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

SHIVAJI UNIVERSITY, KOLHAPUR - 416 004, MAHARASHTRA
www.unishivaji.ac.in, bos@unishivaji.ac.in
शिवाजी विद्यापीठ, कोल्हापूर - ४१६ ००४,महाराष्ट्र
दूरध्वनी - ईपीएबीएक्स - २६०९०००, अभ्यासमंडळे विभाग दुरध्वनी ०२३१—२६०९०९३/९४

SU/BOS/Science/480 **Date: 01/07/2023**

To,

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

Subject: Regarding syllabi of M.Sc. Part-II (Sem. III & IV) as per NEP-2020 degree programme under the Faculty of Science and Technology.

Sir/Madam,

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

M.Sc. Part-II (Sem. III & IV) as per NEP-2020			
1.	Mathematics	8.	Botany
2.	Mathematics (Distance Mode)	9.	Electronics
3.	Mathematics (Online Mode)	10.	Zoology
4.	M.Sc. Tech (Industrial Mathematics With Computer Application)	11.	Agro Chemical and Pest Management (AGPM)
5.	Geography	12.	Alcohol Technology
6.	Statistics	13.	Sugar Technology
7.	Applied Statistics and Informatics	14.	Geology

This syllabus, nature of question and equivalence shall be implemented from the academic year 2023-2024 onwards. A soft copy containing the syllabus is attached herewith and it is also available on university website **www.unishivaji.ac.in**)

The question papers on the pre-revised syllabi of above-mentioned course will be set for the examinations to be held in October /November 2023 & March/April 2024. These chances are available for repeater students, if any.

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

Thanking you,

Dy Registrar
Dr. S. M. Kubal

Copy to:

1	The Dean, Faculty of Science & Technology	8	P.G. Admission/Seminar Section
2	Director, Board of Examinations and Evaluation	9	Computer Centre/ Eligibility Section
3	The Chairman, Respective Board of Studies	10	Affiliation Section (U.G.) (P.G.)
4	B.Sc. Exam/ Appointment Section	11	Centre for Distance Education

SHIVAJI UNIVERSITY, KOLHAPUR

SYLLABUS

For

M. Sc. Applied Statistics and Informatics Part II (Sem III and Sem IV)

As per NEP-2020

To be implemented from June, 2023

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, atomic domain and normalization(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, means, variances, covariances, partial and multiple correlation coefficients, correlations of linear transforms, graphical representation. 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. Wishart matrix and its distribution, properties of Wishart distribution, distribution of generalized variance. 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.

(12L+3T)

Unit 3: Discrimination and classification: Minimum ECM rule, Fisher's discriminant function and likelihood ratio procedure, Bayes discriminant rule, Rao's U statistics and its use in tests associated with discriminant function, classification with several 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, canonical variates and canonical correlation, interpreting population canonical variables, sample canonical variates and sample canonical correlations; Principal component analysis: Introduction, population principal components, summarizing sample variation by principal components, graphing the principal components, large sample inferences; Factor analysis: Introduction, orthogonal factor model, methods of estimation, interpretation of factors, factor rotation and factor score.

(12L+3T)

References:

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

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 Binaymi Branching process. Generating functions and its properties, moments. Probability of ultimate extinction. Distribution of population size and association results. Simulation of branching process. Basic elements of Queuing model. Steady state probabilities and various average characteristics for the models: M/M/1, M/M/1 with balking, M/M/c and M/G/1. (12 L+ 3T)

References:

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

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)

Unit 2: 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)

Unit 3: 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)

Unit 4: 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 functions: identity, sigmoid, double sigmoid, tanh, softmax, loss functions: squared error, cross entropy; optimizers: gradient decent, stochastic and minibatch gradient decent; McCulloch-Pitts AN model, single layer network, multilayer feed forward network model, training methods, ANN & regression models. Support vector machine: Introduction to support vector machine, loss functions, soft margin, optimization hyperplane, support vector classification, support vector regression, linear programming support vector machine for classification and regression.

(12L+3T)

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

(12L+3T)

References:

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

DSE-304: ARTIFICIAL INTELLIGENCE

Unit 1: AI Problem Solving: Historical development of AI, Role of heuristic 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

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

(12 L + 3 T)

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

(12 L + 3 T)

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

(12 L + 3 T)

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

(12L + 3T)

Reference:

1. 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. Consistent Estimator
2. Consistent Asymptotic Normal Estimator
3. Asymptotic and Variance Stabilizing Transformation based Confidence Intervals
4. LRT, Wald and Score test
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, Odds ratio and its interpretation, 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: 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 of parameters 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 and Hall.
2. Hosmer, D. and Lemeshow S. (1999). Applied Survival Analysis: Regression Modeling of Time to Event 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. New
6. York: Springer.
7. 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: Aim, need and ethics of clinical trials, Role of ethics in clinical trials, conduct of clinical trials, preclinical research, phase I-IV trials, multi-center trials, bias and random error in clinical studies, randomization; concept of blinding/masking in clinical trials, Data management: data definitions, case report forms, database design, data collection systems for good clinical practice.

(12L+3T)

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

(12L+3T)

Unit 3: Bioavailability, pharmacokinetics, and pharmaco-dynamics, two compartment model. Design of bio-equivalence trials, Decision rules for bioequivalence, Inference for 2x2 crossover design: Classical methods of interval hypothesis testing for bioequivalence, Bayesian methods, nonparametric methods. 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: aims, 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

Unit1: Introduction, Installation and Working with Python http://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: analyzingtext 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'ReillyMedia, 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.".

CCS 404: CIRCULAR DATA ANALYSIS

Unit-1: Examples of Circular data, Need for appropriate analysis of circular data; Descriptive Statistics, Measure of Center, Circular Distance and Measure of Dispersion, Higher Moments; Probability distributions on the Circle: Uniform Distribution, Cardioid Distribution, Triangular Distribution Circular Normal (CN) Distribution, Offset Normal Distribution, Wrapped Normal (WN) Distribution, Wrapped Cauchy (WC) Distribution.

(12L + 3T)

Unit-2: Sampling Distributions for CN Distribution: Distribution of (C, S), Distribution of *R*, Distribution of *V*, Estimating the Parameters of a CN Distribution, Properties of the MLEs, Confidence Intervals, Tests for Mean Direction, Tests for the Concentration Parameter, Comparing Mean Directions and Approximate ANOVA.

(12L+3T)

Unit-3: One-Sample Problem and Goodness-of-fit: Tests Based on Empirical Distribution Functions; Tests Based on Sample Arc-Lengths; Two-Sample Tests based on Edf's; Wheeler and Watson Test; Tests based on Spacing-Frequencies; Homogeneity Tests in Large Samples.

(12L + 3T)

Unit-4: Circular Correlation Measure *pc*, Rank Correlation, Other Measures of Circular Correlation; Circular-Linear Correlation; Circular-Circular Regression; Estimation of Regression Coefficients; Circular-Linear Regression; Linear-Circular Regression.

(12L + 3T)

References:

1. Jammalamadaka, S. Rao, and A. Sengupta. (2001) Topics in circular statistics. world scientific.
2. Mardia, K. V. (2014). Statistics of directional data. Academic press.
3. Fisher, N. I. (1995). Statistical analysis of circular data. cambridge university press.

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, Shewhart Control charts: basic statistical principles and assumptions, phase I and phase II applications, concept of rational subgroups, performance measures of a control chart, $\overline{X}$, *R, S, p, c* and *D* charts, σ-control limits and probability control limits, Modifications to control chart procedures: warning limits, sensitizing rules, adaptive design parameters, Engineering Process Control.

(12L+3T)

Unit 2: Alternatives to Shewhart control charts for process mean: Moving average control chart, CUSUM chart, EWMA charts, combined Shewhart-CUSUM chart, combined Shewhart–EWMA Chart, SPRT chart, and GLR Chart; Nonparametric control charts for process mean; Multivariate Control Charts: multivariate chart versus individual charts, Hotelling's T^2 control chart.

(12L+3T)

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

(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.

*****