

**CURRICULUM
OF
MASTER OF PHILOSOPHY
IN STATISTICS**

(2026)



**RAWALPINDI WOMEN UNIVERSITY,
RAWALPINDI**

TABLE OF CONTENTS

Program Details	3
ADMISSION CRITERIA M. Phil STATISTICS	4
DIVISION AND DISTRIBUTION OF COURSES:	4
List of Courses for M. Phil Statistics Program	6

DEPARTMENT VISION:

To provide quality education and research innovations in Statistics that empower women, with a focus on producing skilled graduates and researchers equipped to address future challenges across various professional fields.

DEPARTMENT MISSION:

To cultivate academic excellence, encourage impactful research, and equip graduates with robust statistical and analytical skills through innovation, collaboration, practical application, and sustainable professional development.

Program Details

Title of the Program	M. Phil Statistics
Duration of the Program	Minimum two years and maximum four years.
Total No. of Credit Hours	32 Credit Hours. A minimum of 26 Credit Hours of Course Work + 6 Credit Hours for Thesis.
Semester Duration	16 weeks of Teaching (excluding Examinations)

MPhil Statistics Program Objectives

1. To equip students with an in-depth understanding of statistical theories, methodologies, and interdisciplinary approaches, enabling them to analyze and solve complex real-world problems across diverse domains.
2. To train students in the proficient use of modern statistical software, programming languages, and computational tools essential for effective data management, analysis, and visualization.
3. To prepare students to design, conduct, and evaluate independent research using advanced statistical techniques, thereby contributing new insights to the field of statistics.
4. To enable students to effectively communicate complex statistical ideas and results to both technical and non-technical audiences through written, oral, and visual means.
5. To instill a strong sense of professional ethics, critical thinking, and creativity in applying statistical methods to address contemporary challenges responsibly and innovatively.

MPhil Statistics Program Learning Outcomes

By the completion of MPhil Statistics, the graduates will be able to

1. Demonstrate advanced knowledge of theories, concepts, and methodologies in Statistics, integrating interdisciplinary approaches to analyze and address complex problems and conduct high-level data analysis across various domains.
2. Master the use of modern statistical software, computational techniques, and programming languages for efficient data processing, analysis, and visualization.
3. Design and execute original research projects, utilizing advanced statistical methods to generate insights and contribute to the body of knowledge in the field.
4. Present and interpret statistical findings clearly and effectively to diverse audiences, ensuring accurate and meaningful dissemination of results.
5. Integrate ethical considerations and innovative approaches into statistical analysis, addressing real-world challenges responsibly and creatively.

ADMISSION CRITERIA M. Phil STATISTICS

1. At least sixteen years of education in Statistics / Mathematics / Econometrics / Data Analytics/ Data Science / Computer Science and other related fields with minimum CGPA of 2.0 on the scale of 4.00 (semester system) or minimum 50% marks in the annual examination system or 55% marks in the semester system examination where CGPA is not mentioned at terminal degree.
2. NTS/GAT test with 50% score or university admission test with a minimum of 50% score is required to qualify the departmental subject wise test.

DIVISION AND DISTRIBUTION OF COURSES:

M. Phil Statistics will contain a total of 32 credit hours out of which 26 will comprise course work and 6 credit hours for research dissertation.

Frame work for M. Phil Statistics Program

	Courses Details	No. of Credit Hours
1st year	Number of core courses: 04	12
	Number of elective courses: 04	12
2nd Year	Research Dissertation	6
	Understanding of Holy Quran	2
	Total Credit Hours	32

Semester I			
S.N.	Course	No. of Credit Hours	Category
1	Core I	3	Core
2	Core II	3	Core
3	Elective I	3	Elective
4	Elective II	3	Elective
Total Credit Hours = 12			
Semester II			
S.N.	Course	No. of Credit Hours	Category
1	Core III	3	Core
2	Core IV	3	Core
3	Elective III	3	Elective
4	Elective IV	3	Elective
Total Credit Hours = 12			
Semester III			
S.N.	Course	No. of Credit Hours	category
1	Thesis	6	Research
2	Understanding of Holy Quran - I	1	General Education
3	Understanding of Holy Quran - I	1	General Education
Total Credit Hours = 8			

List of Courses for M. Phil Statistics Program

Course Code	Course Title	Credit hours
Core Courses		
STA-730	Linear Models and Regression Analysis	3
STA-731	Computational Statistic	3
STA-732	Advanced Machine Learning Techniques	3
STA-702	Advanced Probability Theory	3
Elective Courses		
STA-703	Advanced Statistical Inference	3
STA-704	Applied Survey Sampling	3
STA-705	Randomized Response	3
STA-706	Applied Categorical Data Analysis	3
STA-707	Complex Survey Analysis	3
STA-708	Applied Multivariate Analysis	3
STA-709	Survival Data Analysis	3
STA-710	Applied Stochastic Models	3
STA-711	Modeling and Simulation	3
STA-712	Spatial Data Analysis	3
STA-713	Applied Operation Research	3
STA-714	Statistical Methods in Quality Control	3
STA-715	Time Series Analysis and Forecasting	3
STA-716	Meta-Analysis	3
STA-717	Applied Experimental Design	3
STA-718	Statistical Genetics	3
STA-719	Environmental Statistics	3
STA-720	Regression Models for Count Data	3
STA-721	Applied Functional Data Analysis	3
STA-723	Classification and Random Forest Technique	3
STA-725	Statistical Methods for Clinical Trials	3
STA-726	Mathematical Demography	3
STA-727	Bayesian Statistical Techniques	3
STA-728	Mixture Distributions	3

STA-729	Recent Developments in Statistics	3
STA-733	Applied Multivariate Statistics	3
STA-734	Applied Research Methodology in Statistics	3
STA-735	Research Design and Biostatistics	2
STA-735L	Research Design and Biostatistics (Lab)	1
STA-798	Research Dissertation	6

Linear Models and Regression Analysis

M. Phil -Statistics

Course Code: STA-730

Credit Hours: 03

Pre-requisite: Nil

Objectives:

This course aims to provide a rigorous and comprehensive understanding of linear modeling techniques and their extensions. It equips students with modern inference strategies, diagnostic tools, and advanced regression techniques applicable in high-dimensional and real-world settings. Emphasis will be placed on theoretical foundations, computational implementation, and application to real data.

Learning Outcomes

- Students will be able to explain the assumptions and inferential structure of the classical linear regression model.
- Students will be able to apply robust and generalized linear model techniques to non-ideal data conditions.
- Students will develop the understanding to implement and interpret results from ridge regression, lasso regression, polynomial regression, and distributed lag models.
- Students will be equipped with the use statistical software (e.g., R, Python) for implementing regression methods.
- Students will be able to explore and analyse real-world datasets using appropriate linear model techniques.

Course Outlines

Unit 1. Introduction to Linear Models

Matrix approach to linear regression. Estimability and identifiability. Gauss-Markov theorem. Estimability and identifiability of parameters. Projection matrices, orthogonality, and geometry of least squares. Gauss-Markov theorem and generalized least squares (GLS). Inference using likelihood methods and Wald tests.

Unit 2. Diagnostics and Model Adequacy

Influence diagnostics: Hat matrix, leverage, Cook's distance, DFFITS. Residuals: studentized, PRESS residuals, and bootstrapped residuals. Multicollinearity: Variance inflation factors (VIF) and condition indices. Heteroskedasticity-consistent standard errors (White's standard errors).

Unit 3. Penalized Regression and High-Dimensional Models

Ridge regression and singular value decomposition (SVD) interpretation. LASSO and solution paths using coordinate descent. Elastic net: theory and implementation. High-dimensional variable selection. Cross-validation and information-theoretic model selection (AIC, BIC, Cp, EBIC).

Unit 4. Generalized and Robust Regression Models

Polynomial and spline regression; natural and B-splines. Generalized linear models (GLMs) and link functions. Distributed lag models and dynamic linear regression. Robust regression: M-estimators, Huber, Tukey's bisquare function. Introduction to quantile regression and its applications.

Unit 5. Advanced Topics in Linear Modeling

Mixed models: random intercepts, random slopes. Bayesian linear regression and credible intervals. Shrinkage priors (ridge/lasso priors). Regularization in time series regression.

Recommended Books/ Reference Material

1. Bingham, N. H., & Fry, J. M. (2010). *Regression: Linear models in statistics*. Springer Science & Business Media. Chapman & Hall/CRC Press.
2. Christensen, R. (2013). *Log-Linear models and logistic regression*. United States: Springer.
3. Christensen, R. (2020). *Plane answers to complex questions: the theory of linear models*. Springer.
4. Fox, J., & Weisberg, S. (2022). *An R companion to applied regression* (4th ed.). Sage Publications.
5. Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2023). *Regression: Models, methods and applications* (2nd ed.). Springer.
6. Harville, D. A. (2023). *Linear models and the relevant distributions and matrix algebra: a unified approach volume 2*. Chapman & Hall/CRC Press.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
8. Faraway, J. J. (2021). *Linear Models with R* (2nd ed.). Chapman & Hall/CRC Press.
9. Michael, K., Nachtsheim, C., Neter, J., & Li, W. (2004) *Applied linear statistical model* (5th ed.). McGraw-Hill.
10. Rao C. R., Toutenberg, H., Shalabh., & Heumann, C. (2007). *Linear models and generalizations: lest squares and alternatives*. Springer.

Computational Statistics

M. Phil -Statistics

Course Code: STA-731

Credit Hours: 03

Pre-requisite: Nil

Objectives:

This course aims to focus on computational techniques used in modern statistics. Students will learn how to implement simulation-based inference, bootstrap methods, Markov Chain Monte Carlo (MCMC), and other numerical techniques. The course emphasizes algorithmic thinking, reproducible statistical computing, and the ability to perform inference in complex models using R or Python.

Learning Outcomes

- Students will acquire programming skills using R / Python.
- Students will be able to develop user-defined functions and packages for statistical computing.
- Students will generate pseudo data and simulate different models using programming with R or Python.
- To introduce core computational methods for statistical inference and data analysis.
- To develop programming skills for implementing statistical algorithms.

Course Outlines

Unit 1. Introduction to Computational Statistics

Role of computation in modern statistics. Importance of computation in modern statistics. Review of matrix operations and numerical accuracy. Floating point arithmetic and numerical errors. Setting up programming environments (RStudio, Jupyter, GitHub). Data types, indexing, and control flow in R and Python. Vectorization vs. loops. Writing clean, readable, and efficient code.

Unit 2 Functions and Statistical Packages

Writing user-defined functions and handling scope. Error handling and testing. Modular programming. Documenting and packaging functions (Roxxygen2 in R, docstrings in Python). Creating and sharing basic packages.

Unit 3. Pseudo-Random Data Generation

Random number generators and setting seeds. Generating data from discrete and continuous distributions. Simulating data for regression, classification, and time series models. Random sampling, resampling, and stratified sampling.

Unit 4. Simulation and Model Evaluation

Monte Carlo simulation basics. Estimating sampling distributions, bias, variance. Bootstrap and permutation tests. Simulating linear regression, logistic regression, and ANOVA models. Evaluating estimators through simulation studies.

Unit 5. Markov Chain Monte Carlo (MCMC) Methods

Introduction to Bayesian computation. Markov chains: convergence and ergodicity. Metropolis-Hastings algorithm. Gibbs sampling. Applications in Bayesian inference.

Recommended Books/ Reference Material

1. Davison, A. C., & Hinkley, D. V. (2021). *Bootstrap methods and their application*. Cambridge University Press.
2. Efron, B., & Hastie, T. (2021). *Computer age statistical inference: Algorithms, evidence, and data science*. Cambridge University Press.
3. Eubank, R. L. (2021). *A practical guide to simulation in R*. Chapman & Hall/CRC.
4. Givens, G. H., & Hoeting, J. A. (2022). *Computational statistics (3rd ed.)*. John Wiley & Sons.
5. Jones, M., & Altman, N. (2022). *Python for data analysis (3rd ed.)*. O'Reilly Media.
6. Marin, J. M., & Robert, C. P. (2022). *Bayesian essentials with R (2nd ed.)*. Springer.
7. Robert, C. P., & Casella, G. (2022). *Monte carlo statistical methods (3rd ed.)*. Springer.
8. Wickham, H., & Grolemund, G. (2023). *R for data science (2nd ed.)*. O'Reilly Media.

Advance Machine Learning Techniques

M. Phil -Statistics

Course Code: STA-732

Credit Hours: 03

Pre-requisite: Nil

Objectives

The goal of this course is to train students with foundational concepts and skills of advanced algorithms and methodologies in machine learning, with emphasis on theoretical understanding, mathematical foundations, and research-level applications. It explores scalable learning algorithms, ensemble methods, deep learning foundations, probabilistic models, and techniques for high-dimensional and structured data. Students will gain hands-on experience with real-world and simulated datasets.

Learning Outcomes

- Students will be able to learn key supervised and unsupervised algorithms for regression and classification.
- Students will gain skills in cross validation, generalization error and model selection.
- Evaluate models using techniques such as cross-validation, ROC analysis, and confusion matrices.
- Students will be able to construct ensemble models and compare their bias-variance trade-offs.
- Students will gain understanding of probabilistic graphical models and Bayesian machine learning.
- Students will be able to apply machine learning concepts to real world problems using programming language.
- Interpret machine learning results for academic, business, and policy-making contexts.

Course Outlines

Unit 1. Foundations of Machine Learning

What is machine learning? Supervised vs. unsupervised learning. Bias-variance trade-off. Overfitting, underfitting, and regularization. Cross-validation and model assessment metrics (MSE, MAE, accuracy, precision, recall, F1-score).

Unit 2. Supervised Learning Algorithms

Supervised learnings. Linear regression and regularized regression (Ridge, LASSO). Logistic regression and softmax for multiclass classification. Decision trees and random forests. Support vector machines (SVMs). k-Nearest neighbors (k-NN).

Unit 3. Unsupervised Learning

Unsupervised learning. K-means and hierarchical clustering. Principal component analysis (PCA). t-SNE and dimensionality reduction. Anomaly detection and density estimation.

Unit 4. Ensemble and Kernel Methods

Bagging, Boosting (AdaBoost, gradient boosting). Random forests and extreme gradient boosting (XGBoost). Kernel methods and support vector machines with kernel trick. Model calibration and ensemble strategies.

Unit 5. High-Dimensional and Sparse Learning

Curse of dimensionality and feature selection. L1 (LASSO), L2 (Ridge), and Elastic Net in high dimensions. Sparse SVM, Group lasso, stability selection. Scalable algorithms for big data (SGD, online learning).

Unit 6. Probabilistic and Bayesian Machine Learning

Gaussian mixture models and em algorithm. Variational inference and expectation propagation. Markov chain monte carlo (MCMC) in ML. Bayesian linear regression and logistic regression.

Recommended Books/Reference Material

1. Bishop, C. M. (2021). *Pattern recognition and machine learning*. Springer.
2. Breiman, L., Friedman, J., Olshen, R., & Stone, C. (2017). *Classification and regression trees*. New York: Wadsworth.
3. Duda, R., Hart, P., & Stork, D. (2000). *Pattern classification (2nd ed.)*. John Wiley & Sons.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning (2nd ed.)*. Springer.
5. James. G., Witten. D., Hastie. T., Tibshirani. R., & Taylor. J. (2023). *An introduction to statistical learning: with applications in Python*. New York: Springer-Verlag.
6. Jo, T. (2021). *Machine learning foundations: supervised, unsupervised, and advanced learning*. Springer.
7. Marsland, M. (2009). *Machine learning: an algorithmic perspective*. Chapman & Hall/CRC Press.

8. Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). *Multivariate statistical machine learning methods for genomic prediction*. Springer.
9. Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press.
10. Zhi-Hua Zhou. (2012). *Ensemble methods: foundations and algorithms* (1st ed.). Chapman & Hall/CRC Press.

Advanced Probability Theory

M. Phil - Statistics

Course Code: STA-702

Credit Hours: 03

Pre-requisite: Nil

Objectives

This course outline aims to provide a comprehensive understanding of recent developments in probability theory, covering both theoretical foundations and practical applications in various fields.

Learning Outcomes

- Students shall be familiarized with advanced concepts and techniques in probability theory research.
- Students will be able to explore applications of recent developments in various fields.
- Students will develop the ability to critically analyze and evaluate recent research papers in probability theory.

Course Outline

Unit 1. Introduction to Recent Developments in Probability Theory

Overview of classical probability theory. Introduction to measure theory and probability spaces. Basic concepts in random variables and expectations.

Unit 2. Expectation and Integral, Weak and Strong Convergence

Expectations and integrals. The inequalities of Markov, Chebyshev and Kolmogorov. Sequence of integrals. The strong law of large numbers. Zero one laws. Lebesgue stieltjes integrals. Weak convergence.

Unit 3. Transforms of Distributions

Characteristic functions, Laplace transforms, central limit theorem. Special distributions.

Unit 4. Stochastic Processes

Introduction to modern stochastic processes. Gaussian processes and their applications. Non-Markovian processes and fractional Brownian motion. Stochastic calculus.

Unit 5. Probabilistic Modeling

Advances in probabilistic graphical models. Deep generative models (e.g., variational autoencoders, generative adversarial networks). Probabilistic programming languages.

Unit 6. Advanced Topics in Probability Theory

Random matrix theory and its applications. Large deviations theory. Extreme value theory. Non-commutative probability.

Unit 7. Applications of Probability Theory

Finance: Stochastic calculus in quantitative finance. Biology: Stochastic modeling in population genetics. Physics: Statistical mechanics and stochastic processes. Computer Science: Probabilistic algorithms and data analysis.

Recommended Books/Reference Material

1. Feller, W. (1968). *An introduction to probability theory and its applications*, Vol 1. John Wiley & Sons.
2. Fristedt, B. E., & Gray, L. F. (2013). *A modern approach to probability theory*. Springer.
3. Galambos, J. (2023). *Advanced probability theory (2nd ed.)*. Chapman & Hall/CRC.
4. Kroese, D. P., & Botev, Z. (2023). *An advanced course in probability and stochastic processes*. Chapman & Hall/CRC.
5. Rohatgi, V. K. (1976). *An introduction to probability theory and mathematical statistics*. John Wiley & Sons.

Advanced Statistical Inference

M. Phil - Statistics

Course Code: STA-703

Credit Hours: 03

Pre-requisite: Nil

Objectives

The objective of this course is to provide a solid understanding of statistical inference theory. It is structured to empower students to effectively articulate the objectives of their analyses, interpret the outcomes, and comprehend the implications of their findings. Furthermore, the course aims to enhance students' ability to optimize model fitting by employing various estimation techniques.

Learning Outcomes

- Students have a solid understanding of probability-based statistical inference.
- Students will be equipped with the knowledge and skills to select and implement appropriate techniques to strike the right balance between variance and bias, ensuring the development of accurate and robust statistical models.
- Students will be equipped to apply statistical inference techniques to real-world datasets and draw insightful conclusions.

Course Outline

Unit 1. Review of Mathematical Statistics

Objective of statistical analysis and theory. Properties of estimators. Asymptotic theory. Concepts of convergence, asymptotic methods including the delta method, measure theory and lebesgue integration.

Unit 2. Point and Interval Estimation

Interval estimation: scalar parameter, scalar parameter with nuisance parameters, vector parameter, confidence region. Point estimation: general considerations on bias and variance, Cramer–Rao inequality, achievement of minimum variance and remove of bias,

estimates of minimum mean squared error, U-statistics, maximum likelihood estimation, the EM-algorithm. Bayes estimation, decision theory.

Unit 3. Randomization Tests

Distribution free and randomization tests, permutation tests, rank test, randomization tests and distance test.

Unit 4. Hypothesis Testing

Some general principles of statistical inference. Discrete problems, composite alternatives, local power, multidimensional alternatives, composite null hypothesis, similar region and invariants tests.

Unit 5. Advanced Topics and Applications

Statistical functional, robustness, bootstrap and Jackknife estimation with dependent data.

Recommended Books/Reference Material

1. Deshmukh, S., & Kulkarni, M. (2021). *Asymptotic statistical inference: A basic course using R*. Singapore: Springer Nature.
2. Hogg, R.V., & Craig, A.T. (1996). *Introduction to mathematical statistics*. New Jersey: Prentice Hall.
3. Illukkumbura, A. (2020). *Introduction to hypothesis testing*. Independently Published.
4. Lindgren, B.W. (1998). *Statistical theory*. Chapman & Hall/CRC.
5. Mood, A.M., Graybill, F.A., & Boes, D.C. (1997). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw Hill.
6. Niewiadomska, B.M., & Bartoszynski, R. (2020). *Probability and statistical inference*. John Wiley & Sons.
7. Rao, C.R. (2009). *Linear statistical inference and its applications*. John Wiley & Sons.
8. Rohatgi, V.K. (2003). *Statistical inference*. John Wiley & Sons.
9. Silvey, S. D. (1975). *Statistical inference*. Chapman & Hall/CRC.

Applied Survey Sampling

M. Phil - Statistics

Course Code: STA-704

Credit Hours: 03

Pre-requisite: Nil

Objectives

This course aims to introduce students a deep understanding of various sampling designs, including simple random sampling, stratified sampling, cluster sampling, probability proportional to size sampling, and multistage sampling and understand their applications and limitations. It is structured to learn about sources of bias and errors in survey sampling, such as nonresponse bias, coverage bias, measurement error, and techniques for minimizing their impact.

Learning Outcomes

- Students will be able to demonstrate a deep understanding of various sampling designs, including simple random sampling, stratified sampling, cluster sampling, and multistage sampling, and their respective advantages, disadvantages, and applications.
- Students will be able to achieve an understanding of the diverse methodological issues arising in sample survey research and the relationships between them.

Course Outline

Unit 1. Complex Estimators

Complex estimators in simple random sampling (SRS). Chain estimators in SRS. Stratified random sampling, construction of strata. Separate ratio, product and regression estimators in stratified sampling. Combined ratio, product and regression estimators in stratified sampling. Methods of removing bias from the estimators. Application in R using survey package.

Unit 2. Multi-Stage Sampling and Successive Sampling

Procedure for construction of estimators. Selection of first and second stage units using SRSWOR sampling. Optimum allocation in two-stage sampling. Optimum allocation of sample in three-stage sampling. Prediction approach to robust variance estimation in two-stage cluster sampling. Effects of survey design on regression analysis. Effects of two-stage sampling on OLS methods. Comparison of domain means in two stage sampling. Multistage sampling. Two-stage sampling with successive occasions. Practice in R.

Unit 3. Probability Proportional to Size Sampling

Probability is proportional to size with replacement (PPSWR) sampling methods, Use of more than one auxiliary variable in PPSWR, Estimation of the correlation coefficient. Probability proportional to size without replacement (PPSWOR) sampling methods. Horvitz and Thomson estimator. Multi-character survey. Using R, describe the samples, weights, quantiles, sampling by size. Practice in R.

Unit 4. Complex Surveys

Complex surveys. Variance estimations in complex Surveys. Categorical data analysis in complex surveys. Regression analysis for complex survey. Effects of survey design on regression analysis. Practice in R.

Unit 5. Non-Response and Errors in Sampling

Non-sampling errors. Observational errors. Incomplete sampling. Nonresponse, effects of non-response. Response and response Variance. Sources of response error. Detection, control and measurement of response error. Practice in R.

Recommended Books/Reference Material

1. Blair, E., & Blair, J. (2014). *Applied survey sampling*. Sage Publications.
2. Cingi, H., & Kadilar, C. (2009). *Advances in sampling theory-ratio method of estimation*. Bentham Science Publishers.
3. Cochran, W.G. (1996). *Sampling techniques*. John Wiley & Sons.

4. Latpate, R., Kshirsagar, J., Gupta, V. K., & Chandra, G. (2021). *Advanced sampling methods*. Springer.
5. Lohr, S. L. (2021). *Sampling: design and analysis*. Chapman and Hall/CRC.
6. Lumley, T. (2011). *Complex Surveys: A guide to analysis using R*. John Wiley & Sons.
7. Mukhopadhyay, P. (2005). *Theory and methods of survey sampling*. Prentice-Hall of India.
8. Sarndal, C. E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. Springer-Verlag.
9. Singh, S. (2003). *Advanced sampling theory with applications-How Michael selected Amy*. Kluwer Academic Publishers.
10. Valliant, R., Dever, J.A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. Springer.
11. Zhang, C. (2024). *Fundamentals of environmental sampling and analysis*. John Wiley & Sons.

Randomized Response

M. Phil- Statistics

Course Code: STA-705

Credit Hours: 03

Pre-requisite: Nil

Objectives

The aim of the course is to teach the student that how to improve the quality and reliability of survey data by protecting respondent privacy, reducing bias, and promoting ethical research practices.

Learning Outcomes

- Students shall be able to learn how to increase respondent's participation in data collection by ensuring them ethical considerations, such as confidentiality and informed consent.
- Students shall be able to understand the importance of data accuracy and reducing bias in survey research.

Course Outline

Unit 1. Introduction to Randomized Response Models

Overview of survey research and privacy concerns. Motivation for using randomized response methods. Historical background and development of randomized response techniques.

Unit 2. Basic Randomized Response Techniques

Simple randomization method. Forced response method. Unrelated question method. Warner's model and its variants Use of unrelated question in randomized response model to estimate unknown parameter.

Unit 3. Polychotomous Population and Multi-attribute Situations

Randomized response models in polychotomous population and multi-attribute situations.

Unit 4. Techniques for Quantitative Characters

Randomized response models for the estimation of quantitative characteristics.

Unit 5. Miscellaneous Topics on Randomize Response Techniques

A bayesian approach, lying Model, direct response and some allied methods for sensitive characters.

Recommended Books/Reference Material

1. Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys*. Chapman and Hall/CRC.
2. Chaudhuri, A., & Mukerjee, R. (2020). *Randomized response: Theory and techniques*. Taylor & Francis.
3. Fox, J. A., & Tracy, P.E. (1986). *Randomized response: A method for sensitive surveys*. Sage Publications.
4. Wayne, W. D. (1993). *Collecting sensitive data by randomized response: an annotated bibliography*. Georgia State University Business Press.
5. Reiber, F. (2021). *Enhancing the applicability of randomized response techniques*. Springer.
6. Latpate, R., Kshirsagar, J., Kumar Gupta, V., & Chandra, G. (2021). *Advanced sampling methods*. Singapore: Springer Nature.

Applied Categorical Data Analysis

M. Phil - Statistics

Course Code: STA-706

Credit Hours: 03

Pre-requisite: Nil

Objectives

The aim of this course is to grasp fundamental concepts in categorical data analysis and to identify various types of categorical data, applying suitable methodologies accordingly. It is structured to equip scholars with the skills to perform statistical analysis using available software and accurately interpret the results generated by computers.

Learning Outcomes

- Students will be capable of implementing basic categorical methods and integrating them for sampling estimation.
- Students will exhibit the expertise to characterize, analyze, and resolve various problems concerning categorical data.
- Utilize statistical software proficiently for the analysis of categorical data.

Course Outline

Unit 1. Review and Introduction of Categorical Data Analysis

A brief overview of categorical data, distributions and Inference for categorical data. Contingency analysis, probability structure of contingency tables, measures of association. Inference for two-way contingency tables.

Unit 2. Maximum Likelihood Analysis and Generalized Linear Models

Principles of likelihood-based inference, maximum likelihood analysis for frequency data. Components of generalized linear models, generalized linear models for binary and count data, statistical inference and model checking, fitting generalized linear models.

Unit 3. Logistic Regression

Interpreting the logistic regression model: odds ratio interpretation. Inference for logistic regression. Logistic regression with categorical predictors. Multiple logistic regression, building and applying logistic regression models. Model checking (likelihood-ratio model comparison tests, goodness of fit and the deviance). Alternative link functions for binary outcome, diagnostics receiver operating characteristic (ROC) curve analysis, hyper-volume under manifold (HUM) analysis, exact methods and conditional logistic regression.

Unit 4. Log Linear and Logit Models

Measure of association: delta method, odds ratio, invariance property of odds-ratio. Estimation and goodness of fit: simple models and pearson chi square statistic, sampling models and maximum likelihood estimation, likelihood-ratio chi-square statistic. Mantel-Haenszel test, iterative weighted least squares, quasi likelihood models, log linear models for multidimensional contingency tables.

Unit 5. Analysis of Deviance

Analysis of deviance and residuals, over dispersion and under dispersion. Implementation of R for analysis.

Recommended Books/Reference Material

1. Agresti, A. (2012). *Categorical data analysis (3rd ed.)*. John Wiley & Sons.
2. Chap, T.L. (1998). *Applied categorical data analysis*. John Wiley & Sons.
3. Hosmer, D.W., & Lemeshow, S. (2004). *Applied logistic regression*. John Wiley & Sons.
4. Liu, X. (2022). *Categorical data analysis and multilevel modeling using R*. SAGE Publications.
5. Ronald, C. (1997). *Log-linear models and logistic regression (2nd ed.)*. Springer.
6. Tang, W., He, H., & Tu, X. M. (2023). *Applied categorical and count data analysis*. Chapman and Hall/CRC.

Complex Survey Analysis

M. Phil - Statistics

Course Code: STA-707

Credit Hours: 03

Pre-requisite: Nil

Objectives

This course is aimed at introducing the specialized software procedures that have been developed for the analysis of complex sample survey data.

Learning Outcomes

- Students shall be able to learn the introduction of variance estimation techniques that have been developed to take into account stratification and cluster sampling that are properties of the multistage sampling designs used in most major survey programs.
- Students shall be able to understand the importance of the software procedures for commonly used analyses, including testing for between-group differences in means and proportions, regression analysis, logistic regression and multilevel modeling.
- Students will also learn the consequences of nonresponse and missing data on survey analysis and methods for dealing with missing data.

Course Outline

Unit 1. Introduction to Complex Sample Designs

Classification of sample designs. Survey estimation and inference for complex designs. Complex sample design effects. Multi-stage area probability designs, stratification, cluster sampling, weighting, item missing data, finite population corrections. Using R, describe the samples, weights, quantiles, sampling by size.

Unit 2. Foundations and Techniques for Design-Based Estimation and Inference

Finite populations and super population models. Variance estimation methods. Sampling error estimation models. Taylor series linearization method. Ultimate clusters. Use specialized software procedures R. Replication methods for variance estimation. Jackknife repeated replication. Balanced repeated replication. Replication methods using R.

Unit 3. Descriptive Analysis for Continuous Variables and Categorical Data

Descriptive analysis of complex sample survey data. Bivariate relationships between two continuous variables. Bivariate analysis for categorical data. Analysis of multivariate categorical data. Estimation and inference for special statistics. Functions of survey estimates. Bivariate analysis. Hypothesis testing for contrasts.

Unit 4. Regression Analysis

Regression analysis of complex sample survey data. Poisson regression. Logistics regression analysis, multinomial and ordinal logistic regression. Generalized linear models for complex sample survey data. Survival analysis for complex sample survey data. Modeling in R using survey package.

Unit 5. Multiple Imputation Methods and Applications

Important missing data concepts. Imputation of item missing data. Nonresponse adjustment and multiple imputation inference for survey data. Multilevel models for complex sample survey data. Small area estimation for complex survey sample data.

Recommended Books/Reference Material

1. Chambers, R. L., & Skinner, C. J. (Eds.). (2003). *Analysis of survey data*. John Wiley & Sons.
2. Heeringa, S.G., West, B.T. & Berglund, P. A. (2020). *Applied survey data analysis*. Chapman and Hall/CRC.
3. Hosmer, D.W., & Lemeshow, S. (2004). *Applied logistic regression*. John Wiley & Sons.
4. Lehtonen, R., & Pahkinen, E. (2004). *Practical methods for design and analysis of complex surveys*. John Wiley & Sons.

5. Lohr, S. L. (2021). *Sampling: design and analysis*. Chapman and Hall/CRC.
6. Lumley, T. (2007). *Complex sample surveys in R*. R Core Development Team.
7. Lumley, T. (2011). *Complex Surveys: A guide to analysis using R*. John Wiley & Sons.
8. Snijders, T. (2001). *Multilevel modeling of health statistics*. John Wiley & Sons.

Applied Multivariate Analysis

M. Phil -Statistics

Course Code: STA-708

Credit Hours: 03

Pre-requisite: Nil

Objectives

This course aims to introduce the multivariate structures, multivariate hypothesis testing, estimation, data reduction and modeling. The purpose of the course is to provide the conceptual and advanced knowledge of multivariate data analysis.

Learning Outcomes

- Students will be able to understand multivariate data and its real-world application.
- Students will gain understanding of how to select appropriate multivariate statistical tool in any given situation and analyze multivariate data and interpret of results.

Course Outline

Unit 1. Multivariate Methods Review

Multivariate descriptive statistics, data displays and pictorial representations, scatter plots, graphs of growth curves, random vectors and matrices, distance, eigen analysis, positive definite matrices, spectral decomposition. Related computation using R.

Unit 2. Multivariate Normal Distribution

Review of the multivariate normal density, its properties and related sampling distributions. Detection of outliers and cleaning data. Transformations to near normality. Inference about the normal population mean vector. Comparisons of several multivariate means.

Unit 3. Multivariate Linear Regression Models

Classification and regression tree (CART). Path analysis. Multivariate regression, Multivariate analysis of variance (MANOVA), Multivariate analysis of covariance (MANCOVA). Related applications using R.

Unit 4. Principal Components Analysis

Population principal components from standardized variables. Principal components for covariance matrices. Graphing the principal components. Monitoring quality with principal components. Computational work using R.

Unit 5. Factor Analysis and Canonical Correlation Analysis

Factor model. Principal factor analysis. Maximum likelihood factor loadings. Visual approaches to finding a partition. The population and the sample canonical variates and canonical correlations. Practice in R.

Unit 6. Classification, Discrimination and Cluster Analysis

Separation and classification for two multivariate normal populations. Evaluating classification functions. Discriminant analysis. Discrimination under estimation. Probabilities of misclassification. Distances and similarity measures. Hierarchical clustering methods. Nonhierarchical clustering methods.

Recommended Books/Reference Material

1. Afifi, A.A., & Clark, V. (1984). *Computer aided multivariate analysis*. California: Lifetime learning publications.
2. Afifi, A.A., May, S., Donatello, R. A., & Clark, V.A. (2021). *Practical multivariate analysis*. John Wiley & Sons.
3. Anderson, T.W. (2003). *An introduction to multivariate statistical analysis*. John Wiley & Sons.
4. Backhaus, K., Erichson, B., Gensler, S., Weiber, R., & Weiber, T. (2021). *Multivariate analysis*. Springer Books.
5. Cleff, T. (2019). *Applied statistics and multivariate data analysis for business and economics*. Switzerland: Springer International Publishing.
6. Denis, D.J. (2020). *Univariate, bivariate, and multivariate statistics using R: Quantitative tools for data analysis and data science*. John Wiley & Sons.
7. Johnson, R.A. & Wichern, D.W. (2007). *Applied multivariate statistical analysis*. New York: Prentice Hall.
8. Morrison, F. (1990). *Multivariate statistical methods*. New York: McGraw-Hill.

Survival Data Analysis

M. Phil -Statistics

Course Code: STA-709

Credit Hours: 03

Pre-requisite: Nil

Objectives

The objective of the course is to deepen students' understanding of survival analysis concepts beyond basic principles, focusing on advanced techniques and models used to analyze time-to-event data. This course enables the students to gain proficiency in implementing advanced survival models using statistical software.

Learning Outcomes

- Students will be able to gain proficiency in implementing advanced survival analysis techniques using statistical software, such as R or Python, including model estimation, interpretation, and diagnostic procedures.

- Students will apply advanced survival analysis techniques to real-world datasets with complex censoring mechanisms, effectively addressing research questions and practical challenges in various fields such as medicine, epidemiology, and social sciences.

Course Outlines

Unit 1. Introduction to Survival Analysis

Basic concepts in survival analysis: censoring, survival function, hazard function, Kaplan-Meier estimator for survival curves and Log-rank test and other methods for comparing survival curves.

Unit 2. Parametric Survival Models

Exponential, Weibull, Lognormal, Gamma, Generalized Gamma and Extended Generalized Gamma Distributions, Log-Logistic and Gompertz distributions for modeling survival data. Maximum likelihood estimation of parametric survival models. Model selection and goodness-of-fit tests.

Unit 3. Semi-Parametric Models: Cox Proportional Hazards Model

Introduction to the cox proportional hazards model, Partial likelihood estimation, assumptions and diagnostics for cox models, extensions and modifications of the cox model (stratified cox model, time-varying covariates).

Unit 4. Parametric Methods for Regression Model Fitting and Identification of Prognostic Factors

Preliminary examination of data, general structure of parametric regression models and their asymptotic likelihood inference, Exponential, Weibull, Lognormal, The Extended Generalized Gamma, Log-Logistic AFT Models, other parametric regression models, model selection methods, identification of risk factors related to survival time: Cox Proportional Hazards Model and identification of prognostic factors related to survival time: non proportional hazard models.

Unit 5. Advanced Topics in Survival Analysis

Cure models for long-term survival analysis, multi-state models for complex event sequences, and Joint models for longitudinal and time-to-event data.

Recommended Books/ Reference Material

1. Aalen, O. O., Borgan, O., & Gjessing. (2012). *Survival and event history analysis*. New York: Springer series.
2. Collet, D. (2014). *Modelling survival data in medical research (3rd ed.)*. Chapman and Hall/CRC.
3. Harrell, Jr. F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Germany: Springer International Publishing.

4. Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data* (2nd ed.). New York: Springer series.
5. Kleinbaum, D.G., & Klein, M. (2012). *Survival analysis: a self-learning text* (3rd ed.). Springer.
6. Lee, E. T., & Wang, J. W. (2013). *Statistical methods for survival data analysis*. (4th ed.). John Wiley & Sons.
7. Machin, D., Cheung, Y. B., & Parmar, M. K. (2006). *Survival analysis: a practical approach* (2nd ed.). John Wiley & Sons.
8. Nag, A. (2021). *Survival analysis with python*. USA: Chapman and Hall/CRC.
9. O'Quigley, J. (2021). *Survival analysis: proportional and non-proportional hazards regression*. Springer International Publishing.

Applied Stochastic Models

M. Phil - Statistics

Course Code: STA-710

Credit Hours: 03

Pre-requisite: Nil

Objectives

This course aims to provide an understanding of stochastic processes and the ability to analyze certain aspects of these processes. Accordingly, the course starts by reviewing probability theory, conditional probability, independence and certain properties of random variables, and continues by examining stationary processes and Ergodic theorem. Furthermore, Markov chains in discrete and continuous time as well as Poisson processes are investigated in detail.

Learning Outcomes

- Students will be able to define probability models, concept and properties of random variables, random processes, Markov processes and Markov chains.
- Explain properties and functions of random processes with stochastic mathematical models, formulate discrete and continuous time random processes, stationary random processes.
- Devise solutions with probability models for Poisson processes, discrete and continuous time Markov chains.

Course Outline

Unit 1. Introduction to Probability and Random Variables

Probability spaces, random variables and stochastic processes, conditional probability, expected value and conditional expectations, probability generating functions.

Unit 2. Stochastic and Bernoulli

Types of stochastic process. Stationary process. Bernoulli process, basic inequalities (The Markov inequality and the Chebyshev inequality). The law of large numbers, compound distributions, simple random walk, branching processes.

Unit 3. Poisson Processes and Markov Chains Processes

Poisson processes, Markov processes, discrete and continuous time Markov chains. Imbedded Markov chains, birth-death process, immigration and immigration processes, immigration-death-processes.

Unit 4. Queuing Theory and Renewal Phenomena

Queuing Theory. Renewal Processes. Markov renewal process. Ergodic theorem. Gaussian Processes and Brownian motion.

Unit 5. Application of Stochastic Models

Time series analysis with stochastic model, spatial stochastic processes and spatial statistics, multivariate stochastic processes and copulas techniques. Application of stochastic model to a real-world problem using R or Python.

Recommended Books/Reference Material

1. Cinlar, E. (2013). *Introduction to stochastic processes*, New York: Dover Publications.
2. Dobrow, R. P. (2020). *Introduction to stochastic processes with R*. John Wiley & Sons.
3. Durrett, R. (2018). *Essentials of stochastic processes*. Springer.
4. Feller, W. (2008). *An introduction to probability theory and its applications*. John Wiley & Sons.
5. Gall, J.F.L. (2022). *Measure theory, probability, and stochastic processes*. Springer Nature.
6. Gallager, R.G. (2014). *Stochastic processes: theory for applications*, New York: Cambridge University Press.
7. Kao, E. P. (2019). *An introduction to stochastic processes*. Courier Dover Publications.
8. Pogorui, A., Swishchuk, A., & Dagnino, R.M. (2021). *Random motions in markov and semi-markov random environments 1: Homogeneous random motions and their applications*. John Wiley & Sons.
9. Ross, S.M. (2008). *Stochastic process*. New York: John Wiley & Sons.
10. Smith, W. (2020). *Stochastic processes in Python: fundamentals and applications*. Chapman and Hall/CRC.

Modeling and Simulation

M. Phil -Statistics

Course Code: STA-711

Credit Hours: 03

Pre-requisite: Nil

Objectives

The objective of this course is to deepen understanding of advanced modeling and simulation concepts and methodologies. Students will explore various types of complex systems and phenomena and learn how to model them effectively through statistical software and programming languages.

Learning Outcomes

- Students will be able to model and simulate complex systems and phenomena across various domains, including social systems, healthcare, transportation, manufacturing, and finance.
- Students will be able to critically evaluate and compare different simulation approaches, algorithms, and methodologies, selecting the most appropriate ones for a given problem based on their strengths and limitations.

Course Outlines

Unit 1. Introduction to Advanced Modeling and Simulation

Overview of advanced modeling and simulation techniques, role of modeling and simulation in decision-making processes, ethical considerations in modeling and simulation studies.

Unit 2. Advanced Modeling Techniques

Discrete-event simulation, agent-based modeling, system dynamics modeling, petri nets and other formal model approaches.

Unit 3. Simulation Optimization and Stochastic Simulation

Introduction to simulation optimization, optimization algorithms for simulation models, response surface methodology, multi-objective optimization in simulation, stochastic simulation, Monte Carlo simulation methods, Latin hypercube sampling, importance sampling and variance reduction techniques and sequential Monte Carlo methods.

Unit 4. Distributed and Parallel Simulation

Parallel and distributed computing architectures, Parallel discrete-event simulation (PDES) techniques. Load balancing and synchronization in distributed simulation, validation and verification of simulation models. Techniques for model validation and verification. Sensitivity analysis and uncertainty quantification. Model calibration and validation against real-world data.

Unit 5. Advanced Topics in Simulation

Agent-based modeling for social systems. Complex adaptive systems and emergent behavior. Hybrid simulation models combining different modeling paradigms and simulation-based decision support systems.

Recommended Books/ Reference Material

1. Awasthi, A. (2022). *Advanced systems modeling and simulation*. USA: Elsevier Science.
2. Ciaburro, G. (2020). *Hands-On simulation modeling with python: develop simulation models to get accurate results and enhance decision-making processes*. India: Packt Publishing.
3. Daniel P. M., & Maynard T. (2006). *Mathematical modeling and computer simulation*. Thomson Brooks/Cole.
4. Fishman, G.S. (1996). *Monte Carlo: Concepts, algorithms and applications*. Springer.
5. Kleijnen, J.P.C. (1974). *Statistical techniques in simulation*. New York: Marcel Dekker.

6. Pritsker, A.A.B. (1984). *Introduction to simulation & SIAM*. New York: Halsted Press.
7. Ross, S.M. (2002). *Simulation* (3rd ed.). Academic Press.
8. Velten, K. (2009). *Mathematical modeling and simulation*. John Wiley & Sons.

Spatial Data Analysis

M. Phil - Statistics

Course Code: STA-712

Credit Hours: 03

Pre-requisite: Nil

Objectives

The purpose of the course is to introduce spatial statistics and to provide students with necessary background to investigate the geographically represented data. This course is designed to develop a deeper understanding of the three main areas of spatial statistics: geo statistical data, lattice (areal) data and point patterns and to develop comprehension in the application of spatial autocorrelation in statistical modeling. Furthermore, to develop the perception and basic skills to apply spatial methods in their own research using statistical software and Geographical Information System (GIS).

Learning Outcomes

- Student will be able to distinguish different types of spatial data and understand how spatial autocorrelation plays a role in statistical modeling.
- Student will be able to discuss new methods in the spatial statistics literature based on an understanding of the basic spatial statistics approaches, principles and main assumptions.
- They will determine which spatial methods to use in their own research and implement them using existing methods to investigate spatial autocorrelation in example data sets provided as exercises.

Course Outline

Unit 1. Overview and Exploring Area Data

Introduction to spatial statistics and data handling, data and spatial data analysis, types of spatial data, spatial data matrix, spatial autocorrelation and tyranny of spatial data. Mapping and geo visualization. Eigen function analysis of areal unit configuration, spatial weight matrix, global measures and tests for spatial autocorrelation, local measures and tests for spatial autocorrelation and spectral analysis.

Unit 2. Modeling Area Data

Spatial regression models, models of spatial autocorrelation, tests for spatial dependence, the spatial durbin model, estimation of spatial regression models and model parameter interpretation

Unit 3. Spatial Autoregressive Model

Spatial series and spatial autoregression, difference of spatial autoregressive models from standard regression models, variable selection for spatial logistic autoregressive models, higher order autoregressive models, inference on higher-order spatial autoregressive models and model and basic assumptions.

Unit 4. Spectral Models and Spatial Data Interpolation

Introduction to global spectral modeling, relationship between autoregressive and spectral models. Kriging: introduction to concepts and applications, kriging interpolation.

Unit 5. Applications of Spatial Data Analysis in R

Case studies and applications in various domains (e.g., environmental science, urban planning, public health). Hands-on projects and presentations.

Recommended Books/Reference Material

1. Bartlett, M. (1975). *Statistical analysis of spatial pattern*. Chapman and Hall/CRC.
2. Cressie, N. (1987). *Statistics of spatial data*. John Wiley & Sons.
3. Fischer, M. M., & Wang, J. (2011). *Spatial data analysis: models, methods and techniques*. Springer Science & Business Media.
4. Griffith, D. (1988). *Advanced spatial statistics*. Bostan: Kluwer.
5. Hijmans, R. J. (2019). *Raster: Geographic data analysis and modeling*. R package version 3.3-7. URL: <https://CRAN.R-project.org/package=raster>
6. Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2021). *Geographic information science and systems*. John Wiley & Sons.
7. Lovelace, R., Nowosad, J., & Muenchow, J. (2019). *Geocomputation with R*. Chapman and Hall/CRC.
8. Pebesma, E., & Bivand, R. (2005). *Classes and methods for spatial data*. R package version 1.4-5. URL: <https://CRAN.R-project.org/package=sp>
9. Ripley, B. (1988). *Statistical inference for spatial processes*. John Wiley & Sons.
10. Shekhar, S., & Chawla, S. (2022). *Spatial databases: a tour*. Springer.
11. Smith, J. (2023). *Spatial data analysis: methods and applications* (2nd ed.). Academic Press.
12. Upton, G. & Fingleton, B. (1985). *Spatial data analysis by example*, Vol. I & II. John Wiley & Sons.

Applied Operation Research

M. Phil - Statistics

Course Code: STA-713

Credit Hours: 03

Pre-requisite: Nil

Objectives

The aim of the course is that the students should understand the advance knowledge of operations research. The course outline is designed to help the student to identify and develop operation research model, its applications and simulation techniques.

Learning Outcomes

- Students shall be able to understand the concept of the decision analysis, and knowledge about the game theory.
- Students will also able to identify the best path for the network of optimization problem.
- They will be able to develop the mathematical tools for getting optimum solutions.

Course Outline

Unit 1. Introduction to Operation Research

Definition of operation research (OR), nature and scope of operation research. Objectives of OR. Major phases of OR study. Development of OR models and its application in various scenarios of business.

Unit 2. Linear Programming

Introduction to linear programming. Feasible and optimal solutions, idea of simplex method, maximization and minimization case, big-m method, two-phase method or artificial variable method, duality problem, primal-dual relationships, optimal solution to dual problem, dual simplex method, Sensitivity analysis. Introduction to advance linear programming model (LPM), properties of advance LPM, formulation of advance LPM, simplex method fundamentals, revised simplex method, boundary variables algorithms, duality, parametric linear programming. Introduction to transportation model, comparison between LPM and transportation model, feasible solution by North-west corner method, least cost cell method, vogel's approximation method, least time model, sensitivity analysis of transportation model, assignment model.

Unit 3. Non Linear Programming

Introduction to non-linear programming, constraint and unconstrained algorithms. Integer linear programming, its application and algorithms. Deterministic dynamic programming, recursive nature of computation in DP. Forward and backward recursion.

Unit 4. Network Analysis

Networking models, its definition and scope, various algorithm for the networking, maximum flow models. CPM and PERT. Deterministic inventory (DI) models.

Unit 5. Decision Analysis, Games and Simulation Modelling

Introduction to decision analysis and games, classification of decisions, steps in decision theory approach, decision making under uncertainty, decision making under risk, criterion of optimism, criterion of pessimism, hurwicz criterion, regret criterion, decision making with and without experimentation. Bayes' decision rule, decision trees. Game theory and optimal solution of two person zero sum games, mixed strategy games. Introduction to simulation, advantages of simulation, types of simulation models. Monte Carlo simulation, generation of random numbers, use of computer packages and R programming for operation research study.

Recommended Books/Reference Material

1. Gupta, P. K., & Hira D. S. (2015), *Operations research (5th ed.)*. India: S. Chand Publications.
2. Hiller F. S., & Liberman, G. J. (2001). *Introduction to operations research (7th ed.)*. McGraw-Hill.
3. Murthy, P.R. (2007). *Operations research (2nd ed.)*. India: New age international publishers.
4. Taha, H. A. (2007). *Operations research: An introduction (8th ed.)*. United States: Pearson Prentice Hall.
5. Goel, A., & Agarwal, R. (2020). *Operation research*. India: Repro Books Limited.
6. Boucherie, R. J., Braaksma, A., & Tijms, H. C. (2021). *Operations research: Introduction to models and methods*. Singapore: World Scientific.

Statistical Methods in Quality Control

M. Phil - Statistics

Course Code: STA-714

Credit Hours: 03

Pre-requisite: Nil

Objectives

This course provides an in-depth exploration of advanced statistical methods used in quality control. The theory and methods of quality monitoring including process capability, control charts, acceptance sampling, quality engineering and quality design is included. Emphasis is placed on practical applications and case studies across various industries.

Learning Outcomes

- Students shall be able to understand the principles and importance of quality control in manufacturing and service industries.
- Students will be able to learn advanced statistical methods for monitoring and improving quality.
- Students will be able to develop skills in applying statistical techniques to real-world quality control problems.

Course Outline

Unit 1. Introduction to Quality Control

Overview of quality control concepts and terminology. Importance of quality control in manufacturing and service industries. Role of statistical methods in quality improvement.

Unit 2. Statistical Process Control (SPC)

Control charts for monitoring process stability and variability. Shewhart control charts for variables and attributes. Cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) charts. Multivariate control charts for monitoring correlated processes.

Unit 3. Multivariate Process Monitoring and Control

Principal component analysis (PCA) for dimensionality reduction. Multivariate control charts (e.g., Hotelling's T² chart). Discriminant analysis for classification and quality prediction. Multivariate analysis of variance (MANOVA) for comparing multiple groups.

Unit 4. Process Design and Improvement with Designed Experiments

Factorial and fractional factorial experiments for process design and improvement. Process optimizations with designed experiments.

Unit 5. Acceptance Sampling

Acceptance sampling plan for attributes and variables. Single, double, multiple and sequential sampling plans. The dodge-Roming sampling plan. Chain sampling, continuous sampling and skip lot sampling plan.

Unit 6. Advanced Topics in Quality Control

Chi square chart. Robustness and nonparametric approaches for process monitoring. Some Bayesian structures for quality control, covariates and process improvement. Total Quality Management (TQM) and Six Sigma methodology. Quality function deployment

(QFD) and failure mode and effects analysis (FMEA). Lean manufacturing principles and practices. Integration of quality control with other business processes (e.g., supply chain management). Taghuchi's methods for quality control.

Recommended Books/Reference Material

1. Alwan, L. C. (2000). *Statistical process analysis*. Boston: Irwin/McGraw-Hill.
2. Gupta, B. C. (2021). *Statistical quality control: using MINITAB, R, JMP and Python*. John Wiley & Sons.
3. Montgomery, D. C. (2020). *Introduction to statistical quality control*. John Wiley & Sons.
4. Oakland, J., & Oakland, R. J. (2018). *Statistical process control*. United Kingdom: Taylor & Francis.
5. Ryan, T. P. (2011). *Statistical methods for quality improvement*. John Wiley & Sons.

Time Series Analysis and Forecasting

M. Phil - Statistics

Course Code: STA-715

Credit Hours: 03

Pre-requisite: Nil

Objectives

The goal of this course is to provide students with a comprehensive understanding of forecasting techniques and modern statistical methods for analyzing time series data. Students will learn both the theory and practical application of three key components i.e., Univariate methods, Regression methods and ARIMA models.

Learning Outcomes

- Students will be able to conduct exploratory data analysis techniques tailored for time series data, including visualization, trend identification, and seasonal decomposition.
- Students will gain proficiency in forecasting future values of time series data using various forecasting methods, including naive methods, exponential smoothing, and ARIMA models.
- Students will explore advanced topics in time series analysis, such as multivariate time series, dynamic regression models, and state space models.
- Students will become proficient in using statistical software packages such as R or Python for time series analysis, including data preprocessing, model estimation, and forecasting.

Course Outline

Unit 1. Review and Basic Concepts

Type of data, components of time series data, objectives of time series analysis, some simple time series models, stationary models and autocorrelation function, estimation and elimination of trend and seasonal components, testing the estimated noise sequence, stationary and non-stationary processes, forms and tests of non-stationarity, purely random processes, random walk models, lag operator, difference equations and their solutions, smoothing and decomposition methods.

Unit 2. Univariate Time Series Analysis

Stationary, nonstationary and seasonal time series models (ARMA, ARIMA, Box-Jenkins approach, ARCH, GARCH etc.). Time series modelling and diagnostic checking. Modeling and forecasting with ARMA, ARIMA and seasonal ARIMA models. Spectral analysis.

Unit 3. Multivariate Time Series

Multivariate time series analysis, granger causality, vector autoregressive models, vector error correction model. Modeling and forecasting with multivariate AR process, co-integration analysis and johansen approach.

Unit 4. State Space Models and Intervention Analysis

State space representation, basis structural model, state space representation of ARIMA models, Use of Kalman filter and generalized space models. Transfer function, prediction based on a transfer function model and intervention analysis.

Unit 5. Applications of Time Series Analysis in R (or Python)

Time series analysis in economics and finance, time series analysis in engineering and environmental science. Case studies and real-world applications in R (or Python).

Recommended Books/Reference Material

1. Anderson, T. W. (2011). *The statistical analysis of time series* (Vol. 19). John Wiley & Sons.
2. Brockwell, P.J., & Davis, R.A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer
3. Chan, K.S., & Cryer, J.D. (2008). *Time series analysis with applications in R*. Springer Publication.
4. Chatfield, C., & Yar, M. (2019). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
5. Diebold, F.X., Kilian, L., & Nerlove, M. (2006). *Time series analysis. in macroeconometrics and time series analysis*. London: Palgrave Macmillan.
6. Jonathan, D. C., & Kung-Sik C. (2008). *Time series analysis with applications in R*. Springer.
7. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
8. Peter, J. B., & Richard, A. D. (2002). *Introduction to time series and forecasting* (2nd ed.). Springer.
9. Ruey S. Tsay. (2014). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons.
10. Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples* (4th ed.). Springer.
11. Vishwas, B., & Patel, A. (2020). *Hands-on time series analysis with Python*. Apress.
12. Woodward, W.A., Sadler, B.P., & Robertson, S. (2022). *Time series for data science: Analysis and forecasting*. Chapman and Hall/CRC.

Meta-Analysis

M. Phil - Statistics

Course Code: STA-716

Credit Hours: 03

Pre-requisite: Nil

Objectives

The primary goal of this course is to familiarize students with both fundamental and advanced techniques for conducting meta-analysis, with a specific focus on utilizing the statistical software R. Additionally, the course aims to provide students with the skills necessary to systematically review observational studies through the lens of meta-analysis.

Learning Outcomes

- Students will gain insight into crucial aspects of conducting a systematic review of observational studies within the context of meta-analysis.
- Students will develop an understanding of standard meta-analytic techniques and methodologies, along with practical applications utilizing the statistical software R.

Course Outline

Unit 1. Review and Introduction

Introduction to meta-analysis, how meta-analysis works, why perform a meta-analysis, systematic reviews and meta-analyses.

Unit 2. Effect size and Precision

Fixed and random effects in meta-analysis, effect sizes based on means, effect sizes based on binary data, effect sizes based on correlation. Differences in treatment effects in meta-analysis, forest plots, funnel plots.

Unit 3. Heterogeneity

Identifying and quantifying heterogeneity, prediction intervals, classification of heterogeneity.

Unit 4. Power Analysis and Systematic Review.

Meta regression, sub-group analysis and meta regression, computational models, multiple comparisons. Meta-analysis methods based on p-values. Publication bias, network meta-analysis and reporting a systematic review.

Unit 5. Meta-Analysis in R

Getting started with R, loading, saving and restoring data, select variables from an R datasets, running scripts, using libraries of additional functions. A first meta-analysis with R, R package for meta-analysis.

Recommended Books/Reference Material

1. Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2021). *Doing meta-analysis with R*. Chapman and Hall/CRC.
2. Julian, H., & Sally, G. (2008). *Cochrane handbook for systematic Reviews of intervention*. John Wiley & Sons.
3. Matthias, E., George, D. S., & Doug A. (2001). *Systematic reviews in health care: meta-analysis in context (2nd ed.)*. London: BMJ Publishing Group.
4. Michael, B., Hedges, L. L., Julian, P. T., & Hannah, R. (2021). *Introduction to meta-analysis (2nd ed.)*. John Wiley & Sons.
5. Schwarzer, G., Carpenter, J. R., & Rucker, G. (2016). *Meta-analysis in R*. Springer.

Applied Experimental Design

M. Phil - Statistics

Course Code: STA-717

Credit Hours: 03

Pre-requisite: Nil

Objectives

The purpose of the course is to equip students to understand and apply the principles of experimental design, incomplete block designs, and balanced incomplete block designs. The course provides basic and advanced skills to understand the advantages and limitations of factorial designs, fractional factorial designs, symmetrical and asymmetrical factorial designs. Also learn to investigate the advanced designs such as nested designs, split-plot designs, lattice designs and response surface designs.

Learning Outcomes

- Students will be able to explore the general theory of complete and incomplete block designs and understand this theory sufficiently to find appropriate designs for specific applications.
- Students will be able to proficiently evaluate designs using common optimality criteria and use them to critically compare competing designs.

Course Outline

Unit 1. Fractional Factorial Designs

Half fractions of 2^k designs. Quarter and higher fractions of 2^k designs. Criteria for choosing generators for 2^{k-p} designs. Mixed level factorials and orthogonal arrays. Brief introduction to R in experimental design.

Unit 2. Incomplete and Confounded Block Designs

General incomplete block designs (IBD). Analysis of incomplete block designs. Construction of balanced incomplete block designs (BIBD). Construction of partially balanced incomplete block designs (PBIBD). Intra-block and Inter-block analysis of IBD.

Unit 3. Nested Designs, Split Plot Designs and Lattice Designs

Crossed and nested factors. Split Plots, Strip Plots, Split-Split Plots, Strip-Split Plots. Square lattice designs. Rectangular lattice designs. Generalized lattice designs. Latinized block designs.

Unit 4. Special Designs for Process Improvement

Assessing quality performance. Statistical designs for process improvement. First and second order response surface designs. Optimal designs and optimality criteria. Taguchi's robust product design. Related applications and computations with R.

Unit 5. Mixture Experiments

Models and designs for mixture experiments. Creating mixture designs in R. Analysis of mixture experiments. Blocking mixture experiments. Mixture experiments in split plot arrangements.

Recommended Books/Reference Material

1. Box, G.E.P, Hunter, J.S., & Hunter, W.G. (2005). *Statistics for experimenters: design, innovation and discovery*. John Wiley & Sons.
2. Cox, D.R., & Reid, N. (2000). *The theory of the design of experiments*. Chapman and Hall/CRC.
3. Cochran, W.G., & Cox, G.M. (1992). *Experimental design*. John Wiley & Sons.
4. Dean, A., Morris, M., Stufken, J., & Bingham, D. (2020). *Handbook of design and analysis of experiments*. Chapman and Hall/CRC
5. Hinkelmann, K. (2012). *Design and analysis of experiments, volume 3: special designs and applications (Vol. 810)*. John Wiley & Sons.
6. Lawson, J. (2014). *Design and analysis of experiment with R*. Chapman and Hall/CRC.
7. Mead, R., Gilmour, S.G., & Mead, A. (2012). *Statistical principles for the design of experiments: applications to real experiments*. Cambridge University Press.
8. Montgomery, D.C. (2017). *Design and analysis of experiments*. John Wiley & sons.
9. Herzong, M.H., Francis, G., & Clarke, A. (2020). *Understanding statistics and experimental design: How to not lie with statistics*. Springer International Publishing.

Statistical Genetics

M. Phil -Statistics

Course Code: STA-718

Credit Hours: 03

Pre-requisite: Nil

Objectives

The course focuses to provide an understanding of the principles of statistical methods for genetic studies. The contents has the sufficiency to obtain knowledge on statistical genetics and fathom skills to analyze data from human/ animal and plant genetics. Furthermore, this course introduces the Microarray Gene Expression data matrix and its evaluation relationship with applied multivariate techniques.

Learning Outcomes

- Students will be able to have insight into the mathematical, statistical, and computational basis of genetic analysis that use genome-scale data sets in systems biology settings.
- Students will have exposure to study micro array gene expression data analysis that will make the students meet the challenges of large complex data sets and contribute to new methodological and computational advances.

Course Outlines

Unit 1. Introduction to Genetic Variation

Overview of genetic variation and inheritance, introduction to DNA sequencing and genotyping technologies and concepts in population genetics and genetic epidemiology, gene expression data analysis.

Unit 2. Linkage Analysis

Principles of linkage analysis for identifying genetic loci associated with Mendelian traits
Genetic mapping methods, including LOD score analysis and parametric linkage analysis
and software tools for linkage analysis, such as Merlin and Allegro.

Unit 3. Association Mapping

Principles of association mapping for identifying genetic variants associated with complex traits
Single-marker association tests (e.g., chi-square test, Fisher's exact test)
Genome-wide association studies (GWAS) and study designs and Software tools for association mapping, such as PLINK and SNPTEST.

Unit 4. Family-Based Studies and Population Genetics

Family-based association tests (FBAT), Transmission disequilibrium test (TDT)
Haplotype-based association tests, Hardy-Weinberg equilibrium and tests of genetic equilibrium, F-statistics and measures of population structure and principal component analysis (PCA) for population stratification.

Unit 5. Meta-analysis in Genetic Studies

Principles of meta-analysis in genetic association studies, Fixed-effects and random-effects meta-analysis models, software tools for meta-analysis, such as METAL and GWAMA.

Unit 6. Advanced Topics in Statistical Genetics

Epistasis and gene-gene interactions, Polygenic risk scores and genetic risk prediction and Mendelian randomization and causal inference in genetics.

Recommended Books/ Reference Material

1. Gibson, G. (2009). *Statistical genetics: gene mapping through linkage and association*. Genes, Brain and Behavior, 8(1), 127-128.
2. Laird, N. M., & Lange, C. (2010). *The fundamentals of modern statistical genetics*. Springer Science & Business Media.
3. Lin, S., & Zhao, H. (2010). *Handbook on analyzing human genetic data computational approaches and software*. Berlin Heidelberg: Springer-Verlag.
4. McLachlan, G., Do, K. A., & Ambroise, C. (2005). *Analyzing microarray gene expression data (Vol. 422)*. John Wiley & Sons.
5. Mills, M. C., Barban, N., & Tropf, F. C. (2020). *An introduction to statistical genetic data analysis*. United Kingdom: MIT Press.
6. Mukhopadhyay, I., & Majumder, P. P. (2023). *Statistical methods in human genetics*. Singapore: Springer Nature.
7. Reilly, C. (2009). *Statistics in human genetics and molecular biology*. Chapman and Hall/CRC.
8. Wu, R., Ma, C., & Casella, G. (2007). *Statistical genetics of quantitative traits: linkage, maps and QTL*. Springer Science & Business Media.

Environmental Statistics

M. Phil - Statistics

Course Code: STA-719

Credit Hours: 03

Pre-requisite: Nil

Objectives

The course is designed to introduce the basic statistical methods necessary to conduct and understand statistical analysis of environmental issues and problems. To understand measurement, descriptive statistics, graphs, basic probability, correlation and regression. To have knowledge of inferential statistics (hypothesis testing, confidence interval construction, effect size calculation).

Learning Outcomes

- Student must be able to apply statistical methods to important problems in environmental sciences, with a focus on understanding and quantifying change in environmental sciences or problems of this nature.
- Student will have the understanding of the conceptual underpinnings of statistics and the key distinctions among statistical methods commonly used in ecology and conservation.
- They will become aware of a wide range of applications of statistics in environmental management, life sciences & decision making.

Course Outline

Unit 1. Collection Environmental Data, Sampling and Monitoring

The role of statistics in environmental science. Review of statistical techniques applied in environmental science. Probabilistic sampling scheme, simple random sampling, ratios and ratio estimators, developments of survey sampling and inaccessible and sensitive data. Statistical monitoring methods for environmental system.

Unit 2. Examining environmental effects, Stimulus response relationships

Regression type models and methods, Simple and generalized transformations, Quantal response and Repeated measures. censored data, change point analysis. Population Dynamics: One species - exponential, logistic and Gompertz models. Two species - competition, coexistence, predator - prey oscillation, Lotka - Volterra equations, isoclines. Leslie matrix model for age structured populations. Survivorship curves - constant hazard rate, monotone hazard rate and bath-tub shaped hazard rates.

Unit 3. Time Series Methods

Space and time effects, descriptive methods, trends and stationary processes, time domain models and methods, frequency domain models and methods. Auto-covariances and spectral analysis, statistical methods for non-stationarity, time series modeling and forecasting, autoregressive moving average (ARMA) processes.

Unit 4. Spatial methods for Environmental Processes

Spatial point process models and methods, general spatial process, spatial sampling and spatial design, spatial temporal models and methods.

Unit 5. Population density, Ecological diversity, Game theory in ecology and Foraging Theory.

Population density estimation: capture- recapture models, nearest neighbor models, line transect sampling. Ecological diversity: Simpson's index, Shannon – Weaver index, diversity as average rarity. Optimal harvesting of natural resources, maximum sustainable yield, tragedy of the commons. Game theory in ecology: concept of evolutionary stable strategy, its properties, simple cases such as Hawk-Dove game. Foraging theory: diet choice problem, patch choice problem, mean variance tradeoff.

Recommended Books/Reference Material

1. Barnett, V. (2005). *Environmental statistics: methods and applications*. John Wiley & Sons.
2. Chandler, R., & Scott, M. (2011). *Statistical methods for trend detection and analysis in the environmental sciences*. John Wiley & Sons.
3. El-Shaarawi, A. H., & Piegorisch, W. W. (2010). *Encyclopedia of environmetrics (2nd ed.)*. John Wiley & Sons.
4. Gelfand, A. E., & Diggle, P. J. (2019). *Handbook of environmental and ecological statistics*. Chapman and Hall/CRC.
5. Manly, B. F. (2008). *Statistics for environmental science and management*. CRC Press.
6. Millard, S. P. (2013). *EnvStats, an R package for environmental statistics*. John Wiley & Sons.
7. Ott, W. R. (1994). *Environmental statistics and data analysis*. Chapman and Hall/CRC.
8. Shaefer, S. J., & Theodore, L. (2007). *Probability and statistics applications for environmental science*. Chapman and Hall/CRC.
9. Shekhar, S., & Chawla, S. (2022). *Spatial databases: a tour*. Springer.
10. Smith, J. (2023). *Spatial data analysis: methods and applications (2nd ed.)*. Academic Press.
11. Woodward, W.A., Sadler, B.P., & Robertson, S. (2022). *Time series for data science: Analysis and forecasting*. Chapman and Hall/CRC.
12. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer.

Regression Models for Count Data

M. Phil -Statistics

Course Code: STA-720

Credit Hours: 03

Pre-requisite: Nil

Objectives

The course aims to understand the count data exclusively other than categorical data and to learn the characteristics and existence form of count data in different fields. The course is designed to enhance student's skills in comprehension and evaluation of statistical methods for count data and to make them able to apply the various discrete and extended discrete probability distributions in real life count data.

Learning Outcomes

- Students will be able to explore various types of count data distributions and learn how to model them effectively.
- They will be able to apply advanced regression models to analyze real-world count data and derive meaningful insights.
- Students will be able to critically evaluate different modeling approaches and select the most appropriate one for a given dataset.

Course Outline

Unit 1. Introduction to Count Data Analysis

Overview of count data and its characteristics, common applications of count data analysis. Introduction to poisson regression and its limitations.

Unit 2. Poisson Regression

Theoretical background of poisson regression. Modeling count data with poisson regression. Inference and interpretation of Poisson regression coefficients, over dispersion and its consequences.

Unit 3. Negative Binomial Regression

Introduction to negative binomial distribution. Negative binomial regression as a remedy for over dispersion. Model estimation and interpretation. Comparison with Poisson regression

Unit 4. Zero-Inflated Models

Introduction to zero-inflated count data, Zero-inflated Poisson (ZIP) model. Zero-inflated negative binomial (ZINB) model, applications and interpretation of zero-inflated models.

Unit 5. Hurdle Models

Introduction to hurdle models for count data, two-component model structure, hurdle Poisson and hurdle negative binomial models and advantages and limitations of hurdle models.

Unit 6. Advanced Topics in Count Data Analysis

Quasi-Poisson regression for over dispersed count data, generalized linear mixed models (GLMMs) for hierarchical count data and Bayesian approaches to count data modeling.

Recommended Books/ Reference Material

1. Chen, D., & Chen, J. K. (2021). *Statistical regression modeling with R: longitudinal and multi-level modeling*. Germany: Springer International Publishing.
2. Colin, A. C., & Trivedi, K. P. (2010). *Micro-Econometrics Using Stata (2nd ed.)*. USA: Stata Press.
3. Colin, A. C., & Trivedi, K. P. (2013). *Regression analysis of count data (2nd ed.)*. UK: Cambridge University Press.
4. Haas, M. G. (2021). *Regression methods for the analysis of count data. generalised linear models for limited dependent variables*. Germany: GRIN Verlag.

5. Hilbe, M. J. (2011). *Negative binomial regression (2nd ed.)*. UK: Cambridge University Press.
6. Winkelmann, R. (2010). *Econometric analysis of count data*. Berlin Heidelberg: Springer Verlag.

Applied Functional Data Analysis

M. Phil -Statistics

Course Code: STA-721

Credit Hours: 03

Pre-requisite: Nil

Objectives

The objectives of applied functional data analysis (FDA) typically encompass both theoretical understanding and practical application and, introduce the students to functional data analysis methods applied to a wide array of application areas. The aim is to illustrate common numerical and estimation routines to perform functional data analysis. The course enables the students to apply functional data analysis techniques to real life problems.

Learning Outcomes

- Students will be able to recognize and work with functional data structures, which represent data as smooth curves or functions rather than discrete points. Students will learn various techniques for preprocessing functional data, such as smoothing noisy curves, interpolating missing values, and aligning curves to a common reference frame.
- Students will learn regression techniques tailored to functional data, allowing them to model relationships between covariates and functional responses. Moreover, they will learn the methods that goes beyond standard multivariate statistical methodology through dimension free approach that accounts for the order in observed variables.
- Students will learn to apply FPCA (Functional Principal Component Analysis) to extract meaningful features and reduce the complexity of functional datasets as FPCA is a key technique in FDA that allows for dimensionality reduction and visualization of high-dimensional functional data.
- Students will be capable to identify the most suitable approach for the problem at hand, and critically discuss the results of analysis obtained by a particular method.

Course Outline

Unit 1. First steps in the Analysis of Functional Data

What are functional data. Multivariate functional data. Functional models for non-functional data. Functional data analysis. Functional linear models.

Unit 2. Analysis of Noisy Data using R Language

Classes and objects and a quick comparison in R. Fourier series for periodic data and functions. Spline series for nonperiodic data and functions. Constant, monomial and other basis. Methods for functional data objects. Smoothing using regression analysis. Bivariate functional data objects.

Unit 3. Mathematical Framework for Functional Data

Incorporating information about derivatives. Random curve-to-curve variation, amplitude, phase variation. Convert functions into basic expansions. The bivariate covariance function. The residual variance covariance matrix. Phase-plane plots of periodic effects. Confidence intervals for curves and their derivatives.

Unit 4. Functional Principal and Canonical Components Analysis

Functional PCA of log precipitation data and residuals. PCA of joint variation. Exploring functional covariance with canonical components analysis. Aligning features for samples of curves.

Unit 5. Linear Models for Scalar and Functional Responses

A scalar response model for log annual precipitation. Estimates of the regression coefficient predicting annual precipitation. Functional responses and analysis of variance model. Functional responses with functional predictors. Permutation tests of functional hypothesis. Some case studies.

Recommended Books/Reference Material

1. Aubin, J. P. (2011). *Applied functional analysis*. John Wiley & Sons.
2. Balakrishnan, A. V. (2012). *Applied functional analysis: A* (Vol. 3). Springer Science & Business Media.
3. Cleophas, T. J., & Zwinderman, A. H. (2021). *Regression analysis in medical research: for starters and 2nd Levelers*. Springer.
4. Febrero-Bande, M., & de la Fuente, M. O. (2012). Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*, 51(4), 1–28. <https://doi.org/10.18637/jss.v051.i04>.
5. Horvath, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.
6. Hsing, T., & Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.
7. Kokoszka, P., & Reimherr, M. (2017). *Introduction to functional data analysis*. Chapman and Hall/CRC.
8. Ramsay, J.O., & Silverman, B. W. (2005). *Functional data analysis*. Springer.
9. Ramsay, J.O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer.
10. Zeidler, E. (2012). *Applied functional analysis: main principles and their applications* (Vol. 109). Springer Science & Business Media.

Classification and Random Forest Technique

M. Phil -Statistics

Course Code: STA-723

Credit Hours: 03

Pre-requisite: Nil

Objectives

Objective of this course is Understand the concept of classification and how it is used to categorize data into different classes or groups based on attributes or features. Knowledge of different types of

classification algorithms, such as decision trees, logistic regression, support vector machines, and random forests.

Learning Outcomes

- Students will have familiarity with the random forest technique and ensemble learning.
- Students will have ability to implement and tune random forest models using programming languages such as Python or R.
- Students will be able to apply classification and random forest techniques to real-world datasets and problems in various industries, such as finance, healthcare, marketing, and e-commerce.

Course Outlines

Unit 1. Types of Learning Algorithms and Basic Terminologies

Supervised learning algorithm, unsupervised learning algorithm, semi-supervised learning. basic terminology/concepts like tree-structured models, decision node, end node, root node, child node, splitting, pruning, over fitting, training and testing data set etc.

Unit 2. Trees and related Methods

History and definition, how to learn a decision tree (explanation by simple example), splitting criteria, defining decision rules, interpreting the results, commonly used decision trees, strength and weakness of a decision tree.

Unit 3. Random Forest

Introduction to Random Forest, difference between classification and regression tree, selection of attributes, concept of impurity, different impurity measures: Gini's coefficient, Entropy etc. stopping criterion, greedy algorithm of CART, problems associated with CART: over fitting, Bias etc. Use of python or R language to draw a CART.

Unit 4. Ensemble Learning

Different ensembles: BOOSTING, BAGGING, RANDOM FOREST, randomness involves in Random Forest, Algorithm of Random Forest, number of trees in Random Forest, Out-of-bag error in Random Forest, Use of R package for Random Forest.

Unit 5. Support Vector Machines

Support Vector Machines for regression and classification, support vector machines and kernels, SVMS and curse of dimensionality, SVM as penalized Method.

Recommended Books/ Reference Material

1. Breiman, L. (1984). *Classification and regression trees*. New York: Routledge.
2. Brownlee, J. (2021). *Ensemble learning algorithms with Python: make better predictions with bagging, boosting, and stacking*. Machine Learning Mastery.

3. Chakraborty, S., Islam, S. H., & Samanta, D. (2022). *Data classification and incremental clustering in data mining and machine learning*. Switzerland: Springer International Publishing.
4. Rokach, L. (2009). *Pattern classification using ensemble methods*. Singapore: World Scientific Publishing Company.
5. Smith, C., & Koning, M. (2017). *Decision trees and random forests: a visual introduction for beginners*. South Africa. Penguin random house.
6. Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer Science & Business Media.

Statistical Methods for Clinical Trials

M. Phil - Statistics

Course Code: STA-725

Credit Hours: 03

Pre-requisite: Nil

Objectives

The aim of the course to enhance the students' awareness and informed usage of modern methods in the design and analysis of randomized control trials. The course outline is designed to improve statistical thinking as applied to clinical research and to provide a foundation for research in statistical methods for clinical trials.

Learning Outcomes

- Students shall be able to plan and apply various clinical study designs.
- Students shall be able to recognize a research objective that would meet through a clinical trial.
- Students shall be able to discuss the relative contributions of clinical judgment and clinical trials in evaluating new medical therapies.
- Students shall be able to apply various characteristics of statistical reasoning to a research objective in a clinical trial setting.

Course Outline

Unit 1. Introduction to Clinical Trials

Introduction to clinical trials, types and aspects of clinical trials; definition, phases and protocol.

Unit 2. Randomization

Design of Clinical Studies; Randomized controlled design, crossover design, cluster randomization, equivalence trial, and large sample trial. Randomization; fixed randomization and adaptive randomization.

Unit 3. Censored Data and Interim Analysis

Introduction Review of Methods of analysis; Randomization tests, stratified analysis, survival analysis. Surrogate endpoints, surrogate versus clinical endpoints and validation of surrogate end points.

Unit 4. Sample Size Determination

Statistical planning; sample size determination. Equivalency testing. Testing for the similarity of treatments. Multiple testing. Multiple comparisons, subgroup analysis, multiple endpoints and covariate adjustment.

Unit 5. Issues in Efficacy Evaluation

Statistical Monitoring; methods of repeated testing of hypotheses over time: Noncompliance / Departure from intended treatment. Intent to treat principle, and efficacy analysis.

Recommended Books/Reference Material

1. Friedman L.M., Furberg C., & DeMets D.L. (2010). *Fundamentals of clinical trials*. New York: Springer Verlag.
2. Rosenberg, W.F., & Lachin, J.M. (2002). *Randomization in clinical trials: theory and practice*. John Wiley & Sons.
3. C. Jennison, & B. W. Turnbull (1999). *Group sequential methods with applications to clinical trials*. Chapman and Hall/CRC.
4. Chow S.C., & Liu J.P. (2003). *Design and analysis of clinical trials: concepts and methodologies (2nd ed.)*. John Wiley & Sons.
5. E. Marubeni, & M. G. Valsecchi (1994). *Analyzing survival data from clinical trials and observational studies*. John Wiley & Sons.
6. Friedman, L. M., Furberg, C. D., & DeMets, D. L. (2022). *Fundamentals of clinical trials*. Springer.
7. Julious, S. A. (2023). *Sample sizes for clinical trials*. United States: Chapman and Hall/CRC.

Mathematical Demography

M. Phil - Statistics

Course Code: STA-726

Credit Hours: 03

Pre-requisite: Nil

Objectives

The objective of the mathematical demography course is to provide students with a comprehensive understanding of demographic concepts and mathematical tools necessary for analysis, forecasting, and policy making.

Learning Outcomes

- Student will be able to apply mathematical and statistical methods to analyze demographic data, including population growth rates, age-specific fertility rates, and mortality rates.
- Student will be able to construct and interpret mathematical models of population dynamics, such as the Leslie matrix model and cohort-component models.
- They will be able to apply mathematical demography principles and methods to address real-world challenges in areas such as public health, environmental sustainability, and social policy and demonstrate proficiency in using software tools R or Python for demographic data analysis and modeling.

Course Outline

Unit 1. Review and Introduction

Overview of demography, importance of mathematical modeling in demography, population size and composition, birth, death, and migration rates, life expectancy and survivorship. Development and application of Lotka Integral Equation.

Unit 2. Population Growth Models and Smoothing of data

Exponential growth mode, logistic growth model, applications and limitations, smoothing of age data by various methods, estimations of age at first marriage, child mortality, adult mortality, and fertility, construction of abridged life tables and decrement life tables. Lagrange estimates.

Unit 3. Migration and Path Analysis

Types of migration, migration rates and patterns, net migration, age-structured models, Leslie matrix model, cohort-component model, applications in population projection. Projections by application of straight line, logistic, exponential, Gompertz and polynomial curves and by component method. Path analysis for decomposition of effect of factors affecting and dependent variable.

Unit 4. Population Policies and Interventions

Family planning programs, immigration policies, aging population challenges, applications of mathematical demography, urban planning and development, health policy and epidemiology, environmental sustainability. Path analysis for decomposition of effect of factors affecting and dependent variable.

Unit 5. Case Studies and Practical Exercises

Analyzing demographic data sets, building and interpreting demographic models, policy analysis exercises, future trends in population dynamics, population aging, urbanization, global population projections.

Recommended Books/Reference Material

1. Boslaugh, S., & Yu, X. (2019). *Demography and health data: a practical guide*. Springer.
2. Coale, A., & Demeny P. (1966). *Regional Model Life Tables and Stable Populations*. Princeton University Press.
3. Garchitorena, A. (2020). *Demography in R*. Independently published.
4. Goerres, A., & Vanhuyse, P. (2021). *Global political demography: The politics of population change*. Palgrave Macmillan.
5. Grow, A., & Bavel, J.V. (2016). *Agent-based modelling in population studies: Concepts, methods, and applications*. Springer International Publishing.
6. Keilman, N., & Kuijsten, A. C. (2021). *Population aging and the generational economy: a global perspective*. Springer.
7. Keyfitz, N., & Caswell, H. (2005). *Applied mathematical demography*. Springer.
8. Nathan, K., & Cawell, H. (2004). *Applied mathematical demography*. Springer Verlag.
9. Preston, S. H., Patrick, H., & Michel, G. (2001). *Demography: measuring and modeling population process*. Oxford: Blackwell Publishers.
10. Rao, D. S. P. (2019). *Demographic research with R*. Chapman and Hall/CRC.
11. Rogers, A. (1995). *Multiregional demography: principles, methods and extensions*. John Wiley & Sons.
12. Rogers, A., & Little, J. (2022). *Population and society: an introduction to demography*. Cambridge University Press.

13. Schoen, R. (2020). *Analyzing contemporary fertility*. Springer International Publishing.
14. Wachter, K. W. (2014). *Essential demography methods*. Harvard University Press.

Bayesian Statistical Techniques

M. Phil - Statistics

Course Code: STA-727

Credit Hours: 03

Pre-requisite: Nil

Objectives

The aim of the course is to provide a comprehensive overview of Bayesian statistical techniques, covering fundamental concepts, computational methods, advanced applications, and ethical considerations. Each module incorporates theoretical foundations, practical implementations, and real-world examples to ensure a well-rounded understanding of Bayesian inference. Additionally, the course emphasizes hands-on learning experiences and encourages students to explore cutting-edge developments and interdisciplinary applications in Bayesian statistics.

Learning Outcomes

- Students shall be able to develop understanding of Bayesian principles: including the concepts of prior, likelihood, posterior, and Bayes' theorem.
- Students shall be able to apply Bayesian inference methods to solve a wide range of statistical problems, including parameter estimation, hypothesis testing, and model selection.
- Students will develop proficiency in various Bayesian computation techniques, including Markov Chain Monte Carlo (MCMC) methods, variational inference, and Bayesian model assessment and selection.
- Students will become proficient in conducting Bayesian linear regression analysis, including simple linear regression, multiple linear regression, and hierarchical modeling.
- Students will be introduced to advanced Bayesian techniques such as Bayesian non parametric, Gaussian processes, and Bayesian neural networks, and understand their applications in contemporary research.

Course Outline

Unit 1. Introduction to Bayesian Statistics

Overview of Bayesian approach. Key differences between Bayesian and frequentist approaches. Historical context and development of Bayesian methods

Unit 2. Probability Theory Review

Basics of probability theory. Conditional probability and Bayes' theorem. Prior, likelihood, and posterior distributions. Joint, marginal, and conditional distributions. Use of unrelated question in randomized response model to estimate unknown parameter.

Unit 3. Bayesian Inference

Bayesian parameter estimation. Maximum a posteriori (MAP) estimation. Bayesian credible intervals. Bayesian hypothesis testing. Bayes factors. Posterior predictive p-values. Comparison with classical frequentist.

Unit 4. Bayesian Linear Regression

Simple linear regression in a Bayesian framework. Multiple linear regression and model selection. Bayesian hierarchical models.

Unit 5. Bayesian Computation

Markov Chain Monte Carlo (MCMC) methods. Metropolis-Hastings algorithm. Gibbs sampling. Hamiltonian Monte Carlo (HMC). Variational inference. Comparison of different computational methods.

Unit 6. Bayesian Model Assessment and Selection

Model checking and diagnostics. Model comparison criteria (e.g., Deviance information criterion, Bayes factor). Sensitivity analysis.

Unit 7. Advanced Bayesian Techniques

Bayesian non parametric. Gaussian processes. Bayesian neural networks.

Recommended Books/Reference Material

1. Albert, J. (2009). *Bayesian computation with R*. Springer.
2. Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer.
3. Bolstad, W. M., & Curran, J. M. (2016). *Introduction to bayesian statistics*. John Wiley & Sons.
4. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Chapman and Hall/CRC.
5. Gelman, A. (2014). *Bayesian data analysis* (3rd ed.). United Kingdom: Taylor & Francis.
6. Kaplan, D. (2023). *Bayesian statistics for the social sciences*. United States: Guilford Publications.
7. Slater, M. (2021). *Bayesian methods in statistics: from Concepts to Practice*. United Kingdom: SAGE Publications.

Mixture Distributions

M. Phil - Statistics

Course Code: STA-728

Credit Hours: 03

Pre-requisite: Nil

Objectives

This course provides a comprehensive examination of mixture distributions, a powerful class of probability distributions that model complex data arising from multiple underlying sources or populations. This course covers foundational concepts, different types of mixture distributions, estimation techniques, model selection, and applications in real-world problems.

Learning Outcomes

- Students shall be able to understand the fundamental principles of mixture distributions.
- Students will be able to explore different types of mixture distributions and their properties.
- Students will be able to apply mixture distributions to real-world problems in various domains.

Course Outline

Unit 1. Introduction to Mixture Distributions

Overview of mixture distributions. Motivation and applications. Examples of data that can be modeled using mixture distributions.

Unit 2. Gaussian Mixture Distribution (GMD)

Definition and properties of GMDs. Parameterization and representation. Expectation-Maximization (EM) algorithm for parameter estimation. Model selection using criteria such as AIC and BIC.

Unit 3. Other Types of Mixture Distributions

Multinomial mixture distributions for categorical data. Poisson mixture distributions for count data. Exponential mixture distributions for survival data. Dirichlet process mixture distributions for nonparametric modeling.

Unit 4. Estimation Techniques

Expectation-Maximization (EM) algorithm and its variants. Maximum likelihood estimation (MLE) and Bayesian estimation. Computational challenges and strategies for efficient estimation.

Unit 5. Model Selection and Evaluation

Information criteria (AIC, BIC) for model selection. Cross-validation techniques. Goodness-of-fit tests and diagnostic tools.

Unit 6. Applications of Mixture Models

Clustering and classification tasks. Density estimation and anomaly detection. Survival analysis and reliability modeling. Latent variable modeling in psychology and social sciences.

Unit 7. Advanced Topics in Mixture Models

Latent dirichlet allocation (LDA) for topic modeling. Hierarchical mixture models. Dynamic mixture models for streaming data. Mixture models with structured dependencies.

Recommended Books/Reference Material

1. Chen, J. (2023). *Statistical inference under mixture models*. Germany: Springer Nature.
2. Everitt, B. (2013). *Finite mixture distributions*. Springer.
3. McLachlan, G. J., Peel, D. (2004). *Finite mixture Models*. John Wiley & Sons.
4. Nguyen, T. T. (2021). *Model selection and approximation in high-dimensional mixtures of experts models: from theory to practice*. France: Doctoral dissertation, Normandie.
5. Titterton, D. M., Smith, A. F. M., Smith, A. F. M. & Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons.

Recent Developments in Statistics

M. Phil - Statistics

Course Code: STA- 729

Credit Hours: 03

Pre-requisite: Nil

This course will delve into specific subjects of contemporary research significance published in academic journals. Students will be required to deliver presentations or seminars based on published articles, as well as compose reports.

Applied Multivariate Statistics

MPhil-Statistics

Course Code: STA-733

Credit Hours: 03

Pre-requisite: Nil

Objectives:

This course provides an introduction to multivariate analysis, focusing on key concepts, methods, and applications. It covers matrix algebra, multivariate probability distributions, and various techniques for analyzing multivariate data, including principal component analysis, factor analysis, discriminant analysis, and cluster analysis. The course also explores advanced topics such as structural equation modeling, multivariate time series analysis, and robust multivariate methods.

Learning Outcomes

- Students will be able to explain and utilize key mathematical tools including matrix algebra and multivariate probability distributions and to formulate and solve multivariate statistical problems.
- Students will be able to apply techniques such as principal component analysis, factor analysis, discriminant analysis, and cluster analysis using appropriate statistical software, and interpret the results in a real-world context.
- Students will be able to critically assess and apply advanced methods such as structural equation modeling, multivariate time series analysis, and robust multivariate techniques to complex datasets and research problems.

Course Outlines

Unit 1. Introduction to Multivariate Analysis

Overview, importance, and key concepts in multivariate data analysis. Basic matrix operations, inverse matrices, eigenvalues, and eigenvectors. Understanding the multivariate normal distribution, covariance, and correlation matrices. Computational work using R/Python.

Unit 2. Principal Components Analysis and Factor Analysis

Techniques for dimensionality reduction, calculating principal components, and interpreting variance explained. Methods for factor extraction, factor rotation, and differences between confirmatory and exploratory factor analysis. Computational work using R/Python.

Unit 3. Canonical Correlation Analysis and Discriminant Analysis

Techniques for analyzing relationships between two variable sets, understanding canonical variables, and interpreting results. Introduction to linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA), including classifier performance evaluation. Computational work using R/Python.

Unit 4. Factor Analysis and Cluster Analysis

Methods for factor extraction, factor rotation, and differences between confirmatory and exploratory factor analysis. Methods such as K-means clustering and hierarchical clustering, and techniques for evaluating clustering solutions. Computational work using R/Python.

Unit 5. Multidimensional Scaling and Correspondence Analysis

Approaches to dimensionality reduction and stress evaluation in MDS. Techniques for analyzing categorical data, creating correspondence plots, and applications of correspondence analysis. Computational work using R/Python.

Unit 6. Structural Equation Modeling and Advanced Regression Techniques

Concepts of path analysis, measurement models, and goodness-of-fit indices. Handling multicollinearity, model selection strategies, and diagnostics for multivariate regression. Computational work using R/Python.

Recommended Books/ Reference Material

9. Afifi, A.A., & Clark, V. (1984). *Computer aided multivariate analysis*. California: Lifetime learning publications.
10. Afifi, A.A., May, S., Donatello, R. A., & Clark, V.A. (2021). *Practical multivariate analysis*. John Wiley & Sons.
11. Anderson, T.W. (2003). *An introduction to multivariate statistical analysis*. John Wiley & Sons.
12. Backhaus, K., Erichson, B., Gensler, S., Weiber, R., & Weiber, T. (2021). *Multivariate analysis*. Springer Books.
13. Barbara, M. B. (2020). *Multivariate analysis: Techniques for educational and psychological research*. Routledge.
14. Cleff, T. (2019). *Applied statistics and multivariate data analysis for business and economics*. Switzerland: Springer International Publishing.
15. Denis, D.J. (2020). *Univariate, bivariate, and multivariate statistics using R: Quantitative tools for data analysis and data science*. John Wiley & Sons.
16. Johnson, R.A. & Wichern, D.W. (2022). *Applied multivariate statistical analysis* (7th ed.). New York: Prentice Hall.
17. Morrison, F. (1990). *Multivariate statistical methods*. New York: McGraw-Hill.

Applied Research Methodology in Statistics

Course Code: STA-734

Credit Hours: 03

Pre-requisite: Nil

Objectives

This course develops advanced skills in statistical research, enabling students to formulate research problems and hypotheses, design experimental and survey-based studies, construct valid questionnaires, and apply appropriate sampling and statistical modeling techniques. It emphasizes causal inference, research ethics, and plagiarism prevention, while preparing students to write professional theses using LaTeX and MS Word and to publish research following ethical standards and impact metrics.

Learning Outcomes

- Formulate statistically sound and researchable problems grounded in theory.
- Design valid experimental, observational, and survey-based studies.
- Develop reliable and valid questionnaires for data collection.
- Determine appropriate sampling strategies and compute justified sample sizes.
- Apply and interpret advanced statistical models for different types of data.
- Address bias, confounding, and causal relationships in research.
- Prepare professionally formatted research proposals and MPhil theses using Latex and MS Word.
- Understand and apply principles of research ethics, academic integrity, and plagiarism prevention.
- Evaluate journal quality using impact metrics and avoid predatory journals.
- Submit manuscripts to indexed journals following ethical and scholarly standards.

Course Outlines

Unit 1: Foundations of Research Design

Nature and philosophy of scientific research; quantitative, qualitative and mixed methods approaches; identification of research problems; formulation of objectives and hypotheses; literature review strategies; conceptual and theoretical frameworks; operational definitions of variables; validity and reliability concepts; replication and reproducibility in statistical research.

Unit 2: Survey Design, Questionnaire Development and Sampling Techniques

Types of surveys (cross-sectional, longitudinal, panel, cohort); primary and secondary data sources; questionnaire design principles; scaling techniques (Likert, semantic differential, ranking, rating scales); pilot testing; measurement errors; reliability and validity of instruments; probability and non-probability sampling; stratified, cluster, multistage and systematic sampling; sample size determination methods; power analysis; non-response bias and survey quality control.

Unit 3: Experimental Design and Advanced Statistical Modeling

Principles of experimental design; randomized controlled trials; completely randomized design, randomized block design and factorial designs; quasi-experimental designs; model selection strategies; regression models; generalized linear models; count data models; multilevel and mixed models; diagnostics and model validation; introduction to causal inference and causal machine learning concepts.

Unit 4: Plagiarism, Research Ethics and Academic Integrity

Principles of research ethics; ethical approval processes; informed consent; confidentiality and data protection; responsible data management; plagiarism types and detection; similarity index interpretation; citation styles and referencing systems; authorship ethics; research misconduct; replication crisis; open science practices.

Unit 5: Thesis and Proposal Writing (LaTeX and MS Word)

Structure of research proposals and theses; writing problem statements and methodology chapters; presenting statistical results; tables, figures and referencing; formatting standards; use of LaTeX for thesis preparation; document classes and packages; bibliography management (BibTeX); formatting in MS Word; citation management software; thesis defense preparation.

Unit 6: Journal Publication and Research Dissemination

Journal selection strategies; peer review process; manuscript preparation; responding to reviewers; impact factor, h-index and citation metrics; indexing databases; predatory journals; open access publishing; research visibility; conference publications; grant writing basics.

Recommended Books/ Reference Material

1. Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
2. Lohr, S. L. (2021). *Sampling: Design and analysis* (3rd ed.). Chapman and Hall/CRC.
3. Fox, J., & Weisberg, S. (2019). *Applied regression analysis and generalized linear models* (3rd ed.). Sage Publications.
4. Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hall/CRC.
5. Kothari, C. R. (2022). *Research methodology: Methods and techniques* (4th ed.). New Age International.
6. Mittelbach, F., & Goossens, M. (2004). *The LaTeX companion* (2nd ed.). Addison-Wesley.
7. Day, R. A., & Gastel, B. (2016). *How to write and publish a scientific paper* (8th ed.). Cambridge University Press.

Research Design and Biostatistics

Course Code: STA-735

Credit Hours: 02

Pre-requisite: Nil

Objectives

This course is designed to help students develop the skills needed to plan and conduct high-quality research in Zoology. It guides learners in identifying meaningful research problems, setting clear objectives, and developing testable hypotheses. Students learn how to choose suitable experimental designs and sampling methods, and how to apply bio statistical tools to analyse and interpret biological data. The course also introduces the use of regression and count data models in zoological research. In addition, it strengthens students' ability to critically review scientific literature and to clearly and effectively communicate their research findings.

Learning Outcomes

By the end of this course, the students will be able to:

- Understand and apply advanced research methodologies, including experimental design, Statistical analysis, and data interpretation in Zoological studies.
- Develop and execute independent research projects, formulating hypotheses, conducting Literature reviews, and employing appropriate methodologies.
- Critically evaluate research papers and methodologies, and effectively communicate research findings through scientific writing and presentations

Course Outline

Unit 1. Fundamentals of Research Design

Introduction to Research Design, Nature and importance of research in Zoology. Scientific method and research process. Types of research: experimental, observational, and analytical. Research ethics, biosafety, and plagiarism. Identification of research problems. Research objectives and questions

Unit 2. Literature Review, Experimental Design & Sampling

Null and alternative hypotheses, Literature search techniques, Reference management tools. Principles of experimental design: control, randomization, replication, CRD, RBD, and factorial designs. Sampling techniques, Sample size determination and bias.

Unit 3. Data Types, Data Management & Descriptive Biostatistics

Types of data: qualitative, quantitative, count data. Measurement scale Data coding, cleaning, and validation. Measures of central tendency and dispersion. Frequency distributions and graphical presentation

Unit 4. Inferential Biostatistics

Probability concepts and distributions. Estimation and confidence intervals. Hypothesis testing framework. t-test, Chi-square test, ANOVA. Non-parametric tests (Mann–Whitney, Kruskal–Wallis)

Unit 5. Regression and Modeling in Zoology

Correlation analysis. Simple and multiple linear regression. Logistic regression (introduction). Introduction to Generalized Linear Models (GLM). Biological interpretation of regression models.

Unit 6. Count Data Models & Scientific Writing

Nature of count data in Zoology (parasite counts, species abundance, disease cases) Poisson regression. Negative Binomial regression. Structure of research papers and thesis. Proposal writing and critical review of scientific literature.

Recommended Books/Reference Material

1. Quinn, G. P., & Keough, M. J. (2023). Experimental design and data analysis for biologists (2nd ed.). Cambridge University Press.
2. Triola, M. M., Triola, M. F., & Roy, J. (2017). Biostatistics for the biological and health sciences (2nd ed.). Pearson.
3. Zuur, A. F., Ieno, E. N., & Smith, G. M. (2021). Analysing ecological data (2nd ed.). Springer.
4. Cameron, A. C., & Trivedi, P. K. (2022). Regression analysis of count data (2nd ed.). Cambridge University Press.
5. Field, A. (2022). Discovering statistics using IBM SPSS statistics (6th ed.). SAGE Publications.
6. Day, R. A., & Gastel, B. (2020). How to write and publish a scientific paper (8th ed.). Cambridge University Press.

Research Design and Biostatistics (LAB)

Course Code: STA-735

Credit Hours: 01

Pre-requisite: Nil

Objectives

Students will learn to manage, clean, and analyse zoological data using SPSS/R, selecting appropriate statistical methods for different types of biological data. They will apply descriptive and inferential statistics, including regression and count data models, and interpret results in their biological and ecological context. They will also develop skills to present and report statistical findings in a professional format suitable for research papers and M.Phil. Thesis.

Learning Outcomes

- Students will be able to efficiently manage, clean, and analyse zoological data using SPSS/R.

- Students will be able to select and apply appropriate statistical techniques, including regression and count data models, for different types of biological data.
- Students will be able to interpret statistical results in biological and ecological contexts and present them in a professional format suitable for research and thesis writing.

Course Outline

- Use of SPSS / R.
- Data entry and management.
- Descriptive and inferential statistics.
- Regression and count data analysis.
- Experimental designs. Non Parametric tests.
- Mini research proposal preparation

Recommended Books/Reference Material

1. Quinn, G. P., & Keough, M. J. (2023). Experimental design and data analysis for biologists (2nd ed.). Cambridge University Press.
2. Triola, M. M., Triola, M. F., & Roy, J. (2017). Biostatistics for the biological and health sciences (2nd ed.). Pearson.
3. Zuur, A. F., Ieno, E. N., & Smith, G. M. (2021). Analysing ecological data (2nd ed.). Springer.
4. Cameron, A. C., & Trivedi, P. K. (2022). Regression analysis of count data (2nd ed.). Cambridge University Press.
5. Field, A. (2022). Discovering statistics using IBM SPSS statistics (6th ed.). SAGE Publications.
6. Day, R. A., & Gastel, B. (2020). How to write and publish a scientific paper (8th ed.). Cambridge University Press.