Rayat Shikshan Sanstha's

Sadguru Gadage Maharaj College, Karad

(An Autonomous College)

Department of Statistics

Syllabus For

M.Sc. II Statistics

Semester Pattern

Sem III & IV

Choice Based Creadit System

(CBCS)

To be implemented From

June 2023 onwards

M.Sc. Part II Statistics Syllabus: w.e.f. 2023-24

A. Ordinance and Regulations: - (as applicable to degree/ programme)

B. Sadguru Gadage Maharaj College, Karad, New/Revised Syllabus for Master of Science

1. Title of the Program: **M. Sc. (Statistics)**

2. Faculty of Science

3. Year of Implementation: Part-I: 2022-2023 ; Part-II: 2023-2024

4. Preamble:

5. General Objectives of the Program:

- **Program Outcomes**

Post Graduates of the M.Sc. Statistics program will be able to:

a) Have a broad background in Statistics, an appreciation of how its various sub disciplines are inter-related, acquire an in-depth knowledge about topics chosen from those offered through the department,

b) Be familiar with a variety of real-life situations where statistics helps accurately explain the underlying abstract or physical phenomena and able to recognize and appreciate the connections between theory and applications;

c) Develop **the ability** to effectively and aptly use techniques from different sub- disciplines in a broad range of real-life problem solving.

d) Be statistically and numerically **literate**. i.e. graduates will: recognize the importance and value of statistical thinking, training

e) Have the **versatility** to work effectively in a broad range of companies (including R&D sectors of financial, pharmaceutical, market research, software development companies, consultancy etc) , or analytic, scientific, government, financial, health, teaching and other positions or continue for higher education.

f) Be able to independently read statistical literature including survey articles, scholarly books, and online sources;

g) Be life-long learners able to independently expand their statistical expertise when needed, or out of own interest.

h) Exhibit ethical and professional behavior in team work.

- **Program Specific Outcomes**

  After completion of M.Sc. Statistics program the student will be able to:

  a) Develop stochastic models for studying real life phenomenon in diverse disciplines.

  b) Efficiently interpret and translate the outcomes obtained from analysis of stochastic models to an environment understandable to a layman.

  c) Effectively use necessary statistical software and computing environment including R, MS-EXCEL among others

  d) Apply statistical techniques to optimize and monitor real life phenomena related to industry and business analytics etc.

6. Duration: 2 Years
7. Pattern: CBCS
8. Fee Structure:
9. Eligibility criteria for Admission: B. Sc. with Statistics as principal subject at degree level.
10. Medium of Instruction: English
11. Structure of the Program:

Semester I

	Course code	Title of the course
CGPA	NS22-101	Real Analysis
	NS22-102	Linear Algebra
	NS22- 103 a	Distribution Theory
	NS22- 103 b	Population Studies
	NS22-104	Estimation Theory
	NS22-105 a	Statistical Computing
	NS22-105 b	Optimization Technique
	NSP22_-106	Practical I
Mandatory Non-CGPA	Compulsory	AEC: Communicative English-I
	Compulsory	Diploma Course in Python Software

Semester II

	Course code	Title of course
CGPA	NS22-201	Probability Theory
	NS22-202	Theory of Testing of Hypothesis
	NS22-203	Linear Models and Regression Analysis
	NS22-204	Design and Analysis of Experiments
	NS22-205	Sampling Theory and Official Statistics
	NSP22-206	Practical II
Mandatory Non-CGPA	Compulsory	SEC: Fundamentals of Information Technology-I

Semester III

	Course code	Title of course
CGPA	NS22-301	Asymptotic Inference
	NS22-302 a	Survival Analysis
	NS22-302 b	Statistical Quality Control
	NS22-303	Multivariate Analysis
	NS22-304 a	Data Mining
	NS22-304 b	Artificial Intelligence
	NS22-305 a	Clinical Trials
	NS22-305 b	Biostatistics
	NSP22-306	Practical III
Mandatory Non-CGPA	Compulsory	AEC: Communicative English-II
	Compulsory	Advanced Diploma on Python Software
	Compulsory	EC:SWYAM /MOOCS

Semester IV

	Course code	Title of course
CGPA	NS22-401	Optimization Techniques
	NS22-402 a	Elementary Stochastic Process
	NS22-402 b	Functional Data Analysis
	NS22-403 a	Time Series Analysis
	NS22-403 b	Econometrics
	NS22-404 a	Planning and Analysis of Industrial Experiments
	NS22-404 b	Generalized Linear Model
	NS22-405 a	Actuarial Statistics
	NS22-405 b	Bayesian Inference
	NSP22-406	Practical IV
Mandatory Non-CGPA	Compulsory	SEC: Fundamental of Information Technology -II
		GE-Generic Elective

12. Scheme of Teaching and Examination:

M.Sc. (Statistics) Programme structure (CBCS PATTERN) (2022-23)M.Sc. Part – I

SEMESTER-I (Duration- Six Month)												
	Sr. No.	Course Code	Teaching Scheme			Examination Scheme					
			Theory and Practical			University Assessment (UA)			Internal Assessment (IA)		
			Lectures (Per week)	Hours (Per week)	Credit	Maximum Marks	Minimum Marks	Exam. Hours	Maximum Marks	Minimum Marks	Exam. Hours
CGPA	1	NS22-101	4	4	4	80	32	3	20	8	1
	2	NS22-102	4	4	4	80	32	3	20	8	1
	3	NS22-103	4	4	4	80	32	3	20	8	1
	4	NS22-104	4	4	4	80	32	3	20	8	1
	5	NS22-105	4	4	4	80	32	3	20	8	1
	6	NSP22-106	--	12	4	100	40	*	--	--	--
Total (A)			**--**	**--**	**24**	**500**	**--**	**--**	**100**	**--**	**--**
Non-CGPA	1	NAEC-I	2	2	2	--	--	--	50	20	2
	2	NDCP-I	**Diploma Course in Python Software**								
SEMESTER-II (Duration- Six Month)											
CGPA	1	NS22-201	4	4	4	80	32	3	20	8	1
	2	NS22-202	4	4	4	80	32	3	20	8	1
	3	NS22-203	4	4	4	80	32	3	20	8	1
	4	NS22-204	4	4	4	80	32	3	20	8	1
	5	NS22-205	4	4	4	80	32	3	20	8	1
	6	NSP22-206	--	12	4	100	40	*	--	--	--
Total (B)			**--**	**--**	**24**	**500**	**--**	**--**	**100**	**--**	**--**
Non-CGPA	1	NSEC-I	2	2	2	--	--	--	50	20	2
Total (A+B)					**48**	**1000**	**--**	**--**	**200**	**--**	**--**

M.Sc. (Statistics) Programme structure (CBCS PATTERN) (2023-24)M.Sc. Part – II

SEMESTER-III (Duration- Six Month)												
	Sr. No.	Course Code	Teaching Scheme			Examination Scheme					
			Theory and Practical			University Assessment (UA)			Internal Assessment (IA)		
			Lectures (Per week)	Hours (Per week)	Credit	Maximum Marks	Minimum Marks	Exam. Hours	Maximum Marks	Minimum Marks	Exam. Hours
CGPA	1	NS22-301	4	4	4	80	32	3	20	8	1
	2	NS22 -302	4	4	4	80	32	3	20	8	1
	3	NS22-303	4	4	4	80	32	3	20	8	1
	4	NS22-304	4	4	4	80	32	3	20	8	1
	5	NS22 -305	4	4	4	80	32	3	20	8	1
	6	NSP22-306	--	12	4	100	40	*	--	--	-
Total (C)			**--**	**--**	**24**	**500**	**--**	**--**	**100**	**--**	
Non-CGPA	1	NAEC-307	2	2	2	--	--	--	50	20	2
	2	NADP-I	**Advanced Diploma on Python Software**								
	3	NEC (SWM MOOC)-I	Number of lectures and credit shall be as specified on SWAYAM MOOC								
SEMESTER-IV (Duration- Six Month)											
CGPA	1	NS22-401	4	4	4	80	32	3	20	8	1
	2	NS22 -402	4	4	4	80	32	3	20	8	1
	3	NS22-403	4	4	4	80	32	3	20	8	1
	4	NS22-404	4	4	4	80	32	3	20	8	1
	5	NS22-405	4	4	4	80	32	3	20	8	1
	6	NSP22-406	--	12	4	100	40	*	--	--	--
Total (D)			**--**	**--**	**24**	**500**	**--**	**--**	**100**	**--**	**--**
Non-CGPA	1	NSEC-II	2	2	2	--	--	--	50	20	2
	2	NGE	2	2	2	--	--	--	50	20	2
Total (C+D)					**48**	**1000**	**--**	**--**	**200**	**--**	**--**

• Student contact hours per week :**32** Hours (Min.)	• Total Marks for M.Sc.-II :**1200**
• Theory and Practical Lectures :**60** Minutes Each	• Total Credits for M.Sc.-II (Semester III & IV) : **48**
• S22-Core Course • SP22-Core Course Practical • DSE-Discipline Specific Elective • NAEC-Mandatory Non-CGPA compulsory Ability Enhancement Course • NSEC- Mandatory Non-CGPA compulsory Skill Enhancement Course • NEC (SWM MOOC) - Non-CGPA Elective Course • NGE-Generic Elective	• Practical Examination is annual. • Examination for CCPR-306 shall be based on Semester III Practical. • Examination for CCPR-406 shall be based on Semester IV Practical. • *Duration of Practical Examination as per respective BOS guidelines • ***Separate passing is mandatory for Theory, Internal and Practical Examination***

	M.Sc.-I	**M.Sc.-II**	**Total**
Marks	**1200**	**1200**	**2400**
Credits	**48**	**48**	**96**

I. CGPA course:

1. There shall be 12 Core Courses (S22) of 48 credits per Programme.
2. There shall be 06 Core Course Specialization (CCS) of 24 credits per Programme.
3. There shall be 02 Discipline Specific Elective (DSE) courses of 08 credits per Programme.
4. There shall be 4 Core Course Practical (SP22) of 16 credits per Programme

II. Mandatory Non-CGPA Courses:

1. There shall be 02 Mandatory Non-CGPA compulsory Ability Enhancement Courses (NAEC) of 02 credits each per Programme.
2. There shall be 01 Mandatory Non-CGPA compulsory Skill Enhancement Course (NSEC) of 02 credits per Programme.
3. There shall be one Elective Course (NEC) (SWAYAM MOOC). The credits of this course shall be as specified on SWAYAM MOOC.
4. The total credits for Non-CGPA course shall be of 04 credits + 2-4 credits of EC as per availability.
5. The credits assigned to the courses and the Programme are to be earned by the students and shall not have any relevance with the work load of the teacher.

13. Standard of passing:
14. Nature of Question paper and Scheme of marking

(Unit wise weightage of marks should also be mentioned)

- **Nature of the theory question papers:-**
  a) There shall be 7 questions each carrying 16 marks.
  b) Question No.1 is compulsory. It consists of 8 questions for 2 marks each.
  c) Students have to attempt any 4 questions from question No. 2 to 7.
  d) Question No. 2 to 6 shall contain 2 to 4 sub-questions.
  e) Question No. 7 shall contain 4 short note type questions, each carrying 4 marks.

- **Practical Paper:-**

a) Semester I, II, III "Practical SP22-106, SP22-206 and SP22-306"

1. There shall be 20 marks for day-to-day performance and journal.
2. Examination (60): Practical Examinations of practical I and II will be conducted at the endthe of respective year and practical III and IV will be conducted at the end of the respective year.
3. Each exam will be of 3 hrs. duration carrying 60 marks. There shall be 8 questions each of12 marks, of which a student has to attempt any 5 questions.
4. Practical VIVA will be for 20 marks.

b) Semester IV, Practical SP22-406

1. There shall be 10 marks for day-to-day performance and journal.
2. Examination (30): Practical exam will be of 1.5 hrs. (90 Min.) duration carrying 30 marks.There shall be 5 questions each of 10 marks, of which a student has to attempt any 3 questions.
3. Practical VIVA will be for 20 marks.
4. Project work carries 40 marks. Project work consists of understanding the domain of the problem, formulation of the problem, collection of the relevant data, Analysis of the data and report writing. They are expected to use software for which they are trained. 20 marks are reserved for project based VIVA. Project report will be evaluated for 20 marks. The project work should be preferably based on field work or problem in industry

Syllabus and the nature of question paper for each of the courses of M.Sc. (Statistics) programme mentioned in the first column of the following table is exactly same as that ofthe course of M. Sc. (Applied Statistics and Informatics) programme mentioned against itin the second column. The question papers of such two matched courses will be commonin all examinations of M.Sc. (Statistics) and M. Sc. (Applied Statistics and Informatics) programs.

Semester I

	Course code	Title of the course
CGPA	NS22-101	Real Analysis
	NS22-102	Linear Algebra
	NS22- 103 a	Distribution Theory
	NS22- 103 b	Population Studies
	NS22-104	Estimation Theory
	NS22-105 a	Statistical Computing
	NS22-105 b	Optimization Technique
	NSP22_-106	Practical I
Mandatory Non-CGPA	Compulsory	AEC: Communicative English-I
	Compulsory	Diploma Course in Python Software

Semester II

	Course code	Title of course
CGPA	NS22-201	Probability Theory
	NS22-202	Theory of Testing of Hypothesis
	NS22-203	Linear Models and Regression Analysis
	NS22-204	Design and Analysis of Experiments
	NS22-205	Sampling Theory and Official Statistics
	NSP22-206	Practical II
Mandatory Non-CGPA	Compulsory	SEC: Fundamentals of Information Technology-I

Semester III

	Course code	Title of course
CGPA	NS22-301	Asymptotic Inference
	NS22-302 a	Survival Analysis
	NS22-302 b	Statistical Quality Control
	NS22-303	Multivariate Analysis
	NS22-304 a	Data Mining
	NS22-304 b	Artificial Intelligence
	NS22-305 a	Clinical Trials
	NS22-305 b	Biostatistics
	NSP22-306	Practical III
Mandatory Non-CGPA	Compulsory	AEC: Communicative English-II
	Compulsory	Advanced Diploma on Python Software
	Compulsory	EC:SWYAM /MOOCS

Semester IV

	Course code	Title of course
CGPA	NS22-401	Optimization Techniques
	NS22-402 a	Elementary Stochastic Process
	NS22-402 b	Functional Data Analysis
	NS22-403 a	Time Series Analysis
	NS22-403 b	Econometrics
	NS22-404 a	Planning and Analysis of Industrial Experiments
	NS22-404 b	Generalized Linear Model
	NS22-405 a	Actuarial Statistics
	NS22-405 b	Bayesian Inference
	NSP22-406	Practical IV
Mandatory Non-CGPA	Compulsory	SEC: Fundamental of Information Technology -II
		GE-Generic Elective

Equivalence of M. Sc. (Statistics) Programme

Semester I

	Course code	Title of the course	
		Old	**New**
CGPA	NS22-101	Real Analysis	Real Analysis
	NS22-102	Linear Algebra	Linear Algebra
	NS22-103	Distribution Theory	Distribution Theory
	NS22-104	Estimation Theory	Estimation Theory
	NS22T-105	Optimization Technique	Statistical Computing
	NSP22_-106	Practical I	Practical I
Mandatory Non-CGPA	Compulsory	AEC: Communicative English-I	AEC: Communicative English-I
	Compulsory	Diploma Course in Python Software	Diploma Course in Python Software

Semester II

	Course code	Title of course	
		Old	**New**
CGPA	NS22-201	Probability Theory	Probability Theory
	NS22-202	Theory of Testing of Hypothesis	Theory of Testing of Hypothesis
	NS22-203	Multivariate Analysis	Linear Model and Regression Analysis
	NS22-204	Design and Analysis of Experiments	Design and Analysis of Experiments
	NS22-205	Sampling Theory and Official Statistics	Sampling Theory and Official Statistics
	NSP22-206	Practical II	Practical II
Mandatory Non-CGPA	Compulsory	SEC: Fundamentals of Information Technology-I	SEC: Fundamentals of Information Technology-I

Semester III

	Course code	Title of course	
		Old	**New**
CGPA	NS22-301	Asymptotic Inference	Asymptotic Inference
	NS22-302	Survival Analysis	Survival Analysis
	NS22-303	Regression Analysis	Multivariate Analysis
	NS22-304	Data Mining	Data Mining
	NS22-305	Clinical Trials	Clinical Trials
	NSP22-306	Practical III	Practical III
Mandatory Non-CGPA	Compulsory	AEC: Communicative English-II	AEC: Communicative English-II
	Compulsory	Advanced Diploma on Python Software	Advanced Diploma on Python Software
	Compulsory	EC:SWYAM /MOOCS	EC:SWYAM /MOOCS

Semester IV

	Course code	Title of course	
		Old	**New**
CGPA	NS22-401	Optimization Techniques II	Optimization Techniques
	NS22-402	Elementary Stochastic Process	Elementary Stochastic Process
	NS22-403	Time Series Analysis	Time Series Analysis
	NS22-404	Planning and Analysis of Industrial Experiments	Planning and Analysis of Industrial Experiments
	NS22-405	Actuarial Statistics	Actuarial Statistics
	NSP22-406	Practical IV	Practical IV
Mandatory Non-CGPA	Compulsory	SEC: Fundamental of Information Technology -II	SEC: Fundamental of Information Technology -II
	Compulsory	GE-Generic Elective	GE-Generic Elective

Semester III

NS22-301: Asymptotic Inference

Unit 1: Consistency of an estimator, weak and strong consistency, joint and marginalconsistency, invariance property under continuous transformations, methods of constructingconsistent estimators, asymptotic relative efficiency. Consistent and Asymptotic Normal(CAN) Estimators: Definition of CAN estimator for real and vector valued parameters,invariance of CAN property under non-vanishing differentiable transformation. Methods ofconstructing CAN estimators: Method of Moments, method of percentiles, comparison of CAN estimators.

(12+ 3T)

Unit 2: CAN and BAN estimators in one parameter and multi-parameter exponential family of distributions, BAN estimators, super-efficient estimators, Crammer regularity conditions,Cramer – Huzurbazar results.

(12L+3T)

Unit 3: Variance stabilizing transformations; their existence; their applications in obtaininglarge sample tests and estimators. Asymptotic Confidence Intervals based on CAN estimatorsand based on VST, Asymptotic Confidence regions in multi-parameter families.

(12L+3T)

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

(12L +3T)

Reference:

1. Kale B.K. (1999): A first course on parametric inference, Narosa Pub
2. Zacks S. (1971): Theory of statistical inference, Wiley & Sons inc.
3. Rohatagi V.K. and Saleh A. K. Md. E.(2001) : Introduction to Probability Theory and Mathematical Statistics- John Wiley and sons Inc
4. Ferguson, T.S. (1996): A Course in Large Sample Theory. Chapman and Hall
5. Lehmann E L (1999): Elements of Large Sample Theory, Springer.
6. DasGupta A. (2008): Asymptotic Theory of Statistics and Probability, Springer Texts inStatistics.

NS22-302: SURVIVAL ANALYSIS

Unit 1: 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, progressive censoring and random censoring; likelihood function, estimation and testing ofparameters under above types of censoring.

(12L+3T)

Unit 2: Non parametric estimation of survival function: Actuarial Estimator, Kaplan Meier product limit estimator, properties: self-consistency and asymptotic normality, redistribution to the right algorithm. Nelson Aalen estimator, non parametric estimates of the mean, median and percentiles of the survival times; non parametric tests for two-sample problem: Gehen test, Log rank test, Mantel Haenszel test.

(12L + 2T)

Unit-3: The Cox regression model: A regression model for the comparison of two groups; The general proportional hazards model, Models corresponding to the linear component of the model: including a variate, a factor, an interaction, a mixed term. Fitting the Cox regression model in R, Likelihood function for the model, Treatment of ties, Confidence intervals and hypothesis tests for coefficients and for hazard ratios using R. Measures of explained variation, Measures of predictive ability, Model checking using various types of residuals: Cox-Snell; Modified Cox- Snell; Martingale; Deviance; Schoenfeld; Score residuals, plots based on these residuals and their interpretation.

(12L + 2T)

Unit-4: Competing risks: Summarizing competing risks data; Hazard and cumulative incidence functions; Cause-specific hazard function; Cause-specific cumulative incidence function; Likelihood functions for competing risks models; Parametric models for cumulative incidence functions.

(12L + 2T)

References:

1. Collet, D. (2015). Modeling Survival Data in Medical Research. London: Chapman andHall.
2. Hosmer, D. and Lemeshow S. (1999). Applied Survival Analysis: Regression Modeling ofTime to Event Data. New York: Wiley.
3. Breslow, N. and Day, N. (1987). Statistical Methods in Cancer Research, v. 2: The Designand
4. Analysis of Cohort Studies. Lyon: IARC.
5. Therneau T, and Grambsch, P. (2000). Modeling Survival Data: Extending the Cox Model.New
6. York: Springer.
7. Kalbfleish, JD. and Prentice, RL. (2002). The Statistical Analysis of Failure Time Data.New York: Wiley.

NS22-303: MULTIVARIATE ANALYSIS

Unit 1: Exploratory multivariate data analysis, sample mean vector, sample dispersion matrix, correlation matrix, graphical representation, means, variances, covariances, Partial and multiple correlation coefficients. Correlations of linear transforms. Multivariate normal distribution, two definitions and their equivalence, singular and nonsingular normal distribution, characteristic function, moments, marginal and conditional distributions.

(12L+3T)

Unit 2: Maximum likelihood estimators of the parameters of the multivariate normal distribution and their sampling distributions. Hotelling's T^2 Statistic and its null distribution. Applications of T^2 statistics and its relationship with Mahalanobis' D^2 statistic. Confidence region for the mean vector, Wishart matrix and its distribution, properties of Wishart distribution, distribution of generalized variance. (12L+3T)

Unit 3: 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 4: Canonical correlation analysis, Introduction to principal component analysis and relatedresults, Introduction to factor analysis and estimation.
(12L+3T)

References:

1. Kshirsagar A. M.(1972) : Multivariate Analysis. Marcel-Dekker.
2. Johnson, R.A. and Wichern . D.W (2002) : Applied multivariate Analysis. 5thEd.Prentice –Hall.
3. Anderson T. W. (1984) : An introduction to Multivariate statistical Analysis2nd Ed. JohnWiely.
4. Morrison D.F. (1976) : Multivariate Statistical Methods McGraw-Hill.

S22-304: DATA MINING

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 function, 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 RegressionTrees, 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 UniversityPress.
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. Penraminternational.
10. Hastie T, Tibshirani R, Friedmant J, (2009): The elements of statistical Learning, Springer.
11. Chattamvelli, R. (2015). Data mining methods. Alpha Science International.

NS22 -305 CLINICAL TRIALS

Unit-1: Introduction to clinical trials: the need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multi-center trials. Data management: data definitions, case report forms, database design, data collection systems for good clinical practice, Concept of Randomization and blinding. **(12L+3T)**

Unit-2: Design of clinical trials: parallel vs. cross-over designs, cross-sectional vs. Longitudinal designs, review of factorial designs, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials, Active control trials and combination trials, design and monitoring of Phase III trials with sequential stopping. **(12L+3T)**

Unit -3: Design of bioequivalence trials, Classical methods of interval hypothesis, testing for bioequivalence Baysian methods, non-parametric methods, Assesment of inter and intra subject variability, drug interaction study, Dose proportionality steady state analysis, Clinical end points, alpha spending function. **(12L+3T)**

Unit-4: Analysis and Reporting of clinical trials: Concept of sample size and it's calculation, Analysis of categorical outcomes from Phase I - III trials, analysis of survival data from clinical trials**.** **(12L+3T)**

References:

1. S. Piantadosi (1997). Clinical Trials : A Methodologic Perspective, Wiley and Sons.
2. C. Jennison and B. W. Turnbull (1999): Group Sequential Methods with Applications to Clinical Trials, CRC Press.
3. L. M. Friedman, C. Furburg, D. L. Demets (1998). Fundamentals of Clinical Trials,SpringerVerlag.
4. J. L. Fleiss (1989). The Design and Analysis of Clinical Experiments.Wiley and Sons.
5. S.D.Wang & A. Bakhai (2006): Clinical trials : Analysis and Reporting.
6. Todd A Durham & J. Rick Turner :Introduction to Statistics in Pharmaceutical Clinical Trials

Practical III

1. Construction of Consistent/CAN Estimators.
2. Construction of BAN Estimators and confidence interval based on it.
3. Confidence interval based on CAN.
4. Confidence interval based on VST.
5. Parametric analysis of survival data.
6. Estimation of Survival Function.
7. Estimation of parameters in PH model.
8. Analysis of two sample non-parametric problem.
9. Exploratory data Analysis
10. Application of Hoteling's T^2 statistic
11. Cluster Analysis
12. Principal Component Analysis and Factor Analysis
13. Classification.
14. Cluster Analysis.
15. Artificial Neural Network.
16. Support Vector Machine.
17. Pharmokinetics.
18. Confidence interval.
19. Analysis of categorical outcomes.
20. Non-parametric test based on Clinical Trials.

Semester IV

NS22-401: OPTIMIZATION TECHNIQUES

Unit 1: a) Linear programming problem (LPP): Theorems related to the development of Simplex algorithm, theorems related to a basic feasible solution ; Reduction of a feasible solution to a basic feasible solution, Improvement of a basic feasible solution, Existence of unbounded solution, Optimality conditions and other related theorems (statements only), Examples based on these theorems. Revised Simplex Method. b) Artificial variable technique: Two phasemethod, redundancy.

(12L+3T)

Unit 2: a) Concept of Duality, related theorems, complementary slackness property and development of dual simplex algorithm. b) Sensitivity Analysis: Changes in the cost vector, requirement vector and non-basic activity vector; addition of new variables and addition of new constraints.

(12L+3T)

Unit 3: a) Theory of games: two person zero sum games, minimax and maximin principles, Saddle point, mixed strategies; rules of dominance, solution of 2 x 2 game by algebraic method, Graphical method, Reduction of the game problem as LPP, Minimax and Maximin theorem. b) Dynamic Programming: The Recursion Equation Approach, Computational Procedure, Characteristics of Dynamic Programming, Solution of L.P.P. by Dynamic Programming.

(12L+3T)

Unit 4: a) Integer Linear Programming Problem (ILPP): The concept of cutting plane, cuttingplane method for all ILPP and mixed ILLP, Branch and Bound method. b) Quadratic programming: Kuhn-Tucker conditions, methods due toBeale,Wolfe.

(12L+3T)

References:

1. G Hadley, Linear Programming, Addison Wesley, 1969.
2. Taha H. A., Operation Research An Introduction, Macmillan, 1971.
3. Kanti Swaroop & M. M. Gupta, Operations Research, Sultan Chand & P. Gupta, 1985.
4. D. S. Hira, Operation Research, Sultan Chand & Co.ltd,2010
5. J. K. Sharma., Operation Research Theory and Applications, Macmillan, 2003

NS22-402: Elementary Stochastic Processes

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 Poissonprocess and discrete state space Markov processes (12L +3T)

Unit 4: Galton-Watson BinaymiBranching 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 StochasticProcesses. (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 (AcademicPress)

NS22-403: Time Series Analysis

Unit1: (a) Exploratory time series analysis, Exponential, Double exponential and Holt – Winter smoothing and forecasting. (b) Auto-covariance, auto-correlation functions, their properties and characterization (without proof), Partial auto-covariance function, auto–covariance generating function and its applications. First and second order Stationary time series, white noise process, Linear Process, Sample Estimates of mean, auto-covariance, auto-correlation and Partial auto-covariance functions, their asymptotic distributions and confidence intervals (without proof) (12 L + 3 T)

Unit2: Wold representation of linear stationary processes, linear time series models: Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) models. 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 a general computational procedure for ARMA(p,q) process. The need for differencing a time series, The unit root problem, Autoregressive Integrated Moving Average models. (12 L + 3 T)

Unit3: (a) 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. (b) Spectral Representation of the ACVF, Spectral density of an ARMA process, its computation for white noise, AR (1), MA (1), ARMA (1,1) and the ARMA (p, q) models. (12 L + 3 T)

Unit4 (a) Introduction to ARCH and GARCH models. Properties and estimation under ARCH(1) and GARCH (1,1) model. (b) 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)

References:

1. W. S. Wei (2005) Time Series Analysis: Univariate and Multivariate Methods
2. Box, G.E.P and Jenkins G.M. (1970) Time Series Analysis, Forecasting & Control, Holden-Day.
3. Brockwell, P.J and Davis R.A. (1987) Time Series: Theory and Methods, Springer-
4. TsayR. S. Analysis of Financial Time Series, 3rd Ed. (Wil. Ser. in Prob. and Statistics)
5. Kendall, M.G. (1978) Time Series, Charler Graffin
6. Chatfield, C. (2004) The Analysis of Time Series - An Introduction, Sixth edition, Chapmanand Hall.

NS22-404: PLANNING AND ANALYSIS OF INDUSTRIAL EXPERIMENTS

Unit-1: 2^k factorial Experiments: Concepts of main graphical representation, Analysis of full 2k replicated and unreplicated Concept of Confounding: Total and partial Confounding, construction design. effects, interaction, their factorial designs. and analysis confounded

(12 L + 3 T)

Unit-2: 3^k factorial Experiments: Concepts of main effects, interaction, their graphical representation, linear and quadratic components, Analysis of full 3^k replicated and unreplicated factorial designs. Confounding: construction and analysis confounded design, Factorials with mixed levels. (12 L + 3 T)

Unit-3: Fractional Factorial: Fractional replication for symmetric factorials, concept of generator, defining contrasts, aliasing, resolution and minimum aberration, construction and analysis of 2^{k-p} and 3^{k-p} fractional designs, Central composite designs. (12 L + 3 T)

Unit-4: Response surface experiments : linear and quadratic model, test for curvature, stationary point, central ridge systems, Rotatability, Multiple responses. Taguchi methods: Concept of noise and control factors, inner and outer arrays, concept of loss function, S/N ratio, orthogonal arrays, linear graphs, interaction tables, ANOVA. (12 L + 3 T)

References:

1. Montgomery D.C. (2013): Design and Analysis of Experiments, 8^{th} edition, Wiley India PvtLtd.

2. Davies.O.L(1954):The design and analysis of industrial experiment,Oliver and Boyd.

3. Voss, D., Dean, A., and Dean, A.(1999).Design and Analysis of Experiments, Springer verlag Gmbh.

4. Wu, C. F., Hamada M. S.(2000). Experiments: Planning, Analysis and Parameter Design Optimization, 2nd edition, John Wiley & Sons.

5. CochranW.G.andcox,G.M.(1959): Experimental Design.

NS22-405: ACTUARIAL STATISTICS

Unit 1: Introduction to Insurance Business, Concept of risk, types of risk, characteristics of insurable risk ,Risk models for Insurance: Individual and aggregate Risk models for short term, Distribution of aggregate claims, compound Poisson distribution and its applications. Survival function and Life tables: Survival function, Distribution function, Density functions and Force of mortality. Time-until death random variable and Curtate-future lifetime random variable.

(12L+3T)

Unit 2: Life tables, Select and ultimate life tables. Assumptions for fractional ages and some analytical laws of mortality.Life Insurance: Principles of compound interest: Nominal and effective rates of interest and force of interest and discount, compound interest, Insurance payable at the moment of death and at the end ofthe year of death ,Whole life insurance, endowment insurance, term insurance, deferred insurance and varying benefit insurance..

(12L+3T)

Unit 3: Annuities: annuity certain, discrete annuity, monthly annuity, continuous annuity, deferred annuity, present values and accumulated values of these annuities, Continuous life annuity, discrete life annuity, such as whole life annuity, temporary life annuity, n-year certain and life annuity, life annuities with mthly payments, Present value random variables for these annuity payments, their means and variances, Actuarial present value of the annuity. (12L+3T)

Unit 4: Loss at issue random variable, various principles to decide net premiums for insurance products and annuity schemes defined in unit II and III, fully continuous premiums and fully discrete premiums, True monthly payment premiums. Extended equivalence principle to decide gross premiums, Concept of reserve, Fully continuous reserve, Fully discrete reserve. (12L+3T)

References:

1.Deshmukh S. R.,An Introduction to Actuarial Statistics, University Press, 2009

2.Robin Cunningham, Thomas N. Herzog, Richard L. Models for Quantifying Risk, 4th Edition,ACTEX Publications, 2011.

3.Dickson, David C. M., Hardy, Mary R. and Waters, Howard R., Actuarial Mathematics for life contingent risks, International series on actuarial science, Cambridge 2009.

4.Narang, Uma, Insurance Industry in India: Features, Reforms and Outlook, New Century Publications

NSP22- 406 PRACTICAL-IV

1. Revised Simplex method and Dual simplex method
2. Game Theory
3. Quadratic Programming
4. Integer and Dynamic Programming
5. Realization of stochastic process.
6. Classification of t.p.m. and computation of n- step probability matrix.
7. Classification of states: Computations of absorption probabilities.
8. Stationary distribution and recurrence time.
9. Autocovariance and Autocorrelation.
10. Causal and Invertible
11. Smoothing the series
12. Forecasting.
13. Analysis of full replicated unconfounded 2^n and 3^n factorial experiments.
14. Analysis of single replicated 2^n and 3^n factorial experiments.
15. Analysis of confounded 2^n and 3^nfactorial experiments: total and partial confounding.
16. Analysis of response surface 1st and 2nd order experiments.
17. Construction of Life Tables.
18. Computations of benefit premiums for n-year term insurance, whole life insurance, endowment insurance.
19. Computation of Annuities.
20. Computation of Reserve.

MANDATORY NON CGPA COURSE (To be offered by the Statistics Department) COMPULSORY (GE): DATA MANAGEMENT AND ANALYSIS USING MSEXCEL

1. Introduction: Introduction to MSEXCEL, data input, cell formatting, entering survey data, entering data generated through scientific experiment, data editing, sorting, filtering, find andreplace, conditional formatting, text to columns. Preparing data available at secondary sources for further analysis. Formulae: commonly used Mathematical functions, Statistical functions, Text functions, Lookup functions, Reference functions, Error functions, Logical Function, Array and Summarizing functions, Database Functions, Date and Time Functions. (12 L + 3 T)

2. Working with data: Graphical representation of the data, Data validation, data consolidation, what-if analysis, Pivot tables and charts, advanced filter, subtotals and outlines, securing sheets/workbook. Data Analysis using 'Analysis tool pack'. Introduction to MACROs.

(12 L + 3 T)

References:

1. Bissett B. D. (2007). Automated Data Analysis Using Excel. CRC Press.
2. Harvey G.(2011). Excel 2007 For Dummies. John Wiley & Sons.
3. Held B (2010). Microsoft Excel Functions & Formulas. Word ware Publishing, Inc.
4. Liengme B.(2008). A Guide to Microsoft Excel 2007 for Scientists and Engineers.Academic Press.