

Central Department of Statistics  
Institute of Science and Technology  
Tribhuvan University, Kathmandu, Nepal

**Revised Course Structure with Syllabus**  
of  
**M. Sc. / M. A. in Statistics (Semester System)**

**January, 2024**

**Submitted to:**

Dean's Office  
Institute of Science and Technology  
Tribhuvan University  
2024

# Contents

1. Introduction	1
1.1 Enrolment Quota	1
1.2 Eligibility for Student Intake	1
1.3 Working Days and Class Duration	2
1.4 Dissertation and its Alternatives	2
1.5 Completion of Dissertation	3
2. Course Structure	3
3. Course Details	4
First Semester	5
<i>Mathematics for Statistics</i>	6
<i>Probability</i>	8
<i>Statistical Inference</i>	11
<i>Mathematical Demography</i>	13
<i>Sampling Theory</i>	15
<i>R Programming Language</i>	17
<i>Statistical Computing-I</i>	19
Second Semester	20
<i>Multivariate Analysis</i>	21
<i>Stochastic Processes</i>	23
<i>Statistical Simulation</i>	25
<i>Python Programming</i>	27
<i>Econometrics</i>	29
<i>Quality Control and Reliability</i>	31
<i>Operations Research</i>	33
<i>Nonparametric Statistics</i>	35
<i>Statistical Computing-II</i>	37
Third Semester	38
<i>Bayesian Inference</i>	39
<i>Research Methodology</i>	41
<i>Design of Experiments</i>	44
<i>Biostatistics</i>	46
<i>Environmetrics</i>	49
<i>Time Series Analysis</i>	52
<i>Actuarial Statistics</i>	54
<i>Survival Analysis</i>	56
<i>Statistical Computing-III</i>	58
Fourth Semester	59
<i>Meta-analysis</i>	60
<i>Nonparametric Modeling</i>	63
<i>Population and Financial Statistics</i>	65
<i>Geographical Information System</i>	67
<i>Machine Learning</i>	69
4. Course Conduction Guidelines	71
5. Evaluation Scheme	71
5.1 Internal Evaluation	71
5.2 Evaluation Guidelines	71

## **1. Introduction**

Central Department of Statistics (CDS) shifted the M. Sc. / M. A. course of Statistics from annual system to semester system in 2012. The course was revised in 2015 mainly to maintain the uniformity among departments as regards to the allocation of total credits for the program and evaluation scheme, and also keeping in view of the international norms of the semester system. A considerable period of time has elapsed since the revision was made. At present, the need to revise and upgradation of the course was felt to address the current situation and international trends regarding the course contents of Statistics at Master level. Therefore, the course contents of the program have been re-designed, developed and revised thoroughly with the participation of concerned subject experts including external experts and faculty members of the Department. It consists of the core papers, optional papers, computer based statistical computing (practical) papers and dissertation. Some additional courses namely R programming language, Python programming, statistical simulation, etc. have been added to make it relatively more relevant in the present context. The designed course has incorporated and balanced between theoretical and applied aspects of Statistics as far as possible and practicable. The proposed revised course comprises of 67 credits (1675 marks) spread over two academic years equivalent to four semesters.

The re-structured syllabus for the M. Sc. / M. A. in Statistics in Semester System with major changes is forwarded to the Dean's Office, Institute of Science and Technology (IOST), Tribhuvan University (TU) for its approval.

### **1.1 Enrolment Quota**

Considering the available laboratory space and its holding capacity, the Department has decided to fix quota for the number of students for enrolment in each batch. A maximum of eighty students will be admitted in M.Sc. / M. A. in each batch of semester system. The admission of the students will be made on the merit basis through entrance examination and as per the rules and regulations of TU.

### **1.2 Eligibility for Student Intake**

Students for the semester system will be enrolled based upon the rules and regulations laid down by TU. Students having a Bachelor Degree in Statistics or equivalent degree recognized by TU

are eligible to apply for admission in Semester System M. Sc. / M. A. in Statistics program. Each applicant must appear and pass entrance examination conducted by the Central Department of Statistics, IOST. Admission will be made in the merit basis of the entrance examination. Applicants with below cut-off point can be disqualified from enrollment.

### **1.3 Working Days and Class Duration**

#### **A. Working Days**

Total number of working days for a semester will be 96 days equivalent to 16 weeks.

#### **B. Class Duration**

- **Theory Paper**

One credit will be equivalent to 16 teaching hours for theoretical papers in each semester. Consequently, theory papers of three credits will have three lecture hours per week totaling to 48 lecture hours.

- **Statistical Computing (Practical) Paper**

1 practical class hour = 3 theory hours

22 practical class hour = 4 credits

A computing paper of 4 credits will have 21-23 practical days in each semester. For Statistical Computing classes, groups of students will be formed having no more than 20 students in each group if the number of students exceeds 20. Three teachers (and/or instructors) will be allotted to instruct students simultaneously in each computing classes of each semester.

### **1.4 Dissertation and its Alternatives**

Dissertation is not mandatory to all students. The department will provide rules and criteria for the selection of students for dissertation. It will be of 6 credits and is allocated in fourth semester. The department will provide its orientation within the first two weeks from the date of the commencement of the fourth semester. The duration of the dissertation work will be of 96 working days after the completion of orientation. A single student will be facilitated by one supervisor with 1 hour of assistance/ supervision per week. Provision of the co-supervisor is also applicable, if necessary. Regarding an alternative option to the dissertation

work, students will have to take three theory papers of 2 credits each. The details of which are given in the syllabus.

## 1.5 Completion of Dissertation

Dissertation should be completed within 3 months after the end of the fourth semester to be regarded as a regular student.

## 2. Course Structure

The program is divided into four semesters (six months per semester) with a total duration of 2 years. The program contains 13 core courses, 4 optional courses, 3 computing (practical) courses, dissertation and three additional courses as an alternative option for dissertation. The distribution of courses in different semesters is shown below.

**Table 1: Distribution of Courses**

Semester	Nature of the Course					Total Courses		Total Credit	Total Marks
	Core Theory	Optional Theory	Statistical Computing (Practical)	Dissertation or Alternative Papers		Option A	Option B		
				Dissertation (Option A)	Alternative Paper (Option B)				
1 <sup>st</sup>	6	0	1	0	0	7	7	21	525
2 <sup>nd</sup>	4	2	1	0	0	7	7	21	525
3 <sup>rd</sup>	3	2	1	0	0	6	6	19	475
4 <sup>th</sup>	0	0	0	1	3	1	3	6	150
Total Papers	13	4	3	1	3	21	23		
Total Credit	37	12	12	6				67	
Total Marks	925	300	300	150					1675

The details of the program as regards to the distribution of subjects, credit hours, and marks in different semesters are shown in the following table. In total there are 67 credits (49 credits for theoretical papers, 12 credits for computing papers, and 6 credits for dissertation or additional exam-based papers with 1675 marks in total allocated in the program.

**Table 2: Course Structure**

SN	Subject	Code	Credit	Marks
<b>First Semester</b>				
1.	Mathematics for Statistics	STA 511	3	75
2.	Probability	STA 512	3	75
3.	Statistical Inference	STA 513	3	75
4.	Mathematical Demography	STA 514	3	75
5.	Sampling Theory	STA 515	3	75
6.	R Programming Language	STA 516	2	50
7.	Statistical Computing-I	STA 517	4	100
	<b>Total</b>		<b>21</b>	<b>525</b>
<b>Second Semester</b>				
8.	Multivariate Analysis	STA 521	3	75
9.	Stochastic Processes	STA 522	3	75
10.	Statistical Simulation	STA 523	3	75
11.	Python Programming	STA 524	2	50
12.	Optional Paper (Any Two)		3×2 = 6	150
i.	Econometrics	STA 525		
ii.	Quality Control and Reliability	STA 526		
iii.	Operations Research	STA 527		
iv.	Nonparametric Statistics	STA 528		
13.	Statistical Computing-II	STA 529	4	100
	<b>Total</b>		<b>21</b>	<b>525</b>
<b>Third Semester</b>				
14.	Bayesian Inference	STA 631	3	75
15.	Research Methodology	STA 632	3	75
16.	Design of Experiments	STA 633	3	75
17.	Optional Paper (Any Two)		3×2 = 6	150
i.	Biostatistics	STA 634		
ii.	Environmetrics	STA 635		
iii.	Time Series Analysis	STA 636		
iv.	Actuarial Statistics	STA 637		
v.	Survival Analysis	STA 638		
18.	Statistical Computing-III	STA 639	4	100
	<b>Total</b>		<b>19</b>	<b>475</b>
<b>Fourth Semester</b>				
Any one between Dissertation and three Alternative papers				
19.	Dissertation	STA 641	6	150
20.	Alternative Papers (2 credit each, Any Three)		2×3 = 6	150
i.	Meta-analysis	STA 642		
ii.	Nonparametric Modeling	STA 643		
iii.	Population and Financial Statistics	STA 644		
iv.	Geographical Information System	STA 645		
v.	Machine Learning	STA 646		
	<b>Total</b>		<b>6</b>	<b>150</b>
<b>Grand Total</b>			<b>67</b>	<b>1675</b>

### 3. Course Details

# **First Semester**

**Course Title: Mathematics for Statistics**  
**Course Code: STA 511**  
**Nature: Theory**  
**Semester: First**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture Hours: 48**

### **Course Description**

This course covers fundamental concepts and techniques in numerical analysis and real analysis which are useful for statistical analysis. It includes numerical techniques for solving equations to find roots, interpolation, and integration. It also provides the knowledge related to real analysis, including sequences, series, integration, and functions of multiple variables. Emphasis will be given on understanding the principles behind these methods, their geometrical interpretation and their applications in various mathematical contexts.

### **Learning Outcomes**

After completion of this course, the students will be able to

- understand and apply various method of finding a root of an equation,
- be familiar with different methods finding a function for a set of data,
- find an approximate integral by using different methods,
- check the convergence of various series,
- understand the different theorems related integral and multiple integral,
- find the extreme values of the function of several variables.

### **UNIT 1: Numerical Analysis**

**[18 LH]**

Finding a Root: Algebraic and Transcendental Equations, Bisection method, Secant method, Fixed point Iteration method, Newton Raphson method, geometric interpretation of Bisection and Secant Method

Finding a Function: Interpolation and Extrapolation: Finite differences (forward, backward and central), Newton's formula for interpolation (forward, backward) and its application, Lagrange's interpolation formula

Finding an Integral: General formula for numerical differentiation and integration, Trapezoidal rule, Simpson's 1/3 and 3/8 rules, their application, Ordinary Differential Equations: Solution of Taylor's series, Picard's method, Euler's method, Runge- Kutta method

### **UNIT 2: Real Analysis**

**[30 LH]**

Sequences and Series: Sequences and series of functions, Point wise and uniform convergence, Cauchy general principle of convergence for sequence, Limit superior and limit inferior.

Power series: Radius of convergence, Convergence of power series, applications

Fourier series: Periodic function and its properties, Sum of Fourier series, Fourier series of even and odd functions, applications

Integration: Review of Riemann integral, Riemann-Stieltjes integral, Condition of integrability, Mean value theorem, Integration by parts, Improper Integral, Convergence of improper integrals, Convergence of Beta and Gamma function

Function of several variables: Definition of Limit & Continuity, Partial Derivative, Euler's Theorem, Jacobian, Maxima and Minima, Multiple integral, Dirichlet's theorem, Liouville's expansion to Dirichlet's theorem, Parametric integration

## References

1. Apostol, T. M. (2002). *Mathematical analysis*. Narosa Publishing House.
2. Bartlett, R. G., & Sherbet, D.R. (1994). *Introduction to real analysis*. John Wiley and Sons.
3. Chatterjee, D. (2005). *Real analysis*. Prentice-Hall of India.
4. Malik, S. C., & Arora, S. (1992). *Mathematical analysis*. New Age International.
5. Sastri, S. S. (2003). *Introductory methods of numerical analysis*. Prentice-Hall of India.

**Course Title: Probability**

**Course Code: STA 512**

**Nature: Theory**

**Semester: First**

**Full Marks: 75**

**Pass Marks: 37.5**

**Total Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

This course explores into probability theory through a measure-theoretic approach, covering fundamental concepts like field, sigma field and Borel field. It covers key aspects like discrete and continuous random variables and its probability distributions and expectations. It also discusses on different probability space, distribution functions, cumulative distribution functions, quintile functions, and characteristic functions. Additionally, the course focuses on theoretical aspects and applications of different advanced level probability distributions such as multinomial distribution, generalized power series distributions, distribution of order statistics, some non-central distributions, extreme value distribution, compound probability distributions, prior and posterior distributions, mixed-type distributions, and some inverse distributions.

### **Learning Outcomes**

After completion of this course, the students will be able to

- understand the concepts related to class of sets such as fields, sigma fields, Borel fields, measure theory of probability, and solve related problems,
- apply key concepts of probability, including discrete and continuous random variables, probability distributions, conditioning, independence, expectations, and variances,
- know the tools to describe a random variable, such as the probability density function, the cumulative distribution function, the quintile function and the characteristic function,
- be equipped with a solid theoretical understanding and practical skills related to multinomial distribution and the generalized power series distribution,
- demonstrate a comprehensive understanding of distribution of order statistics and non-central chi-square, t and f distribution,
- understand and apply concepts related to extreme value distributions, compound probability distributions, prior and posterior distributions, mixed-type distributions, and some inverse distributions in real-world scenarios.

### **UNIT 1: Sets, Fields, Measure and Random Variable**

**[10 LH]**

Limit and field: Event, algebra of sets, limit of sequence of sets, limit superior and limit inferior, field, Sigma field, minimal field, monotone field, Borel field

Function, measure and random variable: definition of function, set function, inverse function, measure, measure space and measurable function, probability measure, random variable

**UNIT 2: Probability Space, Functions and Expectation** [12 LH]

Probability Space: Axiomatic definition of probability, properties of probability measure including monotonicity and continuity, types of probability space -discrete, finite, countable and general probability spaces with examples

Conditional Probability Space: Conditional probability measure and independence of events, Distribution Function: Definition and its properties, continuous, discrete and mixed distribution functions, decomposition theorem, distribution function of random vectors

Quantile functions: Definition and their properties, quantile functions as random variables and their distribution functions

Expectation: Definition, existence and finiteness of expectations, expectation in univariate and multivariate distributions and independent random variables, conditional expectation, expectation of linear combinations, relation between expectation and cumulative distribution function

Characteristic function: definition and its properties, inversion formula and uniqueness theorem, examples of use of inversion formula

**UNIT 3: Discrete Distribution** [6 LH]

Multinomial Distribution: Probability mass function, moment generating and characteristic function, moments, covariance and correlation, distribution fitting and examples

Generalized Power Series Distribution: Unified PMF, its special cases (binomial, Poisson, negative binomial)

**UNIT 4: Continuous Distribution** [8 LH]

Extreme Value Distributions: Concept of extreme value distribution and generalized extreme value distribution, Gumbel Distribution, Fréchet Distribution and Weibull Distribution - Probability densities and distribution functions, moments, properties, examples and applications

Inverse Gamma Distribution and Inverse Gaussian Distribution: Concept, Density function, distribution function, moments, examples and applications

**UNIT 5: Sampling Distribution** [6 LH]

Distribution of Order Statistics: Distribution of kth order statistics, joint and marginal distributions of order statistics, distribution of sample median, sample range, maximum values and minimum values, problems and examples

Sampling Distributions of Statistics: Non-central chi-square, non-central t and F distribution, and their general features

**UNIT 6: Some other Probability Distributions** [6 LH]

Compound Probability Distribution: Concept, density function, distribution function, mean and variance, compound negative exponential distribution and its moments

Prior and Posterior Distributions: Concept, examples including cases binomial, poisson and normal likelihood

Mixed Type Distribution: Mixed random variable, meaning and examples, computation of moments of mixed random variables

## References

1. Bhat, B. R. (1999). *Modern probability theory - An introductory textbook* (3rd ed.). New Age International.
2. Biswas, S. (1991). *Topics in statistical methodology*. Wiley Eastern Limited.
3. Bolstad, W. M. (2007). *Introduction to Bayesian statistics*. John Wiley & Sons, Inc.
4. Hoel, P.G., Port, S. C., & Stone, C.J. (1971). *Introduction to probability theory*. Houghton Mifflin Company.
5. Hogg, R.V., & Tanis, E. A. (2001): *Probability and statistical inference*. Pearson Education.
6. Johnson N. L., Kotz S., & Balakrishnan N. (1994). *Continuous univariate distributions, volume I* (2<sup>nd</sup> ed.). John Wiley.
7. Johnson N.L, Kotz S., & Balakrishnan N. (1995). *Continuous univariate distributions, volume II* (2<sup>nd</sup> ed.). John Wiley.
8. Krishnamoorthy, K. (2016). *Handbook of statistical distributions with applications* (2<sup>nd</sup> ed.). CRC Press.
9. Meyer, P. L. (1970). *Introductory probability and statistical applications*. Addison-Wesley.
10. Rohatgi, V. K., & Saleh, A. K. M. E. (2015). *An introduction to probability and statistics*(3<sup>rd</sup> ed.). John Wiley & Sons, Inc.

**Course Title: Statistical Inference**

**Course Code: STA 513**

**Nature: Theory**

**Semester: First**

**Full Marks:75**

**Pass Marks: 37.5**

**Total Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

Statistical Inference course is designed to equip students with the theoretical and applied foundation of statistical methods as estimation, testing of hypothesis and interval estimation. This course explores both point estimation and interval estimation methods, equipping students with the theoretical knowledge and practical skills to make informed estimates and draw meaningful conclusions. Hypothesis testing is a fundamental aspect of statistical analysis, allowing students to make informed decisions based on sample data. This unit covers the theoretical foundations of hypothesis testing, various test statistics, and practical applications

### **Learning Outcomes**

After completion of the course the students will be able to

- apply point estimation methods, including the method of moments and maximum likelihood estimation. Evaluate the properties of maximum likelihood estimators,
- construct confidence intervals and shortest length CI for population parameters, such as the mean, proportion, and variance,
- conduct hypothesis tests for one-sample situations, including tests for means and proportions,
- understand the principles of matched pairs and independent samples. Carry out likelihood ratio test for composite hypothesis.

### **UNIT 1: Estimation**

**[12 LH]**

Minimal sufficient statistics, completeness, exponential families and complete statistics, Fisher information for one and several parameters, necessary and sufficient conditions for Minimum Variance Unbiased Estimator (MVUE), Locally MVUE, Uniformly MVUE (with Proof), lower bound to variance of estimators, information inequality, Consistent Asymptotic Normal (CAN) estimator

Method of estimation, maximum likelihood estimation with application to standard statistical distributions (with alternative methods), asymptotic properties of maximum likelihood estimator (MLE is sufficient, consistent and efficient with proof), moment matching estimator

### **UNIT 2: Interval Estimation**

**[13 LH]**

Method of finding interval estimators by pivotal approach, confidence interval for difference between population mean by pivotal approach, confidence interval for population proportion and recovery of population proportion appears at the end point of inequality, confidence intervals for matched pair differences

Confidence interval for large samples based of central limit theorem, central and non-central confidence interval, construction of shortest expected length confidence interval, shortest length confidence interval for Normal, Gamma, Beta and Uniform distribution, Uniformly Most Accurate Unbiased confidence sets

### **UNIT 3: Testing of Hypothesis**

**[11 LH]**

Concept of composite hypothesis, one sided and two-sided composite alternative hypothesis, families of distribution with monotonic likelihood ratio, Uniformly Most Powerful (UMP) test, UMP test via Neymann Pearson's Lemma, UMP test via monotonic likelihood ratio, UMP test application to standard statistical distribution, unbiased test, UMP unbiased test

### **UNIT 4: Likelihood Ratio Test**

**[12 LH]**

Likelihood ratio test with derivation, derivation of likelihood ratio test for testing mean and variance in exponential families with one sample and two samples problem, Comparing the means: the partial difference t method, Likelihood ratio test for the correlation coefficient, Test for the variances, Behren's Fisher problem, Asymptotic properties of likelihood ratio test

### **References**

1. Casella, G., & Berger R.L, (2002) *Statistical inference* (2<sup>nd</sup> ed.). Cengage Learning India Private Limited.
2. Hoel, Port, & Stone (1971) *Introduction to probability theory*. Houghton Mifflin Company.
3. Hogg R.V., & Criag A. T, (1978) *Introduction to mathematical statistics* (3<sup>rd</sup> ed.), Academic Press.
4. Rohatgi, V. K., & Saleh, A. K. Md. E. (2005) *An Introduction to probability and statistics* (2<sup>nd</sup> ed.) John Wiley.
5. Hogg, R.V., Tanis E. A., & Zimmerman D.L (2015) *Probability and statistical inference* (9<sup>th</sup> ed.). Pearson's Education.
6. Mukhopadhyay, N. (2006) *Probability and statistical inference*. CRC Press.
7. Kale, B. K. (1999). *A first course on parametric inference*. Narosa Publishing House.
8. Lehmann E. L. (1986). *Theory of point estimation*. John Wiley and Sons.
9. Lehmann E. L. (1986). *Testing statistical hypotheses*. John Wiley and Sons.

**Course Title: Mathematical Demography**

**Course Code: STA 514**

**Nature: Theory**

**Semester: First**

**Full Marks: 75**

**Pass Marks: 37.5**

**Total Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

This course broadly introduces the science of demography starting from the study of age distribution to indirect techniques of demographic estimation. The course is organized by comprising different facets of demography like migration, marital rate, nuptiality models, fertility models, and some indirect techniques for measuring fertility and mortality levels.

### **Learning Outcomes**

After the completion of the course, students will be able to

- know the basic features of age distribution and execute the necessary adjustments in the reported age data,
- execute measures of migration using direct and indirect methods along with the fundamentals of migration,
- understand and derive the marital rate, nuptiality models, and fertility models with underlying theoretical concepts,
- recognize the importance of indirect techniques in demographic estimation and computation of some demographic indicators using such methods.

### **UNIT 1: Age Distribution**

**[5 LH]**

Sources of errors in age statement, population pyramid, ageing index, age adjustment: method of parabola, adjustment of error in age group 0-4 years, adjustment of error in age group 5-9 years, adjustment of error in the age group 75 and above

### **UNIT 2: Migration**

**[9 LH]**

Background, causes of migration, Neo-Marxist theory, migration in Nepal, measurement of migration: Direct methods (based on birthplace, based on place of residence, based on the duration of residence, based on residence on a fixed prior data), Indirect methods (residue method, survival ratio method, net migration of children, national growth rate method, net reproduction method), population redistribution with reference of Nepal

### **UNIT 3: Marital Rate**

**[6 LH]**

Risk for first marriage in demographic sense, measuring singulate mean age at marriage (SMAM): Hajnal method, cohort method, stable population method, Van de Walle method

**UNIT 4: Nuptiality Models****[4 LH]**

Nuptiality and its trends in Nepal, models for the assessment of nuptiality; Coale's Gompertz curve, Coale and McNeil's extension of Gompertz curve, T. James Trussell's extension of Gompertz curve

**UNIT 5: Fertility Models****[5 LH]**

Fertility levels and its status in Nepal, some fertility models; Parabolic model, Gompertz model, Coale and Trussell's model, Brass model

**UNIT 6: Indirect Techniques in Measuring Fertility Levels****[12 LH]**

Requirement of indirect techniques, nature of data on children ever born (CEB), errors in fertility data, El-Badry correction for the data on CEB, estimation of ever-fertile women, estimation of number of women with known parity, Coale- Demeny's method of estimation of total fertility rate, adjustment of ASFR: Brass's P/F method and Coale- Trussell's method, comparing period fertility rates with a hypothetical cohort, Estimation of age-specific fertility from the increment of cohort parities between two surveys; cohort parity increment method, Estimation of birth rates and death rates from stable population model; ten years survival ratio method

**UNIT 7: Indirect Techniques in Measuring Mortality Levels****[7 LH]**

Brass's method of estimation of infant and child mortality, ten-year survivorship method for estimation of death rates from stable population, indirect method of estimation of IMR, James M. Cann's method of estimation of life expectancy at birth, Vig's relation between life expectancy at birth and death rates

**References**

1. Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The methods and materials of demography* (2<sup>nd</sup> ed.). Elsevier Academic Press.
2. Cox, P. R. (1976). *Demography* (5<sup>th</sup> ed.). Cambridge University Press.
3. Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Blackwell Publishing.
4. Weinstein, J., & Pillai, V. K. (2015). *Demography: The science of population* (2<sup>nd</sup> ed.). Rowman & Littlefield.
5. United Nations (1983). *Indirect techniques for demographic estimation (Manual X)*.
6. Bhende, A., & Kanitkar, T. (2006). *Principles of population studies*. Himalaya Publishing House.
7. Singh, M. L., & Sayami, S. B. (1997). *An introduction to mathematical demography*.

**Course Title: Sampling Theory**

**Course Code: STA 515**

**Nature: Theory**

**Semester: First**

**Full Marks: 75**

**Pass Marks: 37.5**

**Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

The course is designed to impart knowledge on random sampling methods used widely in sample surveys. Specifically, the course includes techniques adopted in multivariate ratio method of estimation, super population, cluster sampling, multistage sampling, double sampling, sampling methods in complex surveys and variance estimation.

### **Learning Outcome**

After completion of the course the students will be able to

- apply sampling methods in order to execute sample surveys with theoretical understanding,
- use specific sampling methods namely cluster sampling, multistage and multiphase sampling in sampling-based studies,
- perform computational works related to estimation of parameters, sampling errors and related statistics,
- design and execute sampling design suitable for a specific sample survey.

### **UNIT1: Ratio Estimator**

**[8 LH]**

Concept of multivariate ratio estimator, multivariate ratio estimator for two auxiliary variables, comparison of multivariate ratio estimators with customary ratio estimator having single auxiliary variable and with simple random sampling without replacement, combined and separate ratio estimators, product estimator

### **UNIT 2: Cluster Sampling**

**[12 LH]**

Cluster sampling of equal cluster size, estimation of population mean, total, variance of mean and total, efficiency of cluster sampling with respect to SRS, comparison with stratified random sampling, intra-cluster correlation coefficient, estimation of optimum cluster size

Cluster sampling of unequal cluster size: Probability proportional to size (PPS) sampling, PPS sampling with and without replacement: Estimation of population mean, total and their variances, gain in efficiency, Hansen-Hurwitz estimator, Horvitz Thompson estimator and estimation of variance, Yates and Grundy estimator of variance, illustrative examples of estimating population mean, total and standard error

### **UNIT 3: Sub-Sampling**

**[8 LH]**

Concept of sub-sampling, two-stage sampling with equal first stage units: Estimation of the population mean, unbiased estimate of sampling variance; Three- stages sampling with equal

first and second stages: estimation of mean and variance, illustrative examples of estimating population mean, total and standard error

#### **UNIT 4: Multiphase Sampling**

**[6 LH]**

Concept of multiphase sampling, double sampling for stratification, optimum allocation, estimate of variances in double sampling for regression, illustrative examples

#### **UNIT 5: Sample Design**

**[10 LH]**

Introduction, complex surveys, sample designs in complex surveys, self-weighting design, weighting in sample designs, estimation of sample size in complex surveys, design effect, post stratification, analytic domains, examples of sample designs in complex surveys with SRS, stratified sampling, cluster sampling, PPS sampling, systematic sampling, etc.; Sample designs in nationally representative survey of Nepal: Nepal Demographic and Health Survey (NDHS), Nepal Living Standards Survey (NLSS), Nepal Multiple Indicator Cluster Survey (NMICS), Household Budget Survey

#### **UNIT 6: Variance Estimation**

**[4 LH]**

Concept, variance estimation in complex survey, methods of variance estimation: method of linearization, Jackknife and bootstrap methods, illustrative examples

#### **References**

1. Cochran, W. G. (1977). *Sampling techniques*. Wiley-Eastern.
2. Mukhopadhyay, P. (1998). *Theory and methods of survey sampling*. Prentice Hall of India.
3. Lohr, S. L. (1999). *Sampling: Design and analysis*. Duxbury Press.
4. Raj, D., & Chandhok, P. (1999). *Sample survey theory*. Narosa Publishing House.
5. Wolter, K. M. (1985). *Introduction to variance estimation*. Springer-Verlag.
6. Chaudhuri, A. (2010). *Essentials of survey sampling*. PHI Learning Pvt.
7. Sampath, S. (2005). *Sampling theory and methods*. Narosa Publishing House.

**Course Title: R Programming Language**

**Course Code: STA 516**

**Nature: Theory**

**Semester: First**

**Full Marks: 50**

**Pass Marks: 25**

**Total Credits: 2**

**Total Lecture Hours: 32**

### **Course Description**

This is an introductory course on statistical concepts and applied techniques needed for data preparation, data modeling by using R. This course is intended as a guide to data analysis with the R system for statistical computing. This course covers an introduction to R, the R basics, logical operator, data structure, data management in R, matrix operation, data visualization and graphics in R, and basic statistical functions and inferences.

### **Learning Outcomes**

After completion of this course, the students will be able to

- use real data sets and perform analysis using R,
- write programs using R for analyzing data,
- make graphs and charts by using R,
- use R for various statistical computations and inferences.

### **UNIT 1: An Introduction to R and Data Structure**

**[6 LH]**

Purpose of using R software, installing R and R Studio, The R environment and working with R, The R packages (Meaning and purpose of packages, installing, loading and learning Packages), Basic math, variables, data types, basic R function, List, Vectors, Matrices, Arrays, Data Frames, Factors

### **UNIT 2: Conditional and Control Flow**

**[3 LH]**

Relational Operators and vectors, Logical operators (AND, OR, NOT), Logical operators and vectors, Conditional statement

### **UNIT 3: Data Entering, Reading and Management**

**[3 LH]**

Data entering, Data reading from external files (CSVs, Excel, SPSS, Stata files), creating new variables, recoding variables, renaming variables, missing values, sorting data, merging data sets (Adding columns and rows to the data frame), selecting variables, dropping variables, selecting observations

### **UNIT 4: Matrix Operation**

**[3 LH]**

Matrix operations, creating matrix from vector, combine vectors or matrices, matrix addition, transpose matrix, Find the dimension of a matrix or dataset, matrix multiplication, finding the inverse of a matrix, component-wise multiplication, create a submatrix, create a diagonal matrix,

a vector of diagonal elements, create a vector from a matrix, calculate the determinant, find eigenvalues and eigenvectors, find the singular value decomposition

**UNIT 5: Graphics, Visualization and Summary Measures** [5 LH]

Bar chart, Pie chart, Box plot, Histogram, Line graph, Density plots, Normal Q-Q plot, Scatter plot, Basics of ggplot, Measures of central tendency, Measures of variation, measures of skewness and Kurtosis

**UNIT 6: Probability and Probability Distributions** [4 LH]

Random Number, Random number generator, binomial distribution, Poisson distribution, exponential distribution, negative binomial distribution, normal distribution, plot for testing normality

**UNIT 7: Statistical Inference** [8 LH]

Inferences in the one sample case, Student's t distribution, comparing two variances (Fisher's F test), comparing two sample means with normal errors (Student's t-test), correlation coefficient, correlation matrix, testing for independence in contingency tables using chi-squared, linear regression, assumptions, standardized coefficient, hypothesis testing

**References**

1. Lander, J. P. (2014). *R for everyone: Advanced analytics and graphics*. Pearson Education.
2. Crawley, M. J. (2015). *Statistics: An introduction using R* (2<sup>nd</sup> ed.). John Wiley & Sons Ltd.
3. Fox, J. and Weisberg, S. (2019). *An R companion to applied regression* (3<sup>rd</sup> ed.). Sage Publication.
4. Dalgaard, P. (2008). *Introductory statistics with R* (2<sup>nd</sup> ed.). Springer.
5. Kabacoff, R. I. (2015). *R in action* (2<sup>nd</sup> ed.). Manning.

**Course Title: Statistical Computing-I**

**Nature: Practical**

**Course Code: STA517**

**Semester: First**

**Total Lectures: 21-23 Practical Days (3 Hours/Lecture)**

**Full Marks: 100**

**Pass Marks: 50**

**Total Credits: 4**

**Total Duration: 64 hours**

**Table 3: Subject-wise allocation of number of practical**

SN	Subject Area	No. of Practicals (Minimum)
1	Mathematics for Statistics	4
2	Probability	4
3	Statistical Inference	4
4	Mathematical Demography	4
5	Sampling Theory	4
6	R Programming Language	10
	Total	30

# **Second Semester**

**Course Title: Multivariate Analysis**  
**Course Code: STA 521**  
**Nature: Theory**  
**Semester: Second**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture Hours: 48**

### **Course Description**

This course is developed to provide fundamental knowledge of multivariate normal distribution and associated multivariate statistical methods with their applications. The course will impart theoretical as well as applied knowledge of advanced multivariate statistical methods namely principal components analysis (PCA), factor analysis, multivariate analysis of variance (MANOVA), discriminant analysis and canonical correlation analysis along with their applications using statistical computer software.

### **Learning Outcomes**

After completion of this course, the students will be able to

- understand the characteristic features of multivariate normal distribution along with statistical inference and apply multivariate statistical methods to observed data,
- perform PCA and factor analysis with theoretical understanding,
- classify data using discriminant function with theoretical knowledge,
- understand and apply MANOVA and canonical correlations to observed data wherever applicable and suitable.

#### **UNIT 1: Multivariate Normal Distribution (MVND) [8 LH]**

Density function and characteristic function of MVND, distribution of linearly transformed multivariate normal random vector, marginal and conditional distribution of MVND, Necessary and sufficient condition for independence in MVND

#### **UNIT 2: Estimation of Multivariate Normal Distribution Parameters [9 LH]**

Concept about sampling from MVND, maximum likelihood estimators of mean vector and dispersion matrix, properties and distributions of maximum likelihood estimators, Wishart distribution and its properties, maximum likelihood estimators of simple, partial and multiple correlation coefficient and their distributions

#### **UNIT 3: Hypothesis Testing in Multivariate Normal Distribution [7 LH]**

Hotelling's  $T^2$  statistic as a generalization of square of Student's  $t$  statistic, defining Hotelling's  $T^2$  statistic from likelihood ratio test, distribution of Hotelling's  $T^2$  statistic and its invariance property, applications of  $T^2$  statistic in hypothesis testing (one sample and two sample problems), distance between two populations, Mahalanobis  $D^2$  statistic

**UNIT 4: Principal Components Analysis (PCA)****[5 LH]**

Model formulation, number of components and component structure, extraction of principal components, maximum likelihood estimators of principal components and their variances, applications

**UNIT 5: Factor Analysis****[6 LH]**

Orthogonal factor model, estimation of factor loadings, communalities, factor extraction, factor rotation, factor scores and their applications, maximum likelihood estimators for random orthogonal factors, tests of hypothesis in factor models

**UNIT 6: Discriminant Analysis****[5 LH]**

Separation and classification for two populations, applications of discriminant analysis, classification with two multivariate normal populations, Fisher's Discriminant function, classification with several populations,

**UNIT 7: Multivariate Analysis of Variance (MANOVA)****[4 LH]**

Multivariate one-way analysis of variance model, Wilks test, Roy's Test and its application

**UNIT 8: Canonical Correlation Analysis****[4 LH]**

Introduction, applications, population canonical correlations, sample canonical correlations, properties of canonical correlations, tests of hypotheses

**References**

1. Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Wiley-Interscience.
2. C. Radhakrishna Rao. (2009). *Linear statistical inference and its applications*. John Wiley & Sons.
3. Johnson, R. A., & Wichern, D. W. (2019). *Applied multivariate statistical analysis*. Pearson.
4. Wolfgang Karl Härdle, & Simar, L. (2013). *Applied multivariate statistical analysis*. Springer Science & Business Media.
5. Tenko Raykov, & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Routledge.

**Course Title: Stochastic Processes**

**Course Code: STA 522**

**Nature: Theory**

**Semester: Second**

**Full Marks: 75**

**Pass Marks: 37.5**

**Total Credit: 3**

**Total Lecture Hours: 48**

### **Course Description**

This course introduces the theory and applications of random processes required in various disciplines like finance, economics, physics, engineering, telecommunications, mathematical biology, medicine, environmental science, and many more. The overall course is designed across different topics including random walk, branching process, Markov chain, discrete and continuous Markov processes in continuous time, renewal theory, and queuing theory.

### **Learning Outcomes**

After completion of the course, the students will be able to

- understand the fundamental concepts, classifications, and importance of stochastic processes,
- recognize and derive the theoretical relations of the random walk and branching process,
- know the theoretical background of the Markov chain and compute the different probabilities related to Markov chain problems,
- distinguish and understand the discrete and continuous Markov processes in continuous time,
- understand the basics of renewal theory and queuing theory along with their applications in different areas.

### **UNIT 1: Introduction to Stochastic Processes**

**[13 LH]**

Classification of stochastic processes according to state space and time domain, probability generating function (PGF) and its properties, martingales, types of martingales, some illustrations of martingale property

Random walk: Introduction, transition probabilities, return to origin probability, gambler's ruin problem, expected duration of the game

Branching Process: Galton-Watson branching process, PGF of generation size, mean and variance of generation size, probability of ultimate extinction, total progeny

### **UNIT 2: Markov Chain**

**[9 LH]**

Introduction, homogeneous Markov chain, transition probability, Lemmas on transition probabilities, absolute probability, two-state Markov chain, spectral analysis of two-state Markov chain, classification of states of Markov chain, theorems related to communicating, recurrent and transient properties of states of Markov chain, countable state Markov chain

### **UNIT 3: Discrete and Continuous Markov Processes in Continuous Time [14 LH]**

Discrete Markov Processes in Continuous Time: Poisson process & its properties, probability mass function of Poisson process, homogenous birth and death processes (pure birth process, general birth process, simple linear death process, simple birth and death process, general birth and death process)

Continuous Markov Processes in Continuous Time: Introduction to continuous Markov process, diffusion process, Brownian motion process, Kolmogorov forward and backward diffusion equations, correlation coefficient for a Wiener process

### **UNIT 4: Renewal Theory [6 LH]**

Introduction to renewal theory, general renewal theory, renewal function, integral equation of renewal theory, stopping time and Wald's equation, spent and residual time distribution, lattice random variable, elementary renewal theorem

### **UNIT 5: Theory of Queue [6 LH]**

Basics of queue theory, structure of a basic queue system, M/M/1 queue system, steady state solution of M/M/1 queue, waiting time distributions in M/M/1 queue system, operating characteristic of M/M/1 queue system, M/M/s queue system, operating characteristics of M/M/s queue system

### **References**

1. Pinsky, M. A. (2011). *An introduction to stochastic modeling* (4<sup>th</sup> ed.). Academic Press.
2. Durrett, R. (2016). *Essentials of stochastic processes* (3<sup>rd</sup> ed.). Springer.
3. Resnick, S. I. (1992). *Adventures in stochastic processes*. Birkhauser Boston.
4. Ross, S. M. (1983). *Stochastic processes* (2<sup>nd</sup> ed.). Wiley.
5. Taylor, H. M., & Karlin, S. (1994). *An introduction to stochastic modeling*. Academic Press.
6. Tijms, H. C. (2003). *A first course in stochastic models*. Wiley.
7. Dobrow, R. P. (2016). *Introduction to stochastic processes with R*. Wiley.
8. Hoel, P. G., Port, S. C., & Stone, C. J. (1972). *Introduction to stochastic processes*. Houghton Mifflin Company.
9. Shrestha, H. B. (2009). *Stochastic processes, An introductory text*. Ekta Books.

**Course Title: Statistical Simulation**

**Course Code: STA 523**

**Nature: Theory**

**Semester: Second**

**Full Marks: 75**

**Pass Marks: 37.5**

**Total Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

This course is designed for master's level students aiming to provide students with a solid foundation in simulation methodology, enhance their computational skills, foster critical thinking abilities, and offer practical experience. This course covers various resampling methods, Monte Carlo, MCMC, and many other simulation methods and some special algorithms. This equips the students to effectively apply simulation techniques in statistical analysis and decision-making using R and Python and improves their critical thinking.

### **Learning Outcomes**

After completion of the course, the students will be able to

- learn how to use statistical simulation effectively to solve various types of statistical problems and gain a strong grasp of its principles,
- gain proficiency in using software like R and Python for statistical simulations and improve their critical thinking abilities for complex data analysis,
- develop critical thinking abilities, through hands on projects, to handle complex statistical problems and apply simulation in real-world scenarios or research projects.

### **UNIT 1: Basic Simulations Techniques**

**[16 LH]**

Introduction to Statistical Simulation: Motivation and importance of statistical simulation, Overview of simulation Techniques and their applications, Advantages and limitations of simulation in statistical modeling and analysis

Random Number Generation: Basics of random number generation and random variates, Pseudo-random number generators, Sampling from probability distributions, Testing randomness and evaluating quality of random number generators

Resampling Methods: Bootstrapping: Basics, principles, types, applications, Jackknife resampling for bias estimation, Permutation tests and their use in hypothesis testing, Cross-validation and resampling techniques for model selection

Monte Carlo Simulation: Basics of Monte Carlo simulation and Monte Carlo integration, Sampling from probability distributions, Estimation and inference using MC method, Variance reduction techniques: importance sampling, rejection sampling

### **UNIT 2: MCMC and Advanced Simulation Methods**

**[16 LH]**

Markov Chain Monte Carlo (MCMC) Methods: Introduction to MCMC, its roles and properties, Metropolis-Hastings algorithm, Gibbs sampling and other MCMC algorithms, Sampling from complex posterior distribution, Convergence diagnostics (mixing, burn-in, and thinning)

Simulation-Based Inference: Simulation-based hypothesis testing and model comparison, Simulation for constructing confidence intervals, Assessing model fit using simulation-based techniques, Simulation-based computational Statistics and Bayesian computation  
Advanced Simulation Techniques: Discrete event simulation, Sequential Monte Carlo methods, Rare event simulation, Parallel and distributed computing for simulation

### **UNIT 3: Other Algorithms and Applications of Simulation [16 LH]**

Some Important Principles and Algorithms: Occam's Razor, Random Forest, Kalman Filter, EM Algorithms and Support Vector Machines

Applications, Case Studies and Project: Real-world examples /case studies demonstrating the use of simulation methods, (from epidemiology, environmental science, medicine, reliability etc.), Hands-on exercises/ projects applying simulation techniques to analyze datasets using R and /or Python

### **References**

1. Ross, S. (2006). *Simulation* (4<sup>th</sup> ed.). Academic Press.
2. Rubinstein, R., & Kroese, D. (2016). *Simulation and the Monte Carlo method*. John Wiley.
3. Gelman, A., Jones, G., Brooks, S., & Meng, Xi. (2011). *Handbook of Markov Chain Monte Carlo* (Ed.). CRC Press.
4. Robert, C., & Casella, G. (1999). *Monte Carlo statistical methods*. Springer.
5. Robert, C. P., & Casella, G. (2010). *Introducing Monte Carlo methods with R*. Springer Science & Business Media.
6. Jones, O, Maillardet, R., & Robinson, A. (2014). *Introduction to scientific programming and simulation using R* (2<sup>nd</sup> ed.). CRC Press.

**Course Title: Python Programming**

**Course Code: STA 524**

**Nature: Theory**

**Semester: Second**

**Full Marks: 50**

**Pass Marks: 25**

**Total Credits: 2**

**Total Lecture Hours: 32**

### **Course Description**

This course covers different concepts of Python programming including basic language features, control statements, functions, object-oriented concepts, exception handling, file handling, modules and packages, and some common Python libraries.

### **Learning Outcomes**

After completion of the course, the students will be able to

- provide knowledge to student with different concepts and object-oriented features of Python programming language,
- provide knowledge of some common Python libraries such as Numpy, Pandas, and Matplotlib.

### **UNIT 1: Introduction**

**[3 LH]**

Python Introduction; Why Python; Installing Python; Using Interactive Shell from Console; Running Python Scripts from Console; Using Interactive Shell from IDLE; Using IDEs; Installing Third Party Libraries; Virtual Environment; Writing Comments; Indentation; Variables; Operators

### **UNIT 2: Built-In Data Types**

**[5 LH]**

String; Integer; Float; Complex; List; Tuple; Range; Set; Dictionary; Boolean; Indexing and Slicing; Comprehensions

### **UNIT 3: Conditionals and Iterations**

**[3 LH]**

Introduction; The if Statement; The match-case Statement; The for Loop; The while Loop; The break and continue Statements

### **UNIT 4: Functions**

**[3 LH]**

Introduction; Why use Functions? Defining Functions; Passing Arguments; Return Values, Recursive Function

### **UNIT 5: Object-Oriented Programming**

**[5 LH]**

Object-oriented Programming Concepts; Creating Class and Object; Inheritance; Method Overriding Modules and Packages

### **UNIT 6: Exception and File Handling**

**[3 LH]**

Exceptions; Handling Exceptions; Reading and Writing Files

## **UNIT 7: Python Libraries**

**[10 LH]**

Numpy: Introduction; Creating Array; Dimensions; Array Attributes, Indexing and Slicing; Array Copy and View; Creating Array from Numerical Range; Array Broadcasting; Iterating Over Array; Sorting and Searching; Statistical Functions

Pandas: Series and Data Frames; Creating Data Frames; The head and tail Functions; Attributes; Working with Missing Data; Indexing, Slicing, and Subsetting; Merging and Joining Data Frames; Working with CSV Data

Matplotlib: Introduction; pyplot; The plot() function; Marker; Line; Color; Label; Grid Lines; Subplot; Scatter Plot; Bar Graph; Histogram, Pie Chart; Box Plot

### **Laboratory Work:**

The laboratory work includes writing computer programs using Python programming language covering all the concepts studied in each unit of the course.

### **References**

1. Summerfield, M. (2009). *Programming in Python 3: A complete introduction to the Python language*. Addison-Wesley Professional.
2. Lampart, K. A. (2018). *Fundamental of Python: First program*, (2<sup>nd</sup> Ed.), Cengage Learning Publishing.
3. Jackson, C. (2018): *Learn programming in Python with cody Jackson*. Packt Publishing.
4. Romano, F., & Kruger, H. (2021). *Learn Python programming – An in-depth introduction to the fundamentals of Python*, (3<sup>rd</sup> Ed.), Packt Publishing.

**Course Title: Econometrics**  
**Course Code: STA 525**  
**Nature: Theory**  
**Semester: Second**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture Hours: 48**

### **Course Description**

Econometrics use statistical methods to answer economic questions. But also, it is equally important and applicable in every field of research. This course introduces the concept of econometrics and discusses on the necessary tools to conduct empirical research with a theoretical basis. A special focus will be given on explaining that available data is usually not experimental and what consequences this has on the empirical model. The main topic of this course will be linear regression models. Qualitative response variable regression modelling will also be discussed. Both theoretical and practical examples will be part of the course.

### **Learning Outcomes**

After completion of the course, the students will be able to

- know statistical foundations of regression analysis with OLS,
- understand the causality concept in an empirical context,
- detect and treat violations of OLS assumptions, such as omitted variables, multicollinearity, heteroscedasticity, and serial correlation,
- learn of econometric modelling,
- understand qualitative response regression,
- find estimation and testing with OLS,
- know forecasting with regression modelling,
- read and critically discuss empirical papers,
- use statistical computer software for empirical analysis.

### **UNIT 1: Multiple Linear Regression (Matrix Approach)**

**[13 LH]**

Review of simple linear regression model with OLS and some matrices algebra The k variable regression model, model specification, assumptions in matrix notation, OLS estimation of the parameter vector, variance-covariance matrix of the estimated parameter vector, standard error of estimates, properties of vector estimate, unadjusted and adjusted multiple coefficients of determinations, correlation matrix, hypothesis testing of regression coefficients, testing of goodness of fit by analysis of variance, prediction. Problems and examples

### **UNIT 2: Violation of Assumptions**

**[15 LH]**

Multicollinearity: The nature of multicollinearity, estimation, detection, consequences and remedial measures., Heteroscedasticity: The nature of heteroscedasticity, OLS estimation in the presence of heteroscedasticity, detection of heteroscedasticity: Goldfeld-Quandt test, Breusch-Pagan-Godfrey test, White's test, treatment of heteroscedasticity: method of weighted least

squares, White's Heteroscedasticity - consistent variance and standard errors, Autocorrelation and residual analysis: The nature of autocorrelation, detection and consequences of autocorrelation, Durbin Watson test, and remedial measures: Changing the functional form, Cochrane-Orcutt iterative procedure

### **UNIT 3: Econometric Modeling**

[5 LH]

Model specification criteria, Type of specification error, Consequence of model specification (overfitting and underfitting a model) Durbin Watson d Statistics for model misspecification. Distributed Lag Models: Model specification, estimation: Koyck approach and Almon lag

### **UNIT 4: Qualitative Response Regression**

[15 LH]

Binary Logistic Regression Model: Categorical dependent variable in regression, binary logistic regression: model specification, assumptions, estimation and interpretation of model parameters, odds ratio, Multinomial Logistic Regression Model: Model specification, assumptions, estimation of parameters (derivation not required) with interpretations, examples of fitted models with model adequacy tests, Ordinal Logistic Regression Model: Model specification, assumptions, estimation of parameters (derivation not required) with interpretations, examples of fitted models with model adequacy tests

### **References**

1. Maddala, G. S. (2001). *Introduction to econometrics*. John Wiley and Sons.
2. Johnstone, J. (2007). *Econometric methods*. McGraw Hill Book Company.
3. Gujarati, D. N., Porter, D. C., & Gunasekar, S. (2011). *Basic econometrics*. McGraw Hill Education.
4. Bhumik, S. K. (2015). *Principles of econometrics*. Oxford University Press.
5. Wooldridge, J. M. (2019). *Introductory econometrics, A modern approach* (7<sup>th</sup> ed.). CengageLearning.

**Course Title: Quality Control and Reliability**  
**Course Code: STA 526**  
**Nature: Theory**  
**Semester: Second**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Credits: 3**  
**Total Lecture hours: 48**

### **Course Description**

This course explores advanced topics in quality systems and reliability analysis in three units. The first unit covers quality management foundations, customer-centric approaches, and recent developments. It also includes offline quality control strategies, a review of DOE in quality, and performance optimization using response surface methodology. The second unit focuses on statistical process control methods, including control chart evaluation, short run SPC techniques, and economic design of x-bar chart. It covers sampling inspection plans, OC functions, advanced plans like sequential, chain, and skip-lot sampling, corrective sampling plans, and methods for estimating sample sizes and constants. The third unit delves into reliability analysis, life testing models, system reliability, maintainability, and handling censored and truncated data in life testing experiments.

### **Learning Outcomes**

After completion this course, students will be able to

- comprehend quality management principles, monitoring strategies, and recent developments,
- design effective quality experiments, incorporating Taguchi's strategies and response surface methodology,
- utilize proper SPC procedures for general and short runs, along with control chart design,
- implement various sampling inspection plans and interpret their performance, including corrective sampling plans,
- increase knowledge of reliability fundamentals and estimation techniques for different life testing models,
- interpret system reliability, maintainability, and handling censoring and truncated cases in life testing experiments.

### **UNIT 1: Quality Management and Designing Quality Experiments [16 LH]**

Foundations of Quality Management; Philosophies behind quality and quality management, Quality Gurus and their key contributions, Quality management principles, Deming's circle and other continuous quality improvement strategies; Process capability improvement methods and recent developments; Customer-centric approaches: Kano Model, Quality Function Deployment Designing Quality Experiments: Offline quality control and its significance, Taguchi's offline quality control strategies: Robust design strategy (system design, parameter design, tolerance design), variance reduction strategy, performance measure strategy, productivity enhance strategy, customer satisfaction strategy; Designing offline quality experiments using full factorial

and fractional factorial designs; Response surface methodology and optimization; Taguchi's orthogonal arrays

## **UNIT 2: Statistical Process Control and Sampling Inspection [20 LH]**

SPC Strategy Analysis: Review of SPC and control chart procedures, OC functions, ARL and process capacity. Short run SPC techniques: Individual measurement and moving range chart, Moving average chart, EWMA chart, CUSUM (Tabular and V mask) charts, Multivariate control charts, Modified control chart, Acceptance control charts; Economic design of  $\bar{x}$ -bar chart, Duncan's Model Application of quality control principles, methods, and control charts for non-manufacturing applications and research (healthcare, banking, education, customer service, hospitality, transportation, insurance etc.)

Sampling Inspection: Review of single and double acceptance sampling plans and their OC, AQL, LTPD, AOQ, AOQL, ASN, and ATI Multiple, sequential, chain, and skip-lot sampling plans; sampling plans for variables, Corrective sampling plan: Rectifying inspection program, Curtailed and semi-curtailed inspection plan, Method for estimating  $n$  and  $c$  using large sample for single, double sampling plans

## **Unit 3: Reliability Analysis [12 LH]**

Introduction to Reliability: Concepts, importance and applications of reliability analysis; Failure function, Reliability function, Hazard rate, Mean time to failure, Failure time distribution, General equation of failure rate distribution and related problems

Life testing models and reliability estimation: Exponential, Gamma, Weibull, Lognormal, Rayleigh and Bath-tub models, Reliability estimation techniques and interpretation for each model. System Reliability and Maintainability: Series, parallel and series-parallel systems and their reliability, Reliability of coherent system, Maintainability and Availability, System availability, Preventive maintenance. Censored and Truncated cases in life testing, and their reliability with some examples

## **References**

1. Montgomery, D. C. (2004). *Introduction to Statistical Quality Control*. Willey.
2. Mitra, A. (1998). *Fundamentals of quality control and improvement* (2<sup>nd</sup> ed.). Prentice Hall.
3. Grant, E. L., & Leavenworth, R. S. (2004). *Statistical quality control*. Tata McGraw Hill.
4. Biswas, S. (1997). *Statistical quality control*. New Age India.
5. Sinha, S. K., & B. K. Kale (1980). *Life Testing and Reliability Estimation*. Willey Eastern.
6. Khatiwada, R. P. (2013). *An introduction to statistical quality control and reliability*. Quest Publication.

**Course Title: Operations Research**  
**Course Code: STA 527**  
**Nature: Theory**  
**Semester: Second**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture Hours: 48**

### **Course Description**

This course provides an introduction to fundamental principles, mathematical and analytical optimization techniques which are useful for solving simple as well as complex decision-making problems. It focuses on applications of operation research in various fields like business, management, economics, engineering, industry, finance and logistics.

### **Learning Outcomes**

After completion of this course, the students will be able to

- be familiar with concepts of feasible, infeasible, basic solution, optimum solutions of linear programming problem,
- understand how to formulate and solve LP problems related different sectors, the concepts of duality of LP problem and importance of duality,
- measure the sensitivity of changing the price vector and requirement vector,
- find the integer solution of LP problem,
- know the importance of location planning, factors associated to it and different methods used for location planning,
- know the importance of capacity and aggregate planning, different methods of aggregate planning,
- be familiar with transformation process, importance of product design and concepts of product design,
- understand the concepts of crashing the project.

### **UNIT 1: Linear Programming**

**[5 LH]**

General Nature of programming problems, Scope and limitation, concepts of feasible, infeasible, optimum solutions, effective, ineffective, simultaneous linear equations, basic solutions, linear transformations, point sets, lines and hyper planes, convex cones

### **UNIT 2: Formulation and solution of Linear Programming Problems**

**[10 LH]**

Formulation of linear programming problems related to different sectors, Reviews of Simplex Method for LPP with maximization / minimization objective function. Dual linear programming problems, fundamental properties of dual problems, complementary slackness, unbounded solution in the primal, dual-simplex algorithm, importance of duality in LPP

**UNIT 3: Post Optimal/ Sensitivity Analysis****[5 LH]**

Post-optimality problems, changing the price vector, changing the requirement vectors, adding and removing variables or constraints, upper and lower bounds

**UNIT 4: Integer Programming****[3 LH]**

Introduction, Application of integer programming, methods of integer programming: Branch and bound algorithm, Gomory Fractional cut algorithm

**UNIT 5: Location Planning****[7 LH]**

Introduction to facility location planning, problems due to improper location planning, strategies for multiple facility planning, factors affecting for location planning, methods used for location selection: Centre of gravity, factor rating, cost volume analysis, transportation technique

**UNIT 6: Capacity and Aggregate Planning****[8 LH]**

Introduction to capacity, measure of capacity, types of capacity, utilization of capacity, aggregate planning, types of aggregate planning: level production plan, chase production plan, varying utility plan

**UNIT 7: Designing a Product****[3 LH]**

Transformation process, defining a product, types of product: goods vs services, importance of designing a product, concepts of product design, life cycle of a product

**UNIT 8: Network Analysis****[7 LH]**

Introduction of network diagram, critical path method, crashing of project

**References**

1. Hadle, G. (1978). *Linear programming*. Edision-Wesley Publishing Co.
2. Gupta, P. K., & Hira, D. S. (2007). *Operations research* (4th ed.). S. Chand & Company Ltd.
3. Vohra, N. D. (2006). *Quantitative techniques in management*. TATA McGraw Hill.
4. Bedi, K. (2013). *Production and operations management* (3rd ed.). Oxford University Press.
5. Heizer, J. & Render, B. (2015). *Operations management* (11th ed.). Pearson.

**Course Title: Nonparametric Statistics**  
**Course Code: STA 528**  
**Nature: Theory**  
**Semester: Second**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture Hours: 48**

### **Course Description**

This course deals with statistical inference when parametric distributions are not assumed and presents the theory and procedures of decision making in absence of rigid distributional assumptions.

### **Learning Outcomes**

After completion of this course, the students will be able to

- know about order statistics,
- know about distribution free statistics,
- know about different nonparametric statistics and their applications.

### **UNIT 1: Order Statistics**

**[10 LH]**

Probability integral transformation, Joint and marginal distribution of  $r^{\text{th}}$  order statistics, Moments of order statistics, Distribution of median and range, Asymptotic distribution of order statistics, Confidence interval estimates for population quintiles, Hypothesis testing for population quintiles

### **UNIT 2: Distribution-Free Statistics**

**[10 LH]**

Distribution-free statistics over a class, Counting statistics, Ranking statistics, U-statistics: one sample and two sample U-Statistics and their asymptotic properties, Asymptotically distribution-free statistics

### **UNIT 3: Non-Parametric Tests**

**[28 LH]**

One sample test: Binomial test, Tests based upon runs, Exact null distribution of R, Moments of the null distribution of R, Asymptotic null distribution of R, Sign test, Wilcoxon signed rank test, Tests of goodness of fit: Chi-square test, Kolmogorov-Smirnov test  
Two sample tests: Wald-Wolfowitz runs test, Kolmogorov-Smirnov two sample test, Median test, Mann-Whitney U test  
Several sample tests: Kruskal-Wallis one way ANOVA test, Measures of association: Kendall's tau coefficient, Spearman's coefficient, Contingency coefficient, Coefficient of concordance, Friedman's two-way analysis of variance by ranks

## References

1. Gibbons, J. D. (1985). *Nonparametric statistical inference*. Marcel Dekker.
2. Randles, R. H., & Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics*. John Wiley and Sons.
3. Rohatgi, V. K., & Saleh, A. K. Md. E. (2005). *An introduction to probability and statistics*. John Wiley and Sons.
4. Conover, W. J. (1980). *Practical nonparametric statistics*. John Wiley and Sons.

**Course Title: Statistical Computing-II**

**Nature: Practical**

**Course Number: STA 529**

**Semester: Second**

**Total Lectures: 21-23 Practical Days (3 Hours / Lecture)**

**Full Marks: 100**

**Pass Marks: 50**

**Total Credits: 4**

**Total Duration: 64 hours**

**Table 4: Subject-wise allocation of number of practicals**

SN	Subject Area	No. of Practical (Minimum)
1	Multivariate Analysis	4
2	Stochastic Theory	4
3	Statistical Simulation	4
4	Python Programming	10
5	Two optional Papers of Semester-II	8
	Total	30

# **Third Semester**

**Course Title: Bayesian Inference**  
**Course Code: STA 631**  
**Nature: Theory**  
**Semester: Third**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture hours: 48**

### **Course Description**

This course is designed for master's level students and it serves as an introduction to the Bayesian methodology for conducting statistical inference in various practical scenarios. It covers the fundamentals of Bayesian inference theory and places significant emphasis on practical data analysis, primarily using the statistical software R and WinBUGS.

### **Learning Outcomes**

After completion of this course, the students will be able to

- grasp the differences in inferential procedure between Bayesian and frequentist methods and know how to deal with uncertainty through probabilities,
- acquire proficiency in conducting Bayesian analysis, Bayesian inference, and prediction using appropriate software,
- learn to construct and analyze Bayesian models, ranging from single-parameter models to hierarchical and generalized linear models and able to diagnose models, perform sensitivity analysis, and address model misspecification.

### **UNIT 1: Concepts and Fundamentals of Bayesian Statistics**

**[16 LH]**

Foundations of Bayesian Thinking: Probabilistic Thinking, Uncertainty and probability, Probability in different viewpoints Probability and Bayes' Theorem, Probability in Bayesian Perspective, Conditional probability and interpretation of Bayes' theorem, prior, likelihood and posteriors

Fundamentals of Bayesian Statistic: Basics of Bayesian paradigm, Motivating things for Bayesian Statistics, Difference in Bayesian and Frequentist methods, Basics of Bayesian Analysis, Maximum likelihood, Asymptotic, Invariance, Ancillarity, Independence and Exchangeability, Kernel and Sufficiency, Pre-posterior, Posterior and Predictive distributions, Principle of prior selection, objective and noninformative priors, standard reference priors; informative & conjugate priors. Posterior distribution using various non-conjugate priors, reference priors, and empirical priors; Conjugate analysis of binomial, Poisson, and Gaussian sample; Predictive analysis, Asymptotic properties of the posterior. (Practices using R/WinBUGS)

### **UNIT 2: Bayesian Inference and Computational Methods:**

**[16 LH]**

Bayesian Inferential Methods: Bayesian Inference and prediction, Bayesian estimation, Credible sets, HPD; Bayes Factor and Hypothesis testing; Empirical Bayes inference; Decision theory, Frequentist Risk, Bayes Risk, Bayesian Expected Loss; Lindley's paradox

Computational Methods: Review of sampling methods and MCMC: Rejection Sampling, Importance Sampling, Discrete Markov Chains, MH algorithm, MCMC with Gibbs Sampling, MCMC diagnostics

**UNIT 3: Bayesian Modeling and Model Assessment:**

**[16 LH]**

Bayesian Modelling: Single and multiparameter Models, Hierarchical models, Nuisance parameters: Bayesian linear regression and robust regression (using JAGS in R); Generalized linear models and examples (logistic, probit, Poisson), using STAN in R  
Bayesian hierarchical models: DAG, Hierarchical linear regression models, random effect BHM, Poisson-gamma and beta-binomial hierarchical models  
Model Assessment: Bayesian Model Specification, Model diagnosis, Sensitivity Analysis, test of Model Misspecification; Bayesian Model Averaging

**References**

1. Hoff, Peter D. (2009). *A First course in Bayesian statistical methods*. Springer.
2. Ntzoufras, I. (2009). *Bayesian modelling using WinBUGS*. Wiley.
3. Lee, P. M. (2012). *Bayesian statistics: An introduction* (4<sup>th</sup> ed.). Wiley.
4. Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
5. Gelman, A., Carlin, J., Stern, H., Dunson, D. B., Vehtari, A., & Rubin, D., (2013). *Bayesian data analysis* (3rd ed.). CRC Press.
6. Bansal, A. K. (2007). *Bayesian parametric inference*. Narosa Publishing House.

**Course: Research Methodology**

**Course Code: STA 632**

**Nature: Theory**

**Semester: Third**

**Full Marks: 75**

**Pass Marks: 37.5**

**Total Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

This course is a Three Credit Unit (48 hours) course aimed at equipping students with knowledge and skills of research philosophy and scientific research methodology for the purpose of designing research study. It also covers scientific research approach, process, research design, types of research, development of tools/techniques, measurement scales, collection and analysis of data, reports and presentations. Finally, this course is also premised on provision of sufficient information and knowledge to enable students acquire the skills to be able to formulate a relevant and acceptable research problem, make an educated choice of method(s), implement this/these method(s) and finally to write a scientific output.

### **Learning Outcomes**

After completion of this course, the students will be able to

- critically analyze a situation and formulate relevant research problems and also can be expressed, and defined in a professional way,
- make an informed choice of methods from the relevant research paradigm(s) associated to the specified research problem and also develop skills to make effective use of the library and e-resources,
- develop a research proposal/project plan and justify the theory with methodological decisions, including, sampling and measurements, and also gain experience with tools/techniques, data collection and their analysis, interpretations and presentations,
- understand relationship between theory and research, and limitations of particular research methods, and develop advanced critical thinking skills and also carry out research work independently.

### **UNIT 1: Scientific, Approach of Research Methodology**

**[11 LH]**

Concept and Nature of Scientific Research, Process and Approach of Scientific Research, Framework of the Philosophical Assumptions, Research Philosophy (positivism, critical realism, interpretivism, post modernism and pragmatism, Research Approach (inductive, deductive and abductive), Methodological Choice (qualitative, quantitative and mixed methods), Research Strategies (experimental, survey, case study, ethnography, narrative, grounded theory, action research, and archival), Time Horizon (cross-sectional and longitudinal) , Data Collection and Analysis

**UNIT 2: Research Design and Types of Research** [11 LH]

Concept and Nature of Research Design, Selection of Appropriate Research Design, Quantitative Research Designs (Experimental Design, Correlational Design, Survey Design, Action Research Design and Mixed Method Research Design)., Qualitative Research Designs (Grounded Theory Design, Ethnographic Design, Narrative Research Design, Action Research Design and Mixed Method Research Design)., Types of Research, Quantitative, Qualitative and Mixed Research, Longitudinal and Cross-sectional Research, Action and Participatory Action Research, Survey and Case Study Research, Field Studies, Field Experimental and Laboratory Research, Exploratory, Explanatory, Casual, Comparative and Policy Research, Theory-testing and Theory-building Research

**UNIT 3: Techniques/Tools Development and Data Collection Methods** [5LH]

Sampling Plans and Strategies, Development of Techniques/Tools for Data Collection, Review of Data Collection Methods (Observation, Interview, Questionnaire, Schedules, etc.)., Nominal Group Technique, Delphi Method, Snowball Sampling Method, Focus Group Discussion, Rural Rapid Appraisals and Participatory Rural Appraisals and In-depth-interview

**UNIT 4: Measurements and Scales** [12 LH]

Concept of Measurement and Scale, Review of Measurement scales (Nominal, Ordinal, Interval and Ratio scales, Reliability and Validity, Concept of Reliability and Validity, Test and Measure of Reliability and Validity, Content Validity, Criterion Validity and Construct Validity, Estimation of True Score of the Test, Applications and Construction of Social Scales, Concept and Nature of Social Scales, Scales used in Measuring- Mental Health, Stress and Life Changes Experiences; Social Conflict; Epidemiological Depression and Social Reaction, Scales used in Measuring social phenomena (Likert Scale, Bradley-Terry Model, Thurstone Model, Bogardus Scale and Guttman Scale, Theory of Sociometry, Semantic Differentials and Q-sort

**UNIT 5: Ethical Concerns in Research** [2 LH]

Research Ethics, Plagiarism, Intellectual Property Rights

**UNIT 6: Scientific Writing and Presentations** [7 LH]

Overview on Research Activities, Research Proposals and Grants Research Proposals, Research Report, Academic Thesis/Dissertation, Scientific Research Paper, Presentations of Research Documents

## References

1. Abbas, T., & Charles, T. (2002). *Handbook of mixed methods in social and behavioral research*. Sage Publications.
2. Aryal, T. R. (2008). *Research methodology*. Paluwa Prakashan Limited.
3. Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE publication.
4. Donna, M., & Pauline, E. G. (2008). *The handbook of social research ethics*. Sage Publications.
5. Drapper, N., & Smith, H. (1968). *Applied regression analysis*. John Wiley and Sons.
6. Day, R. A., & Gastel, B. (2006). *How to write and publish a scientific paper* (6<sup>th</sup> ed.), Cambridge University Press.
7. Gronmo, S. (2019). *Social research methods: Qualitative, quantitative and mixed methods approaches*. Sage Publications.
8. Kerlinger, F. N. (1983). *Foundations of behavioral research*. Surjeet Publication.
9. Kumar, R. (2019). *Research methodology: A step-by-step guide for beginners* (5<sup>th</sup> ed.). Sage Publications.
10. Mishra, S. B., & Alok, S. (2017). *Handbook of research methodology*. Edu-creation.
11. Novikov, A. M., & Novikov, D. A. (2012). *Research methodology: From philosophy of science to research design*. CRC Press.
12. Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Rowman & Littlefield.
13. Richardson, J. (2002). *Handbook of qualitative research methods for psychology and the social sciences*. Blackwell Publishing Co.
14. Singh, M. L. (1999): *Understanding research methodology*. Kathmandu.
15. Strauss, A., & Corbin, C. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publication.
16. Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. (4-8<sup>th</sup> eds.). Pearson.
17. Thomas, C. G. (2021). *Research methodology and scientific writing*. Springer.

**Course Title: Design of Experiments**

**Course Code: STA 633**

**Nature: Theory**

**Semester: Third**

**Full Marks: 75**

**Pass Marks: 37.5**

**Total Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

The course is designed to impart and enhance knowledge and understanding (theoretical and applied), and computational methods of experimental designs and analysis along with their applications. The course mainly covers topics in various aspects of basic experimental designs, nested and split plot designs, balanced incomplete block designs (BIBD), fractional factorial designs and response surface methodology

### **Learning Outcomes**

After completion of the course the students will be able to

- understand, design and apply experimental designs to real data with enhanced computational skill,
- carryout analysis of data generated from execution of experimental designs,
- confidently apply nested, split-plot, BIBD, fractional factorial, and response surface method with theoretical understanding.

### **UNIT 1: Basic Designs**

**[7 LH]**

Linear Models: Fixed effects, random effects and mixed effects models, computation of expected sum of squares, construction of ANOVA table, multiple comparisons, hypothesis testing and confidence intervals with contrasts, orthogonal contrasts

### **UNIT 2: Nested and Split Plot Designs**

**[10 LH]**

Nested Design: The two stage nested design, crossed and nested factors, design layout and model specification, estimation of model parameters, computation of sum of squares, construction of ANOVA table, applications of nested design

Split Plot Design: Design layout and model specification, least square estimation of parameters, computation of sum of squares, construction of ANOVA table, applications of split design

### **UNIT 3: Incomplete Block Designs**

**[10 LH]**

Balanced Incomplete Block Design: Design layout and model specification, intra-block estimation, computation of sum of squares (adjusted and unadjusted), recovery of inter-block information

Other Incomplete Block Designs: Connectedness, partially balanced incomplete design and Youden Squares

**UNIT 4: Two-level Factorial and Fractional Factorial Designs****[15 LH]**

$2^k$  factorial design and its analysis: Confounding in  $2^k$  Design, Confounding  $2^k$  design in two, four and  $2^p$  blocks, construction of blocks, partial and complete confounding

Fractional Factorial Designs: Two-level fractional factorial design and its analysis, alias structure and design resolutions, one-half and one-quarter fractions of the  $2^k$  design, the general  $2^{k-p}$  fractional factorial design

**UNIT 5: Response Surface Methodology****[6 LH]**

The first and second order response surface models, method of steepest ascent, location of stationary point, fitting and analysis of response surfaces

**References**

1. Montgomery, D. C. (2003). *Design and analysis of experiments*. John Wiley and Sons.
2. Charles, R. Hicks, & Turner, K. V. Jr. (1999). *Fundamental concepts in the design of experiments*. Oxford University Press.
3. Cochran, W. G., & Cox, G. M. (1992). *Experimental designs* (2<sup>nd</sup> ed.). John Wiley and Sons, Inc.

**Course Title: Biostatistics**  
**Course Code: STA 634**  
**Nature: Theory**  
**Semester: Third**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture Hours: 48**

### **Course Description**

The course focuses on history and development of epidemiology, different measures of health and illness, different study designs used in epidemiological and clinical studies with their relative advantages and disadvantages. It also deals with the major epidemiological issues such as bias, interaction, confounding, and causation of disease, methods of evaluation of confounding and interaction effects. Further, different aspects of executing RCT are incorporated. Besides these, different survival analysis techniques including Proportional Hazards (CPH) model are also discussed. Detail discussions about the use of appropriate statistical analysis for epidemiological and clinical data with the help of appropriate statistical software are also be done.

### **Learning Outcomes**

After completion of this course, the students will be able to

- be familiar with history & development of epidemiology, different measures of health & illness,
- recognize major epidemiological issues such as bias confounding and interaction, and discuss the different measures to control them,
- identify necessary and sufficient causes, causal pathways, factors in causation, and process of establishing the causes of a disease,
- apply appropriate biostatistical tools and models compatible to the study designs in epidemiology,
- know how to execute RCT following standard protocol,
- handle time to event data for estimating survival function & hazard functions, and apply CPH model,
- have hands on proficiency for performing statistical analysis of epidemiological and clinical data.

### **UNIT 1: Introduction to Epidemiology**

**[8 LH]**

Introduction: Basic epidemiologic concepts and principles, history and development of epidemiology, modern epidemiology, role of epidemiology in public health

Measures of Health and Illness: Measures of occurrence of disease and other health related events, point and period prevalence, incidence, relationship among prevalence, incidence and duration of disease, morbidity, attributable risk, population attributable risk, risk difference, person time

## **UNIT 2: Study Designs and Statistical Analysis in Epidemiological Research [20 LH]**

Study Design in Epidemiological Research: Observational and experimental study design, case report, case series, cross sectional studies, case control studies, selection of cases and controls, matched case control studies, nested case control studies, cohort studies, historical cohort studies, ecological studies, advantages and disadvantages of each study design, choice of a study design  
Issues in Epidemiological Studies: Systematic error, random error, selection bias, measurement bias, confounding, interaction, methods of controlling confounding, Simpson's Paradox, stratification, internal validity, external validity

Causation in Epidemiology: Basic concept of cause, necessary and sufficient causes, causal pathways, factors in causation, process of establishing the causes of a disease, Hill's criteria, guidelines for causation

Statistical Analysis for Epidemiological and Clinical Studies: Use of appropriate statistical analysis according to study design; measures of exposure effects: Odds Ratio (OR), Risk Ratio (RR), respective confidence interval, and their problem specific interpretation; evaluation of confounding and interaction: Mantel Hanszel (MH) OR & MH RR with respective confidence interval, application of regression models in epidemiological and clinical data

## **UNIT 3: Clinical Trials**

**[7 LH]**

Introduction: Experimental study design and its importance, importance of Randomized Controlled Trial (RCT), field trial, community trial

Type and Phases of Clinical Trials: Therapeutic trials, prophylactic trials, controlled and uncontrolled trials, randomized and non-randomized trials, superiority and inferior trials, equivalence trials, phases of trial

Design of RCT: parallel design (block randomization and stratified randomization), cross-over design, factorial design and adoptive design, concept of blinding, masking, concealment, issue of size of RCT

Other Aspects of Clinical Trial: Ethical issues in clinical trials, preparation of trial protocol, compliance, registration of RCT, conduct of clinical trials - single centric and multi-center trials, Consolidated Standards of Reporting Trials (CONSORT) guidelines, role of data safety and monitoring board (DSMB) in conducting clinical trials

## **UNIT 4: Survival Analysis**

**[13 LH]**

Introduction: History and development of survival analysis, need and importance of survival analysis over the standard statistical analysis techniques, concept of event, censoring, right censoring, left censoring, interval censoring, reasons of censoring, structure of time to event data.  
Estimation of Survival Function and Hazard Function: survival function, probability density function, hazard function, relationship between them, general features of exponential, Weibull and lognormal survival distribution, bathtub curve, Kaplan-Meier(K-M) estimate of survival function, life table estimate of survival function, Nelson-Aalen estimate of the survival function, K-M estimate of the hazard function, estimation of median & percentiles of survival times with

standard error, construction of K-M survival curves and interpretations, applications of these techniques in clinical data

Comparison of Survival Curves: Comparison of survival experiences between two or more than two groups of survival data and test of statistical significance- Log-rank test, Gehan's generalized Wilcoxon test, Tarone-Ware test, Peto test, and their applications in analyzing clinical data

Regression Modeling for Time to Event Data: Need of regression analysis in handling time to event data, Cox Proportional Hazards (PH) Model with one and more than one independent covariates, interpretation of Hazard Ratios (HR), testing the proportionality of hazards assumptions, applications of survival analysis in clinical data

## References

1. Collett, D. (2003). *Modeling survival data in medical research*. Chapman and Hall/CRC.
2. Friedman, L.M., Furburg, C., & Demets, D.L. (1998). *Fundamentals of clinical trials*. Springer Verlag.
3. Hennekens, C.H., & Buring, J.E. (1987). *Epidemiology in medicine* (L.M. Sherry, Ed.). Little, Brown and Company Boston.
4. Hosmer, D.W., & Lemshow, S. (1999). *Applied survival analysis: Regression modeling of time to event data*. John Wiley and Sons.
5. Mathews, J. N. S. (2006). *Introduction to randomized controlled clinical trials*. Chapman and Hall/CRC.
6. Lee, E. T., & Wang, J.W. (2003). *Statistical methods for survival data analysis*. John Wiley & Sons.
7. Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology*. Lippincott Williams and Wilkins.

**Course Title: Environmetrics**

**Course Code: STA 635**

**Nature: Theory**

**Semester: Third**

**Full Marks: 75**

**Pass Marks: 37.5**

**Total Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

The course is designed to acquaint students with specific statistical methods useful to address and analyze environmental health phenomena and associated problems which are often complex in nature. In order to understand and capture such complex phenomena, the course includes topics on introduction of Environmetrics, environmental health related variables and study designs, models based upon generalized least squares, nonlinear models based upon nonlinear least squares, exponential family of distributions and generalized linear models, and some specific environmental impact assessment models widely applied in environmental and environmental health studies.

### **Learning Outcomes**

After completion of the course the students will be able to

- perform specific statistical analysis concerning generalized least squares (GLS), nonlinear models, generalized linear models (GLMs) and impact assessment models,
- understand and derive theoretically to obtain GLS, nonlinear least squares (NLS) and iteratively reweighted least squares (IRLS),
- apply specific environmetric methods and techniques to environmental health related data,
- execute analysis manually from spreadsheet (Excel) without support of programmed software through enhanced computational skills, and verify results through programmed software (SPSS).

### **UNIT 1: Introduction, Environmental and Environmental Health Variables, and Studies [8 LH]**

Environmetrics: Introduction, origin and its historical development, environmental epidemiology and its applications in public health issues

Environmental Pollution: Environmental pollution, pollutants and their impacts on human health, climate change and health risks

Environmental and Environmental Health Variables/Statistics: Risk assessment and outcome assessment variables and statistics, confounders, qualitative and quantitative measures, direct and indirect measures, indicators, biomarkers, environmental monitoring, sources of environmental and related data, national and international standards of pollutant levels

Environmental Health Study Designs: Descriptive studies, analytical studies, ecological studies, cohort studies, cross-sectional and longitudinal studies, experimental or intervention studies,

case-crossover design: case and control times, illustrations related to environmental health phenomena

**UNIT 2: Transformations and Generalized Least Squares** [8 LH]

Transformations in Models: Transformations to skewed distributions, variance stabilizing transformations, transformations to linearize models, Box-Cox transformation

Generalized Least Squares (GLS): Definition and derivation of generalized least squares, use of transformations in GLS, properties

Weighted Least Squares (WLS): Definition, condition under which generalized least square is a weighted least square, choice of weights

Illustrations of applications in environmental health issues

**UNIT 3: Nonlinear Models** [10 LH]

Introduction: Definition, nonlinear functions: exponential, sigmoid, Michaelis-Menten, etc. with their graphical illustrations; differences between linear and nonlinear models, assumptions, model specification, interpretation of parameters

Inference: Nonlinear least squares, use of Taylor's series expansion for nonlinear functions, linearization (Gauss-Newton) method of parameter estimation (with derivation), Marquardt compromise, methods of selection of starting values, interval estimation, bootstrap confidence intervals, construction of ANOVA table; Pseudo  $R^2$ , residual analysis

Illustrations of applications in environmental health issues

**UNIT 4: Generalized Linear Models** [16 LH]

Exponential Family of Distributions: Probability density function, derivation of moments (mean and variance), members of exponential family of distributions: normal, binomial, Poisson, negative binomial, exponential, gamma, etc.

Generalized Linear Models (GLM): Introduction: Definition, transformations versus GLM, canonical link functions: identity, log, logit, inverse, inverse square, models for different canonical links: Poisson and logit models and their uses in environment related studies

Inference: Maximum likelihood estimation, iteratively re-weighted least square (IRLS) estimation (with derivation), interpretation of parameter coefficients, interval estimation, tests of significance, model deviance

Residual Analysis: Types of residuals, raw, deviance and Pearson residuals, deviance and Pearson statistics, over-dispersion, equi-dispersion, under-dispersion, analysis of deviance, Omnibus test, pseudo  $R^2$ , Illustrations of applications in environmental health studies

**UNIT 5: Environmental Pollution Assessment Models** [6 LH]

Dose-response models: Introduction, dose-response curve, functional forms, Richard function, Hill function model: model specification and interpretations, estimation of parameters

Air Dispersion Models: Introduction, Gaussian plume model: model specification, characteristic features, meteorological conditions, dispersion coefficient, plume rise  
Illustrations of applications in environmental health studies

## References

1. Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications. Wiley series in probability and statistics*. John Wiley & Sons, Inc.
2. Cameron, A. C., & Trivedi, P. K. (2008). *Regression analysis of count data*. Cambridge University Press.
3. Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear models*. CRC Press.
4. McCullagh, P., & Nelder, J. (2000). *Generalized linear models*. Chapman & Hall/CRC.
5. Merrill, R. M. (2010). *Environmental epidemiology: Principles and methods*. Jones and Bartlett.
6. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2013). *Introduction to linear regression analysis*. Wiley-Blackwell

**Course Title: Time Series Analysis**

**Course Code: STA 636**

**Nature: Theory**

**Semester: Third**

**Full Marks: 75**

**Pass Marks: 37.5**

**Total Credits: 3**

**Total Lecture Hours: 48**

### **Course Description**

This course provides tools for empirical work with time series data and is an introduction to the theoretical foundation of time series models. In this course, we will concentrate on both univariate and multivariate time series analysis, with a balance between theory and applications. Students are expected to prepare a project report on real-life data.

### **Learning Outcomes**

After completing this course, a student will be able to

- differentiate between stationary and nonstationary time series,
- analyze univariate and multivariate time series data,
- formulate and analyze different time series models with real data,
- analyze panel data model,
- use available software as well as pursue research using time series data,
- apply theory to real (or simulated) data using R.

#### **UNIT 1: Nature of the Time Series and Tests**

**[12 LH]**

Introduction of time series and its components, stationarity and non-stationary time series detection and testing, graphical method, auto-correlation and partial autocorrelation function, random walk, Ljung-Box test, unit root problem, testing for unit root, Dickey-Fuller test and augmented Dickey-Fuller test, transforming non-stationary time series, differencing. Spurious regression problem. Testing for structural change, Chow test

#### **UNIT 2: Time Series Models**

**[16 LH]**

Auto-Regressive (AR) model, Moving Average (MA) model, Auto-Regressive and Moving Average (ARMA) model, Integration of time series, Auto Regressive Integrated Moving Average (ARIMA) model. Model specification, estimation, diagnostic checking and forecasting in ARIMA model, Box-Jenkins methodology

Co-integration: definition and testing, Eagle-Granger (EG) cointegration test

Vector Autoregressive (VAR) model: model specification, estimation and forecasting with VAR model and its application. Causality tests, Granger Causality Test

#### **UNIT 3: ARCH and GARCH Models**

**[10 LH]**

Autoregressive Conditional Heteroscedasticity (ARCH) Model: Auto-Regressive Conditional Heteroscedastic (ARCH) model, ARCH (q) model, model specification, estimation and testing

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Model, GARCH (p, q) model, model specification and estimation

#### **UNIT 4: Panel Data Model**

**[10 LH]**

Definition, usefulness and type of panel data, Panel Data models; the constant coefficients model, the fixed effect model, the random effect model, choosing between FEM and REM: the Hausman test, Estimation of panel regression model and interpretation of results, use statistical software

#### **References**

1. Gujarati, D. N, Porter, D. C., & Gunasekar, S. (2011). *Basic econometrics*. Mc Graw Hill Education.
2. Bhaumik, S. K. (2015). *Principles of econometrics*. Oxford University Press.
3. Enders, W. (2014). *Applied econometric time series* (4<sup>th</sup> ed.). John Wiley and Sons.
4. Ramanathan, B. (2002). *Introductory econometrics with applications*. South-Western Thomson Learning.
5. Madala, G. S. (2001). *Introduction to econometrics*, John Wiley and Sons.

**Course Title: Actuarial Statistics**  
**Course Code: STA 637**  
**Nature: Theory**  
**Semester: Third**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture Hours: 48**

### **Course Description**

The course is designed to impart knowledge, foster understanding, and facilitate the application of concepts related to insurance, as well as statistical and actuarial principles. It encompasses a comprehensive study of life insurance, life annuities, and health insurance. Furthermore, the curriculum delves into critical subjects such as risk assessment, mortality tables, and exposure formulas.

### **Learning Outcomes**

After completion of this course, the students will be able to

- know about the actuarial science, insurance and life insurance,
- know about the actuarial statistics, survival distributions and life tables, risk and mortality table,
- know about the different exposure formulas.

### **UNIT 1: Insurance and Related Statistics**

**[18 LH]**

Introduction of actuarial sciences, nature and functions of insurance, benefits and costs of insurance system to the society; economic theories of insurance; the mathematical basis for insurance; insurable interest

Life Insurance: Essential features of life insurance contract; Risk selection for life insurance; Sources of risk information

Life annuities: Single payment, continuous life annuities, discrete life annuities, life annuities with monthly payments, commutation functions, varying annuities, recursions, complete annuities-immediate and apportionable annuities-due. Health insurance: Types of health insurance coverage; Exclusion in health insurance policies, payment of claim, Other schemes of insurance

### **UNIT 2: Actuarial Statistics**

**[30 LH]**

The economics of insurance, utility theory, application of probability to problems of life and death, de-termination of single premiums for insurances and annuities, theory and practice of pension funding, assumption, basic actuarial functions and population theory applied to private pensions

Survival distributions and life tables, life insurance, life annuities, net premium, premium series, multiple life functions, multiple decrement models, valuation theory for pension plans, the expense function and dividends

Risk and Mortality Table: Mortality tables and its classification; construction of mortality tables; premium calculation of various life policies

Exposure formulas: Techniques of calculating exposures from individual records including consideration involving selection of studies, various observation periods and various methods of tabulating deaths, techniques of calculating exposures from variation schedules, use of interim schedules and variations in observation period or method of grouping deaths and practical aspects of construction of actuarial tables

### **References**

1. Dorfman, M. S. (1991). *Introduction to risk management and insurance*. Prentice Hall.
2. Mishra, M. N. (1989). *Insurance principles and practice*. S. Chand and Company.
3. Atkinson, M. E., & Dickson, D. C. M. (2000). *An introduction to actuarial studies*. Elgar Publishing.

**Course Title: Survival Analysis**  
**Course Code: STA 638**  
**Nature: Theory**  
**Semester: Third**

**Full Marks: 75**  
**Pass Marks: 37.5**  
**Total Credits: 3**  
**Total Lecture Hours: 48**

### **Course Description**

The course deals with the different survival analysis techniques. Details about the cox proportion hazards model, tests of proportionality assumptions, goodness of fit of the model and model diagnostics are incorporated. Further, different parametric survival models and Accelerated Failure Time (AFT) models and their fittings, residual analysis and their applications are also discussed.

### **Learning Outcomes**

After completion of this course, the students will be able to

- know about the survival analysis,
- understand the cox proportional hazards regression model,
- perform parametric proportional hazards model,
- do accelerated failure time models.

### **UNIT 1: Review of Basic Survival Analysis Terms and Techniques [5 LH]**

Need and importance of survival analysis in clinical research, concept of event, censoring, reasons of censoring, estimation of Kaplan-Meier (K-M) survival functions, hazard functions, survival times, K-M survival curves, Log-rank test

### **UNIT 2: Cox Proportional Hazards Regression Model [16 LH]**

Modeling the hazard function, Cox Proportional hazards (PH) model, linear component of Cox PH model, fitting of Cox PH model, confidence intervals and hypothesis tests for regression coefficients, strategy for model selection, interpretation of parameter estimates, estimation of hazards and survival functions, assumptions of Cox PH model, tests of proportionality of hazards assumption- graphical method, test based on Schoenfeld residuals, interaction with time  
Assessment of model adequacy of Cox regression model: Residuals for the Cox regression model, assessment of the model fit, identification of influential observations, overall goodness of fit

Extended Cox regression model: Stratified proportional hazards model, time varying covariates

### **UNIT 3: Parametric Proportional Hazards Model [10 LH]**

Parametric proportional hazards model: Exponential distribution, Weibull distribution; Assessing the suitability of a parametric model, fitting of a parametric model to a single sample, model for the comparison of two groups, Weibull proportional hazards model, fitting of Weibull proportional hazards model, log-linear form of the model, Gompertz proportional hazards

model, tests of proportionality of hazards assumption in parametric PH models, residual analysis, goodness of fit of the model, model selection

#### **UNIT 4: Accelerated Failure Time Models**

**[17 LH]**

Probability distributions for survival data: Log-logistic distribution, lognormal distribution, gamma distribution, inverse Gaussian distribution

Exploratory analysis for the selection of appropriate model

Accelerated Failure Time (AFT) models: Concept of AFT, AFT model for comparing two groups, general AFT model, log-linear form of AFT model, interpretation of parameter estimates and measures, difference between PH metric and AFT metric in survival analysis

Parametric AFT models: Weibull AFT model, Log-logistic AFT model, lognormal AFT model

Residual analysis for parametric models: Standardized residuals, Cox-Snell residuals, deviance residuals, score residuals

#### **References**

1. Collett, D. (2003). *Modeling survival data in medical research*. Chapman and Hall/CRC.
2. Klein J. P., & Moeschberger, M. L. (2003). *Survival analysis techniques for censored and truncated survival data*. Springer Publication.
3. Hosmer D. W., & Lemshow, S. (1999). *Applied survival analysis: Regression modeling of time to event data*. John Wiley and Sons.
4. Therneau, T. M., & Grambsch, P. M. (2001). *Modeling survival data: Extending the Cox model*. Springer Publication.
5. Lawless, J. F. (2003). *Statistical models and methods for lifetime data*. John Wiley & Sons Inc Publication.

**Course Title: Statistical Computing-III**

**Nature: Practical**

**Course Number: STA 639**

**Semester: Third**

**Total Lectures: 21-23 Practical Days (3 Hours/ Lecture)**

**Full Marks: 100**

**Pass Marks: 50**

**Total Credits: 4**

**Total Duration: 64 hours**

**Table 5: Subject-wise allocation of number of practicals**

SN	Subject Area	No. of Practical (Minimum)
1	Bayesian Inference	6
2	Research Methodology	6
3	Design of Experiment	6
4	Two optional papers of Semester III	12
	Total	30

# **Fourth Semester**

**Course Title: Meta-analysis**  
**Course Code: STA 642**  
**Nature: Theory**  
**Semester: Fourth**

**Full Marks: 50**  
**Pass Marks: 25**  
**Total Credits: 2**  
**Total Lecture Hours: 32**

### **Course Description**

The course deals with the principles of systematic review, major steps & stages, standard methods of reporting systematic review & meta-analysis, quality assessment, and the significance of systematic review over narrative review of literature. Need, importance and principles of meta-analysis, measures of summary effect size under different scenario, different methods of assessing between study heterogeneity, fixed effect and random effects meta-analysis, different methods of publication bias are incorporated. At the end of course, discussions on the positive and negative aspects, standard format of reporting of meta-analysis along with the appraisal of relevant published journal articles in systematic review and meta-analysis will also be performed.

### **Learning Outcomes**

After completion of the course students will be able to

- be well acquainted with the principles, steps, stages and reporting of systematic review in a standard format,
- utilize the detail concepts of history, development & importance of meta-analysis,
- apply different methods of measuring summary effect size under different scenario,
- apply different methods of examining study heterogeneity, fixed effect and random effects meta-analysis, and methods of assessing the publication bias,
- analyze both positive and negative aspects of meta-analysis,
- prepare manuscript of systematic review in the relevant field, and develop good foundation to study forward in the area of meta-analysis.

### **UNIT 1: Systematic Review and Meta-analysis**

**[6 LH]**

Systematic review: Introduction, definition, major steps & stages of systematic review: scoping, planning, identification, screening, eligibility, deciding whether to carry out qualitative or quantitative research synthesis, presentation of systematic review; AMSTAR (Assessment of Multiple Systematic Reviews) and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses), quality assessment in systematic review, significance of systematic review over routine review of literature

Meta-analysis: Introduction to meta-analysis, history and development, meta-analysis in different field of research; elements of meta-analysis: individual studies, effect size, precision,

study weights, the summary effect, heterogeneity of effect sizes; the streptokinase meta-analysis, statistical significance vs. subject specific significance, consistency of effects, forest plot.

## **UNIT 2: Measures of Effect Size and Precision** [5 LH]

Introduction: Treatment effects and effect sizes, parameters and estimates, outline of effect size computations

Effect Size based on Means: Raw (unstandardized) mean difference (D), standardized mean difference (d and g), for independent groups and for match groups, response ratios

Effect Size based on Binary Data: Effect size based on Risk Ratio (RR), Odds Ratio (OR), risk difference, choosing an effect size index

Effect sizes based on correlations; factors affecting precision: sample size, study design

## **UNIT 3: Assessment of Study Heterogeneity** [6 LH]

Statistical Test for Assessing Heterogeneity: Cochran's  $Q$  test and interpretation of results based on it, extensions/alternative test ( $I^2$  test) and its interpretation, advantages of  $I^2$  test over Cochran's  $Q$  test

Graphical Assessment of Heterogeneity: Plot of normalized (z) score, forest plot, Radial Plot (Galbraith diagram), L'Abbe plot Possible causes of heterogeneity, methods of investigating and dealing with sources of heterogeneity, changing the scale of outcome variable, exclude studies, use of random and fixed effect models, validity of pooling studies with heterogeneous outcomes

## **UNIT 4: Fixed Effect versus Random Effects Model** [6 LH]

Introduction: Basic concepts of fixed effect model and random effects model in meta-analysis, true effect, observed effect, combined effect Fixed Effect Meta-Analysis: Performing a fixed effect meta-analysis, true effect size, estimating summary effect size

Random Effects Meta-analysis: Performing random effects meta-analysis, true effect size, estimating summary effect size, estimating between study variance ( $\tau^2$ )

Fixed effect vs. random effects, extreme effect size in a large study or a small study, confidence interval, model selection

## **UNIT 5: Publication Bias** [4 LH]

Publication Bias: Introduction, evidence of publication and related bias, seriousness and consequences of publication bias in meta-analysis, most common reasons of publication bias, predictors of publication bias, Methods for Identifying Publication Bias: Funnel plots, caveats of funnel plots, Rosenthal's Fail-safe N and its appropriateness, Orwin's Fail-safe N, Duval and Tweedie's Trim and Fill method, cumulative forest plot

## **UNIT 6: Criticism and Reporting the Results of Meta -Analysis**

**[5 LH]**

Positive and negative aspects of meta-analysis, standard format of reporting a meta-analysis, appraisals of relevant published journal articles on systematic review and meta-analysis, statistical software used in meta-analysis

### **References**

1. Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd.
2. Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.
3. Lipsey, M. W., & Wilson, D. (2000). *Practical meta-analysis*. Sage Publication.
4. Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. John Wiley & Sons, Ltd.

**Course Title: Nonparametric Modeling**

**Course Code: STA 643**

**Nature: Theory**

**Semester: Fourth**

**Full Marks: 50**

**Pass Marks: 25**

**Total Credits: 2**

**Total Lecture Hours: 32**

### **Course Description**

The course is design to impart knowledge, understanding, computation skill and applications of essential techniques of nonparametric modeling related to scatterplot smoothing. It contains basic to complex smoothing methods such as bin smoothers, running mean and running line smoothers, kernel smoothers, locally weighted regression smoothers, local polynomial regression, regression splines and cubic smoothing splines, and cross validation.

### **Learning Outcomes**

After completion of the course the students will be able to

- understand the importance and applications of nonparametric regression methods,
- build nonparametric models with theoretical understanding related to scatterplot smoothing and interpret the results,
- perform nonparametric density estimation,
- create necessary plots and perform model adequacy tests with the support of goodness of fit statistics and cross validation,
- execute analysis manually from spreadsheets (Excel) without the support of programmed software by enhanced computational skills.

### **UNIT 1: Introduction to Nonparametric Modeling**

**[5 LH]**

Parametric versus nonparametric regression: merits and demerits, smoother, its importance and need in data modeling, scatterplot smoother: smoothing parameter, bandwidth, span, the bias and variance tradeoff, goodness of fit statistics, types of nonparametric smoothers: unweighted and weighted smoothers, introduction to generalized additive models, illustrative examples

### **UNIT 2: Basic Unweighted Smoothers**

**[6 LH]**

Bin smoothers, neighborhood: Local, nearest and symmetric neighborhoods; running mean smoother, running line smoother, their local averaging techniques, span, smoothing parameter and goodness of fit tests, illustrative examples and numerical computations using spreadsheet

### **UNIT 3: Weighted Smoothers**

**[10 LH]**

Nonparametric density estimation, Kernel density estimation, Kernel functions: Uniform, Gaussian, triangular, bi-weight, Epanechnikov, bandwidth and its selection

Kernel smoothing: Locally weighted averaging, Kernel functions, bandwidth, weights and smoothing; Locally weighted regression smoother (Loess): k nearest neighborhood, span, tri-

cubic weight function, effective degrees of freedom, numerical computations and verifications using spreadsheet and statistical software, respectively

**UNIT 4: Splines and Local Polynomial Regression** [8 LH]

Regression splines: piecewise polynomials, interior knots, smoothing function, estimation of parameters, local polynomial regression: kernel functions as weights, cubic smoothing splines, penalized least squares, numerical computations and verifications using spreadsheet and statistical software, respectively

**UNIT 5: Cross-Validation** [3 LH]

Selection of span (or smoothing parameter) using cross-validation approach, application of cross-validation for different smoothers, generalized cross validation, numerical computations using spreadsheet

**References**

1. Hastie, T., & Tibshirani, R. J. (1999). *Generalized additive models*. Chapman & Hall/Crc.
2. K. Takezawa. (2005). *Introduction to nonparametric regression*. John Wiley & Sons.
3. Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach*. Chapman & Hall.
4. Wolfgang Hardle. (1999). *Applied nonparametric regression*. Cambridge University Press.

**Course Title: Population and Financial Statistics**

**50 Course Code: STA 644**

**25**

**Nature: Theory**

**Semester: Fourth**

**Full Marks:**

**Pass Marks:**

**Total Credits: 2**

**Total Lecture Hours: 32**

### **Course Description**

This course is a Two Credit Unit (32 hours) course aimed at equipping students with knowledge and skills of population and financial statistics. This course is a course that teaches students how to collect, analyze, interpret and presents the data relating to population and financial sectors using statistical methods/tools. Students may learn about topics such as principles of population and financial statistics, models and indices relating to population and financial statistics and their applications.

### **Learning Outcomes**

After completion of this course, the students will be able to

- pursue careers in the population and finance sectors and also as a stepping stone towards research activities in statistics for population and finance,
- analyze and critically interpret the data, build statistical models of real situation, and use relevant statistical software packages for the purpose of research along with policy interventions independently.

### **UNIT 1: Population Statistics**

**[16 LH]**

Concept and Nature of Population Statistics, Concept and Nature of Population Statistics, Fundamental Objectives and Importance of Population Statistics, Basic Principles of Official Statistics. Some Models and Indices on Population Statistics (fertility, mortality, migration, family planning and human development). Applications in real Data sets and their Interpretations

### **UNIT 2: Financial Statistics**

**[16 LH]**

Concept and Nature of Financial Statistics, Basic Objectives and Significance of Financial Statistics (monetary and financial, Fundamental Principles of Financial Statistics, Some Models and Indices of Financial Statistics in the Market/Business Cycles (Fibonacci Number, Dow Theory, Elliate Wave Theory, Wychoff Method and Ichimoku Cloud Technique)., Applications in real Data sets and their Interpretations

## References

1. Aryal, T. R. (2011). *Fertility dynamics of Nepal*. Ekta Book Distributors.
2. Aryal, T. R. (2010). *Nuptiality*. Gyankunja Prakashan.
3. Aryal, T. R. (2011). *Mortality of Nepal*. Prime Publication.
4. Aryal, T. R. (2008). *Migration and occupational mobility in Nepal*. Paluwa Prakashan.
5. Central Bureau of Statistics. (2014). *Population monograph of Nepal volume I: Demographic analysis*.
6. Central Bureau of Statistics. (2014). *Population monograph of Nepal volume II: Social demography*.
7. Central Bureau of Statistics. (2014). *Population monograph of Nepal volume III: Economic demography*.
8. National Statistics Office. (2023). *National population and housing census 2021*.
9. Dobson, E. D. (1984). *Understanding Fibonacci numbers*. Traders Press Inc.
10. Frost, A. J., & Prechter, R. R. (1997). *Elliott Wave Principle: Key To Market Behavior*. New Classics Library.
11. Lai, T.L., & Xing, H. (2008). *Statistical models and methods for financial markets*. Springer.
12. Miller, M. B. (2013). *Mathematics and statistics for financial risk management*, (2<sup>nd</sup> ed.). John Wiley & Sons, Inc.
13. Nikolai, V. N. (2011). *Fibonacci numbers (Dover books on mathematics)*. Dover Publications Inc.
14. Patel, M. (2011). *Trading with Ichimoku clouds: The essential guide to Ichimoku Kinko Hyo technical analysis*. Wiley, John Wiley & Sons, Inc.
15. Severini, T. A. (2017). *Introduction to statistical methods for financial models*. (17<sup>th</sup> ed.). Routledge N.Y. Publisher.
16. Tripathi, S. (2021). *Dow theory and its applications*.
17. United Nations. (2013). *Fundamental principles of official statistics*. Economic and Social Council.
18. United Nations. (2012). *Guidelines on integrated economic statistics*.
19. Weeks, J. R. (2016). *Population: Introduction to concepts and issues* (12<sup>th</sup> ed.). Cengage Learning publisher.
20. Weis, D. H. (2001). *Trades about to happen: A modern adaptation of the Wychoff methods*. Wiley.
21. Wilmott, P., Howison, S., & Dewynne, J. (1995). *The mathematical of financial derivatives: A student introduction*. Cambridge University Press.

**Course Title: Geographical Information System**

**Full Marks: 50**

**Course Code: STA 645**

**Pass Marks: 25**

**Nature: Theory**

**Total Credits: 2**

**Semester: Fourth**

**Total Lecture hours: 32**

### **Course Description**

The course Geographical Information System (GIS) for MSc level students of statistics provides an introduction to GIS principles, spatial data analysis, GIS software usage, and its applications in statistics. Students will learn about spatial data manipulation, spatial statistics, visualization, and ethical considerations. Practical applications and integration with statistical analysis will be emphasized.

### **Learning Outcomes**

After completion of this course the students will be able to

- gain a solid understanding of GIS Fundamentals and data acquisition techniques,
- develop the ability to perform spatial data analysis, spatial interpolation, and spatial regression for statistical modeling and inference,
- acquire proficiency in using popular GIS software and tools, and develop practical skills,
- able to apply GIS techniques in various statistical research domains.

### **UNIT 1: Geospatial Science Fundamentals**

**[16 LH]**

Introduction to spatial data and GIS: Basics of spatial data its types and formats; Overview of GIS and its applications in statistics; GIS components and data types, (vector vs. raster), Sources of GIS data (satellite imagery, GPS, Surveys), Data acquisition techniques and platforms. Data Management and GIS software: Data storage formats, Introduction to GIS software (ArcGIS, QGIS); Data cleaning and integration techniques, Overview of Scripting and automation in GIS Spatial Analysis and spatial statistics: Spatial operations (buffering, overlay, proximity analysis), Exploratory spatial data analysis (ESDA), Spatial autocorrelation and spatial dependence, Spatial interpolation and smoothing, Point pattern analysis, Geo statistics (e.g., kriging), Concepts of Spatial regression and modeling, Relevance of spatial statistics in statistical analysis

### **UNIT 2: Geospatial Analysis and Applications**

**[16 LH]**

Spatial Data Visualization: Cartography and map design principles, Interactive mapping and web-based GIS, Web-based GIS platforms and applications Remote Sensing and Image Analysis: Basics of remote sensing and satellite imagery, Image processing and analysis in GIS. GIS Applications in Statistics: Examples of spatial epidemiology and disease mapping, Environmental statistics and spatial analysis, Urban and regional planning using GIS

Digitization Basics: Introduction to digitization and overview of digitization techniques in in GIS, Digitization of Point, line and polygon spatial data. Real world applications of GIS: Group projects integrating GIS and statistics, Case studies demonstrating GIS application and Preparation of Enumeration Area (EA) mapping for different surveys

## References

1. Bolstad, P. (2016). *GIS fundamentals: A first text on geographic information systems* (5<sup>th</sup> ed.). Eider Press.
2. De Smith, M. J., Goodchild, M. F., & Longley, P. A. (2018). *Geospatial analysis: A comprehensive guide* (3<sup>rd</sup> ed.). Matador.
3. Fu, P., & Sun, J. (2010). *Web GIS: Principles and applications*. ESRI Press.
4. Brunson, C., Comber, L., & Singleton, A. (2015). *Spatial analysis: Statistics, visualization, and computational methods*. SAGE Publications.
5. Weng, Q. (2015). *Remote sensing and GIS integration: Theories, methods, and applications*. McGraw-Hill Education.

**Course Title: Machine Learning**  
**Course code: STA 646**  
**Nature: Theory**  
**Semester: Four**

**Full Marks: 50**  
**Pass Marks: 25**  
**Total Credits: 2**  
**Total Lecture Hours: 32**

### **Course Description**

This course is highly focused on introducing the concepts of machine learning. It focuses on different machine learning techniques and models. It also emphasizes on evaluating different models as well. The course intends to impart practical knowledge about different learning techniques and models in Python. It focuses on supervised and unsupervised domains. Last but not the least, it also introduces the basic concepts of deep learning.

### **Learning Outcomes**

After completion of this course, the students will be able to

- be familiar with machine learning concepts,
- understand supervised and unsupervised learning,
- preprocess the data using Python,
- implement machine learning algorithms using Python,
- understand architectures of deep learning and implement using Python.

### **UNIT 1: Introduction**

**[2 LH]**

Motivation & Applications of Machine Learning, The Overview of Supervised Learning, Unsupervised Learning

### **UNIT 2: Data Preprocessing**

**[4 LH]**

Scaling Data, Dealing with Categorical Variables, Handling with Missing Value, Methods for Dealing with Imbalanced Data, Reducing Dimensionality: Principal Component Analysis (PCA) as dimensionality reduction, Feature extraction [Note: The Data Preprocessing Operations has to be done using Python]

### **UNIT 3: Mining Frequent Patterns, Associations, Rule Mining**

**[3 LH]**

Basic Concepts, Frequent Itemset Mining Methods: Apriori Algorithm, A Pattern-Growth Approach for Mining Frequent Item sets

### **UNIT 4: Supervised Learning**

**[10 LH]**

Gradient Descent, Batch Gradient Descent, Stochastic Gradient Descent, Coordinate Descent, The Concept of Underfitting and Overfitting, Decision Tree Induction for classification using ID3, Bayesian Belief Network, Basics of Artificial Neural Network, Perceptron and classification by back propagation, Implementation of Linear Regression and Logistic Regression [Note: All the algorithms have to be implemented in python]

**UNIT 5: Unsupervised Learning****[5 LH]**

Basic concepts of Self Organizing Maps, Independent Component Analysis (ICA) algorithm and its application, Implementation of clustering algorithms: K-means, K-medoids, Agglomerative, Divisive, DBSCAN in Python

**UNIT 6: Ensemble Methods for Machine Learning****[4 LH]**

Concept and Implementation of Bootstrap Aggregation, Random Forests Classifier and Regressor, Adaboost, and Gradient Boosting in python

**UNIT 7: Introduction to Deep Learning****[4 LH]**

Basic Concepts, Architectures of LSTM, GRU and their implementation in python

**References**

1. Christopher, B. (2006). *Pattern recognition and machine learning*. Springer.
2. Richard, D., Peter, H., & David, S. (2001). *Pattern classification*. John Wiley & Sons.
3. Tom, M. (1997). *Machine learning*. McGraw-Hill.
4. Andrea, G. (2020). *Applied machine learning with Python*. Logo s.r.l, Borgoricco (Padua), Milano.
5. Morgan, K., Han, J., & Kamber, M. (2022). *Data mining concepts and techniques*. Morgan Kauffmann.
6. Eugene, C. (2019). *Introduction to deep learning*. MIT Press Cambridge.

#### **4. Course Conduction Guidelines**

The course shall be implemented considering different methods of teaching and learning processes. Lecture method, illustrations, students' participation through discussion, assignments, class tests and presentations will be the key methodological aspects in transforming knowledge from teachers to students. Students' participation will be given top priority for building their confidence in understanding the subject matter and its application. Computational exercises will be executed manually through spreadsheets (Excel) as well as using software programs like R, Python, SPSS, STATA, etc. wherever deemed necessary or suitable. Occasionally, guest lectures by experts will be provided for enhancement of their subject knowledge.

#### **5. Evaluation Scheme**

The student performance will be judged through internal evaluation and final board examination. Different modes of evaluation system are given as follows.

##### **5.1 Internal Evaluation**

- Attendance
- Class performance
- Written examinations (mid-term, pre-board examinations)
- Oral (Viva-Voce) examinations
- Presentations for core and optional papers
- Submission and Presentation of assignment work
- Thesis, presentation and Viva-Voce for thesis work

##### **5.2 Evaluation Guidelines**

- A minimum of 80% attendance will be required for students to appear in final examination.
- Internal assessment covers 40% of the total marks for each theory papers, computational papers and dissertation.
- Internal assessment marks include marks of two written exams (mid-term, pre-board), class performance and attendance and at least two of the following: assignment work, class seminar, presentation, oral examination, class test in each of the papers.
- A final board examination of 60% of the total marks will be conducted for each of the papers.

- An initial presentation will be required for the proposed dissertation title.
- A pre-submission seminar will be required for the submission of the dissertation.
- Dissertation will be evaluated through internal assessment and on the basis of external expert examination followed by Viva-Voce.

A breakdown of marks for assessment for each of the courses and dissertation work is given below.

**Table 6: Evaluation Guidelines of Theoretical Papers**

Nature of Evaluation	Examination	Allocated Marks (for 3 Credits)	Allocated Marks (for 2 Credits)
Attendance and Class Performance	Evaluation	5	3
Assessment	Mid-term (written)	5	3
	Pre-board (written)	10	7
	Assignment work /Oral test /class test	10	7
	Presentation/class Seminar		
Total of Internal Assessment		30 (40% of Total)	20 (40% of Total)
Final Examination	Written	45 (60% of Total)	30 (60% of Total)
Total		75	50

**Table 7: Evaluation Guidelines of Computing Papers (4 Credits)**

Nature of Evaluation	Examination	Allocated Marks
<b>Internal</b>		
Practical class Performance + Attendance	Evaluation	10
Submission of practical assignments	Evaluation	30
Total of Internal Evaluation		<b>40</b>
<b>External</b>		
Final Practical Examination	Written	40
	Viva-Voce	20
Total of Final Examination		<b>60</b>
Total of the Practical Evaluation		<b>100</b>

**Table 8: Evaluation Guidelines of Dissertation (for 6 Credits)**

Nature of Evaluation	Examination	Allocated Marks
Internal		
Regularity and Research Work	Evaluation	15
Perseverance	Evaluation	15
Pre-submission Assessment (Presentation, Compilation, Documentation)	Evaluation	30
Total of Internal Evaluation		<b>60</b>
External		
Dissertation Evaluation	Evaluation	45
Viva-Voce	Evaluation	45
Total of External Evaluation		<b>90</b>
Dissertation Evaluation Total		<b>150</b>