Estd. 1962
"A++" Accredited by NAAC (2021)
With CGPA 3.52

SHIVAJI UNIVERSITY, KOLHAPUR - 416004, MAHARASHTRA
PHONE:EPABX-2609000, www.unishivaji.ac.in, bos@unishivaji.ac.in
शिवाजी विद्यापीठ, कोल्हापूर -४१६००४,महाराष्ट्र
दूरध्वनी-ईपीएबीएक्स -२६०९०००, अभ्यासमंडळे विभाग दुरध्वनी ०२३१—२६०९०९४
०२३१—२६०९४८७

SU/BOS/Science/473 **Date: 19/08/2024.**

To,

The Principal, All Concerned Affiliated Colleges/Institutions Shivaji University, Kolhapur	The Head/Co-ordinator/Director All Concerned Department (Science) Shivaji University, Kolhapur.

Subject: Regarding Minor Change syllabi of M.Sc. Part-I & II as per NEP-2020 (2.0) degree programme under the Faculty of Science and Technology.

Ref: SU/BOS/Science/482 Date: 01/07/2023 & 09 Date: 02/01/2024

Sir/Madam,

With reference to the subject mentioned above, I am directed to inform you that the university authorities have accepted and granted approval to the Minor Change in syllabi, nature of question paper and equivalence of M.Sc. Part-I & II as per NEP-2020 (2.0) degree programme under the Faculty of Science and Technology.

M.Sc. Part-I & II as per NEP-2020 (2.0)			
1.	Statistics	2.	Applied Statistics and Informatics

This syllabus, nature of question shall be implemented from the academic year 2024-2025 onwards. A soft copy containing the syllabus is attached herewith and it is also available on university website www.unishivaji.ac.in,NEP-2020@suk(Online Syllabus).

The question papers on the pre-revised syllabi of above-mentioned course will be set for the examinations to be held in October /November 2024 & March/April 2025. These chances are available for repeater students, if any.

You are, therefore, requested to bring this to the notice of all students and teachers concerned.

Thanking you,

Dy Registrar
Dr. S. M. Kubal

Copy to:

1	The Dean, Faculty of Science & Technology	4	P.G Admission / Eligibility Section
2	The Chairman, Respective Board of Studies	5	Computer Centre/ Eligibility Section
3	B.Sc. Exam/ Appointment Section	6	Affiliation Section (U.G.) (P.G.)

SHIVAJI UNIVERSITY, KOLHAPUR

Established: 1962

A^{++} Accredited by NAAC (2021) with CGPA 3.52

Structure and Syllabus in Accordance with

National Education Policy – 2020 (NEP 2.0)

with Multiple Entry and Multiple Exit

of

Master of Science (Applied Statistics and Informatics)

under
Faculty of Science and Technology

(To Be Implemented From Academic Year 2023-24)

1. Preamble

The M.Sc. (Applied Statistics and Informatics) programme at Shivaji University, Kolhapur, is a dynamic postgraduate programme designed to provide students with a comprehensive knowledge of statistical methods and their applications in the realm of informatics. This programme merges statistical analysis with information technology, enabling students to extract valuable insights from complex data sets. Through a blend of theoretical teachings and practical hands-on experience, students acquire proficiency in data analysis, visualization, and interpretation. The curriculum emphasizes real-world problem-solving and equips graduates with the skills to make informed decisions in diverse sectors such as business, healthcare, and technology, where data-driven strategies are pivotal.

Equivalence: This programme is equivalent to M. Sc.(Statistics) programme of Shivaji University as per approvals of UGC, New Delhi and the authorities of Shivaji University, Kolhapur. It covers more than 75% of the core courses of M.Sc.(Statistics) programme of Shivaji University.

2. Duration

- 2 Years

3. Eligibility for Admission

- Three-year B. Sc. Degree with Statistics as principal / major subject

4. Medium of Instruction

- English

5. Programme Structure

Structure in Accordance with National Education Policy - 2020
With Multiple Entry and Multiple Exit Options
M.Sc. (Applied Statistics and Informatics) Part – I (Level - 6.0)

	Course Code	Teaching Scheme			Examination Scheme					
		Theory and Practical			University Assessment (UA)			Internal Assessment (IA)		
		Lectures + Tutorial/ (Hours/ week)	Practical (Hours/ week)	Credit	Maximum Marks	Minimum Marks	Exam. Hours	Maximum Marks	Minimum Marks	Exam. Hours
Semester-I										
Major Mandatory	MMT11	4	--	4	80	32	3	20	8	0.5
	MMT12	4	--	4	80	32	3	20	8	0.5
	MMT13	2	--	2	40	16	1.5	10	4	0.25
	MMP14	--	12	4	--	--	--	100	40	3
Major Elective	MET1	4	--	4	80	32	3	20	8	0.5
Research Methodology	RM	4	--	4	80	32	3	20	8	0.5
Total				22	360			190		
Semester-II										
Major Mandatory	MMT21	4	--	4	80	32	3	20	8	0.5
	MMT22	4	--	4	80	32	3	20	8	0.5
	MMT23	2	-	2	40	16	1.5	10	4	0.25
	MMP24	--	12	4	--	--	--	100	40	3
Major Elective	MET2	4	--	4	80	32	3	20	8	0.5
OJT/FP	OJT/FP	4	--	4	*					
Total				22						
Total (Sem I + Sem II)				44						

<table>
<tr><td>• MMT–Major Mandatory Theory
• MMP–Major Mandatory Practical
• MET–Major Elective Theory
• RM - Research Methodology
• OJT/FP- On Job Training/ Field Project</td><td>• Total Marks for M.Sc.-I : 1100
• Total Credits for M.Sc.-I (Semester I & II) : 44
• Separate passing is mandatory for University and Internal Examinations</td></tr>
<tr><td colspan="2">* Evaluation scheme for OJT/FP shall be decided by concerned BOS</td></tr>
<tr><td colspan="2">• Requirement for Entry at Level 6.0: B. Sc. degree of 3 years with Statistics as Principle/Major Subject.</td></tr>
<tr><td colspan="2">• Requirement for Exit after Level 6.0:
Student can exit after completion of Level 6.0 with Post Graduate Diploma in Applied Statistics and Informatics</td></tr>
<tr><td colspan="2">• Requirement for Entry at Level 6.5: Completion of Level 6.0 with Statistics as Principle/Major Subject.</td></tr>
</table>

Structure in Accordance with National Education Policy - 2020
With Multiple Entry and Multiple Exit Options
M.Sc. (Applied Statistics and Informatics) Part – II (Level-6.5)

	Course Code	Teaching Scheme			Examination Scheme					
		Theory and Practical			University Assessment (UA)			Internal Assessment (IA)		
		Lectures + Tutorial/ (Hours/ week)	Practical (Hours/ week)	Credit	Maximum Marks	Minimum Marks	Exam. Hours	Maximum Marks	Minimum Marks	Exam. Hours
Semester-III										
Major Mandatory	MMT31	4	--	4	80	32	3	20	8	0.5
	MMT32	4	--	4	80	32	3	20	8	0.5
	MMT33	2	-	2	40	16	1.5	10	4	0.25
	MMP34	--	12	4	--	--	--	100	40	3
Major Elective	MET3	4	--	4	80	32	3	20	8	0.5
Research Project	RP			4	#					
Total				22						
Semester-IV										
Major Mandatory	MMT41	4	--	4	80	32	3	20	8	0.5
	MMT42	4	--	4	80	32	3	20	8	0.5
	MMP43	--	12	4	--	--	--	100	40	3
Major Elective	MET4	4	--	4	80	32	3	20	8	0.5
Research Project	RP			6	##					
Total				22						
Total (Sem I + Sem II)				44						

• MMT–Major Mandatory Theory • MMP–Major Mandatory Practical • MET–Major Elective • RP- Research Project	• Total Marks for M.Sc.-II : **1100** • Total Credits for M.Sc.-II (Semester III & IV) : 44 • Separate passing is mandatory for University and Internal Examinations
# Evaluation scheme for Research Project shall be decided by concerned BOS	
## Evaluation scheme for Research Project shall be decided by concerned BOS	

6. Programme Outcomes (POs) and Programme Specific Outcomes (PSOs)

Programme Outcomes (POs)

Post Graduates of the M.Sc. Applied Statistics and Informatics programme will be able to:

- Have sufficient knowledge of theoretical concepts in Statistics for (i) choosing and applying the most appropriate statistical methods/ techniques for collecting and analyzing data (ii) interpreting the results of analyses in relation to given real life situations.
- Have ability to explain the inter-connections between various sub disciplines and apt use of these inter-connections in modelling real life problems.
- Have ability to recognize the importance of statistical thinking and training, and to acquire the state-of-the-art developments in Statistics independently from available resources.
- Develop expertise in statistical research and real life data analysis.

Programme Specific Outcomes (PSOs)

After completion of M.Sc. Applied Statistics and Informatics programme the student will be able to:

- Develop stochastic models for studying, analyzing, interpreting and forecasting real life phenomenon in diverse disciplines.
- Effectively use necessary statistical software and computing environment including C, C++, R, Python, MS-EXCEL among others and have expertise in handling computational systems for data storage, update and retrieval.
- Have the versatility to work effectively in a broad range of establishments (including R&D sectors, analytics, scientific laboratories, government, financial, health, educational) or to continue for higher education, and exhibit ethical and professional behaviour in team work.

7. Course Codes

Semester-I		
Sr. No.	**Major Mandatory**	
1	Distribution Theory	MSU0325MML905G1
2	Estimation Theory	MSU0325MML905G2
3	Statistical Computing	MSU0325MML905G3
4	Practical I	MSU0325MMP905G4
5	Research Methodology in Statistics	MSU0325RML905G
	Major Elective	
6	Fundamentals of Computer Programming	MSU0325MEL905G1
	Statistical Programming using C	MSU0325MEL905G2
Semester-II		
Sr. No.	**Major Mandatory**	
1	Theory of Testing of Hypothesis	MSU0325MML905H1
2	Linear Models and Regression Analysis	MSU0325MML905H2
3	Statistical Programming using Python	MSU0325MML905H3
4	Practical-II	MSU0325MMP905H4
5	On Job Training/Field Project	MSU0325OJP905H/ MSU0325FPP905H
	Major Elective	
6	Advanced Data Structure with C++	MSU0325MEL905H1
	Cloud computing and Data Visualization	MSU0325MEL905H2
	Statistical Programming using C++	MSU0325MEL905H3
	Reliability Theory	MSU0325MEL937H4

Courses

M.Sc. Semester-III			
Sr. No.	**Course Category**	**Course Title**	**Course Code**
1	Major Mandatory	Stochastic Processes	MSU0325MML905I1
2		Theory and Practice of Machine Learning	MSU0325MML905I2
3		Multivariate Analysis	MSU0325MML905I3
4		Practical-III	MSU0325MMP905I4
5	Major Elective*	Data Base Management System	MSU0325MEL905I1
6		Statistical Programming using SAS	MSU0325MEL905I2
7		Artificial Intelligence	MSU0325MEL905I3
8	Research Project	Research Project	MSU0325RPP905I
M.Sc. Semester-IV			
Sr. No.	**Course Category**	**Course Title**	**Course Code**
1	Major Mandatory	Biostatistics	MSU0325MML905J1
2		Time Series Analysis	MSU0325MML905J2
3		Practical-IV	MSU0325MMP905J3
4	Major Elective*	Advanced Topics in Artificial Intelligence	MSU0325MEL905J1
5		Marketing and Financial Analytics	MSU0325MEL905J2
6		Analysis of Big Data	MSU0325MEL905J3
7	Research Project	Research Project	MSU0325RPP905J

*Only one course under this category is to be chosen by the student.

8. Syllabus

M. Sc. Applied Statistics and Informatics (Part I) (Level-6.0) (Semester I)
(NEP-2020)
(Introduced from Academic Year 2023-24)

Title of Course: DISTRIBUTION THEORY
Course Code: MSU0325MML905G1
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the students will be able to:

i) Understand and explain the concept of univatiate and multivariate random variables and related entities
ii) Understand and explain the nature of various probability distributions and perform related computations.
iii) Understand probability models for multivariate data and perform related computations
iv) Understand non-central sampling distribution and able to perform their applications, able to perform computations related to order statistics

Unit 1: Random variable, Cumulative distribution function (CDF) and its properties, continuous and discrete distributions, mixtures of probability distributions, decomposition of mixture CDF into discrete and continuous CDFs, computation of probabilities of events using CDF, expectation and variance of mixture distributions. Quantiles of probability distributions. Transformations of univariate random variables, probability integral transformation. (12L+3T)

Unit 2: Concepts of location, scale and shape parameters of distributions with examples. Symmetric distributions and their properties. Moment inequalities: Basic, Holder, Markov, Minkowski, Jensen, Chebyshev's inequalities, and their applications. Random vectors, joint distributions, Independence, variance-covariance matrix, joint MGF. Conditional expectation and variances. Transformations of bivariate random variables, Convolutions, compound distributions. (12L+3T)

Unit 3: Multivariate normal distribution, two definitions and their equivalence, singular and nonsingular normal distribution, characteristic function, moments, marginal and conditional distributions. Maximum likelihood estimators of the parameters of the multivariate normal distribution and their sampling distributions. Wishart matrix and its distribution (statement only), properties of Wishart distribution, distribution of generalized variance(statement only). Marshall-Olkin bivariate exponential distributions. (12L+3T)

Unit 4: Non-central chi-square, t and F distributions, distributions of linear and quadratic forms involving normal random variables, Fisher Cochran and related theorems (statements only) and their applications. Order Statistics: Distribution of an order statistics, joint distributions of two order statistics, distribution of spacings, normalized spacings with illustration to exponential case, distribution of sample median and sample range. (12L+3T)

References:

1. Casella, G., & Berger, R. L. (2021). Statistical inference. Cengage Learning.
2. DasGupta, A. (2010). Fundamentals of probability: A first course. Springer Science & Business Media.
3. Johnson N. L. & Kotz. S. (1996). Distributions in Statistics Vol-I, II and III. JohnWiley and Sons New York.
4. Kotz, S., Balakrishnan, N., & Johnson, N. L. (2004). Continuous multivariate distributions, Volume 1: Models and applications (Vol. 1). John Wiley & Sons.
5. Rao C. R. (1995). Linear Statistical Inference and Its Applications. John Wiley & Sons.
6. Rohatagi V. K. & Saleh A. K. Md. E.(2001). Introduction to Probability Theory andMathematical Statistics. John Wiley and sons Inc.

Title of Course: ESTIMATION THEORY
Course Code: MSU0325MML905G2
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the students will be able to:

i) Explain the principles of data reduction and obtain sufficient, minimal sufficient, and complete statistics for various families of distributions.
ii) Obtain UMVUE of parameters of various distributions and determine Cramer-Rao and Chapman-Robbins-Kiefer lower bounds for the variances of unbiased estimators.
iii) Apply parametric and nonparametric methods to obtain estimators.
iv) Obtain CAN and BAN estimators.

Unit 1: Principles of data reduction: sufficiency principle; sufficient statistics; factorization theorem; minimal sufficient statistic; minimal sufficient statistic for exponential family, power series family, curved exponential family, and Pitman family; completeness; bounded completeness; ancillary statistics, Basu's theorem and it's applications.

(12L + 3T)

Unit 2: Unbiased estimation: unbiased estimator; uniformly minimum variance unbiased estimator (UMVUE); A necessary and sufficient condition for an estimator to be UMVUE; Rao-Blackwell theorem and Lehmann-Scheffe theorem, and their applications in finding UMVUEs; Fisher information function and Fisher information matrix; Cramer-Rao lower bound; Chapman-Robbins-Kiefer lower bound.

(12L + 3T)

Unit 3: Methods of finding estimators: method of moments estimator; maximum likelihood estimator (MLE), properties of MLE, MLE in nonregular families; method of scoring; method of minimum chi-square, EM algorithm. Nonparametric estimation: degree of an estimable parameter, kernel, U-statistic and its properties.

(12L + 3T)

Unit 4: Consistency of an estimator, weak and strong consistency, joint and marginal consistency, invariance property under continuous transformations, methods of constructing consistent estimators, Comparison of consistent estimators, asymptotic relative efficiency, minimum sample size required by the estimator to attain certain level of accuracy. Consistent Asymptotic Normal (CAN) Estimators: Definition of CAN estimator for real and vector valued parameters, invariance of CAN property under non-vanishing differentiable transformation (delta method). Methods of constructing CAN estimators: Method of Moments, method of percentiles, comparison of CAN estimators. BAN estimators, CAN and BAN estimators in one parameter and multi-parameter exponential family of distributions, Crammer family of distributions, Cramer – Huzurbazar theorem (Statement only).

(12L + 3T)

References:

1. Casella, G., & Berger, R. L. (2021). Statistical inference. Cengage Learning.
2. Deshmukh S., Kulkarni M. (2022). Asymptotic Statistical Inference: A Basic Course Using R. Springer Verlag, Singapor.
3. Dudewicz, E. J., & Mishra, S. (1988). Modern mathematical statistics. John Wiley & Sons, Inc.
4. Kale, B. K., & Muralidharan, K. (2015). Parametric inference: An introduction. Alpha Science International Limited.
5. Lehmann, E. L. (1983). Theory of Point Estimation. John Wiley & sons.
6. Mukhopadhyay, P. (2015). Mathematical Statistics, Books and Allied (p) Ltd.
7. Rao C. R. (1995). Linear Statistical Inference and Its Applications. John Wiley & Sons.
8. Rohatgi, V. K., & Saleh, A. M. E. (2015). An introduction to probability and statistics. John Wiley & Sons.

Title of Course: STATISTICAL COMPUTING
Course Code: MSU0325MML905G3
Total Credits: 02
Course outcomes:
Upon successful completion of this course, the students will be able to:

i) Perform data organization, data manipulation, statistical and mathematical computations, and data analysis using MSEXCEL.
ii) Perform data organization, data manipulation, statistical and mathematical computations, and data analysis using R.

Unit 1: MSEXCEL: Introduction to MSEXCEL. Cell formatting, conditional formatting, Data manipulation using EXCEL: sort and filter, find and replace, text to columns, remove duplicate, data validation, consolidate, what-if-analysis. Working with Multiple Worksheets and Workbooks. Built-in mathematical and statistical functions for obtaining descriptive statistic, computing PMF/PDF, CDF and quantiles of the well-known distributions, rand and randbetween function, Logical functions: if, and, or, not. Lookup functions: hlookup, vlookup, Formula Errors, Creating and Working with Charts, Database functions, Text functions, Date and time functions, Excel add-ins: analysis tool pack, Pivot tables and charts.

(12L+3T)

Unit 2: R-software: Introduction to R, data types and objects, operators, data input, data import and export, built in functions for descriptive statistics, random sampling and computation of pdf, cdf and quantiles of well known distribution. Strings and Dates in R. apply family of functions. Saving work in R. Matrix algebra, graphical procedures, frequencies and cross tabulation, built in functions: lm, t.test, prop.test, wilcox.test, ks.test, var.test, chisq.test, aov. Control statements. Programming, user defined functions, R- packages. R-studio.

(12L+3T)

References:

1. Gardener, M. (2012). Beginning R: the statistical programming language. John Wiley & Sons.
2. Held, B., Moriarty, B., & Richardson, T. (2019). Microsoft Excel Functions and Formulas with Excel 2019/Office 365. Mercury Learning and Information.
3. Herkenhoff, L., & Fogli, J. (2013). Applied statistics for business and management using Microsoft Excel. New York: Springer.
4. Purohit, S. G., Gore, S. D., & Deshmukh, S. R. (2015). Statistics using R. Alpha Science International.
5. Thulin, M. (2021). Modern Statistics with R: From wrangling and exploring data to inference and predictive modelling. BoD-Books on Demand.
6. Weblinks:
   https://support.microsoft.com/en-us/excel
   https://cran.r-project.org/manuals.html

Title of Course: FUNDAMENTALS OF COMPUTER PROGRAMMING

Course Code: MSU0325MEL905G1

Total Credits: 04

Course outcomes:

Upon successful completion of this course, the students will able to:

i) Develop algorithms, flowcharts for simple as well as complex algorithms.
ii) Describe how data represented, manipulated, and stored in a computer.
iii) Understand and apply the control structures while writing program.
iv) Understand and use an arrays, pointers and structures effectively in programming.

Unit 1: Overview of Computer programming, Algorithms: The concept and features of the algorithm, ways of writing the algorithm, writing step by step procedure, Problem redefinition, Flow charts/Decision Trees: various symbols used to develop a flow chart, Advantages and drawbacks of flowcharts, concept of Tracing and Testing of Algorithm/flowchart. Pseudo code generation. Fundamentals of C programming: Features of C language, structure of C program, comments, header files, data types, constants and variables, operators, expressions, evaluation of expressions, type conversion, precedence and associativety, I/O functions.

(12 L + 3 T)

Unit 2: Control Structures: Decision control structures and Loop control structures. Structured programming, Modular programming, TOP DOWN/BOTTAM UP approach, Recursive algorithms, Examples, Illustrating structured program development methodology and use of block structured algorithmic language to solve specific problem. Syntax and semantics, documentation and debugging of a program.

(12 L + 3 T)

Unit 3: Advanced Data Types and their implementation: Arrays, Strings, Records, Pointers, Structures, Union, and Applications in the record keeping of real life system. Dynamic memory allocation using MALLOC, CALLOC and REALLOC.

(12 L + 3 T)

Unit 4: Searching, Sorting and Update of the data: Various methods, Design and Analysis of algorithms - Devide and conqure, Backtracking. Dynamic programming, Emphasis should be given on better programming styles and implementation is expected through C compiler.

(12 L + 3 T)

References:

1. Balagurusamy, E. (2016). Programming In Ansi C.
2. Horowitz & Sahani (1998). *Fundamentals of Computer Algorithms*. GalgotiaPublications,
3. Knuth D. E. (1997). *The Art of Computer Programming*.Volume 1: Fundamental -Narosa Publishing House.
4. Kruse R. L. (2006): *Data structures and program design C*, 2nd edition-PEARSON.
5. Ravichandran D. *Programming with C++,* 3rd edition, McGraw Hill Education.
6. Shukla R. K. (2015). *Analysis and Design of Algorithms, A beginners approach.* .Wiley.

Title of Course: RESEARCH METHODOLOGY IN STATISTICS

Course Code: MSU0325RML905G

Total Credits: 04

Course outcomes:

Upon successful completion of this course, the students will be able to:

i) Understand the concept of research, research process, and research ethics.
ii) Understand and apply various sampling methods for data collection and estimate the parameters.
iii) Understand the concept of simulation and able to simulate real life processes
iv) Estimate bias and standard error of an estimator using resampling techniques, apply, numerical methods to solve systems of linear equations, to obtain the roots of a nonlinear equation, and to solve definite integrals.

Unit 1: Meaning of research, objectives of research, motivation in research, types of research, research approaches, significance of research, research methods vs. methodology, research and Scientific method, research process, criteria of good research, defining research problem, research design, Research Ethics, publication of research, Plagiarism, Intellectual property rights, Patents and its filing procedures. (12L+3T)

Unit 2: Sampling techniques: review of simple random sampling stratified random sampling, systematic random sampling, cluster sampling, two phase sampling, ratio and regression method of estimation. Probability proportional to size sampling: Cumulative total method, Lahiri's method, Hansen-Horwitz estimator and its properties, Horwitz-Thompson estimator, Des Raj estimators for a general sample size. Non-sampling errors, techniques for handling non-response: Hansen–Horwitz and Demings model for the effect of call-backs. Randomised response techniques, dichotomous population, Warners model, MLE in Warners model, unrelated question model. (12L+3T)

Unit 3: Concept and need of simulation, random number generator, true random number and pseudo random number generators, requisites of a good random number generator. Tests for randomness. Congruential method of generating uniform random numbers. Algorithms for generating random numbers from well-known univariate discrete and continuous distributions, generating random vectors from multinomial, bivariate normal, and bivariate exponential distributions, generating random numbers from mixture of distributions (related results without proofs). Acceptance-Rejection Technique. Use of random numbers to evaluate integrals, to study the systems involving random variables, to estimate event probabilities and to find expected value of random variables. Use of random numbers for performance evaluation of estimators and statistical tests. (12L+3T)

Unit 4: Resampling methods: Bootstrap methods, estimation of bias and standard errors, estimation of sampling distribution, confidence intervals. Jackknife method: estimation of bias and standard errors, bias reduction method. Numerical methods for solution to system of linear equations: Jacobi and Gauss-Seidel methods with convergence analysis. Numerical methods for finding roots of nonlinear equation: Newton-Raphson method, bisection method; Newton-Raphson for system of non- linear equations. Numerical integration: quadrature formula, trapezoidal rule and Simpson's rules for single integral. (12L+3T)

References:

1. Atkinson, K. E. (1989). An introduction to numerical analysis, John Wiley and Sons.
2. Chaudhuri, A., & Stenger, H. (2005). Survey sampling: theory and methods. CRC Press.
3. Cochran, W. G. (1977). Sampling techniques. John Wiley & Sons.
4. Devroye L. (1986). Non- Uniform Random Variate Generation. Springer- Verlag New York.
5. Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.
6. Kennedy, W. J., & Gentle, J. E. (2021). Statistical computing. Routledge.
7. Kothari, C. R. (2004). Research methodology: Methods and techniques. New Age International.
8. Morgan, B. J. (1984). Elements of simulation (Vol. 4). CRC Press.

9. Mukhopadhyay, P. (2008). Theory and methods of survey sampling. PHI Learning Pvt. Ltd..
10. Robert, C. P., Casella, G., & Casella, G. (1999). Monte Carlo statistical methods (Vol. 2). New York: Springer.
11. Ross, S. M. (2022). Simulation. Academic Press.
12. Rubinstein, R. Y., & Melamed, B. (1998). Modern simulation and modeling (Vol. 7). New York: Wiley.
13. Singh, D., & Chaudhary, F. S. (1986). Theory and analysis of sample survey designs. John Wiley & Sons.
14. Sukhatme P. V., Sukhatme S. & Ashok C (1984). Sampling Theory of surveys and applications . Iowa university press and Indian society of agricultural statistics, New Delhi.

M. Sc. Statistics (Part I) (Level-6.0) (Semester II)
(NEP-2020)
(Introduced from Academic Year 2023-24)

Title of Course: THEORY OF TESTING OF HYPOTHESES
Course Code: MSU0325MML905H1
Total Credits: 04
Course outcomes:
Upon successful completion of this course, a student will be able to:

i) Understand the concept of testing of hypothesis, test statistic, critical regions, size and power of a test.
ii) Develop and apply MP test, UMP test, UMPU test, similar tests and a test with Neyman structure.
iii) Obtain and interpret different interval estimates of parameters.
iv) Develop and apply large sample tests.

Unit 1: Problem of testing of hypothesis, null and alternative hypotheses, Simple and composite hypotheses, test function, Randomized and non-randomized tests, power function of a test, Most powerful (MP) test, Neyman-PearsonLemma, Monotone likelihood ratio property, Uniformly Most Powerful (UMP) test and its existence, determination of minimum sample size to achieve the desired strength of a test. Concept of p-value.

(12L+3T)

Unit 2: UMP tests for two sided alternatives, examples of their existence and non-existence. Unbiased test, Generalized Neyman Pearson lemma, UMPU tests and their existence in case of exponential families (Statements of the theorems only). Similar tests, test with Neyman structure. Tests for means of multivariate normal distributions based on Hotelling's T^2 Statistic.

(12L+3T)

Unit 3: Interval estimation: Confidence interval, relation with testing of hypotheses problem, Uniformly Most Accurate(UMA) and Uniformly Most Accurate Unbiased(UMAU) confidence intervals, shortest length confidence intervals, Asymptotic Confidence Intervals based on CAN estimators, Variance stabilizing transformations (VST), confidence interval based on VST, Asymptotic Confidence regions.

(12L+3T)

Unit 4: Likelihood ratio test (LRT) and its asymptotic distribution, Wald test, Rao's Score test, Pearson Chi-square test of goodness of fit, Bartlett's test for homogeneity of variances. Large sample tests based on VST. Consistent test, comparison of tests: asymptotic relative efficiency of tests (Pitman and Bahadur efficiency). Performance evaluation (based on simulation) of asymptotic tests and confidence intervals.

(12L+3T)

References:

1. Kale, B. K., &Muralidharan, K. (2015). Parametric inference: An introduction. Alpha Science International Limited.
2. Dudewicz, E. J. and Mishra, S. N. (1988). Modern Mathematical Statistics, John Wiley & Sons.
3. Ferguson, T. S. (2014). Mathematical statistics: A decision theoretic approach. Academic press.
4. Gibbons, J.D., & Chakraborti, S. (2010). Nonparametric Statistical Inference (5th ed.). Chapman and Hall/CRC.
5. Lehman, E. L. (1987). Theory of testing of hypotheses. Students Edition.
6. Randles, R. H., & Wolfe, D. A. (1979). Introduction to the theory of nonparametric statistics. John Wiley.
7. Rohatgi, V. K., & Saleh, A. M. E. (2015). An introduction to probability and statistics. John Wiley & Sons.
8. Zacks, S. (1971). Theory of Statistical Inference, John Wiley & Sons, New York.

Title of Course: LINEAR MODELS AND REGRESSION ANALYSIS
Course Code: MSU0325MML905H2
Total Credits: 04
Course outcomes:
Upon successful completion of this course, a student will be able to:

i) Understand the concept of general linear model and associated inferential procedures.
ii) Understand and develop multiple linear regression models
iii) Identify the problems in developing multiple linear regression models and apply remedies.
iv) Understand generalized linear models and apply them for analyzing real data.

Unit 1: General linear model: definition, assumptions, concept of estimability, least squares estimation, BLUE, estimation space, error space, Gauss Markov theorem, variances and covariances of BLUEs, Distribution of quadratic forms for normal variables: related theorems(without proof), Tests of hypotheses in general linear models. Description of the ANOVA and linear regression models as the particular cases of the general linear model.

(12L+3T)

Unit 2: Multiple linear regression model, Least squares estimates (LSE) of parameters, Properties of LSE, Hypothesis testing, confidence and prediction intervals, General linear hypothesis testing, Model adequacy checking, Dummy variables and their use in regression analysis. Transformations to correct model inadequacies: VST and Box-Cox power transformation.

(12L+3T)

Unit 3: Multicollinearity: Consequences, detection and remedies, ridge regression. Autocorrelation: sources, consequences, detection (Durbin-Watson test) and remedies. Parameter estimation using Cochrane-Orcutt method. Variable Selection Procedures: R- square, adjusted R-square, Mallows' Cp, forward, backward and stepwise selection methods, AIC, BIC. Robust Regression: need for robust regression, M-estimators, properties of robust estimators: breakdown and efficiency. Asymptotic distribution of M-estimator (Statement only).

(12L+3T)

Unit 4: Generalized linear models: concept of generalized linear model, Link function, ML estimation, large sample tests about parameters, goodness of fit, analysis of deviance. Residual analysis, types of residuals: raw, Pearson, deviance, Anscombe, quantile; residual plots. Variable selection: AIC and BIC. Logistic regression: logit, probit and cloglog models for dichotomous data, ML estimation, Odds ratio and its interpretation, hypothesis tests about model parameters. Hosmer-Lemeshow test, multilevel logistic regression, Logistic regression for Nominal response. Poisson regression.

(12L+3T)

References:

1. Birkes, D., & Dodge, Y. (2011). Alternative methods of regression. John Wiley & Sons.
2. Cook, R. D., & Weisberg, S. (1982). Residuals and influence in regression. New York: Chapman and Hall.
3. Draper, N. R., & Smith, H. (1998). Applied regression analysis. John Wiley & Sons.
4. Huber,P.J.andRonchetti,E.M (2011)Robust Statistics,Wiley,2ndEdition.
5. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Wasserman, W. (2004). Applied linear regression models. New York: McGraw-Hill/Irwin.
6. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. 5th Ed. John Wiley & Sons.
7. Seber,G.A.,Wild,C.J.(2003).Non linear Regression,Wiley.
8. Weisberg, S. (1985). Applied Linear Regression, John Wiley & Sons. New York.

Title of Course: STATISTICAL PROGRAMMING USING PYTHON
Course Code: MSU0325MML905H3
Total Credits: 02
Course outcomes:
Upon successful completion of this course, a student will be able to:

i) Develop programs in Python.
ii) Perform data organization, data manipulation, statistical and mathematical computations, and data analysis using Python.

Unit 1: Introduction, installation, keywords, identifiers: variables, constants, literals; comments, Operators, statements and expressions, data types with methods: numbers, string, lists, tuple, dictionary, set; indexing and slicing of each data type, data type conversion, built-in functions, control statements and loops, list comprehensions, user defined functions, anonymous/lambda function, local and global variables, modules: math, stat, random;creating own modules.

(12L+3T)

Unit 2: Concept of library and its working, Data storage, manipulation, visualization and analysis using the libraries: Numpy, Pandas, Scipy, statsmodels, Matplotlip, Seaborn, Regular Expressions (RegEx), Ski-kit learn.

(12L+3T)

Reference:

1. Gowrishankar, S., &Veena, A. (2018). Introduction to Python programming. CRC Press.
2. Guttag, J. V. (2021). Introduction to Computation and Programming Using Python: With Application to Computational Modeling and Understanding Data. Mit Press.
3. Haslwanter, T. (2016). An Introduction to Statistics with Python. With Applications in the Life Sciences. Switzerland: Springer International Publishing.
4. Nelli, F. (2018). Python data analytics with Pandas, NumPy, and Matplotlib.
5. Unpingco, J. (2016). Python for probability, statistics, and machine learning (Vol. 1). Springer International Publishing.
6. VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc.".
7. URLs:
    - https://scikit-learn.org/stable/
    - https://numpy.org/
    - https://scipy.org/
    - https://www.statsmodels.org/stable/index.html
    - https://matplotlib.org/
    - https://pandas.pydata.org/

Title of Course: ADVANCED DATA STRUCTURE WITH C++
Course Code: MSU0325MEL905H1
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the student will be able to:

i) Write, compile, and execute programs in C++ and develop code in C++ by identifying the appropriate features of object oriented programming to solve statistical problems.
ii) Explain the concept of data structures, choose appropriate data structure for a specific problem and implement linear data structures such as stacks, queues, linked lists
iii) Implement tree data structure
iv) Implement graph data structure and search data structures such as hashing

Unit 1: C++ programming basics. Data types, Variables, Statements, Expressions, Control structures, Looping, Functions, Arrays, Pointers. Object oriented programming: Concept of OOP, class and objects, inheritance, polymorphism (function and operator overloading).

(12 L + 3 T)

Unit 2: Data Structures: definitions, operations, implementations and applications of basic data structures. Stack, Applications of Stack, Queue, Priority Queue, circular queue Applications of queue, Linked lists, doubly linked list, circular list, dynamic memory allocation, implementation of linked list, further operations, implementation of sparse matrices.

(12 L + 3 T)

Unit 3: Trees, binary trees, binary search trees, operations on binary search tree, applications of binary trees, threaded binary trees. General trees, using binary trees to represent general trees. AVL tree, operations on AVL tree. 2-3 trees, operations on 2-3 tree. Multi-way search tress, B-trees, B-tree indexing, operations on B trees. Huffman coding.

(12 L + 3 T)

Unit 4: Graphs, representation, operations on graph, Applications of graph: shortest path problem, Dijkstra's algorithm, topological ordering, minimum spanning tree, DFS and BFS spanning tree, Kruskal's algorithm. Tables, Hash tables: hashing techniques, collision resolution techniques, closed hashing, open hashing.

(12 L + 3 T)

References:

1. Decker, R., & Hirshfield, S. (1995). *Object Concept: An Introduction to ComputerProgramming Using C++*. International Thomson Computer Press.
2. Kanetkar Y. (2003). *Data Structures through C++*. BPB Publications.
3. Kruse, R., & Tondo, C. L. (2007). *Data structures and program design in C*. PearsonEducation India.
4. Samanta, D. (2001). *Classic data structures*. Prentice Hall India.
5. Sartaj, S. (1998). *Data structures, algorithms, and applications in C++*. McGraw-HillBoston, MA, USA
6. Savitch, W. J., & Savitch, W. (2003). *Problem solving with C++: The object ofprogramming*. Addison Wesley.
7. Stanley, B., & LIPPMAN, L. (2019). *C++ PRIMER*. Addison-Wesley.
8. Weiss, M. A. (2014). *Data structures and algorithm analysis in C++*. Pearson.

M. Sc. Applied Statistics and Informatics (Part II) (Level-6.5) (Semester III) (NEP-2020)

Title of Course: Stochastic Processes
Course Code: MSU0325MML905I1
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the students will be able to:

1. Identify appropriate stochastic process model for a given real life process.
2. Specify a given discrete time Markov chain in terms of a transition probability matrix and a transition diagram, and calculate higher step transition probabilities and limiting probabilities.
3. Understand and analyses discrete state space continuous time Markov chains and their practical applications
4. Explore the Galton-Watson Binary Branching process and understand the fundamental elements of Queuing models

Unit 1: Definition of stochastic process, classification of stochastic processes according to state space and time domain, finite dimensional distributions. Examples of various stochastic processes. Definition of Markov chain. Examples of Markov chains, Formulation of Markov chain models, initial distribution, transition probability matrix, Chapman-Kolmogorov equations, calculation of n-step transition probabilities. Simulation of Markov Chain.

(12L +3T)

Unit 2: Classification of states, irreducible Markov chain, period of the state, random walk and gambler's ruin problem, first entrance theorem, first passage time distribution. Long-Run proportions and limiting probabilities, relation with mean recurrence time, stationary distribution.

(12L +3T)

Unit 3: Discrete state space continuous time Markov chain, Poisson process and related results. Birth and death processes and associated cases. Renewal and delayed renewal processes, related theorems, key renewal theorem (Without proof) and its application. Simulation of Poisson process and discrete state space Markov processes.

(12L +3T)

Unit 4: Galton-Watson Binaymi Branching process. Generating functions and its properties, moments. Probability of ultimate extinction. Distribution of population size and association results. Simulation of branching process. Basic elements of Queuing model. Steady state probabilities and various average characteristics for the models: M/M/1, M/M/1 with balking, M/M/c and M/G/1.

(12 L+ 3T)

References:

1. Bhat B. R. (2000). Stochastic Models: Analysis and Applications, (New Age International)
2. Cinlar E. (2013): Introduction to Stochastic Process. (Courier Corporation)
3. Feller W.(2008): An Introduction to Probability Theory and Its Applications. (Wiley)
4. Hoel P. G., Port S. C. and Stone C. J. (1987): Introduction to Stochastic Processes. (Waveland Press)
5. Karlin S. and Taylor H. M. (1968): A First Course in Stochastic Process. (Academic Press)
6. Medhi J. (2009): Stochastic Process, (New Age International Publications)
7. Ross S. (1996): Stochastic Processes. (Wiley)
8. Ross S. (2014): Introduction to Probability Models. (Academic Press)
9. Taylor H. M. and Karlin S. (2014): An Introduction to Stochastic Modeling (Academic Press)

Title of Course: Theory and Practice of Machine Learning
Course Code: MSU0325MML905I2
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the students will be able to:

1. Explain supervised learning and construct classifiers namely, decision tree, k-nearest neighbour(s), logistic regression, naïve Bayes, Bayesian Belief Network
2. Compare different classifiers and employ techniques to improve their performance.
3. Use Artificial Neural Network and Support Vector Machine for classification and prediction.
4. Apply clustering techniques and generate association rules using apriori algorithm.

Unit 1: Data understanding and data cleaning, concept of supervised and unsupervised learning. Problem of classification, classification techniques: k-nearest neighbor, decision tree, Naïve Bayesian, classification based on logistic regression, Bayesian belief Network.

(12L+3T)

Unit 2: Model evaluation and selection: Metrics for Evaluating Classifier Performance, Holdout Method and Random Subsampling, Cross-Validation, Bootstrap, Model Selection Using Statistical Tests of Significance, Comparing Classifiers Based on Cost–Benefit and ROC Curves. Techniques to Improve Classification Accuracy: Introduction to Ensemble Methods, Bagging, Boosting and AdaBoost, Random Forests, Improving Classification Accuracy of Class-Imbalanced Data.

(12L+3T)

Unit 3: ANN and SVM: Artificial Neutral Network (ANN): Introduction to ANN, types of activation functions: identity, sigmoid, double sigmoid, tanh, softmax, loss functions: squared error, cross entropy; optimizers: gradient decent, stochastic and minibatch gradient decent; McCulloch-Pitts AN model, single layer network, multilayer feed forward network model, training methods, ANN & regression models. Support vector machine: Introduction to support vector machine, loss functions, soft margin, optimization hyperplane, support vector classification, support vector regression, linear programming support vector machine for classification and regression.

(12L+3T)

Unit 4: Unsupervised learning: Clustering: k-mediods, CLARA, DENCLUE, DBSCAN, Probabilistic model based clustering. Market Basket Analysis: Association rules and prediction, Apriori Algorithm, data attributes, applications to electronic commerce.

(12L+3T)

References:

1. Berson and Smith S.J. (1997) : Data warehousing, Data Mining, and OLAP, McGraw-Hill.
2. Breiman J.H Friedman, R.A. Olshen and stone C.J. (1984) : Classification and Regression Trees, Wadsworth and Brooks / Cole.
3. Han, J. and Kamber, M. and Pei, J. (2012) : Data Mining: Concepts and Techniques. MorganGaufmann.3rd Edition.
4. Mitchell T.M. (1997) : Machine Learning , McGraw-Hill.
5. Ripley B.D. (1996) : Pattern Recognition and Neural Networks. Cambridge University Press.
6. Vapnik V.N. The nature of Statistical learning theory, Springer.
7. Cristianini N. and Shawe-Taylor J. An Introduction to support vectormachines.
8. Data set source: http://www.ICS.uci.edu/~mlearn/MLRepository.html
9. Mehrika, K., Mohan, C., and Ranka (1997) Elements of Artificial neural networks. Penram international.
10. Hastie T, Tibshirani R, Friedmant J, (2009): The elements of statistical Learning, Springer.
11. Chattamvelli, R. (2015). Data mining methods. Alpha Science International.

Title of Course: Multivariate Analysis
Course Code: MSU0325MML905I3
Total Credits: 02
Course outcomes:
Upon successful completion of this course, the students will be able to:

1. Perform exploratory multivariate data analysis, understand and apply discrimination and classification techniques and apply hierarchical and non-hierarchical clustering techniques
2. Understand and apply the Canonical Correlation Analysis, principal component analysis and factor analysis.

Unit 1: Exploratory multivariate data analysis, sample mean vector, sample dispersion matrix, correlation matrix, graphical representation, Partial and multiple correlation coefficients. Discrimination and classification: Fisher's discriminant function and likelihood ratio procedure, minimum ECM rule, Rao's U statistics and its use in tests associated with discriminant function, classification with three populations. Cluster analysis, Hierarchical methods: Single, Complete, average linkage method and non-hierarchical clustering method: k-means clustering.

(12L+3T)

Unit 2: Canonical correlation analysis: Introduction, canonical variates and canonical correlation, interpreting population canonical variables, sample canonical variates and sample canonical correlations; Principal component analysis: Introduction, Population principal components, summarizing sample variation by principal components, Graphing the principal components, large sample inferences; Factor analysis: Introduction, Orthogonal factor model, methods of estimation, factor rotation and factor score

(12L+3T)

References:

1. Kshirsagar A. M.(1972) : Multivariate Analysis. Marcel-Dekker.
2. Johnson, R.A. and Wichern . D.W (2002) : Applied multivariate Analysis. 5th Ad.Prentice – Hall.
3. Anderson T. W. (1984) : An introduction to Multivariate statistical Analysis 2nd Ed. John Wiely.
4. Morrison D.F. (1976) : Multivariate Statistical Methods McGraw-Hill.
5. Bhuyan K. C. (2005): Multivariate Analysis and its applications, New central book agency Ltd. Kolkatta.

Title of Course: Data Base Management System
Course Code: MSU0325MEL905I1
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the students will be able to:

1. Understand and express the fundamental concepts of database.
2. Develop proficiency in designing and interpreting Entity-Relationship (ER) models, applying constraints, and addressing issues related to weak entity sets.
3. Master the basics of SQL and some of the advanced SQL queries
4. Familiar with the basic issues of transaction processing and concurrency control.

Unit 1: Introduction to Databases and Data Models: Concept of database system, purpose of database system, view of data, different sources of databases, relational databases, database architecture, importance of data models, basic building blocks, business rules, the evolution of data models, data abstraction, database users and administrators. Database design and ER Model: Overview, ER-Model, Constraints, ER-Diagrams, ERD issues, weak entity sets

(12L+3T)

Unit 2: Codd's rules, relational schema, introduction to UML relational database model: logical view of data, keys, and integrity rules. Relational Database design: features of good relational database design, atomic domain and normalization (1NF,2NF,3NF, BCNF). Relational algebra: Introduction of RDBMS, Selection and projection, set operations, renaming, Joins, Division, syntax, semantics. Operators, grouping and ungrouping, relational comparison. Calculus: tuple relational calculus, Domain relational Calculus, calculus vs algebra, computational capabilities.

(12L+3T)

Unit 3: Constraints: Concept and types of constrains. Views: Introduction to views, data independence, security, updates on views, comparison between tables and views. SQL: Basics of SQL, DDL, DML, DCL, structure: creation and alteration, defining constraints: primary key, foreign key, unique, not null, check, IN operator, Functions: aggregate functions, built-in functions, numeric, date, string functions, set operations, sub-queries, correlated sub- queries, use of group by, having, order by, join and its types, exist, any, all, view and its types joined relations, Triggers.

(12L+3T)

Unit 4: Transaction management: ACID properties, serializability and concurrency control, lock based concurrency control (2PL, Deadlocks), time stamping methods, optimistic methods, database recovery management, data dictionary. NoSQL: Overview, and history of NoSQL databases, definition of the four types of NoSQL database, introduction to Big Data.

(12L+3T)

References:

1. Abraham Silderschutz, H. Korth and S. Sudarshan: "Database systems concepts ",6th Edition, McGraw Hill Education
2. Peter Rob, Carlos Coronel: Database Systems: Design, Implementation, &Management.
3. Oracle installation and user manual
4. Ivan Bayross: SQL, PL/SQL The programming language of oracle, 4th edition, BPB publication
5. R. Elmasri, S. B. Navate: "Fundamentals of Database Systems", 6th Edition, Pearson.

M. Sc. Applied Statistics and Informatics (Part II) (Level-6.5) (Semester IV) (NEP-2020)

Title of Course: Biostatistics
Course Code: MSU0325MML905J1
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the students will be able to:

1. Explain the need, ethics, and various concepts in clinical trials.
2. Design and perform various phases of clinical trials.
3. Analyze Phase I-III bio-equivalence trials, case-control and cohort designs.
4. Explain the concept of censoring, various types of censoring, and perform inference about survival function.

Unit 1: Introduction to clinical trials: aim, need and ethics of clinical trials, conduct of clinical trials, preclinical research, phase I-IV trials, multi-center trials, bias and random error in clinical studies, randomization, blinding/masking in clinical trials, data management: data definitions, case report forms, database design, data collection systems for good clinical practice.

(12L+3T)

Unit 2: Design of clinical trials: parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, factorial designs, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials, design, and monitoring of Phase III trials with sequential stopping.

(12L+3T)

Unit 3: Bioavailability, pharmacokinetics, and pharmaco-dynamics. Design of bio-equivalence trials, Decision rules for bioequivalence, Inference for 2x2 crossover design: Classical methods of interval hypothesis testing for bioequivalence, Bayesian methods, nonparametric methods, assessment of Inter-and Intra-subject variability, drug interaction studies, reporting and analysis: analysis of categorical outcomes from Phase I - III trials. Epidemiological studies: case-control and cohort study designs. Measures of disease occurrence and association.

(12L+3T)

Unit 4: Introduction to survival analysis, examples of survival data/time to event data, measurement of survival time, concept of censoring, various types of censoring, type-I, type-II, random censoring, likelihood constructions and ML estimation. Non parametric estimation of survival function: Actuarial Estimator, Kaplan Meier product limit estimator, non parametric estimates of the mean, median and percentiles of the survival times; Nelson Aalen estimator, non parametric tests for two-sample problem: Gehen test, Log rank test, Mantel Haenszel test. Cox proportional hazard regression model.

(12L + 2T)

References:

1. Chow, S. C., & Liu, J. P. (2008). *Design and analysis of clinical trials: concepts and methodologies* (Vol. 507). John Wiley & Sons.
2. Clayton, D., & Hills, M. (2013). *Statistical models in epidemiology*. OUP Oxford.
3. Collett, D. (2023). *Modelling survival data in medical research*. CRC press.
4. Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: a foundation for analysis in the health sciences*. Wiley.
5. Deshpande, J. V., & Purohit, S. G. (2015). *Lifetime Data: Statistical Models and Methods* (Vol. 16). World Scientific Publishing Company.
6. Fleiss, J. L. (2011). *Design and analysis of clinical experiments*. John Wiley & Sons.
7. Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M., & Granger, C. B. (2015). *Fundamentals of clinical trials*. springer.
8. Hosmer Jr, D. W., Lemeshow, S., & May, S. (2011). *Applied survival analysis: regression modeling of time-to-event data*. John Wiley & Sons.

9. Jennison, C., & Turnbull, B. W. (1999). *Group sequential methods with applications to clinical trials*. CRC Press.
10. Marubeni .E. and Valsecchi M. G. (1994). *Analyzing Survival Data from Clinical Trials and Observational Studies*. Wiley.
11. Miller R. G. (1981). *Survival Analysis*, McFraw-Hill, New York.
12. Piantadosi, S. (2017). *Clinical trials: a methodologic perspective*. John Wiley & Sons.

Title of Course: Time Series Analysis
Course Code: MSU0325MML905J2
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the students will be able to:

1. Understand the time series; understand the concept of stationarity to the analysis of time series data in various contexts (such as actuarial studies, climatology, economics, finance, geography, meteorology, political science, and sociology).
2. Identify and isolate non deterministic components of observed time series; learn to translate an observed non-stationary series to stationary time series using an appropriate transformation.
3. Model, estimate, interpret and forecast observed time series through ARMA, ARIMA and SARIMA approach and Perform residual analysis for checking model adequacy.
4. Learn basics of time dependent volatility in time series, basics of ARCH and GARCH time series heteroscedastic models and basic of multivariate time series and its modeling

Unit 1: Exploratory time series analysis, Exponential, Double exponential and Holt-Winter smoothing and forecasting, auto-covariance, auto-correlation functions and their properties and characterization (without proof), partial auto covariance function, auto-covariance generating function. Weak and strong stationary time series, white noise process, Linear Process, estimates of mean, auto-covariance, auto-correlation and partial auto-covariance functions.

(12 L + 3 T)

Unit 2: Wold representation of linear stationary processes, linear time series models: autoregressive(AR), moving average(MA), autoregressive moving average models (ARMA). causality and invertibility of ARMA processes, computation of π -weights and ψ- weights, computation of ACVF, ACF and PACF for AR(1), AR(2), MA(1), MA(2), ARMA(1,1) processes and general procedure for ARMA(p,q) process. The need for differencing a time series, autoregressive integrated moving average models(ARIMA).

(12 L + 3 T)

Unit 3: Estimation of ARMA models: Yule-Walker estimation for AR Processes, Maximum likelihood and least squares estimation for ARMA Processes, Residual analysis and diagnostic checking. Minimum mean squared error forecasting for ARMA and ARIMA models, updating forecasts. Introduction to SARIMA models, Spectral Representation of the ACVF, Spectral density of an ARMA process, its computation for simple models.

(12 L + 3 T)

Unit 4: Introduction to ARCH and GARCH models. Properties and estimation under ARCH(1) and GARCH(1,1) model. Vector time-series models: Covariance and Correlation Matrix functions, MA and AR representation of vector processes, Covariance matrix function of the vector AR(1) and MA(1) models.

(12L + 3T)

Reference:

1. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
2. Brockwell, P. J., & Davis, R. A. (2009). *Time series: theory and methods*. Springer science & business media.
3. Chatfield, C. (2004) *The Analysis of Time Series - An Introduction*, Sixth edition, Chapman and Hall.
4. Kendall, M.G. (1978) *Time Series*, Charler Graffin
5. Tsay, R. S. (2005). *Analysis of financial time series*. John wiley & sons.
6. WEI, W. W. (2006). *Time Series Analysis: Univariate and Multivariate Methods*.

Title of Course: Advanced Topics in Artificial Intelligence
Course Code: MSU0325MEL905J1
Total Credits: 04
Course outcomes:
Upon successful completion of this course, the student will be able to:

1. Explain the concept of deep learning and apply deep learning algorithms for modelling of real world data using Python
2. Acquire the fundamental concepts of image processing and apply those for analyzing real world images using Python
3. Use NLP methods to explore and gain a broad understanding of text data using Python
4. Understand the concept of generative artificial intelligence and develop applications based on it.

Unit 1: Overview of neural Network concept, optimizers: Gradient Descent (GD) with momentum NAG, GD with adaptive learning rates: AdaGrad, RMSProp, Adam; introduction to deep learning, understanding different types of layers in sequential method: Dense, Convolutional Layers, Recurrent Layers, Normalization Layers. Deep Neural Network architecture design, optimization for Deep NN, regularization methods for deep NN, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Deep learning Python libraries: tensorflow (Low level) and Keras (High Level).

(12L+3T)

Unit 2: Introduction to Natural Language Processing (NLP), Natural Language Toolkit (NLTK) in Python. Data Preparation: punctuation removal, stop-words removal, numeric value removal, frequent words removal, rare words removal, spelling correction, tokenization, stemming, lemmatization. Feature Engineering:count vectors as features, Term Frequency-Inverse Document Frequency (TF-IDF), TF-IDF vectors as features, word level TF-IDF, N-Gram level TF-IDF, Character level TF-IDF, world cloud, Inverse Document Frequency, word embedding as features, Text/NLP based features, Topic Models as features, word2vec, sentiment analysis.

(12L+3T)

Unit 3: Computer vision: Introduction, OpenCv library in Python, Getting Started with images, Basic Operations on Images, Arithmetic Operations on Images,Image Pre-processing: changing colorspaces, geometric transformations, thresholding, smoothing, morphological transformations, gradients, Canny Edge detection, image pyramids, image segmentation with Watershed algorithm, Feature Detection and Description. Image Detection and recognition examples.

(12L+3T)

Unit 4: Introduction to Generative Artificial Intelligence (AI). Understanding generating AI and the difference from discriminative models. Fundamentals of Generative Models: Overview of Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Boltzmann machine, including working principle and applications. GANs: In-depth study, including generator and discriminator networks, data generation, training challenges, and applications. VAEs: Exploring concepts of encoder and decoder networks, latent space, training process, and real-world applications. Large Language Models (LLMs): Introduction to GPT and Bidirectional Encoder Representations from Transformers (BERT), architectural design, training process, transfer learning in NLP, advantages and limitations. Ethical Considerations: Ethical implications in generative AI – such as deepfakes, privacy concerns, and data bias.

(12L+3T)

References:

1. Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".
2. Goodfellow, I. (2016). Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*.
3. Goodfellow, I., Bengio, Y., Courville, A., &Bengio, Y. (2016). *Deep learning (Vol. 1).* Cambridge: MIT press.

4. Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., &Mathur, I. (2016). *Natural Language Processing: Python and NLTK.* Packt Publishing Ltd.
5. Mueller, J. P., &Massaron, L. (2015). *Python for data science for dummies.* John Wiley & Sons.
6. Shanmugamani R. (2018). *Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras.* "Packt Publishing Ltd"
7. Solem, J. E. (2012). *Programming Computer Vision with Python: Tools and algorithms for analyzing images.* " O'Reilly Media, Inc.".