Savitribai Phule Pune University

(Formerly University of Pune)

T. Y. B. Sc. / T. Y. B. A. Statistics Syllabi in NEP

(Level 5.5/300)

National Education Policy Syllabus

To be implemented from Academic Year 2026-2027

Title of the program: T. Y. B. Sc. / T. Y. B. A. Statistics

Students who have Statistics as a major subject in the second year of the three-year undergraduate program are eligible to offer Statistics as the major subject in the third year of the program.

Students who opt for Statistics as the major subject in the third year of the undergraduate program study both theoretical and applied aspects of statistics. The theory papers include mathematically oriented core subjects such as Distribution Theory, Theory of Estimation, and Regression Analysis, which provide a strong theoretical foundation in statistical methodology. Some papers, such as Sampling Methods, emphasize applied aspects of statistics with comparatively less mathematical complexity. The Design of Experiments course introduces experimental designs commonly used in agriculture and industry.

In addition, **elective courses** such as Actuarial Statistics, Time Series Analysis, Optimization Techniques, and Clinical Trials enable students to explore specialized areas of statistics. The curriculum also includes **skill-oriented courses** such as C Programming and R Software to develop computational and data analysis skills essential for modern statistical practice.

The program includes four practical courses related to the core subjects to strengthen understanding through hands-on application of statistical methods, while some statistical methods and their practical components are also covered under minor subjects. Furthermore, **On-the-Job Training (OJT)** provides exposure to real-world applications of statistics in industries and research organizations, enhancing students' practical skills and employability. A **project component** allows students to analyze real data and develop independent research abilities. In addition, a **study tour** offers exposure to statistical applications in industries or research institutions and encourages interaction with professionals in the field.

Course Implementation criteria for Theory and Practical:

a. Each semester comprises of 15 weeks (12 weeks Actual Teaching + 3 weeks for Continuous Internal Evaluation).

b. One Credit of the Theory is equal to15 clock hours (Teaching 1 hour per week for each credit, 12 hours Actual Teaching + 3 hours Continuous Internal Evaluation which includes Assignments, Tutorials, Practice, Problem solving sessions, Group discussion, Seminars and Unit Tests.

c. One Credit of Practical = 30 clock hours (2 Contact hours per credit per week). One Credit = 30 clock hours (24 hours' Actual Table work + 6 hours for journal completion, and Continuous Internal Evaluation of each practical).

d. Practical for each course comprises of 02 Credits = 60 clock hours. Therefore,

- Minimum 12 laboratory/ Filed sessions of 04 clock hours must be conducted in one semester.
- Each practical of 04 clock hours in the laboratory should consist of: Table performance for concerned practical, careful observations, calculation, writing results and conclusion, and submission of practical in written form.
- Pre-laboratory reading and post laboratory assignments should be given on each practical as a part of continuous internal evaluation.
- For regular practical, 1 Batch = 12 students.

Examination Pattern (For each Semester):

- The examinations will be conducted semester wise for both Theory as well as Practical courses. NEP-2020 T.Y. B.Sc.
- Theory Paper of 02 Credits –

  Internal Exam (20 M) + University Theory Exam (30 M) = Total 50 M

  Duration: For Internal exam = 40 Min. and For University Exam = 02 hours.

Q.1	***(A) Multiple choice questions: 04 questions each of 01 marks.*** ***(B) True or false:02 questions each of 01 marks.***	***[04 Marks]*** ***[02 Marks]***
Q.2	***Attempt any 02 out of 03 questions each of 04 marks.***	***[08 Marks]***
Q.3	***Attempt any 02 out of 03 questions each of 04 marks.***	***[08 Marks]***
Q.4	***Attempt any 01 out of 02 questions each of 08 marks.***	***[08 Marks]***

- Practical Paper of 2 Credits –

  Internal Evaluation (20 M) + University Practical Exam (30 M) = Total 50 M

  Duration: For Internal Evaluation = Performance in each practical + internal viva

  For University Exam = 2.40 hours using scientific calculator

  (In case of MS Excel 2.10 hours).

  Submission of practical records compulsory (Journals).

Note

1. Students must complete all the practical to the satisfaction of the teacher concerned.
2. At the time of practical examination, a student must produce the laboratory journal along with the completion certificate signed by the Practical In-charge and Head of the Department.

3. Structure of evaluation of practical paper at

a) Continuous Internal Assessment (CIA) – 20 Marks

Section	Marks
i) Journal	10
ii) Viva-voce	10

b) End Semester Examination – 30 Marks

i) Set four questions each of Thirteen marks.

ii) Student has to solve any two of these four questions.

iii) Viva-voce carries four marks.

General Instruction

1. The theory question paper for each paper shall cover all the topics in the pertaining syllabus with proportional weightage to the number of hours of instruction prescribed.
2. The practical are to be conducted in batches as per the University norms for the faculty of science.
3. **Study Tour:** In order to acquaint the students with applications of statistical methods in various fields such as industries, agricultural sectors, government institutes, etc. at least one Study Tour for T.Y. B.Sc. Statistics students should be arranged and study report should be attached in the journal.
4. Medium of Instruction: English
5. Examination:

   A) Pattern of examination: Semester wise

   B) Standard of passing: As per norms of university.

   C) Each course exam must be of 50 marks – Internal: 20 marks & External: 30 marks.

Statistics as Major (Core) Subject and any other subject as Minor (each theory / practical paper has 2 credits).

Year /Level	Sem	Code Number	Title of the paper (Theory / Practical)	Credits allotted	Lecture/Practical hours per week
III 5.5/300	V	STS-301-MJ	Distribution Theory	**02**	**02**
		STS-302-MJ	Theory of Estimation	**02**	**02**
		STS-303-MJ	Design and Analysis of Experiments	**02**	**02**
		STS-304-MJ	Regression Analysis	**02**	**02**
		STS-305-MJP	Statistics Practical–V (based on STS-301-MJ and STS-302-MJ)	**02**	**04**
		STS-306-MJP	Statistics Practical–VI (based on STS-303-MJ and STS-304-MJ)	**02**	**04**
		Major Electives (choose any one theory and any one practical)			
		STS-310-MJ	Time Series analysis	**02**	**02**
		STS-311-MJ	Operation Research	**02**	**02**
		STS-312-MJ	Reliability Theory	**02**	**02**
		STS-313-MJP	Statistics Practical VII (based on STS-310-MJ)	**02**	**04**
		STS-314-MJP	Statistics practical VIII (based on STS-311-MJ)	**02**	**04**
		STS-315-MJP	Statistics practical IX (based on STS-312-MJ)	**02**	**04**
	VI	STS-351-MJ	Testing of Hypothesis	**02**	**02**
		STS-352-MJ	Sampling Theory	**02**	**02**
		STS-353-MJ	Statistical process and product control	**02**	**02**
		STS-354-MJ	Survival analysis	**02**	**02**
		STS-355-MJP	Statistics Practical–X (based on STS-351-MJ and STS-352-MJ)	**02**	**04**
		STS-356-MJP	Statistics Practical–XI (based on STS-353-MJ and STS-354-MJ)	**02**	**04**
		Major Electives (choose any one theory and any one practical)			
		STS-360-MJ	Stochastic Processes	**02**	**02**
		STS-361-MJ	Medical Statistics and Clinical Trials	**02**	**02**
		STS-362-MJ	Operations Management	**02**	**02**
		STS-363-MJP	Statistics Practical XII (based on STS-360-MJ)	**02**	**04**
		STS-364-MJP	Statistics Practical XIII (based on STS-361-MJ)	**02**	**04**
		STS-365-MJP	Statistics Practical XIV (based on STS-362-MJ)	**02**	**04**

List of Vocational Skill Courses (VSC):

Year / Level	Sem	Code Number	Title of the paper (Theory /Practical)	Credits allotted	Practical hours per week/ Batch
III 5.5/300	V	STS-321-VSCP	Practical on C-Programming (Turbo C) (Practical Course)	02	04
	VI	STS-371-VSCP	Data Analytics Using Python (Practical Course)	02	04

Field Project (FP) / On Job Training (OJT)/ Community Engagement Project (CEP):

Year / Level	Sem	Code Number	Title of the paper (Theory / Practical)	Credits allotted	Lecture/Practical hours per week
III 5.5/300	V	STS-331-FP/CEP	Field Project / Community Engagement Project	02	
	VI	STS-381-OJT	On Job Training	04	

Statistics as Minor Subject and any other subject as Major (each theory / practical paper has 2 credits):

Year /Level	Sem	Code Number	Title of the paper (Theory /Practical)	Credits allotted	Lecture/Practical hours per week
III 5.5/300	V	STS-341-MN	Statistical Methods	02	02

Semester V

STS-301-MJ: Distribution Theory

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

On Completion of this course, student will be able to:

CO1: Apply specific distributions to solve real-life modelling problems

CO2: Calculate moments, mean, mode, variance and quartiles for various distributions.

CO3: Understand order statistics for univariate distributions.

CO4: Utilize moment generating functions and cumulant generating functions to derive properties of distributions.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Beta Distributions**	**07**	**CO1, CO2**
1.1 Beta distributions of first kind: p.d.f. $f(x) = \begin{cases} \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1}; & 0 \le x \le 1;\ m,n > 0 \\ 0; & elsewhere. \end{cases}$ Notation: $X \to \beta_1(m,n)$, Nature of probability curve, symmetry, mean, variance, properties, r^{th} raw moment, harmonic mean, median for $\beta_1(m,m)$. 1.2 Relation with U(0,1). The probability distributions of $\frac{1}{X}$, X+Y , X-Y , XY , $\frac{X}{Y}$ where X and Y are i.i.d. $\beta_1(1,1)$ 1.3 Beta distributions of second kind: p.d.f. $f(x) = \begin{cases} \frac{1}{\beta(m,n)} \frac{x^{m-1}}{(1+x)^{m+n}}; & x \ge 0;\ m,n > 0 \\ 0; & elsewhere. \end{cases}$ Notation : $X \to \beta_2(m,n)$, Nature of probability curve, symmetry, mean, variance, properties, r^{th} raw moment, harmonic mean, median for $\beta_2(m,m)$. 1.4 Interrelation between $\beta_1(m,n)$ and $\beta_2(m,n)$. 1.5 Distribution of $\frac{X}{Y}, \frac{X}{X+Y}$, etc. when X and Y are independent gamma variates. 1.6 Relation between distribution functions of $\beta_1(m,n)$ and binomial distribution. 1.7 Real life situations and applications.			
2	**Order Statistics**	**05**	**CO3**
2.1 Order statistics for a random sample of size n from a continuous distribution, definition, derivation of distribution function and density function of the i^{th} order statistic $X_{(i)}$, particular cases for $i = 1$ and $i = n$. 2.2 Distribution of $X_{(i)}$ from a random sample following Uniform and exponential distributions. 2.3 Derivation of joint p.d.f. of $(X_{(i)}, X_{(j)})$ 2.4 probability distribution of sample range $X_{(n)} - X_{(1)}$, Distribution of sample median (state the distribution only without proof)			
3	**Cauchy distribution**	**04**	**CO1, CO2, CO4**
3.1 P.d.f. of Cauchy distribution			

<table>
<tr><td colspan="4">

$$f(x) = \begin{cases} \frac{\lambda}{\pi} \frac{1}{\lambda^2+(x-\mu)^2}; & -\infty < x < \infty, -\infty < \mu < \infty, \lambda > 0 \\ 0; & elsewhere. \end{cases}$$

3.2 Nature of the probability curve.
3.3 Distribution function, quartiles, non-existence of moments, distribution of
i) $\frac{1}{X}$ ii) X^2, where $X \sim C(0,1)$ iii) aX+b
3.4 Additive property for two independent Cauchy variates (statement only), statement of distribution of the sample mean, comment on limiting distribution of X .
3.5 Relationship with uniform, Student's-t and normal distributions.
3.6 Real life situations and Applications of $C(\mu, \lambda)$.
</td></tr>
<tr><td>4</td><td>Weibull Distribution</td><td>05</td><td>CO1, CO2, CO4</td></tr>
<tr><td colspan="4">
4.1 P.d.f. of Weibull Distribution

$$f(x) = \begin{cases} \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}; & x \geq 0, \alpha, \beta > 0 \\ 0; & elsewhere. \end{cases}$$

Notation :$X \sim W(\alpha, \beta)$
4.2 Probability curve, location parameter, shape parameter, scale parameter, Distribution function, quartiles, mean and variance, coefficient of variation, relationship with gamma and exponential distribution, Hazard rate, IFR and DFR property.
4.3 Real life situations and applications
</td></tr>
<tr><td>5</td><td>Laplace (Double Exponential) Distribution</td><td>04</td><td>CO1, CO2, CO4</td></tr>
<tr><td colspan="4">
5.1. P.d.f of Laplace Distribution

$$f(x) = \begin{cases} \frac{\lambda}{2} e^{-\lambda|x-\mu|}; & -\infty < x, \mu < \infty, \lambda > 0 \\ 0; & elsewhere. \end{cases}$$

Notation:$X \sim L(\mu, \lambda)$,Nature of the probability curve.
5.2. Distribution function, quartiles, comment on MLE of μ, λ.
5.3. MGF, CGF, moments and cumulants $\beta_1, \beta_2, \gamma_1, \gamma_2$.
5.4. Laplace distribution as the distribution of the difference of two i.i.d. exponential variates with mean $1/\lambda$.
5.5. Applications and real life situations.
</td></tr>
<tr><td>6</td><td>Lognormal Distribution</td><td>05</td><td>CO1, CO2, CO4</td></tr>
<tr><td colspan="4">
6.1 P.d.f.

$$f(x) = \begin{cases} \frac{1}{(x-a)\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2\sigma^2}[\log(x-a)-\mu]^2\right\}; & x > a, -\infty < \mu < \infty, \sigma > 0 \\ 0; & elsewhere. \end{cases}$$

Notation: $X \sim LN(a, \mu, \sigma^2)$, Nature of the probability curve.
6.2 Moments (r^{th} moment of $X - a$), first four moments, β_1 and γ_1 coefficients, quartiles, mode.
6.3 Relation with N (μ, σ^2) distribution.
6.4 Distribution of $\prod_{i=1}^{n} X_i$, where $X_i's$ are independent lognormal variates.
6.5 Applications and real life situations.
</td></tr>
</table>

Reference Books

1. Arora, S., & Lal, B. (1989). New mathematical statistics: A problem-oriented first course. Satya Prakashan.
2. Buyan, K. C. (2010). Probability theory and statistical inference (1st ed.). New Central Book Agency.
3. Cramér, H. (1992). Mathematical methods of statistics. Princeton University Press.
4. Feller, W. (1968). An introduction to probability theory and its applications (3rd ed.). Wiley.
5. Gupta, S. C., & Kapoor, V. K. (2020). Fundamentals of mathematical statistics (12th ed.). Sultan Chand & Sons.
6. Hogg, R. V., McKean, J. W., & Craig, A. T. (2012). Introduction to mathematical statistics (7th ed.). Pearson.
7. Lindgren, B. W. (1993). Statistical theory (4th ed.). CRC Press.
8. Mood, A. M., Graybill, F. A., & Boes, D. C. (2017). Introduction to the theory of statistics (3rd ed.). McGraw-Hill Education.
9. Rohatgi, V. K., & Saleh, A. K. Md. E. (2015). An introduction to probability theory and statistics (3rd ed.). Wiley.
10. Ross, S. M. (2018). A first course in probability (10th ed.). Pearson.

STS-302-MJ: Theory of Estimation

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

On Completion of this course, student will be able to:

CO1: Explain the concepts of parameter, estimator, estimate, MSE in point estimation; construct likelihood functions and obtain parameter estimates using the MLE.

CO2: Apply the Method of Moments to estimate parameters and compare estimators using MSE.

CO3: Analyze properties of estimators such as unbiasedness, variance, and minimum variance unbiased estimation.

CO4: Explain the concepts of sufficiency, completeness, Fisher information, Cramér–Rao inequality, and efficiency of estimators.

CO5: Understand the asymptotic behavior of estimators and apply the concept of consistency in statistical inference.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Point Estimation**	02	CO1
	1.1 Notion of a parameter and parameter space. 1.2 General problem of estimating an unknown parameter(s) by point estimation. 1.3 Point Estimation: Definitions of an estimator and estimate. Distinction between estimator and estimate, illustrative examples. Mean Square Error (MSE) of an estimator.		
2	**Methods of Estimation**	06	CO1, CO2
	2.1 Likelihood Function: Definition of likelihood as a function of unknown parameter(s) for a random sample comes from i) a discrete distribution ii) a continuous distribution. Distinction between the likelihood function and p.d.f. or p.m.f. 2.2 Method of Maximum Likelihood: Maximum likelihood estimator (M.L.E.) for parameter(s) of standard distributions (case of two unknown parameters only for the normal distribution). Invariance property of M.L.E. Maximum likelihood estimator of θ in uniform distribution over i) $U(0,\theta)$ ii) $U(-\theta,\theta)$ iii) $U(m\theta,n\theta)$ $(m < n)$, M.L.E. of θ in $f(x\,;\,\theta) = Exp\,\{-(x-\theta)\}, x > \theta$, M.L.E. of location parameter in Laplace distribution. 2.3 Method of Moments: Moment estimator for unknown parameter(s) for standard discrete and continuous distributions. Illustrations of situations where M.L.E. and moment estimators differ and their comparison using MSE.		
3	**Properties of Estimator I**	12	CO3, CO4
	3.1 Unbiasedness: Definition of an unbiased estimator and biased estimator, positive and negative bias, illustrations and examples (including unbiased and biased estimators for the same parameters). Proofs of the following results regarding unbiased estimators: (i) Two distinct unbiased estimators of θ give rise to infinitely many unbiased estimators. (ii) If T is an unbiased estimator of θ, then $\phi(T)$ is unbiased estimator of $\phi(\theta)$ provided $\phi(.)$ is a linear function. Variance of the estimator: Minimum variance unbiased estimator (MVUE), uniqueness of MVUE whenever it exists. 3.2 Sufficiency: Concept and definition of sufficiency, statement of the Fisher-Neyman factorization theorem with proof for discrete probability distribution. Pitmann – Koopman form and sufficient statistic; Exponential family of probability distributions and sufficient		

	statistic. Proofs of the following properties of sufficient statistics: (i) If T is sufficient for θ, then $\phi(T)$is also sufficient for θ provided ϕ is a one to one and on to function. (ii) If T is sufficient for θ then T is also sufficient for $\phi(\theta)$. Introduction of completeness.		
4	**Properties of Estimator II**	**06**	**CO4**
4.1 4.2 4.3	Fisher information function: Statement regarding information in sample and in a sufficient statistic T. Cramer- Rao Inequality: Statement of Cramer - Rao inequality, Cramer – Rao Lower Bound (CRLB), definition of minimum variance bound unbiased estimator (MVBUE) of $\phi(\theta)$ Efficiency: Comparison of variance with CRLB, relative efficiency of T_1 w.r.t. T_2 for (i) unbiased (ii) biased estimators. Efficiency of unbiased estimator T w.r.t. CRLB.		
5	**Asymptotic Behaviour of an Estimator**	**04**	**CO5**
5.1 5.2	Consistency: Definition of convergence in probability and consistent estimator. Proof of the following theorems: (i) An unbiased estimator is consistent if its variance tends to zero as the sample size tends to infinity ii) A biased estimator is consistent if its bias and variance both tend to zero as the sample size tends to infinity. (iii) If T is consistent estimator of θ and $\phi(.)$ is a continuous function, then $\phi(T)$ is a consistent estimator of $\phi(\theta)$.		

Reference Books

1. Buyan, K. C. (2010). Probability theory and statistical inference (1st ed.). New Central Book Agency.
2. Dudewicz, E. J., & Mishra, S. N. (1988). Modern mathematical statistics. John Wiley & Sons.
3. Hoel, P. G., Port, S. C., & Stone, C. J. (1972). Introduction to statistical theory. Houghton Mifflin Company.
4. Hogg, R. V., McKean, J. W., & Craig, A. T. (2012). Introduction to mathematical statistics (7th ed.). Pearson.
5. Kendall, M. G., & Stuart, A. (1943). The advanced theory of statistics. Charles Griffin & Company Ltd.
6. Lindgren, B. W. (1993). Statistical theory (4th ed.). CRC Press.
7. Mood, A. M., Graybill, F. A., & Boes, D. C. (2017). Introduction to the theory of statistics (3rd ed.). McGraw-Hill Education.
8. Ramachandran, K. M., & Tsokos, C. P. (2009). Mathematical statistics with applications. Academic Press.
9. Rohatgi, V. K., & Saleh, A. K. Md. E. (2015). An introduction to probability theory and statistics (3rd ed.). Wiley.

STS-303-MJ: Design and Analysis of Experiments

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

On Completion of this course, student will be able to:

CO1: Understand basic ideas of Design of Experiments, ANOVA, variation, and principles like randomization, replication, and local control.

CO2: Apply CRD, RBD, and Latin Square designs to analyze data using ANOVA and F-test.

CO3: Compare different designs (CRD, RBD, and LSD) and check which design is more efficient.

CO4: Analyze factorial experiments by finding main effects, interaction effects, and preparing ANOVA tables.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction and basic concepts of Design of Experiments**	**03**	**CO1**
1.1	Historical Background: The Founder (Sir Ronald A. Fisher) and the evolution of the method, Definition of ANOVA and its primary use in research, The difference between t-test/z-test and ANOVA (Multiple comparisons), Hypothesis formulation (Null vs. Alternative), The Mechanics of Variation: Concept of variation between Groups (Treatment effect), Concept of variation within Groups (Error/Residual), Explanation of the F-distribution and its application in ANOVA, Assumptions of ANOVA.		
1.2	Elements of Experimental Design: Experimental Unit, Treatment, Block, Plots, Layout of an experiment.		
1.3	Basic Principles and Contrasts: Replication, Randomization, Local Control, Linear Treatment Contrasts, Orthogonal Treatment Contrasts.		
2	**Standard Experimental Designs**	**18**	**CO2, CO3**
2.1	Completely Randomized Design (CRD): Application of the principles of design of experiment in CRD, Layout, Model: $X_{ij} = \mu + \alpha_i + \varepsilon_{ij}; i = 1,2,\ldots,t, j = 1,2,\ldots,n_i$. Assumptions and interpretations. Breakup of total sum of squares into components. Estimation of parameters, expected values of mean sums of squares, components of variance, preparation of ANOVA table, testing equality of treatment effects, Hypothesis to be tested $H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_t$. Statement of Cochran's theorem. F-test for testing H_0 with justification (independence of chi- square is to be assumed). Equivalence between ANOVA and regression model. Identification of real life situations where this design is used.		
2.2	Randomized Block Design (RBD): Application of the principles of design of experiments in RBD, layout. Model: $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij};\quad i = 1,2,\ldots,t,\quad j = 1,2,\ldots,b$. Assumptions and interpretations. Breakup of total sum of squares into components. Estimation of parameters, expected values of mean sums of squares, components of variance, preparation of analysis of variance table, Hypotheses to be tested $H_{01}: \alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_t, H_{02}: \beta_1 = \beta_2 = \ldots = \beta_b$. F-test for testing H_{01} and H_{02} with justification (independence of chi-squares is to be assumed). Identification of real life situations where this design is used.		
2.3	Latin Square Design (LSD): Application of the principles of design of experiments in LSD,		

	layout, Model : $X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk};\ i,j,k = 1,2,\dots,m,$ (i, j, k) ∈ S, where S represents the set of m^2 values. Assumptions and interpretations. Breakup of total sum of squares into components. Estimation of parameters, expected values of mean sums of squares, components of variance, preparation of analysis of variance table, hypotheses to be tested. $H_{01}: \alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_m, H_{02}: \beta_1 = \beta_2 = \ldots = \beta_m, H_{03}: \gamma_1 = \gamma_2 = \cdots = \gamma_m$ and their interpretation. Justification of F-test for H_{01}, H_{02} and H_{03} Independence of chi-square is to be assumed). Preparation of ANOVA table and F test for H_{01}, H_{02} and H_{03} testing for equality of two specified treatment effects. Identification of real life situations where this design is used. (Numerical problems will be solved in practical's)		
4	**Comparative Design Efficiency**	**02**	**CO3**
3.1	Concept and definition of efficiency of a design.		
3.2	Efficiency of RBD over CRD. Direct formulae and illustrative examples (without proof)		
3.3	Efficiency of LSD over (i) CRD (ii) RBD. Direct formulae and illustrative Examples. (without proof)		
5	**Factorial Experiments and confounding**	**07**	**CO4**
4.1	General description of n^k factorial experiment, 2^2 and 2^3 factorial experiments arranged in RBD.		
4.2	Definitions of main effects and interaction effects in 2^2 and 2^3 factorial experiments.		
4.3	Introduction of confounding, types of Confounding: Total confounding and Partial confounding		
4.4	Yate's procedure, preparation of ANOVA table, test for main effects and interaction effects. (Numerical problems will be solved in practical)		

Reference Books

1. Cochran, W. G., & Cox, G. M. (1992). Experimental designs (2nd ed.). John Wiley & Sons.
2. Dass, M. N., & Giri, N. C. (1987). Design and analysis of experiments (2nd ed.). Wiley Eastern Ltd.
3. Federer, W. T. (1967). Experimental design: Theory and application. Oxford & IBH Publishing Co.
4. Goon, A. M., Gupta, M. K., & Dasgupta, B. (1998). Fundamentals of statistics. The World Press Pvt. Ltd.
5. Gupta, S. C., & Kapoor, V. K. (2006). Fundamentals of applied statistics. S. Chand & Company Ltd.
6. Johnson, R. A., Miller, I., & Freund, J. E. (2010). Probability and statistics for engineers (8th ed.). Prentice Hall India.
7. Kempthorne, O. (1994). The design and analysis of experiments. Wiley.
8. Montgomery, D. C. (2012). Design and analysis of experiments (8th ed.). John Wiley & Sons.
9. Snedecor, G. W., & Cochran, W. G. (1994). Statistical methods (8th ed.). Affiliated East–West Press.
10. Wu, C. F. J., & Hamada, M. (2009). Experiments: Planning, analysis, and parameter design optimization (2nd ed.). John Wiley & Sons.

STS-304-MJ: Regression Analysis

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

On Completion of this course, student will be able to:

CO1: Explain and implement the fitting of simple linear regression models, including estimation and interpretation of parameters.

CO2: Describe multicollinearity and apply ridge regression as a remedial technique.

CO3: Compare residual diagnostics and apply corrective measures.

CO4: Analyse the multiple linear regression and logistic regression models.

CO5: Apply and justify the use of the weighted least squares method under heteroscedastic conditions.

CO6: Determine tests of hypothesis of model parameters, AIC and BIC criteria.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Simple Linear Regression Model**	**08**	**CO1**
1.1	Review of simple linear regression model: $Y = \beta_0 + \beta_1 X + \varepsilon$, where ε is a continuous random variable with $E(\varepsilon) = 0, V(\varepsilon) = \sigma^2$.		
1.2	Estimation of β_0 and β_1 by the method of least squares, Properties of estimators of β_0 and β_1, Estimation of σ^2, Assumption of normality of ε.		
1.3	Tests of hypothesis of β_1, Interval estimation in simple linear regression model.		
1.4	Coefficient of determination, Residual analysis: Standardized residuals, Studentized residuals, residual plots, Detection and treatment of outliers		
2	**Multiple Linear Regression Model**	**14**	**CO2, CO3**
2.1	Review of multiple linear regression model $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon$, where ε is a continuous random variable with $E(\varepsilon) = 0, V(\varepsilon) = \sigma^2$.		
2.2	Estimation of regression parameters $\beta_0, \beta_1, \ldots$ and β_p by the method of least squares, obtaining normal equations, solutions of normal equations, Estimation of σ^2, Assumption of normality of ε.		
2.3	Tests of hypothesis of regression parameters, Interval estimation in multiple linear regression model.		
2.4	Variable selection and model building, Residual diagnostics and corrective measures such as transformation of response variable, weighted least squares method.		
2.5	Introduction of multicollinearity, computation of VIF and brief introduction to ridge regression.		
3	**Logistic Regression Model**	**08**	**CO4, CO5**
3.1	Binary response variable, Logit transform, estimation of parameters, interpretation of parameters.		
3.2	Tests of hypotheses of model parameters, model deviance, LR test.		
3.3	AIC and BIC criteria for model selection.		
3.4	Multiple logistic regression: Model, properties.		

Reference Books

1. Chatterjee, S., & Hadi, A. S. (2012). Regression analysis by example (5th ed.). Wiley.
2. Draper, N. R., & Smith, H. (1998). Applied regression analysis (3rd ed.). John Wiley & Sons.
3. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). Wiley.
4. Kleinbaum, D. G., & Klein, M. (2011). Logistic regression: A self-learning text (3rd ed.). Springer.
5. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (5th ed.). Wiley.
6. Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (2005). Applied linear statistical models (5th ed.). McGraw-Hill/Irwin.

STS- 305-MJP: Statistics Practical –V (based on STS-301-MJ and STS-302-MJ)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60 (48+12)**

Course Outcomes

On Completion of this course, student will be able to:

CO1: Apply simulation techniques to illustrate fundamental probability results such as Chebyshev's inequality, order statistics, the Central Limit Theorem, and the Weak Law of Large Numbers.

CO2: Generate random samples from probability distributions such as Cauchy and Laplace and fit appropriate statistical models including the log-normal distribution.

CO3: Apply Monte Carlo techniques to approximate numerical values of mathematical functions such as Gamma and Beta functions.

CO4: Construct confidence intervals for population parameters such as variance, median, and quartiles using statistical methods and order statistics.

CO5: Compute maximum likelihood estimators and method of moments estimators for standard and truncated distributions.

CO6: Evaluate properties of estimators by computing likelihood functions, Fisher Information, Cramér–Rao Lower Bound (CRLB), and minimum variance unbiased estimators (MVBUE).

List of Practical

Sr. No.	Title of the Experiment	No. of Experiments
1	Chebychev's inequality and Order Statistics.	1
2	Central limit theorem and Weak Law of Large Numbers.	1
3	Model sampling from Cauchy and Laplace distribution.	1
4	Application of bivariate normal distribution.	1
5	Fitting of log-normal distribution.	1
6	Monte Carlo techniques to find approximate numerical value of gamma and beta function with real arguments.	1
7	Construction of confidence interval for variance of normal population (σ^2), when i) μ known ii) μ unknown.	1
8	Construction of confidence interval for population median and quartiles based on order statistics.	1
9	M.L.E and moment estimator of truncated Binomial and truncated Poisson distributions (truncated at zero).	1
10	Computation of likelihood function $L(\theta, \underline{x})$ as a function of θ for standard probability distributions and sketch its graph to determine MLE of θ.	1
11	MLE, Moment estimator and Unbiasedness of parameter(s) for standard distributions.	1
12	Fisher Information Function, CRLB and MVBUE.	1

STS-306-MJP: Statistics Practical –VI (based on STS-303-MJ and STS-304-MJ)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60 (48+12)**

Course Outcomes

On Completion of this course, student will be able to:

CO1: Handle experimental data confidently by performing CRD analysis and comparing treatments using suitable statistical methods.

CO2: Analyze RBD and judge its efficiency compared with CRD.

CO3: Use advanced post-hoc techniques to clearly identify significant differences among treatment means.

CO4: Organize and analyse LSD data and evaluate its efficiency against other designs.

CO5: Apply non-parametric approaches like the Kruskal–Wallis test.

CO6: Investigate and analyze factorial experiments.

List of Practical:

Sr. No.	Title of the experiment	No. of experiments
1	Analysis of CRD: Calculation for equal and unequal replications; Pair-wise comparison of treatments using the Critical Difference (C.D.) method.	1
2	Analysis of RBD and its Efficiency: Analysis of RBD, pairwise comparison of treatments using i) C.D ii) Tukey and Scheffe’s procedure, Relative Efficiency (RE) of RBD over CRD	1
3	Analysis of LSD and its Efficiency: Analysis of LSD, pair wise comparison of treatments using C.D. and computation of its Relative Efficiency (RE) relative to both CRD and RBD.	2
4	Non-Parametric ANOVA: Application of the Kruskal-Wallis H Test for CRD and RBD layouts as an alternative to standard F tests. Ferdman Test	1
5	Analysis of Factorial Experiments: Analysis of 2^2 and 2^3 factorial experiments in an RBD layout.	1
6	Analysis of 2^3 factorial experiments in an RBD layout (total confounding).	1
7	Analysis of 2^3 factorial experiments in an RBD layout (partial confounding).	1
8	Simple linear regression analysis and diagnostics. (Estimation of Parameter, Test for significance of parameters Regression model, Interval Estimation, Interpretation of parameters. Computation of R^2, Residual Analysis by graphical method).	1
9	Multiple linear regression analysis and diagnostics. (Estimation of Parameter, Test for significance of parameters Regression model, Interval Estimation, Interpretation of parameters, Computation of R^2, Adjusted R^2, Residual Analysis by graphical method, Variable selection methods).	1
10	Logistic regression. (Estimation of Parameter, Test for significance of parameters Regression model, Interval Estimation, Interpretation of parameter/(s)).	1
11	Detection of Multicollinearity using VIF and its fitting of Multiple Linear Using Ridge Regression.	1

Note: Regression analysis practicals are performed using MS-Excel or R-software. (Using built in commands/direct commands).

Major Elective

STS-310-MJ: Time Series Analysis

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Explain the concept, components, and graphical representation of time series data.

CO2: Apply classical decomposition techniques to estimate trend and seasonal components of time series.

CO3: Explain the concepts of stationarity and describe basic stochastic processes used in time series modeling.

CO4: Analyze and interpret autocorrelation and partial autocorrelation functions for autoregressive and moving average models.

CO5: Identify appropriate ARIMA/SARIMA models for time series data using the Box–Jenkins framework.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Fundamentals and Decomposition of Time Series**	**06**	**CO1, CO2**
1.1 Definition and examples of time series, graphical representation of time series data, components of time series (trend, seasonal, cyclical, irregular), additive and multiplicative models. 1.2 Estimation of trend by graphical method, moving averages, least squares method for fitting linear trend, exponential smoothing (introductory). 1.3 Measurement of seasonal variations by simple average method, ratio-to-moving-average method, ratio-to-trend method, cyclical variations and business cycles.			
2	**Stationarity and Basic Stochastic Processes**	**05**	**CO3**
2.1 Strict stationarity and weak stationarity, non-stationarity and need for differencing. 2.2 Random walk model, iid noise and white noise, general linear processes. 2.3 Autocorrelation function (ACF), properties and interpretation of ACF, sample ACF, correlogram and its interpretation.			
3	**Autoregressive and Moving Average Models**	**14**	**CO4**
3.1 Partial autocorrelation function (PACF), autoregressive models: $AR(1), AR(2)$, $AR(p)$, basic properties of autoregressive processes. 3.2 Moving average models: $MA(1)$,$MA(2)$, $MA(q)$, autoregressive moving average models $ARMA(p,q)$.			
4	**Model Identification and ARIMA Framework**	**05**	**CO5**
4.1 ACF patterns for autoregressive and moving average processes, model identification using ACF and PACF patterns. 4.2 $ARIMA(p,d,q)$ models, difference operator, Box–Jenkins approach (basic steps), Introduction to seasonal ARIMA (SARIMA) model.			

Reference Books

1. Beri, G. C. (2017). Business statistics (3rd ed.). McGraw Hill Education.
2. Brockwell, P. J., & Davis, R. A. (2002). Introduction to time series and forecasting. Springer.
3. Chatfield, C. (2016). The analysis of time series: An introduction (6th ed.). CRC Press.
4. Gupta, S. C., & Kapoor, V. K. (2020). Fundamentals of applied statistics (4th ed.). Sultan Chand & Sons.
5. Gupta, S. P. (2014). Statistical methods (44th ed.). Sultan Chand & Sons.
6. Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: With R examples (4th ed.). Springer.
7. Srivastava, P. K. (2011). Statistical methods. PHI Learning.

STS- 311-MJ: Operation Research

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Able to formulate and develop mathematical representations from verbal descriptions of real life problems

CO2: Ability to solve LPP using graphical and simplex methods.

CO3: Proficiency in solving transportation and assignment problems to minimize costs or maximize profits.

CO4: Ability to apply network models, such as Critical Path Method (CPM) and Project Evaluation and Review Technique (PERT) for project management.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Linear Programming Problem**	**12**	**CO1, CO2**
1.1. Statement of the Linear Programming Problem (LPP), Formulation of problem as L.P. problem. L.P. Problem in (i) Canonical form (ii) standard form. Definitions of (i) A slack variable (ii) A surplus Variable (iii) a solution (iv) a feasible solution (v) a basic feasible solution, (vi) a degenerate and non –degenerate solution vii) an optimal solution (viii) basic and non-basic variables 1.2. Solution of L.P.P by: I) Graphical Method: solution space, unique and non-unique solutions, obtaining an optimal solution. II) Simplex Method: a) Initial basic feasible solution (IBFS) is readily available: obtaining an IBFS, b) IBFS not readily available: introduction of artificial variable, Big-M method, modified objective function, modifications and application of simplex method to L.P.P. with artificial variables criteria for deciding (i) whether obtained solution is optimal and unique solution (ii) criteria for unbounded solution (iii) no solution, and (iv) more than one solution or alternate solution. 1.3. Duality Theory: i) Writing dual of a primal problem, ii) Solution of a L.P.P. by using its dual problem.			
2	**Transportation Problem**	**06**	**CO3**
2.1 Transportation problem (T.P.), statement of T.P., balanced and unbalanced T.P. 2.2 Methods of obtaining basic feasible solution of T.P: i) North-West corner rule ii) Method of matrix minima (least cost method), iii) Vogel's approximation method (VAM). 2.3 U-V method of obtaining Optimal solution of T.P., degenerate solution, uniqueness and non-uniqueness of optimal solutions. Examples and numerical problems.			
3	**Assignment Problem**	**04**	**CO3**
3.1 Assignment problems: i) Statement of an assignment problem, balanced and unbalanced problem ii) relation with T.P. iii) optimal solution of an assignment problem using Hungarian method. 3.2 Examples and numerical problems.			

4	CPM	04	CO4
4.1	Introduction		
4.2	Basic differences between PERT and CPM.		
4.3	Definition of (i) Events with Merge and Burst Events, (ii) Node, (iii) Activities, Predecessor, Successor, Dummy, (iv)Critical Activity, (v)Project Duration.		
4.4	Construction of network, Definitions of (i) earliest start time (ii) earliest finish time (iii) latest start time (iii) latest finish time for an activity. Types of float: total floats, free float, independent float and their significance. Determination of critical path		
4.5	Examples and numerical problems		
5	**PERT**	**04**	**CO4**
5.1	Construction of network: (i) pessimistic time estimate, (ii) optimistic time estimate (iii) most likely time estimates.		
5.2	Determination of critical path, determination of mean and standard deviation of project duration, computations of probability of completing the project in a specified duration. Examples and numerical problems		

Reference Books

1. Gass, S. I. (2011). Linear programming: Methods and applications (5^{th} ed.). Dover Publications.
2. Gupta, P. K., & Hira, D. S. (1976). Operations research. S. Chand & Company Ltd.
3. Kapoor, V. K. (2012). Operations research: Concepts, problems and solutions. Sultan Chand & Sons.
4. Phillips, D. T., Ravindran, A., & Solberg, J. J. (1987). Operations research: Principles and practice (2nd ed.). John Wiley & Sons.
5. Sasieni, M. W., Yaspan, A., & Friedman, L. (2013). Operations research: Methods and problems. Literary Licensing, LLC.
6. Sharma, J. K. (2012). Mathematical models in operations research (5^{th} ed.). Macmillan India Ltd.
7. Shrinath, L. S. (1982). Linear programming: Principles and applications. Affiliated East-West Press Pvt. Ltd.
8. Shrinath, L. S. (2001). PERT-CPM: Principles and applications. Affiliated East-West Press Pvt. Ltd.
9. Taha, H. A. (2017). Operations research: An introduction (10^{th} global ed.). Pearson.

STS- 312-MJ: Reliability Theory

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Understand and analyze structural properties of coherent systems and structure functions.

CO2: Compute and interpret reliability of coherent systems using different methods.

CO3: Understand ageing properties and analyze failure behaviour using hazard rate models.

CO4: Analyze reliability data and estimate reliability measures using likelihood methods and confidence intervals.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Structural properties of coherent system**	**10**	**CO1**
1.1	Binary system of independent components, order of the system, different types of systems, concept of the structure function, structure function of series system, parallel system, k-out-of-n system, essentially series and parallel system, reliability block diagram, guidelines for construction of reliability block diagram.		
1.2	Coherent structure function (maximum 4 components), relevant component, increasing structure function, pivotal decomposition of structure function, dual of a structure function (proof of dual of series system of order n is parallel system of order n,dual of the parallel system of order n is a series system of order n, dual of k-out -of –n system is (n-k+1)-out of – n system). Path sets, cut sets, minimal path and cut set, representation of coherent system in terms of minimal path sets and cut sets, dual coherent structure function, relative/structural importance of components, module of the coherent system, modular decomposition of coherent system.		
2	**Reliability of coherent system**	**10**	**CO2**
2.1	Reliability (h(p)) of a system of independent components, Basic properties of system reliability (such as reliability function is increasing function, system and component redundancy etc.).		
2.2	Computation of reliability of coherent system by using minimal path and cut set representation, upper and lower bound on system reliability by using exact system reliability, relative importance of a component.		
2.3	Reliability of Binary Components and Systems, Examples to show that systems with higher reliabilities can be constructed using components of low values of reliabilities, i.e. unreliable components.		
3	**Ageing Properties and reliability data**	**10**	**CO3, CO4**
3.1	Definitions: Hazard rates, hazard function, survival function, Concept of distributions with increasing and decreasing failure rate (IFR/DFR), Average Increasing (Decreasing) Failure Rate (IFRA/ DFRA)		
3.2	Relationship between i) Survival function and hazard function ii) Density function and hazard rate		

3.3 Derivations of the following results:

i) Hazard rate of a series system of components having independent life times is summation of component hazard rates.

ii) Lifetime of series system of independent components with independent IFR life times is IFR.

3.4 Introduction to reliability data and related Inference:

Reliability concepts and Reliability Data: Introduction, Examples of reliability data, General models for reliability data, Repairable systems and Nonrepairable units.

Parametric Likelihood: Likelihood for complete samples for exponential distribution, confidence intervals for θ and for functions of θ, Examples in the context of reliability estimation. Construction of confidence intervals for exponential distribution.

Reference Books

1. Barlow, R. E., & Proschan, F. (1981). Statistical theory of reliability and life testing: Probability models. Holt, Rinehart and Winston.
2. Besterfield, D. H., Besterfield-Michna, C., Besterfield, G. H., Besterfield-Sacre, M., & Urdhwareshe, H. (2009). Total quality management (3rd ed.). Pearson Education.
3. Dodge, H. F., & Romig, H. G. (1959). Sampling inspection tables. John Wiley & Sons.
4. Duncan, A. J. (1974). Quality control and industrial statistics (4th ed.). D. B. Taraporevala Sons & Co. Pvt. Ltd.
5. Grant, E. L., & Leavenworth, R. S. (1980). Statistical quality control (5th ed.). McGraw-Hill.
6. Meeker, W. Q., & Escobar, L. A. (1998). Statistical methods for reliability data. Wiley-Interscience.
7. Montgomery, D. C. (2008). Introduction to statistical quality control (6th ed.). John Wiley & Sons.
8. Sinha, S. K. (1987). Reliability and life testing. Wiley Eastern Ltd.
9. Trivedi, K. S. (2001). Probability and statistics with reliability, queuing, and computer science applications. Prentice Hall of India.

STS-313-MJP Statistics Practical – VII (based on STS-310-MJ)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60**

Course Outcomes

By the end of the practical course, students should be able to:

CO1: Construct graphical representations of time series data and identify its major components such as trend and seasonality.

CO2: Estimate and analyze trend in time series data using graphical method, moving averages, least squares method, and exponential smoothing.

CO3: Compute and interpret seasonal indices using simple average, ratio-to-moving-average, and ratio-to-trend methods.

CO4: Examine stationarity of time series data and compute autocorrelation measures including ACF and correlogram.

CO5: Fit and interpret autoregressive, moving average, ARMA, and ARIMA models for time series forecasting.

List of Practical:

Sr. No.	Title of the experiment	No. of experiments
1	Graphical Representation and Components of Time Series and Estimation of Trend by Graphical Method	1
2	Estimation of Trend by Moving Average Method	1
3	Fitting Linear Trend by Least Squares Method	1
4	Estimation of Trend by Exponential Smoothing Method	1
5	Computation of seasonal indices by method of simple averages	1
6	Computation of seasonal indices by using Ratio to moving average method	1
7	Computation of seasonal indices by using Ratio to trend method	1
8	Checking stationarity and differencing for elimination of trend and seasonality	1
9	Computation of sample ACF and drawing a correlogram	1
10	Fitting of AR(1) and AR(2) model	1
11	Fitting of ARMA model	1
12	Fitting of ARIMA Model and Forecasting	1

STS-314-MJP Statistics Practical – VIII (based on STS-311-MJ)

Course type - Theory | **No. of credits – 2** | **No. of contact hours – 60**

Course Outcomes

By the end of the practical course, students should be able to:

CO1: Formulate real-life problems as Linear Programming Problems (LPP) and represent them mathematically.

CO2: Solve Linear Programming Problems using graphical and simplex methods, including Big-M and dual simplex methods.

CO3: Construct and solve the dual of a given linear programming problem and interpret the results.

CO4: Apply optimization techniques to solve transportation and assignment problems for minimization and maximization cases.

CO5: Apply network analysis techniques such as Critical Path Method (CPM) and Project Evaluation and Review Technique (PERT) for project scheduling and management.

List of Practical:

Sr. No.	Title of the Experiment	No. of Experiments
1	Formulation of Linear Programming Problem (LPP)	1
2	Solution of LPP using Graphical Method	1
3	Solution of LPP using Simplex Method	1
4	Solution of LPP using Simplex Method – Big M Method	
5	Solution of LPP using Dual Simplex Method	1
6	Finding initial basic feasible solution using: North-West Corner Rule, Least Cost Method, VAM	1
7	Finding optimal solution of Transportation Problem using U–V Method (MODI method)	1
8	Solution of Assignment Problem using Hungarian Method (balanced case).	1
9	Solution of Unbalanced Assignment Problem and its conversion to balanced form.	1
10	Solution of Transportation and Assignment Problem when Problem is of Maximization type	1
11	Project Evaluation and Review Technique (PERT)	1
12	Critical Path Method (CPM)	1

Note: Practical Based on Operations Research are to be verified using TORA

STS-315-MJP: Statistics Practical – IX (based on STS-312-MJ)

Course type - Theory **No. of credits – 2** **No. of contact hours – 60**

Course Outcomes

By the end of the course, students should be able to:

CO1: Construct structure functions and reliability block diagrams for series, parallel, and k-out-of-n systems.

CO2: Identify minimal path sets, minimal cut sets, and dual structure functions of coherent systems.

CO3: Express coherent system structure functions in terms of minimal path sets and cut sets and determine structural importance of components.

CO4: Compute reliability of coherent systems using minimal path and minimal cut set methods.

CO5: Calculate hazard rate and survival functions for different life distributions.

CO6: Analyze ageing properties and determine hazard rate behavior for series systems of components.

List of Practical:

Sr. No.	Title of the experiment	No. of experiments
1	Construction of Structure Function of Series and Parallel Systems	1
2	Construction of Structure Function of k-out-of-n System	1
3	Drawing of Reliability Block Diagram and Identification of System Structure	1
4	Determination of Dual of Structure Function	1
5	Determination of Minimal Path Sets of a Coherent System	1
6	Determination of Minimal Cut Sets of a Coherent System	1
7	Expression of Structure Function in terms of Minimal Path Sets and Cut Sets	1
8	Calculation of Structural Importance of Components	1
9	Calculation of Reliability of Coherent System using Minimal Path Sets	1
10	Calculation of Reliability of Coherent System using Minimal Cut Sets	1
11	Calculation of Hazard Rate and Survival Function for given Life Distribution	1
12	Calculation of Hazard Rate of Series System and Study of Ageing Properties	1

STS- 321-VSCP: Practical on C Programming (Turbo C)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60**

Course Outcomes

By the end of the course, students should be able to:

CO1: To introduce the foundations of computing, programming and problem- solving using computers with structured programming approaches.
CO2: To develop the ability to analyze a problem and devise an algorithm to solve it.
CO3: To formulate algorithms, pseudocodes and flowcharts for arithmetic and logical problems.
CO4: To implement algorithms in the 'C' language, test, debug and execute programs.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**C- Fundamentals**	**04**	**CO1**
1.1	History of 'C' language, Application areas, structure of a 'C' program.		
1.2	'C' Program development life cycle.		
1.3	Function as building blocks. 'C' tokens, Character set, Keywords, Identifiers, variables, Constants (character, integer, float).		
1.4	Data Types (Built-in and user defined data types), Operators, Expressions, types of operators, Operator precedence and Order of evaluation.		
1.5	Character input and output, String input and output, formatted input and output.		
2	**Control Structures**	**08**	**CO2, CO4**
2.1	Introduction to control structures		
2.2	If, if else, while, do...while, for, switch, goto, break, continue, nested loops, programs using control structures.		
3	**Arrays**	**11**	**CO3, CO4**
3.1	Introduction and Concept of arrays		
3.2	Operation on matrices -declaration, definition, initialization of array, problem using arrays, passing to function.		
4	**Functions**	**07**	**CO4**
4.1	Introduction to function		
4.2	Declaration, definition, recursion, user defined functions, library function, calling a function by reference and by value, local and global variables.		

Reference Books

1. Balgurusami, E. (2004). Programming in ANSI C. Tata McGraw-Hill Education.
2. Gottfried, B. S. (1996). Programming with C (Schaum's Outline Series). McGraw-Hill.
3. Kanetkar, Y. (2008). Let us C. BPB Publications.
4. Kernighan, B. W., & Ritchie, D. M. (1988). The C programming language (2nd ed.). Prentice Hall.
5. Rajaraman, V. (2007). Computer programming in C. Prentice Hall of India.

Programs to be covered based on the above units

1. Converting °C temperature to °F
2. To carry out arithmetic calculations (addition, subtraction, multiplication, division, remainder).
3. To find area and perimeter of triangle, rectangle, square, etc.
4. To find area and circumference of circle.
5. To find roots of quadratic equation and print the output according to discriminant.
6. To check whether given number is odd or even.
7. To check whether given number m is divisible by n or not.
8. To find maximum of 2 numbers or 3 numbers.
9. To check whether integer is prime or not.
10. To do arithmetic operations using do while, switch operations.
11. To find arithmetic mean, geometric mean, harmonic mean, variance and coefficient of variation of frequency distribution (individual observations, ungrouped and grouped frequency distribution).
12. To find correlation coefficient and least square regression line of Y on X for a given bivariate data.
13. To arrange the given data in increasing/decreasing order of magnitude.
14. To obtain median of given n observations (individual observations, ungrouped and grouped frequency distribution).
15. To obtain addition and subtraction of two matrices, multiplication of two matrices.
16. To find factorial of integer number (both recursive and non-recursive).

STS-341-MN: Statistical Methods

Course type - Theory | **No. of credits – 2** | **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Concept of sampling techniques and its application in real life.

CO2: Index number as a Statistical tools and comparative use of it for changes in prices and quantities.

CO3: Use of SQC to monitor and control a process for quality product.

CO4: Components of time series, the methodology used for trend and seasonality.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Sampling Methods**	**09**	**CO1**
1.1	Simple random sampling i) with replacement (SRSWR) and ii) without replacement (SRSWOR),		
1.2	Definitions of population mean, population variance and population total parameters, inclusion probabilities.		
1.3	Sample mean ($\bar{y}$) as an estimator of population mean ($\bar{Y}$), derivation of expectation and standard error of $\bar{y}$, $N\bar{y}$ as an estimator of population total, derivation of expectation and standard error of $N\bar{y}$, Estimator of above standard errors, case of SRSWOR		
1.4	Stratified random sampling as a sample drawn from individual strata using SRSWOR in each stratum. a) $\bar{y}_{st} = \frac{1}{N}\sum N_i \bar{y}_i$ as an estimator of population mean ($\bar{Y}$), derivation of expectation and standard error of $\bar{y}_{st}$. b) $N\bar{y}_{st}$ as an estimator of population total, derivation of expectation and standard error of $N\bar{y}_{st}$.		
1.5	Examples.		
2	**Index Numbers**	**05**	**CO2**
2.1	Introduction and scope of Index Numbers.		
2.2	Various types of Index Numbers like Human Development Index, Happiness Index BSE sensitivity Index.		
2.3	Problems/considerations in the construction of index numbers.		
2.4	Simple and weighted price index numbers based on price relatives.		
2.5	Simple and weighted price index numbers based on aggregates.		
2.6	Laspeyre's, Paasche's and Fisher's Index numbers.		
2.7	Numerical problems		
3	**Statistical Quality Control (SQC)**	**06**	**CO3**
3.1	Introduction - Meaning and purpose of Statistical Process Control (SPC), quality of the product, need of process control, statistical process control		
3.2	Concept of Seven Process Control (PC) Tools of SPC - (i) Check Sheet, (ii) Cause and effect diagram (CED), (iii) Pareto Diagram, (iv) Histogram, (v) Control chart, (vi)Scatter Diagram, (vii) Design of Experiments (DOE).		

3.3 Control Charts for Variables – (X-bar & R Chart), 3 sigma control limits, drawing of control chart. 3.4 Numerical problems			
4	**Time Series**	**10**	**CO4**
4.1 Meaning and utility of time series 4.2 Components of time series: trend, seasonal variations, cyclical variations, irregular (error) fluctuations or noise. 4.3 Exploratory data analysis: Time series plot to (i) check any trend and seasonality in the time series (ii) identify the nature of trend. 4.4 Methods of trend estimation and smoothing: (i) moving average, (ii) linear, parabolic, exponential, Pareto curve fitting by least squares principle (iii) exponential smoothing. 4.5 Choosing parameters for smoothing and forecasting. 4.6 Forecasting based on exponential smoothing. 4.7 Numerical problems			

Reference Books

1. Brockwell, P. J., & Davis, R. A. (2003). Introduction to time series and forecasting (2nd ed.). Springer.
2. Chatfield, C. (2001). The analysis of time series: An introduction. Chapman & Hall/CRC.
3. Cochran, W. G. (2007). Sampling techniques (3rd ed.). Wiley India Pvt. Ltd.
4. Gupta, S. C., & Kapoor, V. K. (2007). Fundamentals of applied statistics (4th ed.). Sultan Chand & Sons.
5. Gupta, S. P. (2002). Statistical methods (31st ed.). Sultan Chand & Sons.
6. Hogg, R. V., Craig, A. T., & McKean, J. W. (2012). Introduction to mathematical statistics (10th impression ed.). Pearson Prentice Hall.
7. Latpate, R., Kshirsagar, J., Gupta, V. K., & Chandra, G. (2021). Advanced sampling methods. Springer.
8. Mukhopadhyay, P. (1999). Applied statistics. New Central Book Agency Pvt. Ltd.
9. Mukhopadhyay, P. (2008). Theory and methods of survey sampling. Prentice-Hall of India.

Field Project (FP)/ Community Engagement Project (CEP)

STS-331-FP/CEP: Field Project/ Community Engagement Project

Course type: Practical **No. of Credits: 2** **No. of Contact Hours: 60 (48+12)**

The *Field Project* course requires students to participate in field-based learning activities or projects under the supervision of faculty members. This curricular component is designed to expose students to practical applications of their academic discipline within society. It aims to bridge the gap between theoretical knowledge and real-life situations by encouraging students to observe, analyze, and address practical problems through field engagement.

The field project enables students to connect classroom learning with real-world contexts and to develop solutions to practical issues relevant to their domain.

Course Outcomes

CO1: Identify and formulate a real-life problem suitable for statistical investigation.
CO2: Design and conduct a field survey using appropriate data collection methods such as questionnaires or interviews.
CO3: Organize, summarize, and analyze the collected data using appropriate statistical techniques and software tools.
CO4: Interpret the results of the analysis and draw meaningful conclusions related to the field problem.
CO5: Prepare and present a structured project report demonstrating the application of statistical methods to real-life situations.

Objectives of the Field Project:

- To provide hands-on experience in applying theoretical knowledge to real-life situations.
- To enhance skills in data collection, data analysis, and scientific investigation.
- To develop teamwork, problem-solving ability, and independent research skills.
  - Enhance ability in **data analysis using statistical tools/software**.
  - Develop **analytical thinking and interpretation of results**.
  - Encourage **teamwork, communication, and report writing skills**.

As per the **NEP** guidelines, undergraduate students are required to complete the Field Project during the **fifth semester starting from the academic year 2026–27**.

The Field Project carries **2 credits**, which will be earned through field engagement activities. The specific mode of implementation, nature of fieldwork, and evaluation methods is given below.

Students may undertake projects such as:

- **Survey-based studies** (primary data collection using questionnaires)
- **Statistical analysis of secondary data** from government reports, websites, or research publications

- **Community-based statistical studies**
- **Institutional or industrial data analysis**

Projects should be conducted **in groups (not more than 5 students)**.

Evaluation Scheme for External Assessment (30 Marks)

Component	Marks
Statistical Analysis and Interpretation	10
Project Report (Structure, Presentation, Methodology)	10
Presentation and Viva Voce	10
Total	**30 Marks**

Internal Assessment (20 Marks)

Component	Marks
Project Proposal and Topic Selection	5
Data Collection / Field Work	5
Progress and Regular Interaction with Supervisor	5
Preliminary Data Analysis / Work Diary	5
Total	**20 Marks**

Semester VI

STS-351-MJ: Testing of Hypothesis

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Formulate and differentiate Statistical Hypotheses

CO2: Apply Neyman-Pearson Lemma to construct most powerful test.

CO3: Construct and apply Likelihood ratio test

CO4: Apply and Interpret the non-parametric tests.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Parametric Tests**	**12**	**CO1**
1.1	Statistical hypothesis, problem of testing of hypotheses. Definition and illustrations of simple hypothesis, composite hypothesis, test of hypothesis, critical region, type I and type II errors.		
1.2	Probabilities of type I error and type II error. The problem of controlling the two types of error.		
1.3	Definition and illustrations of level of significance, observed level of significance (p-value), size of a test, power of a test.		
1.4	Definition of most powerful (M.P.) level α test of simple null hypothesis against a simple alternative. Statement of Neyman - Pearson (N-P) lemma for constructing the most powerful level α test of simple null hypothesis against simple alternative hypothesis. Illustrations.		
1.5	Power function of a test, power curve, definition of uniformly most powerful (UMP) level α test for one-sided alternative. Examples on UMP test. Situations where UMP test fails		
2	**Likelihood ratio tests**	**08**	**CO2**
2.1	Notion of likelihood ratio test (LRT) $\Lambda(x) = \frac{sup_{\theta\in\Theta_0} L(x)}{sup_{\theta\in\Theta_0\cup\Theta_1} L(x)}$		
2.2	Application of Likelihood Ratio Test for Testing $H_0: p = p_0$ against $H_1: p \neq p_0$ in a Binomial Distribution		
2.3	Construction of LRT for $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ for the mean of the normal distribution with i) known σ^2 ii) unknown σ^2		
2.4	Construction of LRT for $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$ for the mean of the normal distribution with i) known variance σ^2 ii) unknown variance σ^2		
2.5	Construction of LRT for $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ for the mean of the normal distribution with i) known variance σ^2 ii) unknown variance σ^2		
2.6	Construction of LRT for $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$ for the variance of the normal distribution with i) known mean μ ii) unknown mean μ		
2.7	Construction of LRT for $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 < \sigma_0^2$ for the variance of the normal distribution with i) known mean μ ii) unknown mean μ		

2.8	Construction of LRT for $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 < \sigma_0^2$ for the variance of the normal distribution with i) known mean μ ii) unknown mean μ		
3	**Non-parametric Tests**	**12**	**CO3, CO4**
3.1	Concept of non-parametric tests. Distinguish between parametric and nonparametric Tests. Concept of distribution-free statistic.		
3.2	One-tailed and two-tailed test procedure of the Sign test.		
3.3	One-tailed and two-tailed test procedure of the Wilcoxon signed rank test.		
3.4	One-tailed and two-tailed test procedure of the Mann-Whitney U test.		
3.5	One-tailed and two-tailed test procedure of the Run test, one-sample and two-sample problem.		
3.6	Empirical distribution function $S_n(x)$. Properties of $S_n(x)$ as an estimator of $F(.)$		
3.7	Kolmogorov- Smirnov test for completely specified univariate distribution (one-sample problem only) for two-sided alternative hypotheses. Comparison with the chi-square test.		

Reference Books

1. Buyan, K. C. (2010). Probability theory and statistical inference (1st ed.). New Central Book Agency.
2. Daniel, W. W. (2000). Applied nonparametric statistics (2nd ed.). Duxbury Press.
3. Dudewicz, E. J., & Mishra, S. N. (1988). Modern mathematical statistics. John Wiley & Sons.
4. Gibbons, J. D., & Chakraborti, S. (2010). Nonparametric statistical inference (5th ed.). CRC Press.
5. Gupta, S. C., & Kapoor, V. K. (2020). Fundamentals of mathematical statistics (12th ed.). Sultan Chand & Sons.
6. Hoel, P.G., Port, S.C., & Stone, C.J. (1972). Introduction to statistical theory. Houghton Mifflin Company.
7. Hogg, R.V., McKean, J.W., & Craig, A.T. (2019). Introduction to mathematical statistics (8th ed.). Pearson.
8. Kendall, M. G., & Stuart, A. (1943). The advanced theory of statistics (Vol. 1). Charles Griffin & Company Ltd.
9. Lindgren, B. W. (1993). Statistical theory (4th ed.). CRC Press.
10. Mood, A. M., Graybill, F. A., & Boes, D. C. (2017). Introduction to the theory of statistics (3rd ed.). McGraw-Hill Education.
11. Rohatgi, V. K., & Saleh, A. K. Md. E. (2015). An introduction to probability theory and statistics (3rd ed.). John Wiley & Sons.

STS-352-MJ: Sampling Theory

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Explain the role of sample surveys in research methodology and design reliable and valid questionnaires for data collection.

CO2: Apply the principles of Simple Random Sampling (SRSWR and SRSWOR) and derive estimators for population mean, total, and proportion.

CO3: Analyze stratified random sampling techniques including proportional, Neyman's, and optimum allocation methods.

CO4: Apply systematic sampling techniques and compare their efficiency with other sampling methods.

CO5: Determine appropriate sample size for estimating population parameters under specified precision and confidence levels.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Role of Sample Surveys in Research Methodology**	**03**	**CO1**
1.1 Objectives of a sample survey. 1.2 Designing a questionnaire, characteristics of a good questionnaire (Questions with codes & scores are to be discussed). Reliability and validity testing by using i) Test–Retest method, ii) Internal Consistency: a) Kuder Rechardson Coefficient (KR-20) b) Cronbach's Coefficient Alpha (α) 1.3 Planning, execution and analysis of a sample survey, practical problems at each of these stages. 1.4 Sampling and non-sampling errors with illustrations.			
2	**Simple Random Sampling**	**08**	**CO2**
2.1 Concept of distinguishable elementary units, sampling units, sampling frame, random sample, Requisites of a good sample, Simple random sampling from finite population of size (N), i) with replacement (SRSWR) ii) without replacement (SRSWOR), 2.2 Definitions of population mean, population variance and population total parameters, inclusion probabilities. 2.3 Sample mean ($\bar{x}$) is an estimator of population mean ($\bar{X}$), derivation of expectation and standard error of $\bar{x}$, N$\bar{x}$ as an estimator of population total, derivation of expectation and standard error of N$\bar{x}$, Estimator of above standard errors, both in case of SRSWR and SRSWOR, confidence interval for population mean, population total, standard error. 2.4 Sampling for proportion as an application of a simple random sampling with X_i as zero or one. Sample proportion (p) as an estimator of population proportion (P) of units possessing a certain attribute, derivation of expectation and standard error of (p), Np as an estimator of total number of units in the population possessing a certain attribute, derivation of expectation and standard error of Np. Estimator of above standard error both in case of SRSWR and SRSWOR.			
3	**Stratified Random Sampling**	**11**	**CO3**
3.1 Stratification, basis of stratification, real life situation where stratification can be used.			

3.2 Stratified random sampling as a sample drawn from individual strata using SRSWOR in each stratum 3.3 $\bar{x}_{st} = \frac{\sum N_i \bar{x}_i}{N}$ as an estimator of population mean problem of allocation, Proportional allocation, Neyman's allocation derivation of the expressions for the standard errors of the above estimators when these allocations are used. 3.4 Gain in precision due to stratification, comparison amongst SRSWOR, stratification with proportional allocation and stratification with Neyman's allocation. 3.5 Cost and variance analysis in stratified random sampling, minimization of variance for fixed cost, minimization of cost for fixed variance, optimum allocation, Neyman's allocation as a particular case of optimum allocation in cost and variance analysis.			
4	**Systematic Sampling (Population size divisible by sample size)**	**04**	**CO4**
4.1 Real life situations where systematic sampling is appropriate. Techniques of drawing a sample using systematic sampling 4.2 Estimation of the population mean and population total, standard error of these estimators. 4.3 Comparison of systematic sampling with SRSWOR. 4.4 Comparison of systematic sampling with SRSWOR and stratified sampling in the presence of linear trend.			
4	**Determination of Sample Size**	**04**	**CO5**
5.1 Determination of the sample size for estimating population mean, population total and population proportion for the given: Margin of error and confidence coefficient. 5.2 Coefficient of variation of the estimator and confidence coefficient.			
Reference Books			

1. Arnab, R. (2017). Survey sampling theory and applications. Academic Press.
2. Cochran, W. G. (2007). Sampling techniques (3^{rd} ed.). Wiley India Pvt. Ltd.
3. Mukhopadhyay, P. (2008). Theory and methods of survey sampling. Prentice-Hall of India.
4. Murthy, M. N. (1977). Sampling theory and methods. Statistical Publishing Society.
5. Sampath, S. (2005). Sampling theory and methods (2^{nd} ed.). Narosa Publishing House.
6. Singh, D., & Chaudhary, F. S. (1986). Theory and analysis of sample survey designs. Wiley Eastern Ltd.
7. Sukhatme, P. V., & Sukhatme, B. V. (1984). Sampling theory of surveys with applications. Indian Society of Agricultural Statistics.
8. Wu, C., & Thompson, M. E. (2020). Sampling theory and practice. Springer Nature Switzerland.

STS-353-MJ: Statistical Process and Product Control

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Understand the concept of quality, SPC and process control tools.

CO2: Build the 3 σ control limits for variables and attributes control charts.

CO3: Compare chance causes and assignable causes of variation. Construct and interpret control charts for variables and attributes.

CO4: Calculate and interpret process capability indices. Assess the stability of a process using control charts and evaluate its capability using capability indices.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction to SPC**	**05**	**CO1**
1.1	Introduction to SPC, Quality of the product (8-dimensions of quality) traditional and modern definition of quality, Chance causes and assignable causes of variation, defect and defective.		
1.2	Seven Process Control (PC) Tools of SPC (i) Check Sheet, (ii) Cause and effect diagram (CED), (iii) Pareto Diagram, (iv) Histogram, (v) Control chart, (vi) Scatter Diagram, (vii) Design of Experiments (DOE).		
2	**Introduction to Control Charts**	**02**	**CO2**
2.1	Chance causes and assignable causes of variation, statistical basis of control charts, exact probability limits, k -sigma limits, justification for the use of 3- sigma limits for normal distribution and using Chebychev's inequality for non-normal distributions		
2.2	Criteria for detecting lack of control situations: At least one point outside the control limits, A run of seven or more points above or below the central line. Presence of a non-random pattern eg. cyclic or linear trends etc.		
2.3	Construction of control charts for (i) standards given, (ii) standards not given		
3	**Control Charts for Variables**	**06**	**CO3**
3.1	R chart and $\overline{X}$ chart: Purpose of R and $\overline{X}$ chart, Equivalence between control chart and testing of hypothesis problem.		
3.2	Construction of R chart when the process standard deviation is specified: control limits, drawing of control chart, plotting of sample ranges. Drawing conclusion - determination of state of control process,		
3.3	Construction of $\overline{X}$ chart when the process average is specified: control limits, drawing of control chart, plotting of sample means. Drawing conclusion - determination of state of control of process, corrective action if the process is out of statistical control.		
3.4	Construction of R chart when the process standard deviation (σ) is not given: control limits, drawing of control chart, plotting sample range values, revision of control limits if necessary, estimate of σ for future use.		
3.5	Construction of $\overline{X}$ chart when the process average μ is not given: Drawing of control chart, plotting sample means, revision of control limits of $\overline{X}$ chart, if necessary.		
4	**Control Charts for Attributes**	**04**	**CO3**
4.1	p – chart: Construction of p-chart when subgroup sizes are same and the standard is known:		

	control limits, drawing of control chart, plotting of sample fraction defectives. Determination of state of control of the process.		
4.2	Construction of p-chart when subgroup sizes are different and the value of the process fraction defective P is not specified with separate control limits, drawing of control chart, plotting sample fraction defectives, determination of the state of control of the process. Interpretation of high and low spots. Identification of real-life situations.		
4.3	C chart: Construction of c-chart when the standard is given; control limits justification of 3 sigma limits, drawing of control chart, plotting number of defects per unit.		
4.4	Construction of c chart when the standard is not given; control limits, explanation for the use of 3 sigma limits, drawing of control chart. Plotting number of defects per unit. Determination of state of control, interpretation of high and low spots.		
5	**Capability Studies**	**05**	**CO4**
5.1	Specification limits, natural tolerance limits and their comparisons, decisions based on these comparisons, estimate of present defective.		
5.2	Shift in the process average, evaluation of probability of detecting a shift (or getting signal) on the first sample or on the subsequent samples after the shift (when process standard deviation is fixed). Average Run Length (ARL) for X chart, Average Time to Signal (ATS). Operating Characteristic (O.C.) curve for X chart, using normality assumption.		
5.3	Capability ratio and capability indices (C_p), capability performance indices C_{pk} with respect to machine and process, interpretation, relationship between (i) C_p and C_{pk}, (ii) defective parts per million and C_p.		
6	**Acceptance of Sampling for Attributes**	**08**	**CO4**
6.1	Concept of sampling inspection plan, comparison between 100 % inspection and sampling inspection. Procedures of acceptance sampling with rectification, single sampling plan and double sampling plan. Explanation of the terms: Producer's risk. Consumer's risk, Acceptable Quality Level (AQL). Lot Tolerance Fraction Defective (LTFD), Average Outgoing Quality (AOQ), Average Outgoing Quality Limit (AOQL), Average Sample Number (ASN), Average Total Inspection (ATI), Operating characteristic (OC) curve, AOQ curve.		
6.2	Single Sampling Plan: Evaluation of probability of acceptance using. (i) Hypergeometric (ii) Binomial (iii) Poisson distributions. Derivation of AOQ and ATI.		
6.3	Double Sampling Plan: Evaluation of probability of acceptance using Poisson approximation. Derivation of AOQ, ASN and ATI (with complete inspection of second sample).		

Reference Books

1. Besterfield, D. H., Besterfield-Michna, C., Besterfield, G. H., Besterfield-Sacre, M., & Urdhwareshe, H. (2009). Total quality management (3rd ed.). Pearson Education.
2. Brase, C.H., & Brase, C.P. (2018). Understandable statistics: Concepts and methods (12th ed.). Cengage Learning.
3. Dodge, H. F., & Romig, H. G. (1959). Sampling inspection tables. John Wiley & Sons.
4. Duncan, A. J. (1974). Quality control and industrial statistics (4th ed.). D. B. Taraporevala Sons & Co. Pvt. Ltd.
5. Grant, E.L., & Leavenworth, R.S. (1980). Statistical quality control (5th ed.). McGraw-Hill.
6. Johnson, N. L., & Kotz, S. (1993). Capability studies. Chapman & Hall.
7. Montgomery, D. C. (1983). Statistical quality control. John Wiley & Sons.
8. Moore, D. S., Notz, W. I., & Fligner, M. A. (2013). The basic practice of statistics (6th ed.). W. H. Freeman and Company.

STS-354- MJ: Survival analysis

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Derive fundamental survival analysis functions and explain their interrelationships.

CO2: Analyze ageing concepts including no-ageing, positive ageing classes and prove their theoretical implications.

CO3: Classify major parametric life distributions and apply key properties of the exponential distribution in survival analysis.

CO4: Apply parameter estimation and nonparametric methods (Actuarial and Kaplan–Meier) under various censoring schemes in survival data.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Basic Concepts in Survival**	**02**	**CO1**
1.1 Survival function, probability density function, hazard function, cumulative hazard rate 1.2 Mean residual life function, equilibrium residual life function, interrelation between all these function.)			
2	**Ageing in Survival Models**	**10**	**CO2**
2.1 No ageing, proof of the properties of no ageing: i) Cauchy functional equation ii) Constant failure rate, iii) Constant mean residual life, iv) Exponential life distribution, v) Exponential Equilibrium distribution 2.2 Definitions of Positive ageing: IFR, IFRA, NBU, NBUE, DMRL, bathtub failure rate. 2.3 Proof of the Implications: IFR → IFRA, IFR → DMRL, IFRA → NBU, NBU → NBUE, DMRL → NBUE. 2.4 Examples			
3	**Parametric Families of life distributions**	**10**	**CO3**
3.1 Classification of following parametric families of life distribution according to ageing: Weibull, Gamma, lognormal, linear failure rate, Makeham, Pareto, Lehman. 3.2 Some Properties of Exponential Distribution: i) If $T_1, T_2, \dots \dots, T_n$ are independent with $T_i \to Exp(\lambda_i)$ for $i = 1,2,\dots,n$ and $T = \{T_1, T_2, \dots \dots, T_n\}$ Then $T \to Exp(\sum_i^n \lambda_i)$ ii) If $T_1, T_2, \dots \dots, T_n$ are independent and identically distributed exponential random variable with parameter λ, then $2\lambda\sum_i^n T_i \to \chi^2_{2n}$			
4	**Censoring Mechanisms and Nonparametric Estimation in Survival Analysis**	**08**	**CO4**
4.1 Examples of life time situations, straggled and aligned entries in the survival time data, concept of censoring, time/type I censoring, order/type II censoring, right random censoring, left random censoring, undersigned censoring. Estimation of parameter of exponential distribution using Type I and Type II censoring. 4.2 Nonparametric estimation of survival function, confidence band on survival function, actuarial estimator of survival function, Greenwood's formula, Kaplan-Meier estimator of survival function in the presence of censored observations.			

Reference Books

1. Deshpande, J.V. and Purohit S.G. (2005). Life Time Data: Statistical Models and Methods, Word Scientific.
2. Barlow, R. E. and Proschan F. (1996). Mathematical Theory of Reliability. John Wiley.
3. Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data, Chapman and Hall.
4. Barlow, R. E. and Proschan F. (1975). Statistical theory of Reliability and Life testing: Probability Models Holt, Rinehart and Winston Inc.
5. Tobias, P.A. and Trindane, D. C. (1995). Applied Reliability. Second edition. CRC Press.

STS-355-MJP: Statistics Practical-X (based on STS-351-MJ and STS-352-MJ)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60**

Course Outcomes:

By the end of the course, students should be able to:

CO1: Understand the concept of the power of a test, Most powerful test and Uniformly Most Powerful (U.M.P.) tests.

CO2: Apply the appropriate non-parametric tests (Sign test, Median test, Wilcoxon signed-rank, Mann-Whitney U, Run test, Kolmogorov-Smirnov test).

CO3: Determine the required sample size for estimating population parameters for both variables and attributes to achieve a desired precision and confidence level.

CO4: Understand parameter estimation and confidence interval construction in SRSWR and SRSWOR

CO5: Understand the concept of stratified random sampling with proportional and Neyman allocations, and perform cost and variance analysis.

Sr. No.	Title of the experiment	No. of experiment
1	Testing of hypotheses (Probability of type I and type II errors, power of a test).	1
2	Problems on Most Powerful (MP) test.	1
3	Uniformly Most Powerful (UMP) test. Plotting the power function of a test.	1
4	Non- parametric tests: Sign test and Wilcoxon's signed rank test.	1
5	Mann-Whitney U test, Run test.	1
6	Kolmogorov-Smirnov test, Median test.	1
7	Determination of sample size for variables and attributes.	1
8	Simple random sampling (estimation of population mean, population total with standard errors), i) with replacement, ii) without replacement. Confidence interval for population mean and population total.	1
9	Simple Random Sampling for Proportions (estimation of population proportion, population total with standard errors), confidence interval for population proportion and population total.	1
10	Stratified Random Sampling: Proportional and Neyman allocation, Comparison with SRSWOR.	1
11	Cost and Variance Analysis in Stratified Random Sampling.	1
12	Ratio and Regression methods of estimation. Comparison with SRSWOR.	1

STS-356-MJP: Statistics Practical-XI (based on STS-353-MJ and STS-354-MJ)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60**

Course Outcomes

By the end of the course, students should be able to:

CO1: Construct and interpret control charts for variables and attributes

CO2: Calculate and interpret process capability indices

CO3: Assess the stability of a process using control charts and evaluate its capability using capability indices

CO4: Estimate parameters of Weibull using Newton–Raphson and Gamma distributions using MLE methods including digamma function. Compute MLE of exponential distribution parameters under Type I and Type II censoring schemes.

CO5: Apply Actuarial estimator to estimate survival function and its standard error in the presence of censored data. Apply Kaplan–Meier estimator to compute and plot the survival function with its standard error against time.

Sr. No.	Title of the experiment	No. of experiment
1	Construction of Variable Control Charts ($\bar{X}$ and R Charts)	1
2	Construction of Attribute Control Charts (p-chart for fixed subgroup size and c-chart)	1
3	Construction of p-chart for varying subgroup size	1
4	Problems on Capability Studies	1
5	Single sampling plan for attributes (OC curve, AOQ, AOQL, ATI using Poisson distribution).	1
6	Double sampling plan for attribute (OC curve, AOQ, AOQL, ATI, ASN using Poisson distribution)	1
7	M.L.E. of parameters of Weibull distribution using Newton-Raphson method.	1
8	M.L.E. of parameters of Gamma distribution using digamma function.	1
9	Estimation of parameter of exponential distribution for type I and type II censored data.	1
10	Actuarial estimator of survival function and its standard error (in the presence of censored observations).	1
11	Kaplan-Meier estimator of survival function and its standard error (in the presence of censored observations). Plot of estimator against t.	1
12	Comparative Survival Analysis for Treatment Efficacy: Evaluating the Log-Rank Test and Deshpande’s J-Test under Non-propotional Hazards	1

Major Elective

STS-360-MJ: Stochastic Process

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcome

By the end of the course, students should be able to:

CO1: Model real-world transitions using Transition Probability Matrices and state-space diagrams.

CO2: Calculate n-step transition probabilities using the Chapman-Kolmogorov equations and matrix powers.

CO3: Determine the long-run stability of a system by finding its unique stationary distribution.

CO4: Concept of gambler's ruin and random walk models using absorption probability techniques.

CO5: Understand counting process and apply Poisson process properties to calculate expected event counts and waiting times in counting processes.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Foundations of Markov Chains**	**07**	**CO1, CO2**
1.1 Introduction: Definition of a Stochastic Process, State Space, Index Set. Classification based on nature of state space and index space (Discrete/Continuous time and space) with real-world examples. 1.2 The Markov Property, Definition of Markov Chains (MC) $\{X_n, n \geq 0\}$. Concept of finite and time-homogeneous Markov Chains. 1.3 Transition Dynamics: One-step transition probabilities and the Transition Probability Matrix (TPM). Properties of a Stochastic Matrix. Graphical representation (State Transition Diagrams). n-step transition probability matrix ($P^{(n)}$). 1.4 Chapman-Kolmogorov equations. Initial distribution and calculation of the joint distribution function of $\{X_0, X_1, \ldots, X_n\}$, Partial sum of i.i.d. random variables as a MC. 1.5 Simple Random Walk, and the Gambler's ruin problem.			
2	**Classification of States**	**08**	**CO2**
2.1 Communication & Relations: Concept of accessible and communicating states with its properties: Reflexive, Symmetric, and Transitive. 2.2 Return Probabilities: First return probability (f_{ii}^{n}) and the probability of ever returning to a state (f_{ii}). 2.3 Persistent (Recurrent) vs. Transient states. 2.4 Mean Recurrence Time, Null Recurrent, and Non-Null Recurrent states. 2.5 Periodicity of states: Periodic vs. Aperiodic states. 2.6 Closed sets, Irreducible vs. Reducible MCs, Absorbing states. 2.7 Ergodicity: Definition of an Ergodic state and Ergodic MC.			
3	**Stationary Distributions & Long-Run Behavior**	**07**	**CO3, CO4**
3.1 Equilibrium Concepts: Definition and elementary properties of a Stationary Distribution(π). Existence of a steady-state distribution for irreducible finite chains. 3.2 Computation: Solving the system of equations $\pi P = \pi$ subject to $\sum_i \pi_i = 1$.			

3.3 Long-run behavior of Markov Chains. Concept of the limiting distribution and its relation to the stationary distribution.			
4	**Counting & Poisson Processes**	**08**	**CO5**
4.1 Introduction to counting process, Definition and Postulates of the Poisson process. Derivation of Probability distribution of $N(t)$ (the number of occurrences in $(0, t]$). 4.2 Properties of Poisson Process: a. Additive property b. Difference of two independent Poisson processes c. Relationship between the Poisson process and the binomial distribution. Distribution of inter-arrival times and waiting times. d. Auto-correlation coefficient between N (t) and N (t + s) 4.3 Statistical Moments: Mean, Variance, and Covariance functions of N(t). 4.4 Definition of a Compound Poisson Process. Mean and variance functions. Applications in insurance (claim totals) and inventory management			
Reference Books			

1. Adke, S. R., & Manjunath, S. M. (1984). An introduction to finite Markov processes. Wiley Eastern.
2. Bhat, B. R. (2000). Stochastic models: Analysis and applications. New Age International.
3. Hoel, P. G., Port, S. C., & Stone, C. J. (1972). Introduction to stochastic processes. Houghton Mifflin Company.
4. Medhi, J. (1982). Stochastic processes. Wiley Eastern Limited.
5. Ross, S. M. (1996). Stochastic processes (2nd ed.). John Wiley & Sons.
6. Ross, S. M. (2000). Introduction to probability models (7th ed.). Academic Press.
7. Srinivasan, S. K., & Mehta, K. M. (1981). Stochastic processes. Tata McGraw-Hill.
8. Taylor, H. M., & Karlin, S. (1984). An introduction to stochastic modeling. Academic Press.

ST 361 MJ: Medical Statistics and Clinical Trials

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes:

By the end of the course, students should be able to:

CO1: Analyse population data and apply statistical models such as the exponential model to study population growth.

CO2: Analyse epidemiological data using statistical measures such as odds ratios and construct confidence intervals for interpretation.

CO3: Perform statistical analysis of clinical trial designs, including parallel and crossover studies.

CO4: Interpret bioavailability and bioequivalence in the context of drug development and medical research.

CO5: Estimation of PK Parameters and Determination of sample size in Clinical trials.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Population Study**	**02**	**CO1**
1.1	India's Population and Census		
1.2	Models for Population Growth. Linear, Exponential, and Sigmoidal Growth models.		
2	**Epidemiology**	**10**	**CO2, CO3**
2.1	Introduction to Epidemiology.		
2.2	Precision and Bias.		
2.3	Risk and Relative Risk.		
2.4	Odds and Odds Ratio(OR). Properties, inference for odds ratios, and log odds ratios. Relationship between odds ratio and relative risk.		
2.5	Estimation of odds ratio, Confidence interval for OR.		
2.6	Case Control Studies.		
2.7	Cohort Studies.		
2.8	Symmetry in square contingency tables, collapsing tables, and Simpson's paradox.		
2.9	McNemar's test for symmetry.		
2.10	Bowker's test for Mirror Image symmetry.		
3	**Clinical Trials**	**06**	**CO4**
3.1	General information on the history of drug discovery.		
3.2	Louis Pasteur (rabies and smallpox)		
3.3	Ronald Ross and malaria.		
3.4	Alexander Fleming and penicillin.		
3.5	Jonas Salk and polio.		
4	**Design of Clinical Trials**	**12**	**CO5**
4.1	Phases of clinical trial, purpose, duration, cost, drug regulatory bodies, ICH, statistical analysis plan, clinical study report.		
4.2	Parallel designs		
4.3	Two-period crossover design.		
4.4	Pharmacokinetics (PK) Parameters: Cmax, Tmax, $t_{1/2}$. Measurement of AUC, Measurement of $t_{1/2}$.		
4.5	Bioequivalence and bioavailability.		

4.6	Statistical Criteria for Assessing Bioequivalence. The 75/25 rule, The 80/20 rule, The ±20 rule. The 90% Confidence Interval rule.
4.7	Determination of Sample size. One sample known variance, Two-sample known variance.
Reference Books	

1. Agresti, A. (1996). An introduction to categorical data analysis. John Wiley & Sons.
2. Friedman, L. M., Furberg, C. D., & DeMets, D. L. (2015). Fundamentals of clinical trials (5th ed.). Springer.
3. Gore, A. P., & Paranjape, S. A. (2000). Course on mathematical and statistical ecology. Kluwer Academic Publishers.
4. Gore, A. P., Paranjape, S. A., & Kulkarni, M. B. (2013). Medical statistics. SIPF Academy.
5. Kulkarni, M. B., & Prayag, V. R. (2004). Introduction to statistical ecology. SIPF Academy.
6. Matthews, J. N. S. (2006). Introduction to randomized controlled clinical trials (2nd ed.). Chapman & Hall/CRC.
7. Selvin, S. (2004). Statistical analysis of epidemiologic data (3rd ed.). Oxford University Press.
8. Shoukri, M. M., & Cihon, C. (1998). Statistical methods for health sciences. CRC Press.
9. Senn, S. (2000). Statistical issues in drug development (2nd ed.). John Wiley & Sons.
10. Steven, P. (2000). Clinical trials: A methodological perspective. John Wiley & Sons.

ST 362 MJ: Operation Management

Course type - Theory **No. of credits – 2** **No. of contact hours – 30**

Course outcomes:

At the end of this course, students are able to understand:
CO1: Replacement, inventory methods and situations where to use it.
CO2: Decision methods in certainty and uncertainty cases.
CO3: Meaning of games and the situations to identify the mixed and pure games with their solution.
CO4: The sequencing and simulation techniques to solve the critical problems.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Replacement Theory**	03	CO1
1.1	Introduction of replacement theory.		
1.2	Types of failures – Gradual failure and sudden failure.		
1.3	Replacement of item whose efficiency deteriorates with time when (a) value of money remains constant during the period (for time as a discrete variable and continuous variable), (b) value of money changes with constant rate during the period.		
1.4	Numerical Problems will be conducted in practical.		
2	**Inventory Theory**	07	CO1
2.1	Introduction, meaning, functional role and reasons for carrying inventory.		
2.2	Inventory Control Methods – (i) Always Better Control (ABC) analysis, (ii) Vital, Essential and Desirable (VED) Analysis, (iii) Fast moving, Nonmoving, Slow moving, Dead (FNSD) Analysis.		
2.3	Description of generalized inventory model.		
2.4	Types of inventory models: (i) The economic lot size model with uniform demand, instantaneous replenishment rate and no shortage (with derivation), (ii) The Economic lot size model with uniform rate of demand, finite replenishment rate and no shortage (with derivation).		
2.5	Numerical problems will be conducted in practical.		
3	**Decision Theory**	07	CO2
3.1	Introduction, Types of Decision making Environment.		
3.2	Decision making under uncertainty - (i) Laplace criterion, (ii) Minimax / Maximax criterion, (iii) Savage minimax regret criterion, (iv) Hurwitz criterion.		
3.3	Decision making under risk: (i) Expected Monetary Value, (ii) Expected Opportunity Loss, (iii) Expected Value of Perfect Information.		
3.4	Numerical problems will be conducted in practical.		
4	**Game Theory**	07	CO3
4.1	Introduction, definitions – number of players, sum of gains or losses, strategies and its type.		
4.2	Two Person Zero-Sum game.		
4.3	Pure and Mixed strategies		
4.4	Principle of Dominance		
4.5	Solution methods for games without saddle point – algebraic, arithmetic, matrix.		
4.6	Numerical problems will be conducted in practical.		

5	**Sequencing Theory**	**03**	**CO4**
5.1 Introduction, notations, terminology, assumptions. 5.2 Processing of n jobs through two machines. 5.3 Processing of n jobs through three machines (reducible to two machines and n-jobs) 5.4 Numerical problems will be conducted in practical.			
6	**Simulation Techniques**	**03**	**CO4**
6.1 Introduction to simulations, merits, demerits, limitations 6.2 Pseudo random number generators: Linear congruential, mid-square method. 6.3 Monte Carlo method of simulation: Statistical applications of simulation in numerical integration such as computation of probabilities of events related to gamma, beta and bivariate normal distribution.			
Reference Books			

1. Gupta, P. K., & Hira, D. S. (2022). Introduction to Operations Research. S. Chand Publishing.
2. Kapoor, V. K. (2011). Operations research. Sultan Chand & Sons.
3. Phillips, D. T., Ravindran, A., & Solberg, J. J. (1987). Operations research: Principles and practice (2nd ed.). John Wiley & Sons.
4. Sasieni, M. W., Yaspan, A., & Friedman, L. (1959). Operations research: Methods and problems. Wiley.
5. Sharma, J. K. (1989). Mathematical models in operations research. Tata McGraw-Hill Publishing Company.
6. Taha, H. A. (1997). Operations research: An introduction (5th ed.). Prentice Hall of India.

STS-363-MJP: Statistics Practical XII (based on STS-360-MJ)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60**

Course Outcome

By the end of the course, students should be able to:

CO1: Construct transition matrices and state diagrams to model and predict the behavior of Markovian systems.

CO2: Classify states and chains based on recurrence and ergodicity to determine the structural properties of a process.

CO3: Calculate stationary and limiting distributions to analyze long-run equilibrium and system convergence.

CO4: Apply Poisson processes to model random arrivals and calculate inter-arrival times and aggregate risks.

List of Practical

Sr. No.	Title of the experiment	No. of experiment
1	Practical based on Markovian Property	1
2	Construction of TPM and State Transition Diagrams.	1
3	Computation of n-step Transition Probabilities.	1
4	Problem based on Joint and Initial Distributions.	1
5	Problems based on Communicating Classes and Periodicity.	1
6	Recurrence and Transience Analysis along with computation of Mean Recurrence Time, Null Recurrent, and Non-Null Recurrent states.	1
7	Determination of Ergodic States and Classification of a Markov Chain	1
8	Simulation of Random Walk and Gambler’s Ruin Problem.	1
9	Finding the Stationary Distribution.	1
10	Study of Limiting Behaviour and Convergence of a Markov Chain.	1
11	Study of Poisson Process and Analysis of Inter-Arrival Times.	1
12	Applications and Problem Solving on Compound Poisson Process.	1

STS-364-MJP: Statistics Practical XIII (based on STS-361-MJ)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60**

Course Outcomes:

By the end of the course, students should be able to:

CO1: Analyse population data and apply statistical models such as the exponential model to study population growth.

CO2: Analyse epidemiological data using statistical measures such as odds ratios and construct confidence intervals for interpretation.

CO3: Perform statistical analysis of clinical trial designs, including parallel and crossover studies.

CO4: Interpret bioavailability and bioequivalence in the context of drug development and medical research.

CO5: Estimation of PK Parameters

CO6: Determination of sample size in Clinical trials

List of Practical

Sr. No.	Title of the Experiment	No. of Experiment
1	Study of Population Growth Using an Exponential Model.	1
2	Estimation and Interpretation of Odds Ratio.	1
3	Construction of a Confidence Interval for Odds Ratio.	1
4	McNemar's Test for Symmetry.	1
5	Bowker's Test of Mirror Image Symmetry.	1
6	Analysis of Parallel group design.	1
7	Analysis of Cross Over Design. (To be carried out in 3 steps. Test for carry-over effect, Test for period effect, Comparison of two treatments)	1
8	Computation of PK parameters Cmax, Tmax, AUC	1
9	Computation of PK Parameter $t_{1/2}$ (Graphical Method and Estimation Using Exponential model)	1
10	Assessment of Bioequivalence (The 80/20 rule, the ± 20% rule, The 90% CI rule).	1
11	Determination of sample size (One sample known variance)	1
12	Determination of sample size (Two-sample with known variance)	1

STS-365-MJP: Statistics Practical XIV (based on STS-362-MJ)

Course type - Practical **No. of credits – 2** **No. of contact hours – 60**

Course Outcomes

At the end of this course, students are able to understand:

CO1: The places where replacement, inventory methods are used and how to solve it.

CO2: The methods to solve certainty and uncertainty cases in decision theory.

CO3: Meaning of games and the situations to identify the mixed and pure games with their solution.

CO4: The sequencing and simulation techniques to solve the critical problems.

List of Practical:

Sr. No.	Title of Experiment	No. of Experiments
1	Problems based on replacement models	1
2	Analysis of ABC, VED, FNSD inventory models for the given situation. (case studies)	1
3	Problems based on inventory model I	1
4	Problems based on inventory model II	1
5	Problems based on decision making under uncertainty	1
6	Problems based on decision making under risk	1
7	Problems based on pure strategies of game theory	1
8	Problems based on mixed strategies of game theory	1
9	Problems based on n-jobs on two machines.	1
10	Problems based on n-jobs on three machines (reducible to two machines and n-jobs)	1
11	Problems based on Pseudo random number generators: Linear congruential, mid-square method	1
12	Monte Carlo method of simulation: computation of probabilities of events related to gamma, beta and bivariate normal distribution.	1

STS-371-VSCP: Data Analytics Using Python

Course type - Practical **No. of credits – 2** **No. of contact hours – 30**

Course Outcomes

By the end of the course, students should be able to:

CO1: Prepare and preprocess raw data for data mining by applying techniques such as data cleaning, transformation, discretization, and visualization to ensure data quality and suitability for analysis.

CO2: Implement and analyze the supervised learning techniques, such as k-nearest neighbour, decision trees, and Naïve Bayesian classifiers, for solving classification problems

CO3: Evaluate classification models using appropriate metrics (e.g., accuracy, precision, recall) and validate their performance using techniques like cross-validation to ensure robustness and generalizability.

CO4: Demonstrate the working principles of advanced modeling techniques, including Artificial Neural Networks (ANN) and Support Vector Machines (SVM), and apply them to suitable datasets.

CO5: Apply unsupervised learning methods, specifically k-means and k-medoids clustering algorithms, and association rule mining using the Apriori algorithm to discover hidden patterns and relationships in data.

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	**Introduction to Data Mining**	**06**	**CO1**
1.1 Data preparation for knowledge discovery: Data understanding and data cleaning tools, Data transformation, Data Discretization, Data Visualization. 1.2 Data Mining Process: CRISP and SEEMA, Supervised and unsupervised learning techniques			
2	**Classification**	**06**	**CO2**
2.1 Problem of classification 2.2 classification techniques: k-nearest neighbour, decision tree, Naïve Bayesian, classification based on logistic regression			
3	**Model Evaluation, Selection and Classification Accuracy**	**09**	**CO2, CO3**
3.1 Model evaluation and selection: Metrics for Evaluating Classifier Performance, 3.2 Concept of training data, testing data and validation of model, Cross-Validation, Bootstrap, Model Selection Using Statistical Tests of Significance, 3.3 Techniques to Improve Classification 3.4 Accuracy: Introduction to Ensemble Methods, Bagging, 3.5 Boosting and AdaBoost, Improving Classification Accuracy of ClassIm balanced Data. 3.6 Examples.			
4	**ANN, SVM and Clustering**	**09**	**CO4, CO5**
4.1 Introduction to Artificial Neutral Network (ANN) and Support Vector Machine (SVM), 4.2 Clustering: k-means, k-mediods, 4.3 Market Basket Analysis: Association rules and prediction, Apriori Algorithm, data attributes, applications to electronic commerce. 4.4 Examples			
Reference Books			
1. Berson, A., & Smith, S. J. (1997). Data warehousing, data mining, and OLAP. McGraw-Hill. 2. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth & Brooks/Cole.			

3. Chattamvelli, R. (2015). Data mining methods. Alpha Science International.
4. Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press.
5. Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
6. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
7. McKinney, W. (2022). Python for data analysis: Data wrangling with pandas, NumPy, and Jupyter (3rd ed.). O'Reilly Media.
8. Mehrika, K., Mohan, C. K., & Ranka, S. (1997). Elements of artificial neural networks. Penram International Publishing.
9. Mitchell, T. M. (1997). Machine learning. McGraw-Hill.
10. Ripley, B. D. (1996). Pattern recognition and neural networks. Cambridge University Press.
11. Vapnik, V. N. (2013). The nature of statistical learning theory (2nd ed.). Springer.

Programs to be covered based on the above units:

1. **Data Understanding and Preparation:** Perform data cleaning and transformation on a raw dataset (e.g., handling missing values, scaling/normalizing features) using libraries like Pandas and Scikit-learn.
2. **Data Exploration and Visualization:** Create various data visualizations (histograms, box plots, scatter plots, correlation matrices) using Matplotlib and Seaborn to understand data distributions and relationships.
3. **Data Discretization:** Implement equal-width and equal-frequency binning to convert continuous data into categorical intervals.
4. **k-Nearest Neighbour (k-NN) Classification:** Implement the k-NN algorithm from scratch or using Scikit-learn to classify data and analyze the impact of different 'k' values and distance metrics.
5. **Decision Tree Classification:** Build a decision tree classifier on a dataset, visualize the tree, and interpret the splitting rules and feature importance.
6. **Naïve Bayes Classification:** Implement a Gaussian or Multinomial Naïve Bayes classifier for a text classification or other suitable problem and evaluate its performance.
7. **Model Evaluation and Cross-Validation:** Split a dataset into training and testing sets and evaluate multiple models using k-fold cross-validation, reporting metrics like accuracy, precision, recall, and F1-score.
8. **Improving Accuracy with Ensemble Methods:** Implement a Bagging classifier (e.g., Random Forest) and a Boosting classifier (e.g., AdaBoost) and compare their performance against a single decision tree on the same dataset.
9. **Handling Class Imbalance:** Apply techniques like Random Oversampling or SMOTE to handle an imbalanced dataset and compare the classification results before and after balancing.
10. **Clustering with k-means:** Implement the k-means clustering algorithm, determine the optimal number of clusters using the elbow method, and visualize the resulting clusters.
11. **Clustering with k-medoids (PAM):** Implement the k-medoids algorithm and compare its clustering results with k-means on a dataset containing outliers.
12. **Market Basket Analysis with a priori Algorithm:** Implement the Apriori algorithm to mine frequent item sets and generate association rules from a transactional dataset (e.g., supermarket sales data), interpreting metrics like support, confidence, and lift.

STS-381-OJT: On Job Training

Course type – OJT **No. of credits – 4** **No. of contact hours – 60**

OJT provides students with practical exposure to real-world applications of statistics in industries, research organizations, government agencies, or other professional environments. Through this training, students gain hands-on experience in data collection, data analysis, statistical software usage, and interpretation of results. The training helps bridge the gap between theoretical knowledge and practical implementation.

There should be a supervisor from the organization/Institute from which the OJT/FP is being done apart from the internal supervisor. The completion certificate as well as the activity report should be signed by both supervisors.

Course Outcomes

CO1: Understand real-life applications of statistical methods in industry or research environments.
CO2: Collect, organize, and analyze real-world data using appropriate statistical techniques.
CO3: Interpret results and prepare statistical reports based on practical data analysis.
CO4: Develop professional skills such as teamwork, communication, and problem-solving.

Objectives

1. To provide practical exposure to real-life applications of statistical methods.
2. To develop skills in data collection, data management, and statistical analysis.
3. To familiarize students with statistical software and data analysis tools used in professional environments.
4. To enhance problem-solving, teamwork, and professional communication skills.

Duration: Minimum 120 hours (3-4 weeks) of training in an industry, research institute, NGO, hospital, financial institution, agricultural research center, or any relevant organization.

During the training period, students may engage in activities such as:

- Data collection and survey design
- Data entry and data cleaning
- Statistical analysis of real datasets
- Use of statistical software
- Preparation of tables, graphs, and statistical reports
- Interpretation of statistical results for decision-making
- Participation in project discussions and teamwork

Evaluation Scheme (Example: 100 Marks)

Component	**Marks**
External Assessment	
Submission of Report	40
Presentation / Viva Voce	20
Internal Assessment (attendance & participation)	40
Total	**100**