trees: Adding regression to trees, Process and Interpretation.
	    1.4 A review of Multiple Linear regression, Logistic Regression
	    1.5 Random Forests for Classification and Regression: Concept, Algorithm and Case Studies
		15

2	Classification Techniques II
	    2.1 Factor Analysis: KMO and Bartlett Test and their significance (without derivation), Scree Plot and other methods of Extraction
	    2.2 Cluster analysis:
	      a) Hierarchical Optimization Methods: Agglomerative Schedule based on linkage Method (Single, Complete and Average), Ward's Method, Centroid Method.
	      b) K-Mean Optimization Methods: Sequential Threshold, Parallel Threshold, Optimizing Partitioning.
	      c) Classification using KNN approach ,Case Study on the KNN algorithm:
	    ✓ Data collection / Transformation / Preparing pre-model data set.
	    ✓ Training a model on the data.
	    ✓ Evaluating model on the data.
	    ✓ Improving model performance.
	    ✓ Differences between k-nearest neighbor algorithm and k-means clustering
	    2.3 Linear class classification and Multi Class Classification-Support Vector Maachines(SVMs)
		15

3	Naïve Baye's Classification and Market Basket Analysis Using Association Rules:	15
	3.1 Naïve Baye's Classification: Introduction , Associated Probabilites, Case Study	
	3.2 Market Basket Analysis Using Association Rules	
	3.3 The apriori algorithm for association rule learning. Measuring rule interest – support and confidence, Building a set of rules with the apriori principle.	
	3.4 Case Studies: Identifying frequently purchased groceries with association rules: Data Collection / Preparation Creating a sparse matrix for transaction data Training a model on the data . Evaluating model performance . **3.5** Calculating support / Confidence / Expected confidence & lift . **3.6** Introduction to Natural Language Processing(NLP)	

Self-Learning topics (Unit wise)

Unit	Topics
1.1	Data Preprocessing
1.4	Regression trees and model tree : Adding regression to trees, Process and Interpretation.
2.1	Procedure for conducting cluster Analysis using ☐ Hierarchical **Clustering** Method, Agglomeration Schedule, vertical lcicle Plot with complete linkage, Dendogram with complete linkage
2.2	Classification using KNN approach:
3.1	Naïve Baye's Classification: Introduction , Associated Probabilites, Case Study

Online Resources

"Data Analytics with Python" by PROF. A RAMESH, Department of Management Studies, IIT Roorkee https://nptel.ac.in/courses/106/107/106107220/ for 1.4 and 2.1
"NOC:Data Mining" by Prof. Pabitra Mitra, Computer Science and Engineering, IIT Kharagpur https://nptel.ac.in/courses/106/105/106105174/   for 1.1 and 2.2
Business Analytics and Data Mining Modeling using R, available on the Swayam portal, GOI by Prof. Gaurav Dixit, IIT Roorkee given on week 10 **https://nptel.ac.in/courses/110/107/110107092/ for unit 3.1**

Course Code: STA512B
Title of Paper: **Demography**

Units	Topic	No. of Hours
1	1.1 Demography as a scientific discipline; Development of demography as a discipline. population theories' Malthus and Demographic transition. 1.2 Structure of Population: Measures of Age and Sex Composition of the Population. Age pyramid, Quality of Age data, Errors in demographic data – Whipple's Index, Myer's Index, UN Joint Score Index. 1.3 Completeness check of registration data with balancing equations and Chandrasekharan-Deming formula, 1.4 Sources of data – Census, Vital Statistics, Sample Surveys, Population registers, Quality of Data – Evaluation and Adjustment of Demographic Data Interpolation and Graduation 1.5 Population Growth: Measures of Population growth – Balancing Equation, Arithmetic, Geometric, Exponential, modify exponential, Gompertz and Logistic. Doubling Time 1.6 Stationary and Stable population	
2	2.1 Measurement of fertility: Crude Birth Rate (CBR), General Fertility Rate (GFR), Specific Fertility Rate (SFR) and Total Fertility Rate (TFR). 2.2. Measures of reproduction rate: Gross Reproduction Rate (GFR) and Net Reproductive Rate (NRR). 2.3 Age patterns of Fertility: Coale and Trussell Fertility Model: Estimating M and m, Bongaarts and Potters Aggregate Fertility Model and its applications 2.4 Mortality Measures – Introduction, Crude and Specific Rates, Standardised Rates Infant Mortality – Infant Mortality Rate, Neo-natal mortality rate, Post neonatal mortality, Peri natal mortality, Foetal Death, Maternal Mortality, Direct and indirect method of Standardization of Death Rates, Morbidity: Prevalence and Incidence Rates, Attributable risk, Relative risk 2.5 Life table: Definition and a Brief History, Types of Life Tables, Basic Assumptions for Preparing a Life Table, Description of a Complete Life Table, Interrelations among the Life Table Functions, Description of an Abridged Life Table 2.6 Methods of Construction of complete life table, Construction of abridged Life Table from ASDRs; Conventional Method, Reed-Merrell Method, Greville's Method, Chiang's Method	
3	**3.1 Measures of Migration** 3.1.1 Concept, data source, theories, and types of migration 3.1.2 Direct estimation of lifetime and inter-censal migration rates from census data. 3.1.3 Indirect measures of net internal migration: Vital Statistics Method, National Growth Rate Method, Census and Life Table Survival Ratio	

	methods.
	3.2 Population Estimates and Projections 3.2.1 Concepts of population projections; population estimates, forecasts and projections, 3.2.2 Methods of interpolation; extrapolation using linear, exponential, polynomial, logistics, Gompertz curves and growth rate models, cohort component method	

Self-Learning topics (Unit wise)

Unit	Topics
1	Sources of data – Census, Vital Statistics, Sample Surveys, Population registers, Quality of Data – Evaluation and Adjustment of Demographic Data Interpolation and Graduation
2	Definition and a Brief History, Types of Life Tables, Basic Assumptions for Preparing a Life Table
3.1	Concept, data source, theories, and types of migration
3,2	Concepts of population projections; population estimates, forecasts and projections,

Online Resources

Method of Construction of Complete Life Tables by Dr. Nick Negovetich on Completing a life table for Dall Sheep Link: https://www.youtube.com/watch?v=hCZveJwP9Rg
Introduction and sources of collecting data on vital statistics, error in census and registration data. Coverage and content error in demographic data https://www.youtube.com/watch?v=JYQzUavh97g
Internal migration and its measurement, concept of international migration. https://www.youtube.com/watch?v=_9Ep_79xTpI
University Grants Commission (UGC) ⬇ **e-PG Pathshala**
M-10. Evaluation and Adjustment of Errors in Age Sex Data https://www.youtube.com/watch?v=O7EyjZxhJbQ
M-12. Measures of Error in Age Sex Data https://www.youtube.com/watch?v=egCgEYZ1O0E&t=1385s
M-13. UN Age Sex Accuracy Index https://www.youtube.com/watch?v=rPe_uDJZ8SM
M-15. Methods of adjustment for age sex data https://www.youtube.com/watch?v=zR-HcNE6Iew

Part -4 Detailed Scheme Practical

Computer Application and Practical of Semester II   Total Credit: <u>04</u>

Paper Code	Title	No. of Hours
STA510D	1. Multivariate Normal Distribution 2. Application of Hoteling T^2 3. Multivariate Regression 4. Application of Wishart Distribution 5. Multiple and Partial Correlation Coeficient 6. Multivariate Analysis of Variance (MANOVA) –I 7. Multivariate Analysis of Variance (MANOVA) –II 8. Discriminant Analysis 9. Principal Component Analysis	
STA511D	1. Simple random sampling 2. Simple Random Sampling for Proportion 3. Stratified random sampling. 4. Systematic Random Sampling –I 5. Systematic Random Sampling –II 6. Ratio and Regression methods of Estimation. 7. Cluster sampling. 8. Two-stage and Two-phase sampling. 9. Varying Probability Sampling 10. Double Sampling 11. Determination of Sample Size	2 Hours per Practical
STA512D	1. Multiple linear Regression: Assumption Checking, Multicolliearity, Selection methods. 2. Regression Diagnostics 3. Multiple linear Regression : Dummy Variable Creation and Analysis 4. Polynomial Regression 5. Logistic Regression 6. Multinomial Logistic Regression 7. Poisson Regression 8. Ridge Regression	
STA515D	1. Basics of Machine Learning Techniques 2. Decision Tree for Classification and Regression 3. Multiple Linear Regression 4. Logistic Regression 5. Random Forests for Classification and Regression 6. Factor Analysis 7. Cluster Analysis 8. Linear class classification -Support Vector Maachines 9. Multi Class Classification-Support Vector Maachines 10. Naïve Baye's Classification 11. Market Basket Analysis	

STA516D	1. Evaluation and adjustment of age sex data 2. Completeness check of registered demographic data 3. Smoothing of demographic data 4. Population growth models 5. Measurement of fertility 6. Measurement of Reproduction rate 7. Age pattern of fertility models 8. Measurement of mortality 9. Construction of complete life table 10. Construction of abridge life table 11. Measurement of migration 12. Population estimation and projection	

*Batch Size of 10 students

References: Semester II
Course Code: STA508B:
1. Johnson Richard A and Wichern D.W.(1998) : Applied Multivariate Statistical Analysis (4th Edition)
2. Johnson Richard A and Wichern D.W. : Applied Multivariate Statistical Methods
3. Anderson T.W.(1958) : An Introduction to Multivariate Statistical Analysis.John Wiley & Sons
4. Dillon William R & Goldstein Mathew (1984) : Multivariate Analysis : Methods and Applications.
5. Giri Narayan C. (1995) : Multivariate Statistical Analysis.
6. Kshirsagar A. M. (1979) : Multivariate Analysis ,Marcel Dekker Inc. New York.
7. Hardle Wolfgang & Hlavka : Multivarite Statistics : Exercise & Solutions
8. Parimal Mukhopadhyay: Multivariate Statistical Analysis.
9. Srivastava, M. S. (2002): Methods of Multivariate Statistics. John Wiley.
10. Malhotra, N. K. (2007). *Marketing research: An applied orientation*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Course Code: STA509B:
1. Cochran W.G.: sampling techniques.
2. Raj, d and chandak: sampling theory.
3. Murthy, m.n.: sampling theory & methods.
4. Mukhopadhyay, p.: Theory and methods of survey sampling.
5. Sukhatme et. Al.: sampling theory of surveys with applications.
6. Bansal A.:Survey Sampling
7. Arijit Chaudhuri :Modern Survey Sampling
8. Guide to current indian official statistics, central statistical office, goi, new delhi.
9. http://mospi.nic.in/

Course Code: STA5010B:
1. Thomas P. Ryan, Modern Regression Methods.
2. Douglas C. Montgomery, Introduction to Linear Regression Analysis.
3. David A. Belsley, Edwin Kuh and Roy E. Welsch, Regression Diagnostics: Identifying Influential Data and Source of Collinearity.

4. Peter J. Rousseeuw and Annick M. Leroy, Robust Regression and Outlier Detection.
5. P. McCullagh and John A. Nelder. Generalized linear models.
6. Pagan, A. and A. Ullah, Nonparametric Econometrics.
7. Kshirsagar A.M.: A course in Linear Models
8. Draper N.R & Smith H: Applied Regression Analysis.
9. Song GUI Wang and S.C Chow: Advanced Linear Models.
10. Damodar Gujarati: Basic Econometricsw
11. George A. F. Seber, Alan J. Lee: Linear Regression Analysis
12. Chaterjee and Price: Regression Analysis with examples
13. Applied Logistic Regression, David W. Hosmer and Stanley Lemeshow, Wiley
14. Ratkowsky, D. A. (1983). Nonlinear Regression Modelling,
15. Seber, G. E. F. and Wild, C. J. (1989). Nonlinear Regression.

Course Code: STA5011B:
1. Lior Rokach & Oded Z. Maimon, Data Mining with Decision Trees - Theory & Applications.
2. Lantz, B (2013), Machine Learning with R, 2nd Ed, PACKT Open Source.
3. Miller, J. D. and Forte, R. M. (2015), Mastering Predictive Analytics with R, 2nd Ed, PACKT Open Source.
4. Babcock, J. (2016), Mastering Predictive Analytics with Python, PACKT Open Source
5. Naresh K. Malhotra(2021), Marketing Research: An Applied Orientation, Pearson, 7th edition.
6. https://analyticsindiamag.com/understanding-naive-bayes-classifier-from-scratch/ 10. https://analyticsindiamag.com/how-to-identify-entities-in-nlp
7. SPSS for Windows Step by Step: A simple Guide and Reference, Darren George and Paul Mallery, Pearson
8. Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications.
9. Brian C. Cronk, How to Use SPSS®: A Step-By-Step Guide to Analysis and Interpretation Paperback

Course Code: STA5012B:
1. Barclay G W Techniques of Population Analysis, New York, John Wiley and Sons, Inc
2. Hinde, Andrew Demographic Methds.London,1998
3. Jaffe A J Hand Book of Statistical Methods for Demographers, Washington, US Govt. Printing Office
4. MISRA B D An Introduction to the Study of Population, Madras, Publishing
5. Pathak, K.B.& F.RAM Techniques of Demographic Analysis, Mumbay, Himalaya Publishing house.
6. Pollard J H Demographic Techniques Australia, Pengamon Press
7. Preston, Samuel H, Patrick Heuveline and Michel Guillot: Demography – Measuring and Modeling Population Processes Ramakumar R Technical Demography, New Delhi, Wiley Eastern Ltd.
8. Shryock, Henry S, Jacob S Seigel and Associates, The Methods and Materials of Demography Vol. 1 & 2, Washington DC US Bureau of the Census
9. SPEEGELMAN M Introduction to Demography Cambridge, Harvard University Press Srinivasan K Basic Demographic Techniques and Applications, New Delhi Sage Publications
10. Srinivasan K Basic Demographic Techniques and Applications, New Delhi Sage

Publications

11. Gupta S.C. and Kapoor, V.K. (2008), Fundamentals of Applied Statistics, 4 th Ediction (Reprint), Sultan Chand & Sons
12. M.L. Jhingan ,B.K. Bhatt and J.N. Desai, Demography ,
13. Asha A. Bhende (Author), Tara Kanitkar (Author), Principles of Population Studies
14. Kumar, R. (1986), Technical Demography, Wiley Eastern Ltd. 6.
15. Benjamin, B. (1969), Demographic Analysis, George, Allen and Unwin.
16. Chiang, C.L. (1968), Introduction to Stochastic Progression.
17. Spiegelman, M. (1969), Introduction to Demographic Analysis, Harvard University Press. 9
18. Wolfenden, H.H. (1954), Population Statistics and Their Compilation, Am Actuarial Society

The Scheme of Teaching and Examination:
The performance of the learners shall be evaluated in two components: Internal Assessment with 40% marks by way of continuous evaluation and by Semester End Examination with 60% marks by conducting the theory examination.

Examination Pattern for First Year Degree as per NEP 2020 Academic Year 2023-2024

1) Evaluation of Major and Minor Subjects

Subject	Formative Assessment (Marks)	Summative Assessment (Marks)
Major Subject	40	60
Minor Subject	40	60
Major (Practical based Subject)	-	25
Minor (Practical based Subject)	-	25

.

FORMATIVE ASSESSMENT:- It is defined as the assessment of the learners on the basis of continuous evaluation as envisaged in the credit based system by way of participation of learners in various academic and correlated activities in the given semester of the programme.

A). Formative Assessment – 40%   **40 marks**

Practical's (internal Components of the Practical Course)

1. For Theory Courses

Sr.No.	Particulars	Marks
1	One class open book test / online examination to be conducted in the given semester/Project	25 Marks
2	Self-Learning Evaluation with Active participation in routine class instructional deliveries	10+5 Marks

2. For Courses with Practicals

Each practical course can be conducted out of 50 marks with 10 marks for internal **component of the Practical** and 40 marks for formative assessment which will be converted to 25 marks.

Practical's (Internal component of the Practical Course)

Sr. No	Evaluation type	Marks
1	Journal	5
2	Viva	5

B). SUMMATIVE ASSESSMENT =SEMESTER END EXAMINATION :-
It is defined as the examination of the learners on the basis ofperformance in the semester end theory / written examinations.

The semester end examination (external component) of 60 % for each course will be as follows:

a. **For Theory Courses**
   i) **Duration – 2 Hours** ii) **Theory Question**

   Paper Pattern:-

i. There shall be four questions each of 15 marks. On each unit there will be one question and the fourth one will be based on entire syllabus.
ii. All questions shall be compulsory with internal choice within the questions. Question may be subdivided into sub-questions a, b, c… and the allocation of marks depend on the weightage of the topic.

b. **For Practical Courses**
   i) **Duration – 2 Hours** ii) **Practical Question Paper Pattern**:-

1. There shall be three questions each of 10 marks. On each unit there will be one question based on the syllabus and the fourth one will be based on entire syllabus.
2. All questions shall be compulsory with internal choice within the questions. Question may be subdivided into sub-questions a, b, c… and the allocation of marks depend on the weightage of the topic.