SHIVAJI UNIVERSITY, KOLHAPUR

SYLLABUS

For

M.Sc. Applied Statistics and Informatics

(Semester Pattern)

Sem. III to IV

Estd. 1962

NAAC 'A' Grade

Choice Based Credit System

(CBCS)

To be implemented From

June, 2020 onwards

M.Sc. Applied Statistics and Informatics Part II Syllabus: w.e.f. June 2020

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

B. Shivaji University, Kolhapur, New/Revised Syllabus for Master of Science

1. Title of the Program: **M. Sc. (Applied Statistics and Informatics)**
2. Faculty of Science
3. Year of Implementation: **M.Sc. part-I: June 2019; M.Sc. part-II June-2020**
4. Preamble:
5. General Objectives of the Program:

- **Program Outcomes**

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

  a) Have a **broad background** in applied Statistics and information technology(IT), 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 and IT helps accurately explain the underlying abstract or physical phenomena and able to recognize and appreciate the connections between theory and applications;

  c) Be computationally, statistically and numerically **literate**. i.e. graduates will: recognize the importance and value of statistical thinking, training and using computers in analysis large data generated through various real life systems.

  d) Develop **the ability** to effectively and aptly use techniques from different sub- disciplines in a broad range of real life problem solving; develop appropriate computer programs (in C, C++, Python etc.) for analysis complex data.

  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 recent statistical and IT related literature including survey articles, scholarly books, and online sources;

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

h) Exhibit ethical and professional behaviour in team work.

- **Program Specific Outcomes**

  After completion of M.Sc. Applied Statistics and Informatics 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 the Database Management System tools for handling large data systems.
  d) Effectively use necessary statistical software and computing environment including R, MS-EXCEL, C, C++, Python among others and develop required computer programs in the same
  e) 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:

Structure of M. Sc. (ASI) Programme

Semester I

	Course code	Title of course
CGPA	CC-101	Fundamentals of Computer Programming
	CC-102	Linear Algebra
	CC-103	Distribution Theory
	CC-104	Estimation Theory
	CC-105	Statistical Computing
	CCPR-106	Practical - I
Mandatory Non- CGPA	Compulsory	AEC: Communicative English-I

Semester II

	Course code	Title of course
CGPA	CC-201	Advanced Data Structure with C++
	CC-202	Theory of Testing of Hypothesis
	CC-203	Regression Analysis
	CC-204	Design and Analysis of Experiments
	CC-205	Sampling Theory and Official Statistics
	CCPR-206	Practical - II
Mandatory Non- CGPA	Compulsory	SEC: Fundamentals of Information Technology-I

Semester III

	Course code	Title of course
CGPA	CC-301	Data Base Management System
	CCS-302	Multivariate Analysis
		Bayesian Inference
	CCS-303	Stochastic Processes
		Functional Data Analysis
	CCS-304	Data Mining
		Artificial Intelligence
	DSE-305	Time Series Analysis
		Statistical Ecology
	CCPR-306	Practical - III
Mandatory Non- CGPA	Compulsory	AEC: Communicative English-II
	Compulsory	EC: SWYAM/ MOOCS

Semester IV

	Course code	Title of course
CGPA	CC-401	Generalized Linear Models
	CCS-402	Survival Analysis
		Actuarial Statistics
	CCS-403	Biostatistics
		Econometrics
	CCS-404	Python for Data Science
	DSE-405	Spatial Data Analysis
		Statistical Quality Control
	CCPR-406	Practical IV and Project
Mandatory Non- CGPA	Compulsory	SEC: Fundamentals of Information Technology-II
	Compulsory	GE: Data Management and Analysis using MSEXCEL

Scheme of Teaching and Examination:

M.Sc. (Applied Statistics and Informatics) Programme structure (CBCS PATTERN) (2019-20)M.Sc. Part – I

	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
SEMESTER-I (Duration- Six Month)											
CGPA	1	CC-101	4	4	4	80	32	3	20	8	1
	2	CC-102	4	4	4	80	32	3	20	8	1
	3	CC-103	4	4	4	80	32	3	20	8	1
	4	CC-104	4	4	4	80	32	3	20	8	1
	5	CC-105	4	4	4	80	32	3	20	8	1
	6	CCPR-106	-	12	4	100	40	*	--	--	--
Total (A)			--	--	**24**	**500**	--	--	**100**	--	--
Non-CGPA	1	AEC-107	2	2	2	--	--	--	50	20	2
SEMESTER-II (Duration- Six Month)											
CGPA	1	CC-201	4	4	4	80	32	3	20	8	1
	2	CC-202	4	4	4	80	32	3	20	8	1
	3	CC-203	4	4	4	80	32	3	20	8	1
	4	CC-204	4	4	4	80	32	3	20	8	1
	5	CC-205	4	4	4	80	32	3	20	8	1
	6	CCPR-206	--	12	4	100	40	*	--	--	--
Total (B)			--	--	**24**	**500**	--	--	**100**	--	--
Non-CGPA	1	SEC-207	2	2	2	--	--	--	50	20	2
Total (A+B)					**48**	**1000**	--	--	**200**	--	--

• Student contact hours per week :**32** Hours (Min.)	• Total Marks for M.Sc.-I :**1200**
• Theory and Practical Lectures :**60** Minutes Each	• Total Credits for M.Sc.-I (Semester I & II) : **48**
• CC-Core Course • CCPR-Core Course Practical • AEC-Mandatory Non-CGPA compulsory Ability Enhancement Course • SEC- Mandatory Non-CGPA compulsory Skill Enhancement Course	• Practical Examination is annual. • Examination for CCPR-106 shall be based on Semester I Practical. • Examination for CCPR-206 shall be based on Semester II Practical. • *Duration of Practical Examination as per respective BOS guidelines • ***Separate passing is mandatory for Theory, Internal and Practical Examination***

M.Sc. (Applied Statistics and Informatics) Programme structure (CBCS PATTERN) (2020-21) 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	CC-301	4	4	4	80	32	3	20	8	1
	2	CCS -302	4	4	4	80	32	3	20	8	1
	3	CCS-303	4	4	4	80	32	3	20	8	1
	4	CCS-304	4	4	4	80	32	3	20	8	1
	5	DSE -305	4	4	4	80	32	3	20	8	1
	6	CCPR-306	--	12	4	100	40	*	--	--	-
Total (C)			**--**	**--**	**24**	**500**	**--**	**--**	**100**	**--**	
Non-CGPA	1	AEC-307	2	2	2	--	--	--	50	20	2
	2	EC (SWM MOOC)-308	Number of lectures and credit shall be as specified on SWAYAM MOOC								
SEMESTER-IV (Duration- Six Month)											
CGPA	1	CC-401	4	4	4	80	32	3	20	8	1
	2	CCS -402	4	4	4	80	32	3	20	8	1
	3	CCS-403	4	4	4	80	32	3	20	8	1
	4	CCS-404	4	4	4	80	32	3	20	8	1
	5	DSE -405	4	4	4	80	32	3	20	8	1
	6	CCPR-406	--	12	4	100	40	*	--	--	--
Total (D)			**--**	**--**	**24**	**500**	**--**	**--**	**100**	**--**	**--**
Non-CGPA	1	SEC-407	2	2	2	--	--	--	50	20	2
	2	GE-408	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**
• CC-Core Course • CCS-Core Course Specialization • CCPR-Core Course Practical • DSE-Discipline Specific Elective • AEC-Mandatory Non-CGPA compulsory Ability Enhancement Course • SEC- Mandatory Non-CGPA compulsory Skill Enhancement Course • EC (SWM MOOC) - Non-CGPA Elective Course • GE-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 (CC) 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 (CCPR) of 16 credits per programme
5. Total credits for CGPA courses shall be of 96 credits per programme

II. Mandatory Non-CGPA Courses:

1. There shall be 02 Mandatory Non-CGPA compulsory Ability Enhancement Courses (AEC) of 02 credits each per programme.
2. There shall be 01 Mandatory Non-CGPA compulsory Skill Enhancement Course (SEC) of 02 credits per programme.
3. There shall be one Elective Course (EC) (SWAYAM MOOC). The credits of this course shall be as specified on SWAYAM MOOC.
4. There shall be one Generic Elective (GE) course of 02 credits per programme. Each student has to take generic elective from the department other than parent department.
5. The total credits for Non-CGPA course shall be of 08 credits + 2-4 credits of EC as per availability.
6. The credits assigned to the courses and the programmes are to be earned by the students and shall not have any relevance with the work load of the teacher.

12. Standard of passing:

13. 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:CCPR-106, CCPR-206 and CCPR-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 end the 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 of 12 marks, of which a student has to attempt any 5 questions.
  4. Practical VIVA will be for 20 marks.

  b) **Semester IV, Practical CCPR-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, Statistical analysis of the data, Conclusions and report writing. They are expected to use software in 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 of the course of M. Sc. (Applied Statistics and Informatics) programme mentioned against it in the second column. The question papers of such two matched courses will be common in all examinations of M.Sc. (Statistics) and M. Sc. (Applied Statistics and Informatics) programs.

Semester I

Course code	Title of course
CC-102	Linear Algebra
CC-103	Distribution Theory
CC-104	Estimation Theory
CC-105	Statistical Computing

Semester II

Course code	Title of course
CC-202	Theory of Testing of Hypothesis
CC-203	Regression Analysis
CC-204	Design and Analysis of Experiments
CC-205	Sampling Theory and Official Statistics

Semester III

Course code	Title of course
CCS-302	Multivariate Analysis
	Bayesian Inference
CCS-303	Stochastic Processes
	Functional Data Analysis
CCS-304	Data Mining
	Artificial Intelligence
DSE-305	Time Series Analysis
	Statistical Ecology

Semester IV

Course code	Title of course
CCS-401	Generalized Linear Models
CCS-402	Survival Analysis
	Actuarial Statistics
CCS-403	Biostatistics
	Econometrics
DCE-405	Spatial Data Analysis
	Statistical Quality Control

14. Equivalence in Accordance with Titles and contents of papers

Table showing correspondence of old and new courses

Course Code		Title of the course	
old	new	old	new
Semester-I			
MASI -101	CC-101	Fundamentals of Computer Programming	Fundamentals of Computer Programming
MASI-102	CC-102	Statistical Mathematics	Linear Algebra
MASI-103	CC-103	Distribution Theory	Distribution Theory
MASI-104	CC-104	Estimation Theory	Estimation Theory
MASI-105	CC-105	Statistical Computing	Statistical Computing
MASI -146	AEC:	Communication Skills	Communicative English-I
MASI -117	CCPR-106	Practical-I	Practical-I
Semester-II			
Course Code		Title of the course	
old	new	old	new
MASI -201	CC-201	Advanced Data Structure with C++	Advanced Data Structure with C++
MASI-202	CC-202	Theory of Testing of Hypotheses	Theory of Testing of Hypothesis
MASI-203	**CCS-302**	Multivariate Analysis	**Multivariate Analysis (semester-III course)**
MASI-204	CC-204	Linear Models and Design of Experiments	Design and Analysis of Experiments
MASI-205	CC-205	Sampling Theory	Sampling Theory and Official Statistics
MASI -216	CCPR-206	Practical-II	Practical - II
Semester-III			
Course Code		Title of the course	
old	new	old	New/substitute
MASI -301	CC-301	Data Base Management System	Data Base Management System
*MASI-302	CC-302	Elementary Stochastic Processes	Stochastic Processes
*MASI-303	--------	Planning and Analysis of Industrial Experiments	Planning and Analysis of Industrial Experiments(MASI-303)
*MASI-304	--------	Design and Analysis of Sample Surveys	Design and Analysis of Sample Surveys (MASI-304)
MASI -316	CCPR-306	Practical-III	Practical - III
MASI-321	-------	Reliability Theory	Reliability Theory (MASI-321)
MASI-322	**CC-203**	Regression Analysis	Regression Analysis (semester-II course)
Semester-IV			
Course Code		Title of the course	
old	new	Old	New/substitute
MASI -401	CCS-404(Statistics)	Optimization Techniques	Optimization Techniques (For Statistics)
MASI-402	CCS-404	Python for Data Science	Python for Data Science
MASI-416	CCPR-406	Practical and Project work	Practical and Project
MASI-421	CC-401	Generalized Linear models	Generalized Linear Models
MASI-422	CCS-402	Survival Analysis	Survival Analysis

MASI-423	DSE-405	Statistical Quality Control	Statistical Quality Control
MASI-424	**DSE-305**	Time Series Analysis	Time Series Analysis (sem-III course)
MASI-427	**CCS-304**	Data Mining	Data Mining (sem-III course)

Semester III

CC-301: DATA BASE MANAGEMENT SYSTEM

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.

(12L+3T)

Unit 2: Database design and ER Model: Overview, ER-Model, Constraints, ER-Diagrams, ERD issues, weak entity sets, 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,atomicdomainandnormalization(1NF,2NF,3NF,BCNF).

(12L+3T)

Unit 3: 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. 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.

CCS-302: 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 related results, 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. 5thAd.Prentice – Hall.
3. Anderson T. W. (1984) : An introduction to Multivariate statistical Analysis2nd Ed. John Wiely.
4. Morrison D.F. (1976) : Multivariate Statistical Methods McGraw-Hill.

CCS-302: BAYESIAN INFERENCE

Unit 1: Basic elements of Statistical Decision Problem. Expected loss, decision rules (non-randomized and randomized). Overview of Classical and Bayesian Estimation. Advantage of Bayesian inference, Prior distribution, Posterior distribution, Subjective probability and its uses for determination of prior distribution. Importance of non-informative priors, improper priors, invariant priors. Conjugate priors, construction of conjugate families using sufficient statistics, hierarchical priors. Admissible and minimax rules and Bayes rules.

(12L + 3T)

Unit 2: Point estimation, Concept of Loss functions, Bayes estimation under symmetric loss functions, Bayes credible intervals, highest posterior density intervals, testing of hypotheses. Comparison with classical procedures. Predictive inference. One- and two-sample predictive problems.

(12L + 3 T)

Unit 3: Bayesian analysis with subjective prior, robustness and sensitivity, classes of priors, conjugate class, different methods of construction of objective priors: Jeffrey's prior, probability matching prior, conjugate priors and mixtures.Posterior robustness: measures and techniques. Bayes factors large sample methods: Limit of posterior distribution, consistency of posterior distribution, asymptotic normality of posterior distribution.

(12L + 3 T)

Unit 4: Bayesian Computations: Analytic approximation, E-M Algorithm, Monte Carlo sampling, Markov Chain Monte Carlo Methods, Metropolis-Hastings Algorithm, Gibbs sampling, examples, convergence issues.

(12L + 3T)

References:

1. Bolstad, W. M. (2007). Introduction to Bayesian Statistics, 2nd Edn. Wiley,
2. Christensen R, Johnson, W., Branscum, A. and Hanson T. E. (2011). Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians, Chapman & Hall.
3. Congdon, P. (2006). Bayesian Statistical Modeling, Wiley
4. Ghosh, J. K., Delampady M. and T. Samantha (2006). An Introduction to Bayesian Analysis: Theory & Methods, Springer.
5. Jim, A. (2009). Bayesian Computation with R, 2nd Edn, Springer.
6. Rao. C.R. and Day. D. (2006). Bayesian Thinking, Modeling & Computation, Handbook of Statistics, Vol. 25. Elsevier

CCS-303: 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 Poisson process 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 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)

CCS-303: FUNCTIONAL DATA ANALYSIS

Unit1: Introduction to functional data, Summary statistics for functional data. Functional means and variances; Covariance and correlation functions; Cross-covariance and cross-correlation functions; Representing functions by basis functions; illustration with real data with R.

(12L + 2T)

Unit2:Smoothing functional data by least squares, illustration with real data with R. various data depths for Functional data, Visualization of functional data using bag plots, rainbow plots and box-plots. Outlier detection. illustration with real data with R.

(12L + 2T)

Unit3: Test for Equality of two mean functions, Test for Equality of two covariance functions, One way functional ANOVA; computation (without any proofs) and application to real data using R.

(12L + 2T)

Unit4: Principal components analysis for functional data, Canonical correlation analysis for functional data; concept, computation (without any proofs) and illustration with real data using R.

(12L + 2T)

References:

1. Ramsay, James O., and Bernard W. Silverman. Functional data analysis,. Springer, 2005.
2. Ramsay, James O., and Bernard W. Silverman. Applied functional data analysis: methods and case studies. Springer, 2007.
3. Horváth, Lajos, and Piotr Kokoszka. Inference for functional data with applications. Springer Science & Business Media, 2012.

CCS-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 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.

DSE-304: ARTIFICIAL INTELLIGENCE

Unit 1: AI Problem Solving: Historical development of AI, Role of heuristics in problem solving. Knowledge representation and inference. Comparison of database knowledgebase. Expert Systems: Expert Problems. Predicate logic. Fact- Table. Rulebase. Fuzzy Logic. Case based reasoning. Design of fuzzy rule base. Construction and implementation of knowledgebase systems.

(12L + 3T)

Unit 2: Artificial Neural Network: Signal processing in biological and artificial neurons. ANN architectures. Perceptron learning. Multilayer Perceptron: Back Propagation Algorithm, XOR Problem, Heuristics, Output Representation and Decision Rule, Training and implementation of a neural network.

(12L + 3T)

Unit 3: Genetic Algorithm: History and evolution of G.A. Modeling a problem for the application of G.A. Representation of data in chromosomes. GA operators: Encoding, Crossover, Selection, Mutation, etc. Fitness function. Reproduction and convergence. Comparison of ANN and GA. Application of G.A.

(12L + 3T)

Unit 4: Natural Language Processing: Text categorization. Text summarization and text elaboration. Vision and perception. Image analysis and pattern matching. Robotics.

(12L + 3T)

Reference:

1. S. Rajsekaran,G.A. Vijaylaxmi Pai: Neural Networks. Fuzzy Logic and Generic Algorithms. Synthesis and Applications. (EEE)
2. David Goldberg: Genetic Algorithms (Addison and Wesley)
3. David Rolston: Principles of AI and Expert System Development (MGII)
4. E. Ritch and K.Knight: Artificial Intelligence (MGII)
5. Mehrika, K., Mohan, C., and Ranka (1997) Elements of Artificial neural networks. Penram International.
6. Chattamvelli, R. (2015). Data mining methods. Alpha Science International.

DSE-305: 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, Chapman and Hall.

CCPR 306: PRACTICAL III

1. SQL programming 1
2. SQL programming 2
3. SQL programming 3
4. Application of Relational algebra and relational calculus

5- onwards: At least FOUR practical each on the optional courses.

Semester IV

CC-401: GENERALIZED LINEAR MODELS

Unit 1: Generalized linear models: concept of generalized linear model, Link function, ML estimation, Quasi-likelihood 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.

(12L+3T)

Unit 2: Logistic regression: logit, probit and cloglog model for dichotomous data with single and multiple explanatory variables, ML estimation, large sample tests about parameters. Hosmer-Lemeshow test, ROC curve. Multilevel logistic regression, Logistic regression for Nominal response: Baseline Category model and ordinal response: Proportional odds model.

(12L+3T)

Unit 3: Poisson regression: ML and Quasi-likelihood estimation of parameters, testing significance of coefficients, goodness of fit, power family of link functions, over dispersion: Types, causes and remedies. Negative Binomial regression: NB-2 model.

(12L+3T)

Unit 4: Generalized linear mixed models (GLMM): Structure of the model, consequences of having random effects, estimation by maximum likelihood, marginal versus conditional models, estimation by generalized estimating equations and conditional likelihood, tests of hypothesis: LRT, asymptotic variance, Wald and score test.

(12L+3T)

References:

1. Hosmer D.W. and Lemeshow S. (2000): Applied Logistic regression, 2nd ED. Wiley New York.
2. Agresti A. (1990) : Categorical Data Analysis. Wiley , New York.
3. R. Christensen (1997) Log-Linear Models and Logistic Regression, Springer. New York.
4. Hilbe, J. (2011): Negative Binomial regression, Cambridge University, Press, 2nd Edition.
5. McCulloch, C. E., & Searle, S. R. (2003). Generalized, linear, and mixed models, Wiley series in probability and statistics, New York.

CCS-402: SURVIVAL ANALYSIS

Unit-1: Various types of censoring: right,left,interval censoring; random censoring; Survivor, hazard and cumulative hazard functions. Estimating the survivor function: Life-table estimate, Kaplan-Meier estimate, Nelson-Aalen estimate; Standard error of the estimated survivor functions; Confidence intervals for values of the survivor function; Non-parametric estimation of the hazard function, the cumulative hazard function; the median and percentiles of survival times; Confidence intervals for the median and percentiles.

(12L + 2T)

Unit-2: The Cox regression model:A regression model for the comparison of two groups; The general proportional hazards model, 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)

Unit3: Competing risks**:**Summarising competing risks data; Kaplan-Meier estimate of survivor function; its properties without proof, 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)

Unit4: a) Comparison of two groups of survival data: The log-rank test; The Wilcoxon test; Comparison of three or more groups of survival data.

b) Introduction to frailty Models: Random effects, Individual frailty, Shared frailty; Frailty distributions: The gamma frailty distribution; Lognormal frailty effects; Testing for the presence of frailty; The shared frailty model; Fitting the shared frailty model; Comparing shared frailty models.

(12L + 2T)

References:

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

CCS-402: ACTUARIAL STATISTICS

Unit-1: Introduction to Insurance Business, Insurance and utility theory, 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, accumulation factor, continuous compounding. Insurance payable at the moment of death and at the end of the year of death, level benefit insurance, Whole life insurance, endowment insurance, deferred insurance and varying benefit insurance. Recursion equations and commutation functions.

(12L+3T)

Unit-3: Annuities: Annuities certain, Continuous and Discrete life annuities. Life annuities with monthly payments and approtionable annuities. Recursion equations. Net premium: Fully continuous and discrete premiums, True monthly payment premiums, apportionable premiums and accumulation type benefits. Insurance model including expenses.

(12L+3T)

Unit-4: Net premium reserves: Continuous and discrete net premium reserve, reserves on a semi continuous basis, reserves based on true monthly premiums, reserves on an apportion able or discounted continuous basis, reserves at fractional durations, allocations of loss to policy years, recursive formulas and differential equations for reserves, commutation functions.

Some practical considerations: Premiums that include expenses-general expenses types of expenses, per policy expenses. Claim amount distributions, approximating the individual model, stop-loss Insurance.

(12L+3T)

References:

1. Robin Cunningham, Thomas N. Herzog, Richard L. Models for Quantifying Risk, 4th Edition, ACTEX Publications, 2011.
2. Browers, Newton L et al., Actuarial Mathematics 2nd. Society of Actuaries, 1997.
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. Deshmukh S. R., An Introduction to Actuarial Statistics, University Press, 2009
5. Narang, Uma, Insurance Industry in India: Features, Reforms and Outlook, New Century Publications

CCS-403: BIOSTATISTICS

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 blinding/masking in clinical trials. Bioavailability, pharmacokinetics and pharmaco-dynamics, two compartment model.

(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, design and monitoring of Phase III trials with sequential stopping.

(12L+3T)

Unit 3: Design of bio-equivalence trials, Inference for 2x2 crossover design: Classical methods of interval hypothesis testing for bioequivalence, Bayesian methods, nonparametric methods. Reporting and analysis: analysis of categorical outcomes from Phase I - III trials, analysis of survival data from clinical trials.

(12L+3T)

Unit 4: Epidemiological studies: case-control and cohort designs. Measures of disease occurrence and association, variation and bias, identifying non-causal association and confounding, communicating results of epidemiological studies, ethical issues in epidemiology. Causal Inference.

(12L+3T)

References:

1. C. Jennison and B. W. Turnbull (1999): Group Sequential Methods with Applications to Clinical Trials, CRC Press.
2. Chow S.C. and Liu J.P. (2004). Design and Analysis of Clinical Trials. 2nd Ed. Marcel Dekkar.
3. Chow S.C. and Liu J.P.(2009). Design and Analysis of Bioavailability and bioequivalence. 3rd Ed. CRC Press.
4. Clayton, D. and Hills, M. (2013). Statistical methods in Epidemiology, OUP.
5. Daniel, W. W. and Cross, C. L. (2012). Biostatistics: A Foundation for Analysis in the Health Sciences, 10th Edition, Wiley.
6. J. L. Fleiss (1989). The Design and Analysis of Clinical Experiments.Wiley and Sons.
7. L. M. Friedman, C. Furburg, D. L. Demets (1998). Fundamentals of Clinical Trials, Springer Verlag.
8. Marubeni.E. and Valsecchi M. G. (1994). Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley.
9. S. Piantadosi (1997). Clinical Trials: A Methodologic Perspective, Wiley and Sons.

CCS-403: ECONOMETRICS

Unit 1: Introduction: Definition and scope of econometrics, Methodology of econometrics. Nature and source of Data for econometric analysis, Types of data: cross section, time series, panel data, dummy variable, instrumental variable. Basic concepts of estimation: Review of general linear model, Ordinary least squares, generalized least squares.

(12L+ 3T)

Unit 2: Heteroskedasticity: consequences and tests: White test, Goldfeld-Quandt test; Estimation: estimation with grouping of observations, estimation of the heteroskedasticity relation. Linear regression with stochastic regressors, Instrumental variable estimation, Errors in variables. Autocorrelation, Autoregressive linear regression, Distributed lag models.

(12L+ 3T)

Unit 3: Simultaneous linear equations model. Example, Identification problem, Restrictions on structural parameters-rank and order conditions. Estimation in simultaneous equations model. Recursive systems. Two-Stage Least Squares estimators. Limited information estimators.

(12L+ 3T)

Unit 4: Definition of causality, Granger causality, Granger test for causality. Application of econometric methods: estimation of demand and supply function – production and cost function, consumption and investment functions.

(12L+ 3T)

References:

1. Apte, P.G. (1990): Text book of Econometrics. Tata McGraw Hill.
2. Gujarati, D.N. (2003): Basic Econometrics, McGraw Hill.
3. Johnston, J. (2006). Econometric Methods, third edition, McGraw Hill
4. Marno Verbeek, (2012): A guide to Modern Econometrics, 4/e,Wiley and Sons.
5. Nachane, D. M. (2006). Econometrics: Theoretical Foundations and Empirical Perspective, Oxford University Press
6. Ramanathan, R. (2002). Introductory Econometrics with applications, 5/e, Thomson Asia Private Limited
7. Wooldridge, J. (2012). Introductory Econometrics: A Modern Approach, 5/e, South-Western.

CCS-404: PYTHON FOR DATA SCIENCE

Unit 1: Introduction, Installation and Working with Pythonhttp://stanford.edu/class/cme193/python_course/lectures/lecture1.pdf, variables, Operators, data types input/output, data types, lists, dictionaries, data import and export, tuples, operations Functions, Functional Programming, Control Structures, Classes, Objects, Inheritance, basic statistical analysis, Libraries: Numpy, Pandas, Scipy, matplotlip.

(12L+3T)

Unit 2: Overview of neural Network concept, optimizers (Stochastic gradient descent, Adaptive Gradient Algorithm (AdaGrad)), 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 3: 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 4: Machine learning using scikit-learn library: Classification, Regression, and Clustering. 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)

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., Bengio, Y., Courville, A., &Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.
3. Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., &Mathur, I. (2016). Natural Language Processing: Python and NLTK. Packt Publishing Ltd.
4. Lutz, M. (2013). Learning Python: Powerful Object-Oriented Programming. " O'Reilly Media, Inc.".
5. McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.".
6. Mueller, J. P., &Massaron, L. (2015). Python for data science for dummies. John Wiley & Sons.
7. Shanmugamani R. (2018). Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras. “Packt Publishing Ltd”
8. Solem, J. E. (2012). Programming Computer Vision with Python: Tools and algorithms for analyzing images. " O'Reilly Media, Inc.".

9. VanderPlas, J. (2016). Python data science handbook: essential tools for working with data. " O'Reilly Media, Inc.".

DSE-405: SPATIAL DATA ANALYSIS

Unit 1: Spatial data and their types, the components of spatial data, spatial data models, spatial autocorrelation, modeling spatial autocorrelation, measures of spatial autocorrelation, tests for spatial autocorrelation, a spatial random process and its components.

(12L + 3T)

Unit 2: Spatial sampling: design-based and model-based approaches to spatial sampling, sampling plans.

Preprocessing of spatial data: quality of attribute data, spatial interpolation procedures, spatial rectification and alignment of data, Exploratory spatial data analysis.

(12L + 3T)

Unit 3: Regression models for spatially autocorrelated data: Detecting spatial autocorrelation in a regression model, Models for spatial processes: the spatial lag model and the spatial error model, Determining the appropriate regression model, Fitting the spatial lag and spatial error models, The conditional autoregressive model

(12L + 3T)

Unit 4: Analysis of spatiotemporal data: Spatiotemporal data interpolation, representing spatiotemporal data, the spatiotemporal variogram, interpolating spatiotemporal data, spatiotemporal process models, finite state and time models.

(12L + 3T)

References:

1. Plant, R. E. (2019). Spatial data analysis in ecology and agriculture using R, second edition. CRC Press.
2. Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., & Pebesma, E. J. (2013). Applied spatial data analysis with R, second edition. New York: Springer.
3. Fischer, M. M., & Wang, J. (2011). Spatial data analysis: models, methods and techniques. Springer Science & Business Media.
4. Haining, R. P., & Haining, R. (2004). Spatial data analysis: theory and practice. Cambridge University Press.

DSE-405: STATISTICAL QUALITY CONTROL

Unit 1: Quality Improvement Tools: affinity diagram, interrelationship digraph, tree diagram, prioritization matrix, matrix diagram, process decision program chart, activity network diagram, stem-and-leaf display, dot diagrams, boxplot, and normal probability plot, Engineering Process Control. Shewhart Control charts: basic statistical principles and assumptions, phase I and phase II applications, benefits from the use of control charts, concept of rational subgroups, performance measures of a control chart, $\overline{X}$, *R, S,* S^2*, p, c* and *D* charts, σ-control limits and probability control limits, over dispersion. Modifications to control chart procedures: warning limits, sensitizing rules, adaptive design parameters, integration of two charts. Concept of economic design of a control chart.

(12L+3T)

Unit 2: Alternatives to Shewhart control charts: CUSUM and EWMA charts, Shewhart–EWMA Chart. Multivariate Control Charts: multivariate chart versus individual charts, Hotelling's T^2 control chart, multivariate CUSUM Charts, multivariate EWMA charts, Regression adjustment. Other Control Charts: SPRT chart, GLR Chart, charts for autocorrelated data, nonparametric control charts, Bayesian control charts. The change point model for process monitoring.

(12L+3T)

Unit 3: Process capability Analysis: process capability, process capability indices (C_p, C_{pk}, C_{pm}, C_{pmk}), point and interval estimation of C_p and C_{pk}, Nonparametric Capability Indices: robust capability indices, capability indices based on fitted distributions, data transformation, capability indices computed using resampling methods. Multivariate Process Capability Indices. Six Sigma Methodology: components of a Six Sigma methodology, the DIMAC process, Six Sigma applications, Six Sigma concept for customer satisfaction, Six Sigma training, Lean Six Sigma.

(12L+3T)

Unit 4: Acceptance sampling plans for attributes: single sampling plan, double and multiple sampling plans, sequential sampling. Performance measures of sampling plans. Acceptance sampling plans for variables: Advantages and Disadvantages of Variables Sampling, Sampling inspection plans by variables for one or two sided specifications, Sequential Sampling by Variables, Rectifying inspection of lots, the Deming inspection criterion, Continuous sampling plans, skip-lot sampling plans.

(12L+3T)

References:

1. Guenther, W. C. (1977). Sampling Inspection in statistical quality control. Macmillan.
2. Kenett, R. S. and Zacks, S. (2014). Modern Industrial Statistics with applications in R, MINITAB and JMP. John Wiley & Sons.
3. Montgomery, D. C. (2010). Statistical Quality Control: A Modern Introduction, 6th Edition. Wiley India Pvt Ltd.
4. Ryan, T. P. (2011). Statistical Methods for Quality Improvement. John Wiley & Sons.

CCPR 406: PRACTICAL IV and PROJECT

1. Logistic Regression
2. Poisson Regression
3. Negative Binomial Regression
4. Generalized linear mixed models

5- onwards: At least FOUR practical each on the optional courses.

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 and replace, 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.