KANNUR UNIVERSITY

Syllabus

M.Sc. Statistics

(Effective from 2023 admission)

Choice Based Credit and Semester System For Post Graduate Programmes in Affiliated Colleges

(OBE – Outcome Based Education – System)

(KUCBCSSPG 2023)

CREDIT AND MARK DISTRIBUTION FOR M.Sc STATISTICS

Semester	Course Code	Course	Marks			Credit
			Internal	External	Total	
I	MSSTA01C01	Course 1. 1	15	60	75	4
	MSSTA01C02	Course 1. 2	15	60	75	4
	MSSTA01C03	Course 1. 3	15	60	75	4
	MSSTA01C04	Course 1. 4	15	60	75	4
	MSSTA01C05	Course 1. 5	15	60	75	4
Total			75	300	375	20
II	MSSTA02C06	Course 2. 1	15	60	75	4
	MSSTA02C07	Course 2. 2	15	60	75	4
	MSSTA02C08	Course 2. 3	15	60	75	4
	MSSTA02C09	Course 2. 4	15	60	75	4
	MSSTA02C10	Course 2. 5	15	60	75	4
Total			75	300	375	20
III	MSSTA03C11	Course 3. 1	15	60	75	4
	MSSTA03C12	Course 3. 2	15	60	75	4
	MSSTA03E- -	Course3. 3	15	60	75	4
	MSSTA03O- -	Course 3. 4	15	60	75	4
	MSSTA03C13	Course 3. 5	15	60	75	4
Total			75	300	375	20
IV	MSSTA04E - -	Course 4. 1	15	60	75	4
	MSSTA04E- -	Course 4. 2	15	60	75	4
	MSSTA04C14	Course 4. 3	15	60	75	4
	MSSTA04C15	Course 4. 4	20	100	120	6
	MSSTA04C16	Course 4. 5	10	20	30	2
Total			75	300	375	20
Grand Total			300	1200	1500	80

Course Code	Course Title	Hours per Week	Credits	Marks
I Semester (Total Credits: 20)				
MSSTA01C01	Measure and Probability	5	4	75
MSSTA01C02	Mathematical Methods for Statistics	5	4	75
MSSTA01C03	Distribution Theory	5	4	75
MSSTA01C04	Sampling Methods and Applications	5	4	75
MSSTA01C05	Data Analytics Using R - I (Practical)	5	4	75
II Semester (Total Credits: 20)				
MSSTA02C06	Stochastic Processes	5	4	75
MSSTA02C07	Estimation Theory	5	4	75
MSSTA02C08	Regression Methods	5	4	75
MSSTA02C09	Design and Analysis of Experiments	5	4	75
MSSTA02C10	Data Analytics Using R – II (Practical)	5	4	75
III Semester (Total Credits: 20)				
MSSTA03C11	Multivariate Analysis	5	4	75
MSSTA03C12	Testing of Hypotheses	5	4	75
MSSTA03E- -	Elective -I	5	4	75
MSSTA03O- -	Open Elective (Multi Disciplinary)	5	4	75
MSSTA03C13	Data Analytics Using Python (Practical)	5	4	75
IV Semester (Total Credits:20)				
MSSTA04E- -	Elective - II	5	4	75
MSSTA04E- -	Elective–III	5	4	75
MSSTA04C14	Data Analytics Using SAS (Practical)	5	4	75
MSSTA04C15	Project Work	10	6	120
MSSTA04C16	Viva Voce	0	2	30

Syllabi of Courses Offered in Semester III

MSSTA03C11- Multivariate Analysis (4 Credits)

Course Outcomes:

CO 1: Understanding the theoretical foundations of multivariate statistical techniques.

CO2 : Learning various multivariate analysis methods such as principal component analysis, factor analysis and discriminant analysis.

CO3: Developing skills to apply multivariate techniques to analyze complex data sets with multiple variables.

CO4: Gaining proficiency in using software packages for conducting multivariate analyses

Unit 1: Multivariate Normal distribution

Definition, non singular and singular multivariate normal distributions, properties of multivariate normal distribution, characteristic function, marginal and conditional distributions, independence, reproductive property. Independence of a linear and a quadratic form, independence of two quadratic forms, distribution of quadratic form of a multivariate normal vector.

Unit 2: Estimation of mean vector and covariance matrix

Maximum likelihood estimation of mean vector and dispersion matrix, distribution of sample mean vector, inference concerning the mean vector when the dispersion matrix is known for single and two populations. Wishart distribution, characteristic function of Wishart distribution, properties, generalized variance, distribution of sample generalized variance.

Unit 3: Tests based on Likelihood ratio criterion

Likelihood ratio test, one sample Hotelling's T^2 statistic, distribution and properties, two sample Hotelling's T^2 statistic, testing significance of a mean vector and equality of mean vectors, Mahalanobis' D^2 statistics, uses of T^2 and D^2 in testing problems, testing independence of sets of variates, proportionality of covariance matrix and equality covariance matrices, sphericity test, Fisher-Behren problem.

Unit 4: Classification problem

Classification to one of two multivariate normal populations when the parameters are known and unknown. Extension of this to several multivariate normal populations, Fisher's discriminant function, principal components - definition and derivation, canonical correlation - definition and derivation, factor analysis.

Books for Study:

1. Anderson, T. W. (2009). ***An Introduction to Multivariate Statistical Analysis, 3rd edition***

Wiley.

2. Johnson, R. A. AND Wichern, D. W. (2012). ***Applied Multivariate Statistical Analysis***, ***6th Edition.*** Prentice Hall.

3. Rao, C. R. (2009). ***Linear Statistical Inference and Its Applications***, ***2nd Ed***. John Wiley & Sons.

Books for Reference:

1. Barbera, G, Tabachnick and Linda, S. Fidell. (2006). ***Using Multivariate Statistics, 5th Edition,*** Harper & Row

2. Johnson, N.L. & Kotz, S.(2000). ***Continuous Multivariate Distributions,*** Wiley-Interscience.

3. Morrison, D. F. (1990). ***Multivariate Statistical Methods,*** McGraw Hill.

4. Takeuchi, K., Yanai, H. and Mukherjee, B. N. (1983). ***The Foundations of Multivariate Analysis.*** Wiley

5. Giri, N. C. (1996). ***Multivariate Statistical Analysis.*** Marcel Dekker. Inc., New York.

MSSTA03C12- Testing of Hypotheses (4 Credits)

Course Outcomes:

CO 1:. Understand the principles and concepts of hypothesis testing

CO 2: A detailed study of Fisher Neyman method of testing of hypothesis,

likelihood ratio test, non - parametric testing of hypothesis and sequential testing.

CO 3: Masterising various test statistics and their applications

CO 4: Developing skills in choosing appropriate tests for different situations

Unit 1: Neymann- Pearson approaches

Problem of testing of hypotheses- Classical and Neymann- Pearson approaches, parametric and nonparametric tests, simple and composite hypotheses, size and power of a test, p-value, randomized and non randomized tests. The Neymann Pearson lemma and its generalization. MP tests and UMP tests, one sided and two sided tests.

Unit 2: Likelihood ratio tests

UMPU tests- completeness and bounded completeness, similar regions, Neyman's structure and its applications. Likelihood ratio tests and their properties. Asymptotic distribution of likelihood ratio tests.

Unit 3: Sequential probability ratio tests

Sequential probability ratio tests- OC and ASN functions, properties of SPRT, Wald's identity and its applications. Tests based on Binomial, Poisson and normal distributions. Ranking and selection procedures.

Unit 4: **Non parametric tests**

Non parametric tests- Chi-square tests, Kolmogorov-Smirnov test, sign test, signed rank test, Wald-Wolfowitz run test, median test, Wilcoxon test, Mann Whitney U-test, one and two sample cases. Nonparametric confidence interval. Kruskal Wallis test and Friedmann test

Books for Study :

1. Gibbons, J. D. and Chakraborti, S. (2014). ***Nonparametric Statistical Inference, 4th edition***. Taylor and Francis

2. Kale, B. K.(2005). ***A First Course on Parametric Inference, 2nd edition,*** Narosa

3. Rohatgi, V. K. Saleh, A.K.M.E. (2015). ***Introduction to Probability Theory and Mathematical Statistics***. John Wiley & Sons.

Books for Reference:

1. Lehmann.E and Romano, J. P. (2010). ***Testing Statistical Hypotheses***. Springer

2. Kendall, M. G. & Stewart, A. (2008). ***The Advanced Theory of Statistics Vol. 2 , 2nd Ed***. John Wiley and Sons

3. Mood, A. M.& Graybill, F. A.(2009). ***Introduction to the Theory of Statistics***. Tata McGraw Hill 4. Rao, C. R(2009). ***Linear Statistical Inference and Its Applications.*** John Wley & Sons, New York.

5. Siegal, S. and Castellan, N.J. (1988). ***Nonparametric Methods for the Behavioural Sciences, 2nd Ed.*** McGraw-Hill.

6. Zacks, S. (1971). ***Theory of Statistical Inference***. John Wiley & Sons.

MSSTA03C13 – Data Analytics Using Python (Practical) (4 Credits)

Course Outcomes:

CO 1: Understand the basics of Python programming.

CO 2: Analyze various object oriented concepts.

CO 3: Apply Python tools for statistical analysis

CO 4: Demonstrate the use of graphical representations for data analytics.

Unit l: Basics of Python

Installing Python - basic syntax - interactive shell - editing, saving and running a script. The concept of data types - variables- assignments - mutable type - immutable types - arithmetic operators and expressions - comments in the program - understanding error messages - Control statements - Operators.

Unit 2: Introduction to Pandas

Pandas data series - Pandas data frames - data handling - grouping Descriptive statistical analysis and Graphical representation. Hypothesis testing

Unit 3: Graphics using Python

Line graph - Bar chart - Pie chart - Heat map - Histogram - Box plot - Density plot - Cumulative frequencies - Error bars - Scatter plot - 3D plot.

Unit 4: - Data modeling

Linear regression models - Logistic regression model - Principal Component Analysis - K Mean clustering.

Books for Study:

1 . Lambert. K. A. (2018). ***Fundamentals of Python: First Programs.*** Cengage Learning.

2. Haslwanter T. (2016). ***An Introduction to Statistics with Python*****.** Springer International Publishing.

Books for Reference:

1. Perkovie, L. (2011). ***Introduction to Computing Using python: An Application Development Focus.*** Wiley Publishing

Syllabi of Courses Offered in Semester IV

MSSTA04C14- Data Analytics Using SAS (Practical) (4 Credits)

Course Outcomes:

CO 1: Understand the basics of SAS programming.

CO 2: Analyze various object oriented concepts and apply SAS tools for statistical analysis

CO 3: Equipped with different theoretical methods in applied statistics to achieve the objectives.

CO 4: Enhanced with the basic concepts of statistical theories besides developing their ability to handle real world problems with large scale data.

Unit 1: Basic Concepts in data analysis

Variables, values, quantitative variables versus classification variables, observational units, scale of measurements, basic approach for research, descriptive versus inferential statistical analysis, hypothesis testing. Introduction to SAS programs – What is SAS?, three types of SAS files; Data input – Inputting questionnaire data versus other types of data, inputting data using the DATALINES statement, inputting a correlation or covariance matrix.. Working with variables and observations in SAS – manipulating, subsetting, concatenating and merging data.

Unit 2: Simple descriptive data analysis

Introduction, PROC MEANS, creating frequency table with PROC FREQ, PROC PRINT, PROC UNIVARIATE, test for normality, stem-and-leaf plot, skewness. Analysis of bivariate data –

significance tests versus measures of association, levels of measurement, appropriate statistics, scattergrams with PROC GPLOT, Pearson correlation with PROC CORR, options used with PROC CORR, Spearman correlations with PROC CORR, two way classification table, tabular versus raw data, assumptions underlying Pearson correlation coefficient, Spearman correlation coefficient and chi square test of independence

Unit 3 : Small sample tests

Two types of t-test, independent samples t-test, independent variable and dependent variable, writing the SAS program, interpreting and summarizing the results. The paired samples t test, paired versus independent samples, problems with the paired samples approach, difference score variable, interpreting and summarizing the results, assumptions underlying the t tests.

Unit 4: One way ANOVA with one between subjects factor

Basics of one way ANOVA, between subjects design, multiple comparison procedures, statistical significance versus the magnitude of the treatment effect, writing the SAS program, interpreting and summarizing the results. Factorial ANOVA with Two Between Subject Factors – Introduction, Factorial Design Matrix, significant main effects and significant interaction effects, writing the SAS program, interpreting and summarizing the results.

Books for Study :

l. Norm O'Rourke, Larry Hatcher and Edward J.Stepanski (2005): ***Using SAS for Univariate and Multivariate Statistics*.** SAS Institute Inc. and Wiley

2. Lora D. Delwiche and Susan J. Slaughter(2012). ***The Little SAS Book: A programming Approach***. SAS Institute Inc.and Cary

Books for Reference:

1. Der, G. and Everitt, B.S. (2006). **A Handbook of Statistical Analysis Using SAS**, CRC Press.

Practical is to be done using a computer. The question paper for the external examination will be set by the external examiners in consultation with the chairman. The practical will be evaluated on the same day the examination is carried out and the mark sheet will be given to the chairman on the same day.

MSSTA04C15 - Project Work (6 Credits)

As a part of the course work, during the fourth semester each student has to complete a research project in collaboration with any of the authorized research institutions located within or outside the state or within their own Department. The topic could be a theoretical work or data analysis type. At the end of the fourth semester the student shall prepare a report/dissertation which summarizes the project work and submit to the HoD. of the parent department positively before the deadline suggested in the academic calendar. 120 marks shall be for project work (CE =20, ESE =100).

MSSTA04C16 - Viva Voce (2 Credits)

General viva-voce based on all the core and elective papers in the four semesters. The external viva -voce shall be conducted by a board of examiners consisting of at least two external experts appointed by the University. The viva-voce shall cover all the courses undertaken in the two year programme and carries 30 marks.(CE = 10, ESE = 20)

Selection of Elective Courses:

For selection of elective courses, the department may choose one course in semester III and two courses in semester IV from the lists of options being offered by the Department.

Elective I shall be chosen from the following list.

LIST OF ELECTIVE COURSES - III SEMESTER			
Sl No	**Course Code**	**COURSE TITLE**	**CREDIT**
1	MSSTA03E01	Time Series Analysis	4
2	MSSTA03E02	Queueing Theory	4
3	MSSTA03E03	Advanced Distribution Theory	4
4	MSSTA03E04	Applied Regression Analysis	4
5	MSSTA03E05	Analysis of Clinical Trials	4
6	MSSTA03E06	Analysis of Longitudinal Data	4

Elective II and Elective III shall be chosen from the list given below

LIST OF ELECTIVE COURSES - IV SEMESTER			
Sl No	**Course Code**	**COURSE TITLE**	**CREDIT**
1	MSSTA04E01	Operations Research	4
2	MSSTA04E02	Reliability Modeling	4
3	MSSTA04E03	Lifetime Data Analysis	4
4	MSSTA04E04	Mixture Regression Models	4
5	MSSTA04E05	Statistical Machine Learning	4
6	MSSTA04E06	Econometrics	4
7	MSSTA04E07	Survival Analysis	4
8	MSSTA04E08	Advanced Bayesian Computing with R	4
9	MSSTA04E09	Biostatistics	4
10	MSSTA04E10	Demography	4
11	MSSTA04E11	Data Mining	4

Syllabi of Elective Courses

MSSTA03E01: Time Series Analysis (4 Credits)

Course Outcomes:

CO 1: Apply moving average and exponential smoothing techniques to estimate and eliminate trends and seasonality in time series data, enabling accurate forecasting.

CO2: Develop and analyze autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) models to understand and predict linear stationary time series data.

CO 3: Employ Yule-Walker estimation, maximum likelihood, and least squares methods to estimate ARMA models and perform accurate forecasting with diagnostic checking.

CO 4: Understand and interpret the spectral density and periodogram of stationary time series, and gain introductory knowledge of seasonal ARIMA, ARCH, and GARCH models.

Unit 1 :

Time series, components of time series, additive and multiplicative models, trend and seasonality, estimation and elimination of trend and seasonality, moving average, simple exponential smoothing, Holt's exponential smoothing, Holt-Winters exponential smoothing, forecasting based on smoothing.

Unit 2 :

Time series as a discrete parameter stochastic process, auto-covariance and auto-correlation functions and their properties, stationary processes, Wold representation of linear stationary processes, detailed study of the linear time series models: autoregressive, moving average, autoregressive moving average and autoregressive integrated moving average models.

Unit 3:

Estimation of ARMA models: Yule-Walker estimation for AR Processes, maximum likelihood and least squares estimation for ARMA Processes. Choice of AR and MA periods, forecasting using ARIMA models, residual analysis and diagnostic checking.

Unit 4 :

Spectral density of a stationary time series and its elementary properties, Periodogram, Spectral density of an ARMA process. Seasonal ARIMA models (basic concepts only). Introduction to non-linear time series: ARCH and GARCH models (basic concepts only).

Books for Study:

1. Abraham,B and Ledolter,J.C(1983). ***Statistical Methods for Forecasting*.** Wiley

2. Box G.E.P, Jenkins G.M. and Reinsel G.C. (2008) ***Time Series Analysis: Forecasting and Control,*** *Fourth Edition*. Wiley.

3. Brockwell,P. J and Davis, R.A(1987). ***Time Series. Theory and Methods.*** Springer-Verlag

Books for Reference:

1. Anderson, T. W(1971). ***Statistical Analysis of Time Series*,** Wiley

2. Fuller, W. A(1978). ***Introduction to Statistical Time Series***, John Wiley.

3. Kendall, M. G(1978). ***Time Series,*** Charler Graffin

4. Tanaka, K(1996). ***Time Series Analysis,*** Wiley Series.

5. Cryer, J. D. and Chan, K. (2008). ***Time Series Analysis with Applications in R,*** *Second Edition.* Springer-Verlag.

MSSTA03E02 : Queueing Theory (4 Credits)

Course Outcomes:

CO 1: Understand various Markovian queueing models and their analysis

CO 2: Understand transient behavior of queueing models and analysis of advanced Markovian models with bulk arrival and bulk service

CO 3 : Understand various queueing networks and their extensions

CO 4: Understand various non Markovian queueing models and their analysis.

Unit 1: Markovian Queueing Models

Introduction to queueing theory, Characteristics of queueing processes, Measures of effectiveness, Markovian queueing models, steady state solutions of the M/M/I model, waiting-time distributions, Little's formula, queues with unlimited service, finite source queues.

Unit 2: Advanced Markovian Models

Transient behavior of M/M/1 queues, transient behavior of M/M/1. Busy period analysis for M/M/1 and M/M/c models. Advanced Markovian models. Bulk input M[X] /M/1 model, Bulk service M/M[Y]/1 model, Erlangian models, M/Ek/1and Ek/M/1. A brief discussion of priority queues.

Unit 3 : Queueing Networks

Series queues, open Jackson networks, closed Jackson network, Cyclic queues, Extension of Jackson networks. Non-Jackson networks.

Unit 4: Non Markovian Queueing Models

Models with general service pattern, The M/G/1 queueing model, The Pollaczek-Khintchine formula, Departure point steady state systems size probabilities, ergodic theory, Special cases M/Ek/1 and M/D/1, waiting times, busy period analysis, general input and exponential service models, arrival point steady state system size probabilities.

Books for Study :

1. Gross, D. and Harris, C.M. (1985): ***Fundamentals of Queueing Theory,*** *2nd Edition,* John Wiley and Sons, New York.

Books for Reference:

1. Ross, S.M. (2010). ***Introduction to Probability Models.*** *10th Edition,* Academic Press, New York.

2. Bose, S.K. (2002). ***An Introduction to Queueing Systems,*** Kluwer Academic / Plenum Publishers, New York.

MSSTA03E03 : Advanced Distribution Theory (4 Credits)

Course Outcomes:

CO 1: Understand different systems of distributions as generalizations of various standard continuous distributions.

CO 2: Understand various characterizations of probability distributions.

CO 3: Understand the concept of generalized power series distribution as generalization of various discrete distributions.

CO 4: Understand the idea of mixture distributions and non-parametric density estimation.

Unit 1: Systems of Distributions

Pearson system of frequency curves, determination of parameters, the kappa criterion, properties and extensions, estimation of parameters. The Johnson's system, Burr's system, distributions based on series expansion, Edgeworth series, Gram Charlier series.

Unit 2 : Characterization of Probability Laws

Characterization of probability distribution, Exponential and Geometric law, lack of memory property, normal law - characterization based on independence of linear forms and quadratic forms and regression.

Unit 3 : Generalized Power Series Distributions

power series and compound distributions. Generalized Poisson distribution, Hyper Poisson family, distributions derived from Poisson and other generalizations.

Unit 4: Mixture Distributions and Non - parametric density estimation:

Finite and infinite mixtures, identifiability of mixtures, examples of non-identifiable mixtures, finite normal mixtures and estimation, normal mixture regression models. Density estimation, histogram and nave estimate, Kernel density estimate and properties.

Books for Study :

1. Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). ***Continuous Univariate Distributions, Vol.I & Vol.II*****,** John Wiley and Sons, New-York.

2. MacLachlan, P. and Peel, D. (2000). ***Finite Mixture Models***. John Wiley& Sons, New York

3. Silverman, B. (1986). ***Density Estimation for Statistics and Data Analysis***. Chapman & Hall.

Books for Reference:

1. Johnson, N.L., Kotz. S. and Kemp. A.W. (1992). ***Univariate Discrete Distributions,*** John Wiley and Sons, New York.

2. Stuart, A. Ord, A. (1994). ***Kendall's Advanced Theory of Statistics,*** Distribution Theory, 6thEdition. Wiley-Blackwell.

3. Kagan A.M., Linnik, Y.V.and Rao C.R. (1975). ***Characterization Problems in Mathematical Statistics***. John Wiley.

MSSTA03E04 : Applied Regression Analysis (4 Credits)

Course Outcomes:

CO 1: Understand various regression models including logistic regression models and simultaneous equation models.

CO 2: Understand consequences of multicollinearity, heteroscedasticity, autocorrelation, their detection and remedial measures.

CO 3: Apply statistical techniques to model relationships between variables and make predictions.

Unit 1 : Linear Regression Models

The simple linear regression models, least square estimation, statistical assumptions and properties of estimators, standard error of estimates, tests of significance and confidence intervals for the parameters, error and residual plots.

Unit 2 : Regression Diagnostics

Multicollinearity, heteroscedasticity, autocorrelation: their nature, consequences, detection, remedial measures and estimation in the presence of them.

Unit 3 :Multiple regression models and Nonparametric regression

Multiple regression models, OLS and ML estimators, testing and prediction. Nonparametric regression - Nonparametric regressions and concept of spline smoothing.

Unit 4 :Non Linear Regression

Polynomial regression in one and several variables. Linearization transforms, Diagnostic checks and correction. Generalized linear models. Logistic regression.

Books for Study :

1. Draper, N. R. and Smith, H. (1998): ***Applied Regression Analysis, 3rd Edition.*** John Wiley.

2. Montgomery, D .C, Peek, E. A. and Vining, G. G. (2006): ***Introduction to Linear Regression Analysis,*** John Wiley.

Books for Reference:

1. Gujarati, D.N. (2007): ***Basic Econometrics (Fourth Edition),*** McGraw- Hill, New York.

2. Goon, Gupta, Das Gupta (2001): ***An Outline Series in Statistics Vol 1I,*** World Press.

3. Hosmer, D.W. and Lemeshow, S. (1989): ***Applied Logistic Regression,*** John Wiley.

MSSTA03E05 : Analysis of Clinical Trials (4 Credits)

Course Outcomes:

CO 1: Understand basics of clinical trials

CO 2: Understand design of clinical trials

CO 3: Understand sample size determination in clinical trials

CO 4: Understand the concept of meta-analysis in clinical trials

Unit 1: Basics of Clinical Trials

Introduction to clinical trials, the need and ethics of clinical trials, bias and random error in clinical studies, Protocols, conduct of clinical trials, overview of Phase I-IV trials, Data management-data definitions, standard operating procedure, informed consent form, case report forms, database design, data collection systems for good clinical practice.

Unit 2 : Design of Clinical Trials

Design of clinical trials- Different phases, Comparative and controlled trials, Random allocation, Randomization, response adaptive methods and restricted randomization. Methods of Blinding, Parallel group designs, Crossover designs, Symmetric designs, Adaptive designs, Group sequential designs, Zelen's designs, design of bioequivalence trials. Outcome measures.

Unit 3: Sample Size Determination and Testing

Sample size determination in one and two sample cases, comparative trials, activity studies, testing and other purposes, unequal sample sizes and case of anova. Surrogate endpoints-selection and design of trials with surrogate endpoints, analysis of surrogate end point data. Reporting and Analysis-Interpretation of result, multi-center trials.

Unit 4: Meta-Analysis

Meta-analysis in clinical trials-concept and goals, fixed and random effect approaches. Bioassay: Direct and indirect assays, Quantal and quantitative assays, Parallel line and slope ratio assays, Design of bioassays.

Books for study:

1. Friedman, L. M., Furburg, C. D. Demets, L. (1998). ***Fundamentals of Clinical Trials,*** Springer Verlag.

2. Jennison and B. W. Turnbull (1999). ***Group Sequential Methods with Applications to Clinical Trials,*** CRC Press.

3. Kulinskaya E, Morgeathaler S, Staudte R G (2008). ***Meta-analysis,*** Wiley.

Books for Reference:

1. Das, M. N. and Giri (2008). ***Design of Experiments,*** New Age, India

2. Fleiss, J. L. (1989): ***The Design and Analysis of Clinical Experiments,*** Wiley.

3. Marubeni, E. and M. G. Valsecchi (1994): ***Analyzing Survival Data from Clinical Trials and Observational Studies,*** Wiley and Sons.

4. Piantadosi S. (1997): ***Clinical Trials: A Methodological Perspective***. Wiley.

5. W Rosenberger, J Machin (2002): ***Randomization in Clinical Trials Theory and Practice,***Wiley.

MSSTA03E06 : Analysis of Longitudinal Data (4 Credits)

Course Outcomes:

CO 1: Conduct analysis of longitudinal data.

CO 2: Apply statistical techniques to model longitudinal data and make predictions.

CO 3: Understand analysis of longitudinal data with missing data.

CO 4: Understand analysis of longitudinal data with time-dependent covariates.

Unit 1: Linear Model for Longitudinal Data

General Linear Model for Longitudinal Data. ML and REML estimation, EM algorithm: General linear mixed-effects model, Inference for; the random effects, BLUPs, Empirical Bayes, Bayes, Shrinkage Model building and diagnostic, relaxing parametric assumptions: generalized additive mixed model.

Unit 2 : Generalized Linear Model for Longitudinal Data

Generalized Linear Model for Longitudinal Data, Marginal models, for binary, ordinal, and count

data: Random effects models for binary and count data:Transition models: Likelihood- based models for categorical data; GEE; Models for mixed discrete and continuous responses.

Unit 3 : Longitudinal Data with Missing Data

Classification missing data mechanism; Intermittent missing values and dropouts; Weighted estimating equations; Modeling the dropout process (Selection and pattern mixture models).

Unit 4 : Time-dependent Covariates and Special Topics

Dangers of time dependent covariates, Lagged covariates; Marginal Structural models; Joint models for longitudinal and survival data; Multivariate longitudinal data; Design of randomized and observational longitudinal studies.

Books for Study :

1. Diggle, P.J., Heagerty, P., Liang, K.Y and Zeger. S.L (2003). ***Analysis of Longitudinal Data,*** *2nd Edn*. Oxford University press, New York.

2. Fitzmaurice, G.M., Laird, N.M and Ware, J.H. (2004). ***Applied Longitudinal Analysis***, John Wiley & Sons, New York.

Books for Reference:

1. Crowder, M.J. and Hand, D.J. (1990). ***Analysis of Repeated Measures***. Chapman and Hall/CRC Press, London.

2. Davidian, M. and Giltinan, D.M. (1995). ***Nonlinear Models for Repeated Measurement Data.*** Chapman and Hall/CRC Press, London.

3. Hand,D and Crowder, M. (1996). ***Practical Longitudinal Data Analysis.*** Chapman and Hall/CRC Press, New York.

4. Little, R. J. A and Rubin, O.B. (2002). ***Statistical Analysis with Missing Data****, 2nd Edition*, Wiley,New York.

5. Mc Cullagh,P. and Nelder. J. A. (1989). ***Generalized Linear Models.*** *2nd Edition,* Chapman and Hall/CRC Press, London.

6. Weiss, R.E. (2005). ***Modeling Longitudinal Data.*** Springer, New York

MSSTA04E01 : Operations Research (4 Credits)

Course Outcomes:

CO 1: Identify and develop operational research models from the verbal description of the real system.

CO 2: Understand the mathematical tools that are needed to solve optimization problems.

CO 3: Understand various methods in Integer programming and Game theory

Unit 1: Algebra of linear programming problems

Introduction to linear programming problem (LPP), graphical solution, feasible, basic feasible, and optimum basic feasible solution to an LPP. Analytical results in general LPP, theoretical development of simplex method.

Unit 2 : Duality and dual simplex method

Artificial variables, Big-M method, two phase simplex method, duality, duality theorems, dual simplex methods and revised simplex method.

Unit 3: Integer programming, Non linear programming and network analysis

Integer programming: Cutting plane methods, branch and bound technique. Non linear programming. Network analysis, Critical path analysis, -CPM, PERT, distinction between CPM and PERT.

Unit 4: Game theory

Game theory, pure and mixed strategies, conversion of two-person zero gain to a linear programming problem. Solution to game through algebraic, graphical and linear programming methods.

Books for Study:

1. K.V. Mital and Mohan, C (1996). ***Optimization Methods in Operations Research and Systems Analysis,*** *3rd Edition,* New Age International (Pvt.) Ltd.

2. Kanti Swamp, Gupta, P.K. and John, M. M. (1985): ***Operations Research,*** Sultan Chand & Sons.

Books for Reference:

1. Hadley, G. (1964). ***Linear Programming,*** Oxford & IBH Publishing Co, New Delhi.
2. Taha. H. A. (1982): ***Operation Research, An Instruction,*** Macmillan.
3. Hiller F. S. And Lieberman, G.J. (1995). ***Introduction to Operations Research,*** McGraw Hill

MSSTA04E02 : Reliability Modeling (4 Credits)

Course Outcomes:

CO 1: Understand reliability concepts and measures

CO 2: Understand various lifetime Probability distributions and their structural properties

CO 3: Understand univariate and bivariate shock models and reliability estimation based on failure times

CO 4: Understand Maintenance and Replacement Policies

Unit 1: Basic Reliability Concepts

Reliability Concepts and Measures; components and systems;coherent systems; reliability of coherent systems; cuts and paths; modular decomposition; bounds on reliability; structural and reliability importance of components.

Unit 2: Life Distributions and Properties

Life distributions; reliability function; hazard rate; common life distributions-exponential, Weibull, Gamma etc. Estimation of parameters and tests in these models. Notions of ageing; IFR, IFRA, NBU, DMRL, and NBUE Classes and their duals; closures or these classes under formation of coherent systems, convolutions and mixtures.

Unit 3: Shock Models

Univariate shock models and life distributions arising out of them; bi-variate shock models; common bivariate exponential distributions and their properties. Reliability estimation based on failure times in variously censored life tests and in tests with replacement of failed items; stress-strength reliability and its estimation.

Unit 4: Maintenance and Replacement Policies

Repairable Systems,replacement policies, modeling of a repairable system by a non-homogeneous Poisson process. Reliability growth models; probability plotting techniques; Hollander-Proschanand Deshpande tests for exponentiality; tests for HPP vs. NHPP with repairable systems. Basic ideas of accelerated life testing.

Books for Study :

1. Barlow R.E. and Proschan.F(1985). **Statistical Theory of Reliability and Life Testing**; Holt,Rinehart and Winston.

2. Zacks, S. (1992). **Introduction to Reliability Analysis: Probability Models and Statistics Methods**. New York:Springer-Verlag.

Books for Reference :

1. Bain L.J. and Engelhardt (1991). **Statistical Analysis of Reliability and Life Testing Models;** Marcel Dekker.

2. Aven,T.andJensen,U. (1999).**StochasticModelsinReliability**, Springer Verlag, New York,Inc.

3. Nelson, W (1982).**Applied Life Data Analysis;** JohnWiley.

MSSTA04E03 : Lifetime Data Analysis (4 Credits)

Course Outcomes:

CO 1: Understand various lifetime probability distributions and their structural properties

CO 2: Understand different methods for the estimation of survival function.

CO 3: Conduct analysis of time data

CO 4: Apply statistical techniques to model lifetime data and make predictions.

Unit 1: Lifetime Distributions

Lifetime distributions, Important parametric models, Exponential Weibull, Log-normal, Log-logistic, Gamma, Inverse Gaussian distributions, Log-location scale models and mixture models. Censoring and statistical methods.

Unit 2: Estimation of Survival Function

The product-limit estimator and its properties. The Nelson-Aalen estimator, interval estimation of survival probabilities, asymptotic properties of estimators, descriptive and diagnostic plots, estimation of hazard function, methods for truncated and interval censored data, Life tables.

Unit 3: Inference Procedures

Inference under exponential model - large sample theory, type-2 censored test plans, comparison of two distributions; inference procedures for Gamma distribution; models with threshold parameters, inference for log- location scale distribution: likelihood based methods: Exact methods under type-2 censoring; application to Weibull and extreme value distributions, comparison of distributions.

Unit 4: Regression Models

Log-location scale (Accelerated Failure time) model, proportional hazard models, methods for continuous multiplicative hazard models, semi-parametric maximum likelihood estimation of continuous observations, incomplete data; rank test for comparing distributions, log-rank test, generalized Wilcox on test. A brief discussion on multivariate lifetime models.

Books for Study :

1. Lawless, J.F. (2003): ***Statistical Methods for Lifetime*** *(Second Edition)*, JohnWiley & Sons Inc., New Jersey.

2. Kalbfiesche, J. D. and Prentice, R.L. (1980): ***The Statistical Analysis of Failure Time Data*****,** John Wiley & Sons Inc. New Jersey.

Books for Reference:

1. Miller, R.G. (1981): ***Survival Analysis***, John Wiley & Sons Inc.

2. Bain, L.G. (1978): ***Statistical Analysis of Reliability and Life Testing Models***, Marcel Decker.

3. Cox, D.R and Oakes, D. (1984): ***Analysis of Survival Data***. Chapman and Hall.

MSSTA04E04 : Mixture Regression Models (4 Credits)

Course Outcomes:

CO 1: Understand concept of finite mixture regression models with emphasis on its applications.

CO 2: Understand the problem of non-identifiability of mixture models

CO 3: Understand the EM algorithm for the estimation of parameters of mixture regression models and generalized linear mixture models.

CO 4: Work with various R-packages for the analysis of mixture models.

Unit 1: Mixture Distributions

Finite and infinite mixtures, location and scale mixtures, non-identifiable mixtures, examples of non-identifiable mixtures, condition for identifiability when the components belong to the power series family.

Unit 2 : Simulation and Estimation

Finite normal, Poisson and negative binomial mixtures, simulation of random samples from mixtures, applications of mixture models. Estimation of parameters of mixture models, method of moments, maximum likelihood estimation, EM algorithm.

Unit 3: Mixture Regression

Normal mixture regression, Poisson mixture regression, estimation of parameters, examples using real and simulated data, r packages, FlexMix, Mixtools and CAMAN.

Unit 4: Generalized Linear Mixture Models

Exponential family, generalized linear models, examples, generalized linear mixture models, logistic and mixture logistic models, concomitant variables and varying parameter cases.

Books for Study:

1. McLachlan, G.J. and Peel, D. (2000). ***Finite Mixture Models***. John Wiley & Sons, INC, New York.

Books for Reference:

1. Schlattmann, P. (2009). ***Medical Applications of Finite Mixture Models***. Springer Verlag Berlin Heidelberg.

2. Titterington, D. M., Smith, A. and Makov, U. (1985). ***Statistical Analysis of Finite Mixture Distributions.*** New York: Wiley

3. Leisch, F. (2004). **Flex Mix: A general framework for finite mixture models and latent class regression in R.** Journal of Statistical Software,11(8), 1-18. http://www.jstatsoft.org/

4. Wang, P. et.al. (1996). **Mixed Poisson regression models with covariate dependent rates.** Biometrics,52, 381-400.

5. Sapatinas, T. (1995). **Identifiability of mixtures of power-series distributions and related characterizations.** Ann. Inst. Statist. Math.,47 (3), 447-459.

6. McLachlan, G. J. and Krishnan, T. (1997). ***The EM algorithm and Extensions***. New York: Wiley.

MSSTA04E05: Statistical Machine Learning (4 Credits)

Course Outcomes:

CO 1: Understanding principles and theories of statistical machine learning.

CO 2: Gain proficiency in various machine learning algorithms.

CO 3: Develop the ability to implement machine learning models using programming languages like R or Python.

CO 4: Acquire skills in data preprocessing, and handling real world dat sets to prepare them for model training.

CO 5 : Apply machine learning techniques to solve real world problems.

Unit 1: Supervised learning and function approximation, A Statistical Model for the Joint Distribution of input and output vectors, Function Approximation, Structured Regression Models, Linear Methods for Regression: Least squares, Subset selection, Shrinkage Methods, Methods using derived input directions,Multiple outcome shrinkage and selection, Lasso and related path algorithms.

Unit 2: Linear methods for classification using linear regression of an indicator matrix, linear discriminant analysis, logistic regression and separating hyperplanes. Basis expansions and regularizations: Piecewise polynomials and splines, Automatic Selection of the Smoothing Parameters, Nonparametric Logistic Regression, Multidimensional Splines.

Unit 3: One-Dimensional Kernel Smoothers, Selecting the band width of the Kernel, Structured Local Regression Models in Rp, Local Likelihood and Other Models, Kernel Density Estimation and Classification: Kernel Density Estimation, Kernel Density classification and the Naïve Bayes classifier. Mixture Models for Density Estimation and Classification.

Unit 4: Bias, Variance and Model Complexity, The Bias–Variance Decomposition, Optimism of the Training Error Rate, Estimates of In- Sample Prediction Error, The Bayesian Approach and BIC, Minimum Description Length, Cross-Validation, Bootstrap Methods, Conditional or Expected Test Error, introducing Model Inference and averaging: Local regression in IR, The EM Algorithm, MCMC for Sampling from the Posterior, Bagging, Model Averaging and Stacking, Stochastic Search: Bumping

Books for Study

1. Hastie, T., Tbishirani, R. and Friedman, J. (2017). **The Elements of Statistical Learning : Data Mining, Inference and Prediction, 2 nd Edition**. Springer, New York.

2. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). **An Introduction to Statistical Learning with Applications in R.** Springer, New York.

Books for Reference

1. James, G., Witten,D., Tibshirani, R. and Hastie, T. **Neural Networks and Deep Learning: A Textbook.**

MSSTA04E06 : Econometrics (4 Credits)

Course Outcomes:

CO 1: Understand the key concepts of production theory, various production functions, Euler's theorem and elasticity of substitution.

CO 2: Grasp the concepts of linear regression models, execute least square estimates, create confidence intervals, conduct tests of significance, and utilize regression analysis.

CO 3: Understand distributed lag models, including Almon and Koyck lag models; estimate parameters; utilize non-linear regression models; and evaluate the forecasting power of econometric models.

CO 4: Acquire knowledge of various econometric problems like autocorrelation, multicollinearity, and heteroscedasticity.

CO 5: Comprehend the concepts of simultaneous equation models, which encompass structural, reduced form, and recursive models, as well as the concept of identification problems.

CO 6: Understand the estimation techniques for simultaneous equations, including indirect least squares, instrumental variables, two-stage least squares, as well as both limited and full information maximum likelihood methods.

Unit 1:

The theory of production, variable proportions, returns to scale, producer's equilibrium, homogeneous production functions, Euler's theorem, Cobb-Douglas production function, elasticity of substitution, C.E.S production function.

Unit 2:

Simple and multiple linear regression models, assumptions, least square estimation, properties of estimators, confidence intervals and tests of significance. The adjusted coefficient of determination. regression and analysis of variance. Distributed lag models and estimation of parameters, Almon and Koyck lag models, non linear regression models, forecasting power of an econometric model.

Unit 3:

Econometric problems, autocorrelation, sources and tests for autocorrelation, Durbin-Watson test, estimation of parameters, multicollinearity, sources, tests and estimation, heteroscedasticity, sources, tests and estimation, idea of Ridge regression and properties, application of principal components, errors in variables.

Unit 4:

Simultaneous equation models, structural, reduced form and recursive models, the problem of identification, rank and order conditions, identifying restrictions, estimation of simultaneous equations, indirect least squares, instrumental variable technique, two stage least squares, limited information and full information maximum likelihood methods.

Books for Study :

1. Gujarathi, D. and Sangeetha, S.(2007). ***Basic Econometrics,*** Mc Graw Hill

2. Johnston, J.(2009). ***Econometric Methods, 4th edition,*** Mc Graw Hill

3. Judge, G. J, Grifiths, W. E & et al.(1985). ***Theory and Practice of Econometrics, 2nd edition,*** John Wiley.

Books for Reference :

1. Edward Dowling.(2000). ***Introduction to Mathematical Economics,*** Shaums Outline Series.

2. Koutsoyiannis.(2000). ***Theory of Econometrics***, Palgrave.

3. Maddala, G. S.(1979). ***Econometrics*****,** Tata Mc Graw Hill

4. Salvatore Dominic.(1981). ***Statistics and Econometrics,*** Schaum's Outline Series

MSSTA04E07 : Survival Analysis (4 Credits)

Course Outcomes:

CO1: Gain an understanding of the core principles and concepts of survival analysis, which include survival functions, hazard functions, and mean residual life functions.

CO2: Develop an understanding of the features of standard lifetime distributions and their significance in the field of survival analysis.

CO3: Grasp the basic concepts of non-parametric tests.

CO4: Comprehend the concepts of censoring and truncation along with their diverse classifications. Also gain insight into estimating survival rates using extensive datasets such as DHS, NFHS, DLHS, etc., and comparing survival curves.

CO5: Acquire proficiency in advanced survival analysis techniques, such as mastering Kaplan-Meier estimation, life tables, the Mantel-Haenszel test and interval estimation of survival probabilities

CO6: Understand the concept of Cox-Proportional hazard model.

Unit 1: **Basics of survival analysis**

Discrete and continuous time models, survival function, hazard rate function, probability density function, mean residual life time. Aging classes-IFR, IFRA and their duals, bathtub failure rate.

Unit 2: **Life distributions**

Exponential, Weibull, lognormal and gamma distributions, characterizations. Concepts of censoring mechanism –Type-I, Type-II and random censoring, progressive censoring, truncation, methods for truncated and interval censored data.

Unit 3: **Estimation of survival curves**

Likelihood construction and estimation of Censored and Truncated Data. Estimating survival rates using large scale data like DHS, NFHS, DLHS, etc. Comparing survival curves

Unit 4: **Kaplan-Meier estimation technique**

Life tables, Mantel-Haenszel test. Interval estimation of survival probabilities. Introduction to survival regression. Cox proportional hazard model.

Books for study :

1. Lawless, J.F. (2003): ***Statistical Methods for Lifetime*** (Second Edition),John Wiley & Sons Inc., NewJersey.

2. Kalbfleisch, J. D. and Prentice, R.L. (1980): ***The Statistical Analysis of Failure Time Data,*** John Wiley& Sons Inc. NewJersey.

3. Moore, D.F. (2016): ***Applied Survival Analysis Using R*****,** Springer

Books for Reference :

1. Klein J.P. and Moeschberger M.L. (2003) ***Survival Analysis –Techniques for Censored and Truncated Data***, Second Edition, Springer- Verlag, New York

2. Miller,R.G.(1981):***Survival Analysis***, John Wiley & Sons Inc.

3. Bain,L.G.(1978):***Statistical Analysis of Reliability and Life testingModels***, MarcelDecker.

4. Cox,D.RandOakes,D.(1984):***Analysis of Survival Data*****.** ChapmanandHall. NewYork.

5. Fraser,D.A.S.(1957): ***Nonparametric Method in Statistics*****,** Wiley.

6. Elandt - Johnson, R.E. Johnson N.L. (1980): ***Survival models and Data Analysis*****,** John Wiley and Sons

7. Hosmer D.W, Lemeshow S, May S (2008): ***Applied Survival Analysis***,Wiley.

MSSTA04E08 : Advanced Bayesian Computing with R (4 Credits)

Course Outcomes:

CO 1: Understand the advantageous Bayes estimation over that based on frequentist approach.

CO 2: Understand the LearnBayes package for various Bayesian computations

CO 3: Understand MCMC methods in various situations in which the exact computation is difficult.

CO 4: Understand Gibbs sampling to generate random samples from a multivariate distribution.

Unit 1: Bayesian Inference:

Statistical decision problem, randomized decision rule, decision principle, standard loss functions, Prior information, subjective determination of prior density, non-informative priors, maximum entropy priors, conjugate priors, discrete prior. Parametric family and likelihood, exponential family, Bayes' theorem for inference, prior and posterior densities.

Unit 2: Single and multi-parameter models:

Single parameter models, normal distribution with known variance and unknown mean, normal with known mean and unknown variance, Poisson model, normal distribution with both parameters unknown, multinomial model, Dirichlet prior, Bioassay experiment, comparing two proportions, predictive distribution, beta-binomial distribution, multivariate normal distribution, introduction to Learn Bayes package, Examples using Learn Bayes package.

Unit 3: Bayesian Computation

Computing integrals using Monte-Carlo simulation, approximation based on posterior mode, importance sampling, Markov Chain Monte Carlo methods, Metropolis-Hastings algorithm, random walk, Gibbs sampling.

Unit 4: Model Comparison and Regression models

Hierarchical models, shrinkage estimators, posterior predictive model checking, comparison of hypotheses, Bayes factor, one sided test for normal mean, two-sided test for normal mean, normal linear regression model, prediction of future observations, examples and R codes, introduction to Win-BUGS package.

Books for Study :

1. Jim Albert (2007). ***Bayesian Computation with R,*** New York: Springer Verlag.

2. Berger, O. J. (1985). ***Statistical decision Theory and Bayesian Analysis,*** Second Edition, Springer Verlag.

3. Bensal, A. K. (2008). ***Bayesian Parametric Inference,*** New Age, Delhi.

Books for Reference:

1. Ferguson, T.S. (1967). ***Mathematical Statistics: A Decision Theoretic Approach*****,** Academic Press, New-York.

2. Bolstad, W. (2004). ***Introduction to Bayesian Statistics,*** Hoboken, N J: John Wiley.

3. Gelman, A., Carlin, J., Stern, H.and Rubin, D. (2003). ***Bayesian Data Analysis***, New York: Chapman andHall.

4. Gilks, W. R., Richardson, S and Spiegelhalter, D.J. (1996). ***Markov Chain Monte Carlo in Practice*****.** Chapman & Hall/CRC, New York.

5. Robert, C. and Casella, G. (2004). ***Monte Carlo Statistical Methods*****,** New York: Springer

MSSTA04E09 : Biostatistics (4 Credits)

Course Outcomes:

CO 1: understand the essential statistical concepts and methodologies used in biostatistics.

CO 2: gain skills in analyzing biological and health data using statistical software like R, SPSS and Python

CO 3: Learn the principles of designing, analyzing and interpreting various biological and medical research studies.

CO 4: Apply regression techniques.

CO 5 :develop the ability to communicate statistical findings effectively and understand the ethical considerations in biological research.

Unit 1:

Functions of survival time, survival distributions and their applications viz. exponential, gamma, weibull, Rayleigh, lognormal, death density function for a distribution having bath-tub shaped hazard function. Tests of goodness of fit for survival distributions (WE test for exponential distribution, W-test for lognormal distribution, chi-square test for uncensored observations). Parametric methods for comparing two survival distributions viz. L.R. test, Cox's F-test.

Unit 2:

Type I, Type II and progressive or random censoring with biological examples, estimation of mean survival time and variance of the estimator for type I and type II censored data with numerical examples. Non-parametric methods for estimating survival function and variance of the estimator viz. Actuarial and Kaplan-Meier methods.

Unit 3:

Competing risk theory, indices for measurement of probability of death under competing risks and their inter-relations. Estimation of probabilities of death under competing risks by maximum likelihood and modified minimum chi-square methods. Theory of independent and dependent risks. Bivariate normal dependent risk model. Conditional death density functions.

Unit 4:

Basic biological concepts in genetics, Mendel's law, Hardy-Weinberg equilibrium, random mating, distribution of allele frequency(dominant/co-dominant cases), Approach to equilibrium for X-linked genes, natural selection, mutation, genetic drift, equilibrium when both natural selection and mutation are operative, detection and estimation of linkage in heredity.

Books for Study :

1. Biswas, S. (1995). ***Applied Stochastic Processes. A Biostatistical and Population Oriented Approach,*** Wiley Eastern Ltd.

2. Cox, D. R. and Oakes, D. (1984). ***Analysis of Survival Data,*** Chapman and Hall.

3. Elandt, R. C. and Johnson(1975). ***Probability Models and Statistical Methods in Genetics,*** John Wiley & Sons.

Books for Reference:

1. Ewens, W. J.(1979). **Mathematics of Population Genetics**, Springer Verlag.

2. Ewens, W. J. and Grant, G. R.(2001). **Statistical Methods in Bioinformatics: An Introduction,** Springer.

3. Friedman, L. M., Furburg, C. and DeMets, D. L.(1998). ***Fundamentals of Clinical Trials,*** Springer Verlag.

4. Gross, A. J. and Clark, V. A.(1975). ***Survival Distribution; Reliability Applications in Biomedical Sciences,*** John Wiley & Sons.

5. Lee, Elisa, T.(1992). ***Statistical Methods for Survival Data Analysis,*** John Wiley & Sons.

6. Miller, R. G.(1981). ***Survival Analysis,*** John Wiley & Sons.

MSSTA04E10 : Demography (4 Credits)

Course Outcomes:

CO 1: Gain knowledge of fundamental demographic theories, principles, and concepts including population structure and dynamics.

CO 2: Learn and apply various demographic data collection, analysis and interpretation techniques using statistical tools.

CO 3: analyze and interpret population trends and patterns to understand the implications of birth rates, death rates , migration and aging.

CO 4: Evaluating population policies.

Unit 1:

Definitions and concepts used in Demography- Interface between Statistics and Demography- Sources of Demographic data: Census,Vital Registration System,Sample surveys. Population Composition and Structure- Age, Sex, Religion, Education, Income, Dependency etc., Population

pyramid. Concepts of Fertility, Nuptiality, Mortality, Morbidity, Migration and Urbanisation. Determinants and consequences of population change, population distribution.

Unit 2:

Measurement of mortality and morbidity, Force of mortality. Measurement of fertility- TFR, GRR, NRR. Standardization of rates- Concept of life tables- Various types of life tables- Multiple decrement and multi-state life tables- Working life table- mortality models- model life tables- U.N Coale & Demeny, Leaderman's system, Brass' Logit system, U.N. tables for developing countries- Stable population models - database and application- Uses of life table approach in Demography- Birth Interval Analysis.

Unit 3:

Structure of population- Lotka's stable population theory: concepts, assumptions and properties. Stationery and quasi-stable population, population momentum, population waves. Population estimation and projection. Population growth- exponential, logistic- different methods of population estimation and projection- Mathematical and component methods.

Unit 4:

Stochastic models for population changes- Pure birth and death process- birth, death, migration models- Markov chain- Renewal process.

Books for study:

1. Henry, S. Shryock and Jacob, S. Siegel (1976). ***Methods and Materials of Demography,*** Academic Press, New York.

2. Ramkumar, R. and Gopal, Y. S. (1996). ***Technical Demography,*** Wiley Eastern Limited.

3. Srinivasan, K.(1998). ***Basic Demographic Techniques and Applications***; Sage Publications, New Delhi.

Books for Reference:

1. Asha, A. Bhende and Tara Kanitkar () ***Population Studies (5th revised edition),*** Himalaya Publishing House, New Delhi.

2. Krishnan Namboodiri and C. M. Suchindran (1987). ***Life Table Techniques and Their Applications*****,** Academic Press, London.

3. Saxena, P. C. and Talwar, P. P. (1987). ***Recent Advances in the Techniques for Demographic Analysis***, Himalaya Publishing House.

4. UNDP (2003). ***Human Development Report.***

MSSTA04E11 : Data Mining (4 Credits)

Course Outcomes:

CO 1: Understand the fundamental concepts and techniques used in data mining,

CO 2: Students will develop practical skills in using data mining tools and software.

CO 3: able to analyze large data sets, extract meaningful patterns and interpret results.

CO 4: Apply the data mining techniques in various fields.

CO 5 : students will enhance their problem solving and critical thinking skills by tackling complex data mining problems, formulating solutions and making decisions

Unit-1: Review of classification methods from multivariate analysis; classification and decision trees.Clustering methods from both statistical and data mining viewpoints; vector quantization.

Unit-2: Unsupervised learning from univariate and multivariate data; Dimension reduction and feature selection. Supervised learning from moderate to high dimensional input spaces;

Unit-3: Artificial neural networks and extensions of regression models, regression trees. Introduction to databases, including simple relational databases.

Unit-4: Data warehouses and introduction to online analytical data processing. Association rules and prediction; data attributes, applications to electronic commerce.

Books for Study

1.Berson, A. and Smith, S.J. (1997). ***Data Warehousing, Data Mining, and OLAP.*** McGraw-Hill.

2.Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). ***Classification and Regression Trees*.** Wadsworth and Brooks/Cole.

3.Han, J. and Kamber.M. (2000). **Data Mining; Concepts and Techniques.** (MorganKaufmann.)

Books for reference

1. Mitchell, T.M. (1997). **Machine Learning.** (McGraw-Hill.)
2. Ripley, B.D. (1996). **Pattern Recognition and Neural Networks**. (CambridgeUniversity Press).

Open Elective Courses:

Open elective courses mean an elective course which is available for students of all other programmes in the University. Students of other departments may opt these courses subject to fulfilling of eligibility criteria as laid down by the department offering the course. Students can join for the open course depending on their choice and availability of seats in the departments offering such courses. Open Elective Courses can be opted in the third Semesters.The open elective shall be chosen from the following list.

LIST OF OPEN ELECTIVE COURSES - III SEMESTER			
Sl No	**Course Code**	**COURSE TITLE**	**CREDIT**
1	MSSTA03O01	Statistical Data Analysis Using R	4
2	MSSTA03O02	A First Course on Statistics Using Excel	4
3	MSSTA03O03	Basic Statistics Using SPSS	4
4	MSSTA03O04	Research Methodology	4
5	MSSTA03O05	A First Course on LaTex for Scientific Documentation.	4

Syllabi of Open Elective Courses

MSSTA03O01 : Basic Statistics Using R (4 Credits)

Course Outcomes:

CO 1: Understand various built-in functions in R programming for statistical data analysis.

CO 2: Understand different functions in R programming for writing computer programmes and develop computer programmes for different problems.

CO 3: Understand different statistical test using R software

Unit 1 : Introduction to R

Introduction to R- Objects and their classes, operators, vectors and matrices, list and data frames, indexing and accessing data, importing and exporting data. Common built-in functions. Simple applications – Descriptive statistics.

Unit 2 :R-Graphics

R-Graphics- Histogram, Box-plot, Stem and leaf plot, Scatter plot, Q-Q plot. Looping- for loop, repeat loop, while loop, if command, if else command.

Unit 3 :Basic probability and distribution

Basic concepts of probability and random variables, Probability distributions (Binomial, Poisson, Geometric, Uniform, Normal, Gamma, Beta), Plotting of CDF and PDF for different values of the parameters of standard distributions. Generations of random samples from standard distributions.

Unit 4 :Descriptive statistics

The Descriptive statistics, the comparison of means, ANOVA, non-parametric tests, correlation and regression procedures.

Books for study:

1. Maria D.U., Ana F.M. and Alan T.A. (2008): ***Probability and Statistics with R*****. CRC Press.**

2. Dalgaard, P. (2008): ***Introductory Statistics with R (Second Edition)*****,** Springer.

Books for Reference:

1. Purohit, S. G, Ghore, S.D and Deshmukh, S. R.(2004): ***Statistics Using R.***

2. Maria L. Rizzo (2019). ***Statistical Computing with R, Second Edition.*** Chapman & Hall, CRC Press.

MSSTA03O02 : A First Course on Statistics Using Excel (4 Credits)

Course Outcomes:

CO1: Demonstrate proficiency in navigating the Excel interface, organizing data effectively, and utilizing a wide range of Excel functions and features for data manipulation and analysis.

CO2: Develop advanced data analysis skills, including the ability to perform complex calculations, analyze statistical measures, and interpret data trends using Excel's built-in functions and tools.

CO3: Acquire proficiency in creating various types of charts and graphs in Excel to visually represent and communicate data insights effectively.

CO4: Leverage Excel as a tool for informed decision-making by analyzing data, identifying patterns and trends, and deriving actionable insights to support organizational or personal goals

Unit 1: Excel Introduction, Basic Navigation Tab, Concept of Cell and Cell address , row Column concept, Basic mathematical and statistical functions in Excel.

Unit 2: Min, Max, Trim, Lower, Upper, Proper, Left, Right, Mid Exact, Randbetween, Rand, Len (Length of character) Paste special, SQRT, If function with Example of IF, More function like And, OR with their example, Conditional Formatting basic and advance level with OR, AND, Nested IF function, Index, Offset, Match.

Unit 3: Graphics in excel-pie chart, bar chart, multiple bar diagram, sub-divided bar diagram, histogram, line chart, scatter diagram, box plot.

Unit 4: Median, Mode, Standard Deviation (SD), Correlation, Large, Small, Pivot Table, Pivot Charts, Slicing, Sparkling.

Books for study:

1. Linoff, Gordon S (2015). ***Data analysis using SQL and Excel***. John Wiley & Sons.
2. Guerrero, Hector, Rauscher Guerrero, and Rauscher (2019). ***Excel Data Analysis.*** Springer International Publishing.

Books for Reference:

1. Hector Guerrero (2019) ***Excel data Analysis :modeling and simulation*** **.** Springer International Publishing

MSSTA03O03 : Basic Statistics Using SPSS (4 Credits)

Course Outcomes:

CO1: Navigate the SPSS interface, input data, and perform basic operations for data analysis.

CO2: Conduct descriptive statistics analysis to summarize and interpret data distributions in SPSS.

CO3: Perform regression analysis in SPSS, interpret output, and assess model fit.

CO4: Conduct hypothesis tests using SPSS and interpret results effectively.

Unit 1: Overview of SPSS software, Opening SPSS, Layout of SPSS,Structure of SPSS Exiting SPSS, inputting data, An overview of SPSS.

Unit 2: Exploring data distributions using descriptive statistics, Creating frequency distributions and summary tables, Generating basic visualizations (e.g., histograms, box plots) in SPSS.

Unit 3: Understanding correlation and covariance, Performing correlation analysis in SPSS, Introduction to linear regression and its application in SPSS,Understanding hypothesis testing principles, Conducting hypothesis tests in SPSS, Interpreting SPSS output for hypothesis testing.

Unit 4: Generating various types of charts and graphs in SPSS, Customizing visualizations for clarity and impact, Exploring the SPSS Chart Builder tool.

Books for Reading

1. Landau, S., & Everitt, B. S. (2003). ***A Handbook of Statistical Analyses Using SPSS.*** Chapman and Hall/CRC.

2. Tukey, J. W. (1977). ***Exploratory Data Analysis (Vol. 2, pp. 131-160).***

3. Aldrich, J. O. (2018). ***Using IBM SPSS statistics: An Interactive Hands-on Approach.*** Sage Publications.

MSSTA03O04 : Research Methodology (4 Credits)

Course Outcomes:

CO 1: Understand to explain the objectives and types of research, define and select research problems, conduct literature reviews, and identify research gaps.

CO 2: Understand Research design and methods

CO 3: Understand statistical methods in research design

CO 4: Understand the Article and thesis writing guidelines

Unit 1 :

Objectives and types of research – Defining and formulating the research problemsImportance of literature review, Identifying research gaps from literature review – Definition and types of research. The scientific method: Observation – questions – Hypothesis – Experimentation – Critical communication – Formulation of research problem – Research ethics.

Unit 2 :

Research design and methods: Basic Principles- Features of good designImportant concepts relating to research design- Hypothesis: Development for working hypothesis, Procedure for hypothesis testing- Hypothesis testing techniques-power and limitations of Hypothesis testing.

Unit 3 :

Basic statistical measurements: Measures of Central Tendency- Arithmetic Mean, Median, Mode, Geometric Mean, and Harmonic Mean. Measures of Variation-Range Mean Deviation, Quartile Deviation, and Standard Deviation. Data Collection and Analysis: Execution of Research-Observation and collection of data- Data collection methodsSampling methods- Data processing and analysis strategies-Data analysis with statistical packages (R programming)- Hypothesis Testing- Generalization and interpretation.

Unit 4 :

Article and thesis writing guidelines: Structure and components of scientific reports-Types of reports-technical reports and thesis. Guidelines for writing abstract, introduction, methodology, results and discussion. Conclusion structure and language of typical reports-illustration and table-bibliography, reference and footnotes-oral presentation planning and preparation-importance of effective communication-use of visual aids.

Books for study:

1. Kothari, C.R.(1990). **Research Methodology: Methods and Techniques**. New Age International. 418p.

Books for Reference:

1. Garg, B.L., Karadia, R., Agarwal, F. and Agarwal, U.K. (2002). **An introduction to Research Methodology,** RBSA Publishers.

2. Sinha, S.C. and Dhiman, A.K. (2002). **Research Methodology,** EssEss Publications.

3. Michael P Marder (2011). **Research Method for Science.** Cambridge University press edition

MSSTA03O05 : A First Course on LaTex for Scientific Documentation.

Course Outcomes

CO1: Introduce a software that is being widely used for scientific typesetting.

CO2: To make students know the importance of this software for publishing research articles, letters, project reports and books.

CO3: Beamer/slide presentation and thereby help them to be comfortable with the software food presentation .

Unit 1: Installation of Kile and MikeTeX. Class and packages. Latex programming and commands, sample packages. Error messages, Some sample errors, list of LaTeX error messages.

Unit 2: Fonts, symbols, indenting, paragraphs, line spacing, word spacing, titles and subtitles. Document class, page style, parts of the documents, table of contents. Command names and arguments, environments, declarations. Theorem like declarations, comments within text.

Unit 3: Mathematical environments, math mode, mathematical symbols. Graphic package, multivalued functions, drawing matrices. Tables, tables with captions. References to figures and tables in text.

Unit 4: Picture environments. Extended pictures, other drawing packages. Preparing book, project report in LaTeX, LaTeX Beamer for Technical Presentations.

Books for Reading:

1. Kottwitz, S. (2021). ***LaTeX Beginner's Guide: Create Visually Appealing Texts, Articles, and Books for Business and Science Using LaTeX***. United Kingdom: Packt Publishing.

2. Lamport (1994). ***Latex: A Document Preparation System, 2/E.*** India: Pearson Education.

3. Kopka, H., Daly, P. W. (2003). **Guide to LaTeX.** United Kingdom: Pearson Education