Savitribai Phule Pune University

(Formerly University of Pune)

REVISED SYLLABUS

OF

T. Y. B. A. STATISTICS

(General and Special)

Choice Based Credit System Syllabus

With Effect from June 2021

Preamble:

Statistics is used in different ways in different contexts. For example, the manager of a manufacturing unit, Statistics may be the information about the production/manufacturing process. For a medical researcher, investigating the effects of a new drug, Statistics is the evidence of research efforts. For a college student, Statistics shows the grades or marks scored in a course. For a cricket fan, Statistics is the information about runs scored or wickets taken by a player. Thus, in all these examples, Statistics refers to quantitative data in the area under study. Statistics as a subject is an important branch of knowledge and is devoted to various techniques of collection, presentation, analysis and interpretation of data. It is a science of learning from data. The subject provides tools formaking decisions when conditions of uncertainty prevail. Hence Statistical tools and techniques are used in almost all fields such as agriculture, business, management, economics, finance, insurance, education, sports, biotechnology, medical science, etc. For the last two decades, large amount of data has been collected with the help of computers and more sophisticated statistical techniques are needed for the effective analysis and meaningful conclusions from these data. Knowledge of different aspects of Statistics has become crucial in the present technologically advanced scenario. There is a continuous demand for statisticians in fields of education, industry, software and research. The syllabi of three-year B.A. (special) degree course in Statistics are framed in such a way that the students at the end of the course are thorough in basic statistical techniques and are ready to pursuea Master degree and/or simultaneously able to seek jobs involving statistical analysis related to a variety of data sets in order to arrive at some valid conclusions.

Note:

(1) A student of the Three-Year B.A. Degree Course offering 'Statistics' at the special level must offer 'Mathematical Statistics' as a General level subject in all the three years of the course. Further, students of the three-year B.A. Degree Course are advised not to offer 'Statistics' as the General level unless they have offered 'Mathematical Statistics' as a General level subject in all the three years of the course.

(2) A student of three-year B.A. Degree Course offering 'Statistics' will not be allowed to offer 'Applied Statistics' and 'Statistical Pre-requisites' in any of the three years of the course.

(3) A student offering 'Statistics' at the Special level must complete all practicals in Practical paper to the satisfaction of the teacher concerned.

(4) He/She must produce the laboratory journal along with the completion certificate signed by the Head of the Department at the time of Practical Examination.

(5) Preparation by Internal Examiner for practical (Calculator / Computer based):

Keep at least 15 computers with latest configuration ready with battery backup and necessary software, printers, scientific calculators, necessary statistical tables, normal probability paper during the practical examination in the Statistics Laboratory.

(6) Structure of evaluation of practical paper at T.Y.B.A. Statistics:

A) Continuous Internal Assessment (CIA):

Section	Marks
i) Journal	20
ii) Viva-voce	10
iii) Project individual with one credit	*
Total of CIA	30

*Project for one credit will be evaluated as per SPPU norms.

B) End of Semester Examination (ESE):

Section	Nature	Marks	Time
I	Using Calculator / Computer Note: Attempt any two of the four questions (each of 30 marks): Q1, Q2, Q3 and Q4:	60	*****3 hours And ^2hrs 30 minutes
III	Viva-voce	10	10 minutes
	Total of B	**70**	#

Foot note:

* For calculator user

^ For computer user

Total examination time 3 hours 10 minutes for calculator user and 2 hours 40 minutes for computer user.

(7) Structure of evaluation of theory paper at T.Y.B.A. Statistics (special and general), Applied statistics and Statistical pre-requisites:

A) Continuous Internal Assessment (CIA) for theory subjects:

Section	Marks
i) Theory examination	20
ii) Home assignment	05
iii) Seminar/class test etc.	05
Total of CIA	30

B) End Semester Examination (ESE) for theory subjects:

Question	Nature	Marks
1	a) Choose correct alternative: i) ii) iii) iv) v) (each with four multiple choice answer A, B, C, D) b) True or false: i) ii) iii) iv) v)	5 5
2	Attempt any four of the following: a) b) c) d) e) f)	20
3	Attempt any four of the following: a) b) c) d) e) f)	20
4	Attempt any two of the following: a) b) c) d)	20
	Total ESE	**70***

Foot note:
* Numerical problem should not exceed 40% of total marks with option questions.

8. Evaluation for additional credit:

The evaluation for Skill Enhancement Courses (SEC), Generic courses (GE) and one credit added in Discipline Specific Elective Courses (DSE) at T.Y.B.A. special students will be done as per SPPU norms.

9. Instruction for Examination:

1. Theory question paper for each subject shall cover all the topics in the pertaining syllabus with proportional weightage to the number of hours of prescribed.
2. The practicals are to be conducted in batches as per the University norms.
3. Medium of Instruction: English
4. Examination:

A) Pattern of examination: Semester wise

B) Standard of passing: As per norms of University

10.Objectives:

1. To study various continuous probability distributions with real life situations.
2. To identify the appropriate probability model that can be used in real life.
3. To formulate real life situation by set linear constraints and optimize the aim by using various techniques of Operations Research.
4. To test various hypotheses of significance like means, proportions, independence of attributes, variance etc. included in theory (using calculators, software).
5. To study applications of statistics in the field of design of experiments, estimations, operations research, etc.

13. Outcomes: At the end of course, students are able

1.To apply various continuous probability distributions with real life situations.

2. To apply the appropriate probability model that can be used in real life.

3. To solve the real life situation by set linear constraints and optimize the aim by using various techniques of Operations Research.

4. To test various hypotheses of significance like means, proportions, independence of attributes, variance etc. included in theory (using calculators, software).

5. To study the applications of statistics in the field of design of experiments, estimations, operations research, etc.

T.Y.B.A. STATISTICS Syllabus
For Choice Based Credit System-2019 pattern
To be implemented from the Academic year 2021-2022
Structure of the course:

Table with code and Title for CBCS 2019 pattern:

Subject	**SemesterV**	**SemesterVI**	**Sem. Credits**		Marks	
			V	VI	CIA	ESE
Compulsory English	---	---	3	3	30	70
Statistics General-III	ST-33875: Design and Analysis of Experiments **(CC 1E)**	ST-33876: Operations Research **(CC 1F)**	3	3	30	70
Statistics Special-III	ST-33885: Distribution Theory–I **(DSE 1C)**	ST-33886: Distribution Theory – II **(DSE 1D)**	**3+1***	**3+1***	30	70
Statistics Special-IV	ST-33895: Statistics Practical **(DSE 2C)**	ST-33896: Statistics Practical **(DSE 2D)**	**3+1***	**3+1***	30	70
Mathematical Statistics General-IV	ST-33275:Theory of Estimation **(CC 2E)**	ST-33276: Testing of Hypothesis **(CC 2F)**	3	3	30	70
Skill Enhancement Course **(SEC)**	**General subject other than Statistics (SEC 1C)**	**General subject other than Statistics (SEC1D)**	3	3	30	70
Skill Enhancement Course **(SEC)**	Medical Statistics and Clinical Trials **(SEC 2C)**	Data Analytics **(SEC 2D)**	2	2	*	*
Generic Elective	Time Series Analysis **(GE-1)**	Operations Management **(GE-2)**	2	2	*	*
Total credit for third year for Statistics special students **(48)**			**24**	**24**		

*Evaluation will be done as per SPPU norms

Equivalence for old courses (2014-15 to 2019-20)
with
New Courses (2019-20on wards) as CBCS pattern in Statistics

Annual pattern	CBCS semester pattern
Papers in Old Courses (2014-15to2019-20)	Equivalent papers in New Course (2019-20onwards)
Statistics (General) Title: Design of Experiments & Operations Research	No equivalent paper*
Statistics (Special-III) Title:DistributionTheory	No equivalent paper*
Statistics(Special- IV) Title:Practical	No equivalent paper*
Mathematical Statistics (General) Title: Statistical Inference	No equivalent paper*
Applied Statistics (General)	No equivalent paper*
Statistical Prerequisites (General)	No equivalent paper*

*The Examination of the papers having 'No equivalent papers' will be conducted as per University provisions prescribed in such cases.

SYLLABUS FOR SEMESTER-V

Subject: Statistics (General-III)
ST-33875: Design and Analysis of Experiments (CC 1E)

Objectives: The main objective of this course is to introduce the basic concepts of design of experiments. After successfully completion of this course students are expected,

- To plan and conduct experiments.
- To identify significant factors in the experiment.
- To check interaction effect of factors.
- To analyze the data using statistical software
- To interpret results obtained from statistical software.

Outcomes: At the end of course,

- Students will familiar with basic concepts of design of experiments, ANOVA, factorial design, etc.
- Students will get idea regarding a use of design of experiments tools in real life situations.
- Students will be able to plan and conduct smaller experiments within given time frames and also present the planning, implementation and analysis of a conducted experiment, in oral and written form.
- Students will be able to describe the purpose of factorial experiments and how it is applied in experimental design
- Students will be able to analyze experimental data with suitable software

Contents:

1. Introduction of Design of Experiments (3 L)

1.1 Basic terms of design of experiments: Experiment with types, experimental material experimental unit (plot), treatment, block, layout of an experiment, experimental error and precision. Concept and definition of efficiency of a design.

1.2 Basic principles of design of experiments: Replication, Randomization and Local Control. Choice of size and shape of plots for uniformity trials, the empirical formula for the variance per unit area of plots.

1.3 Classification of design: One-way classification, two-way classification, three-way classification,etc.

1.4 Analysis of variance (ANOVA): Only concept, assumptions and techniques.

2. Completely randomized design (CRD) (7L)

2.1 Application of the principles of design of experiment in CRD, Layout, Model, Assumptions and Interpretations. Breakup of total sum of squares into components. Breakup of total degrees of freedom into components.

2.2 Hypothesis to be tested $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$. preparation of (ANOVA) table Comparison of treatment means using box plot techniques. Statement of Cochran's theorem. F test for testing H_0 with justification (independence of chi- square is to be assumed).Testing for equality of two specified treatment effects, comparison of treatment effects using critical difference.

2.3 Estimation of parameters, expected values for various mean sums of squares.

2.4 Identification of real life situations where the CRD is used

3. Randomized block design (RBD) **(10L)**

3.1 Application of the principles of design of experiments in RBD, layout, model, Assumptions and interpretations. Breakup of total sum of squares into components. Breakup of total degrees of freedom of components.

3.2 Hypotheses to be tested $H_{01}: \alpha_1 = \alpha_2 = \alpha_3 = \cdots = \alpha_t = 0$ and $H_{02}: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_b = 0$.Preparation of analysis of variance table, F test for testing H_{01} and H_{02} with justification (independence of chi- squares is to be assumed), testing for equality of two specified treatment effects, comparison of treatment effects using critical difference.

3.3 Estimation of parameters, expected values of various mean sums of squares.

3.4 Identification of real life situations where the RBD is used.

4. Latin square design (LSD) **(12L)**

4.1 Application of the principles of design of experiments in LSD, layout, Model, Assumptions and interpretations. Breakup of total sum of squares into components. Breakup of total degrees of freedom of components.

4.2 Hypotheses to be tested: $H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0 H_{02}: \beta_1 = \beta_2 = \cdots = \beta_m = 0$ $H_{03}: \gamma_1 = \gamma_2 = \cdots = \gamma_m = 0$. F test for H_{01}, H_{02} and H_{03} with Justification (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, comparison of treatment effects using critical difference.

4.3 Estimation of parameters, expected values of various mean sums of squares.

4.4 Identification of real life situations where the LSD is used.

5. Efficiency of Design **(6L)**

5.1 Concept and definition of efficiency of a design

5.2 Efficiency of RBD over CRD

5.3 Efficiency of LSD over CRD

5.4 Efficiency of LSD over RBD when

i) Rows are taken as blocks ii) Columns are taken as blocks

6. Factorial Experiments **(10L)**

6.1 General description of m^n factorial experiment,2^2 and 2^3 factorial experiments arranged in RBD.

6.2 Definitions of main effects and interaction effects in 2^2 and 2^3 factorial experiments.

6.3 Yate's procedure, preparation of ANOVA table, test for main effects and interaction effects.

6.4 General idea of confounding in factorial experiments.

6.5 Construction of layouts in total confounding and partial confounding in 2^2 and 2^3 factorial experiments.

6.6 Total confounding (confounding only one interaction) ANOVA table, testing main effects and interaction effects.

6.7 Partial confounding (confounding only one interaction per replicate); ANOVA table, testing main effects and interaction effects.

References

1. Cochran W.G. and Cox, C.M. (1968) Experimental Design, John Wiley and Sons, Inc., New York.
2. Das, M.N. and Giri, N.C. (1986) Design and Analysis of Experiments, II Edition Wiley Eastern Ltd., New Delhi
3. Federer, W.T. (1967) Experimental Design : Oxford and IBH Publishing Co., New Delhi
4. Goon, A.M., Gupta,M.K. and Dasgupta, B. (1998). Fundamentals of Statistics, Vol.II, The world Press Pvt. Ltd. Kolkatta

5. Gupta, S.C. and Kapoor, V.K. (2006). Fundamentals of Applied Statistics, S.Chand Sons, New Delhi
6. Johnson, R.A., Miller, I. and Freund, J.(2010). Probability and Statistics for engineers, Prentice Hall, India.
7. Kempthorne, O. (1952). Design of Experiments, Wiley Eastern Ltd., New Delhi.
8. Montgomery, D.C. (2001). Design and Analysis of Experiments, John Wiley and sons Inc., New Delhi.
9. Snedecor, G.W. and Cochran, W.G. (1994). Statistical Methods, 8th edition, Affiliated East – West Press, New Delhi
10. Wu, C.F.J. and Hamda, M. (2009). Experiments, Planning, Analysis and Parameter Design Optimization, John Wiley & Sons, Inc., Hoboken, New Jersey.

Subject: STATISTICS (SPECIAL- III)
ST-33885: Distribution Theory–I (DSE 1C)

Objectives: The main objective of this course is to acquaint students with concepts in Statistics. They will be introduced to some univariate continuous distributions. Students should get acquainted with central limit theorem and different type of convergence.

Outcomes: At the end of this course students are able:

- To evaluate various univariate continuous distributions.
- To obtain the distributions of order statistics.
- To apply Chebychev's theorem to evaluate upper bound for different discrete and continuous distributions.
- To understand the Pareto distribution with its scope in Economics.
- To apply Central Limit Theorem and Weak Law of Large Numbers in real life situations.

Contents:

1. Beta distribution **(14L)**

1.1 **Beta distribution of first kind**

p.d.f. $f(x) = \frac{1}{B(m,n)}$ $x^{m-1}(1-x)^{n-1}$, $0 \le x \le 1$, m, n >0

= 0, elsewhere

Notation: X~ β_1 (m, n)

Nature of probability curve, Derivation of mean, variance, r^{th} raw moment, harmonic mean, mode. Symmetry of the distribution.

1.2 Relation with U (0, 1). Probability distributions of $\frac{1}{X}$, X+Y, X-Y, XY, $\frac{X}{Y}$, where X and Y are iid β_1 (1,1),

1.3 **Beta distribution of second kind**

p.d.f $f(x) = \frac{1}{B(m,n)} \cdot \frac{x^{m-1}}{(1+x)^{m+n}}$, $x \ge 0$, m ,n >0

= 0, elsewhere

Notation: X~ β_2 (m,n)

Nature of probability curve, Derivation of mean, variance, rth raw moment, harmonic mean, mode.

1.4 Derivation of interrelation between β_1 (m, n) and β_2 (m, n).

1.5 Derivation of distribution of $\frac{X}{Y}, \frac{X}{X+Y}$, when X and Y are independent gamma variates.

1.6 Statement of relation between distribution function of β_1 (m, n) and binomial distribution.

2. Parato distribution **(6L)**

2.1 p.d.f.

$$f(x) = \frac{\lambda}{X^{\lambda+1}}, \; x \geq 1, \; \boldsymbol{\lambda} > 0$$
$$= \quad 0 \quad , \text{otherwise}$$

2.2 Nature of p.d.f. curve

2.3 Mean, variance and moments. Existence of moments for different values of **λ**. Symmetry, CDF

2.4 Applications in the field of Economics.

3. Order statistics **(10L)**

3.1 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 statistics $X_{(i)}$, particular cases for i = l and i = n.

3.2 Distribution of $X_{(i)}$ for a random sample from uniform and exponential distribution.

3.3 Definition of p-th sample quantile $X_{([np]+1)}$. Distribution of sample median for a random sample from uniform distribution.

4. Chebychey's Inequality **(8L)**

4.1 **Chebychev'stheorem:** If g (x) is a non–negative function of r.v.X such that $[g(X)] < \infty$ then,

$$P[g(x) \geq k] \leq \frac{E[g(X)]}{K}$$, where k is positive real number.

4.2 Chebychev's inequality for discrete and continuous distributions (with proof) in the forms

$$P[|X-\mu| \geq k] \leq \frac{\sigma^2}{K^2}$$, where $k>1$,

$$P[|x-\mu| \geq k\sigma] \leq \frac{1}{K^2}$$, $k > 1$ where $\mu = E(x)$ and $\sigma^2 = Var(x) < \infty$

4.3 Applications of Chebychev's inequality in control charts, statistical inference.

5. Central Limit Theorem and Weak Law of Large Numbers **(10 L)**

5.1 Sequence of r.v.s., convergence of sequence of r.v. in a) probability b) distribution with simple illustrations.

5.2 Statement and proof of the central limit theorem for i.i.d. r.v.s. (proof based on MGF).

5.3 Weak law of large numbers (WLLN).

5.4 Applications of CLT and WLLN.

6. Research project/ Field work/etc. **(1 credit)**

References:

1. H. Cramer: Mathematical Methods of Statistics, Asia Publishing House, Mumbai.
2. Mood, A.M. Graybill, F.Bose,D.C: Introduction to Theory (3rd Edition) Mc-Graw Hill Series.
3. B.W.Lindgren: Statistical Theory(3rd Edition) Collier Macmillan International Edition, Macmillan Publishing Co .Inc. New York.
4. Hogg,R.V.and Craig A.T.: Introduction to Mathematical Statistics (3rd Edition), Macmillan Publishing Company, Inc.86634dAvenue,NewYork, 10022.
5. Sanjay Arora and Bansi Lal: New Mathematical Statistics (1st Edition), Satya Prakashan16/17698, New Market, New Delhi,5.
6. S.C.Gupta and V.K.Kapoor: Fundamentals of Mathematical Statistics, Sultan Chand and Sons,88,Daryaganj,New Delhi,2.
7. V.K.Rohatgi: An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd. New Delhi.

8. J. Medhi: Stochastic Processes, Wiley Eastern Ltd. New Delhi.
9. Hoel, Port and Stone: Introduction to Stochastic Processes, Houghton Miffin.
10. FellerW.: An Introduction of Probability Theory and Its Applications, Vol-I, Wiley Eastern Ltd. Mumbai.
11. Sheldon Ross: A first course in probability (6^{th} edition): Pearson Education.

Subject: STATISTICS (SPECIAL-IV)
ST-33895: Statistics Practical (DSE 2C)

Objective: To able the student to apply various statistical tools in real life problems.
Outcomes: At the end of the course, the student able to,

- Conduct and analyze the data by using various designs of experiment.
- Construct various confidence interval for various population parameters.
- Obtain the estimations of various parameters.

Sr. No.	Title of the Experiment	No. of practical
1.	Analysis of CRD (equal and unequal replications) pair wise comparison of treatments, using critical difference (C.D) Check normality using normal probability plot.	2
2.	Analysis of R.B.D. pair wise comparison of treatments using C.D. Efficiency of RBD w.r.t. CRD	2
3.	Analysis of L.S.D., Pairwise comparison of treatments using C.D. Efficiency of LSD w.r.t i) CRD and ii)RBD	2
4.	Analysis of covariance in CRD, testing $\beta=0$.	1
5.	Analysis of covariance in RBD, testing $\beta=0$.	1
6.	Analysis of 2^2 and 2^3 factorial experiments in RBD.	2
7.	Analysis of 2^3 factorial experiments in RBD (partial confounding)	1
8.	Analysis of 2^3 factorial experiments in RBD (total confounding)	1
9.	Chebychev's inequality and Order Statistics.	1
10.	Central Limit Theorem and Weak law of large numbers.	1
11.	Construction of confidence interval for population mean(μ) for normal Distribution when σ^2 known and σ^2 unknown.	1
12.	Construction of confidence interval for Population variance(σ^2) for Normal distribution when μ known and μ unknown.	1
13.	Construction of confidence interval for population median and quartiles based on order statistics.	1
14.	M.L.E and moment estimator of truncated Binomial and truncated Poisson distributions (truncated at zero).	1
	Total Practicals	18
15.	**Research project/Field work/etc. based on Semester V for individual student**	**1 credit**

NOTE: Mathematical Statistics can be offered only as a general level subject. A student of three Year B.A. Degree Course opting Mathematical Statistics will not be allowed to opt Applied Statistics in any of the three years of the course.

Subject: Mathematical Statistics (General-IV)
ST-33275: Theory of Estimation (CC 2E)

Objectives: The students shall get

- various methods of estimation of parameters.
- Comparison between the various types of estimators.
- applications of different methods on estimations on a different standard probability models.

Outcomes: At the end of the course the students are expected to:

- Apply different methods of estimations to real life data sets.

Contents:

1. Point Estimation (4L)

1.1 Notion of a parameter, parameter space, sample space as a set of all possible values of (X1, X2, . . , Xn),general problem of estimating an unknown parameter by point and interval estimation.

1.2 Point Estimation: Definition of an estimator, distinction between estimator and estimate, illustrative examples.

1.3 Mean Square Error (MSE) of an estimator.

2. Methods of Estimation (10L)

2.1 **Method of moments:** Derivation of moment estimators for standard distributions. Illustrations of situations where M.L.E. and moment estimators are distinct and their comparison using mean square error.

2.2 Definition of likelihood as a function of unknown parameter, for a random sample from i) discrete distribution ii) continuous distribution, distinction between likelihood function and p.d.f./ p.m.f.

2.3 **Method of maximum likelihood:** Derivation of maximum likelihood estimator (M.L.E.) for parameters of only standard distributions (case of two unknown parameters only for normal distribution). Use of iterative procedure to derive M.L.E. of location parameter μ of Cauchy distribution. Invariance property of M.L.E.

2.4 a) M.L.E. of θ in uniform distribution over i) $(0, \theta)$, ii) $(-\theta, \theta)$, iii) $(m\theta, n\theta)$ $(m<n)$

b) M.L.E. of θ in $f(x ; \theta)= Exp \{-(x- \theta)\}$, $x > \theta$.

c) M.L.E. of location parameter in Laplace distribution.

3. Criteria/Properties of Estimation (28L)

3.1 **Unbiasedness:** Definition of an unbiased estimator, biased estimator, positive and negative bias, illustrations and examples (these should include unbiased and biased estimators for the same parameters). Proofs of the following results regarding unbiased estimators:

a) Two distinct unbiased estimators of (θ) give rise to infinitely many estimators.

b) If T is an unbiased estimator of θ, then $ø(T)$ is unbiased estimator of $ø(\theta)$, provided $ø(.)$ is a linear function.

Notion of the Best Linear Unbiased Estimator and uniformly minimum variance unbiased estimator (UMVUE), uniqueness of UMVUE 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 $ø(T)$ is also sufficient for θ provided $ø$ is a one to one and onto function.

ii) If T is sufficient for θ then T is also sufficient for ø (θ).
iii) M.L.E. is a function of sufficient statistic.

3.3 **Efficiency:** Fisher information function. Amount of information contained in statistic T = T (X1, X2, …, Xn). Statement regarding information in sample and in a sufficient statistic T.

3.4 **Cramer- Rao Inequality:** Statement and proof of Cramer –Rao inequality, Cramer – Rao Lower Bound(CRLB), definition of minimum variance bound unbiased estimator (MVBUE) of ø (θ). Proofs of following results:

a) If MVBUE exists for θ then MVBUE exists for ø (θ) where ø (.) is a linear function.
b) If T is MVBUE for θ then T is sufficient for θ.

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.

3.5 **Consistency:** Definition, proof of the following theorems:

a) An estimator is consistent if its bias and variance both tend to zero as the sample size tends to infinity.
b) If T is consistent estimator of θ and φ (.) is a continuous function, then Φ(T) is a consistent estimator of φ (θ).

4. Interval Estimation (6L)

Notion of interval estimation, definition of confidence interval (C.I), length of C.I., Confidence bounds, confidence coefficient. Definition of pivotal quantity and its use in obtaining confidence intervals. Interval estimation for the following cases:

i) Mean (μ) of normal distribution (σ2 known and σ2 unknown).
ii) Variance (σ2) of normal distribution (μ known and μ unknown).
iii) Median, quartiles using order statistics.

References:

1. Dudeweitz, E.J. and Mishra, S.N. (1988). Modern Mathematical Statistics, John Wiley and Sons, Inc.
2. Hoel, P.G. Port, S. and Stone, C.(1972). Introduction to Statistical Theory, Houghton Mifflin Company (International) Dolphin Edition.
3. Hogg, R.V. and Craig, A.T. (1978). Introduction to Mathematical Statistics (fourthedition), Collier Macmillan International Edition, Macmillan Publishing Co. Inc.New York
4. Kendall, M. and Stuart, A. (1943). The advanced Theory of Statistics, Vol 1, Charles and Company Ltd., London
5. Lindgren, B.W.(1976) Statistical Theory (third edition) Collier Macmillan International Edition, Macmillan Publishing Co., Inc. New York
6. Mood, A.M., Graybill, F. and Bose, D.C. (1974). Introduction to the theory of Statistics (third edition) International Student Edition, McGraw Hill.
7. Rohatagi , V.K. (1975). An introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd., New Delhi.
8. Ramchandran, K.M. and Tsokos C. P. (2009). Mathematical Statistics with Applications, Academic Press.

Subject: Skill Enhancement Course (SEC)
SEC 2C: MEDICAL STATISTICS AND CLINICAL TRIALS

Objectives: Students shall get aquatinted with application of Statistics in clinical trials.
Outcomes: At the end of course, students are able

- To set the model for population growth.
- To decide various factors related to epidemiology.
- To design and analyze clinical trial data.

1. Population study (6L)

1.1 India's population and census.
1.2 Population growth and models for population growth.
1.3 Birth and death rates.

2. Epidemiology (10 L)

2.1 Introduction to Epidemiology
2.2 Odds, odds ratio, relative risk.
2.3 Estimation of odds ratio (OR), Confidence interval for OR.
2.4 Symmetry in square contingency tables, collapsing tables and Simpson's paradox.

3. Clinical trials (10 L)

3.1 General information on history of drug discovery including Louis Pasteur (rabies and smallpox), Ronald Ross and malaria, Alexander Fleming and penicillin, Jonas Salk and polio, cholera, asthma, diabetes, blood pressure, heart attack, arthritis.
3.2 Phases of clinical trial, purpose, duration, cost, drug regulatory bodies, ICH, statistical analysis plan, clinical study report.

4. Design of clinical trials (10L)

4.1 Phases of clinical trial, purpose, duration, cost, drug regulatory bodies, ICH, statistical analysis plan, clinical study report.
4.2 Parallel designs, case control studies, longitudinal studies, safety studies
4.3 Treatments, 2-periods cross-over design.
4.4 Bioequivalence and bio-availability, non-inferiority trial
4.5 Practice based medical research, evidence-based medicine

References:

1. Course on mathematical and statistical Ecology: Kluwer publishing Holland, A.P. Gore and S.A, Paranjape (2000)
2. "Introduction to Statistical Ecology: M.B. Kulkarni, V.R. Prayag, SIPF Academy, Nasik (2004)
3. Introduction to Categorical Data Analysis: Alan Agrasti, John Wiley (1996) for part–II epidemiology (mainly odds, odds ratios and inference)
4. Introduction to Randomized Controlled clinical Trials: J.N.S.Matthews: Chapman and Hall(2006)
5. Statistical Issues in drug Development: Stephenson (John Wiley)2000
6. Clinical Trials–A methodological perspective: Steven Diantadosi (JohnWiley2000)
7. Fundamentals of Clinical Trials: L.M.Friedmon, C.D.Forbes, D.L.Demats (TT) Spinner
8. Epidemiologic Analysis: Steveselvin: (Oxford 2000)
9. Statistical Methods for Health Sciences: M.M. Shoukni, C.A. Pavse (1999) CRC Press.
10. Statistical Analysis of Epidemiologic Data: Steve Salvin, Ph.D: Oxford 1999)
11. Lecture Notes on Medical Statistics: A.P.Gore, S.A.Paranjpe and M.B. Kulkarni

Subject: Generic Elective Course (GE)

GE 1: TIME SERIES ANALYSIS

Objective: Students shall get the knowledge about time series and various methods to analyze time series data.

Outcomes: At the end of course, students are able

- To estimate various effects occurs in time series data.
- To analyze the time series data by using regression analysis.
- To apply Box Jenkins Techniques.
- To analyze real life time series data.

Contents:

1. Features of Time Series Data (6L)

1.1 Trend seasonality, correlation and changing variability

1.2 Test for randomness of a series against trend and seasonality.

1.3 Study given series by plots and histograms**.**

2. Population Estimation (12L)

2.1 Moving average and exponential smoothing.

2.2 Forecasting based on smoothing.

2.3 Double exponential smoothing.

2.4 Choosing parameters for smoothing and forecasting.

2.5 Estimating mean square error of forecasting. Prediction Intervals based on normality assumption.

3. Time Series Analysis Through Regression Analysis (6L)

3.1 Regression models for trend and seasonality.

3.2 De-trending and de-seasonalizing of a time series.

3.3 Analysis of residuals

4. Introduction to Box Jenkins Techniques (6L)

4.1 Stationary time series: Differing and seasonal differencing.

4.2 Transformation of data: Transformation and differencing.

4.3 Simple AR models and forecasts based on AR models, MSE of forecasts.

5. Data Analysis of Real Life Time Series. (6L)

5.1 Price index series, share price series, economic time series, Sales tax series, market price of daily consumables.

5.2 Weather related time series: Temperature and rainfall time series, wind speed time series, pollution levels.

Note: Most of the calculations needed are available in MS-EXCEL and through regression analysis.

References

1. Montgomery, D.C. and Johnson L.A.(1976): Forecasting and Time Series Analysis, McGraw Hill.
2. Farmum, N. R. and Stantorr, L.W.(1989):Quantitative Forecasting Methods, PWS-Kent Publishing Company, Boston.
3. Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci (2015): Introduction to Time Series Analysis and Forecasting, Wiley.
4. Galit Shmueli and Kenneth C. Lichtendahl Jr (2016): Practical Time Series Forecasting with R: A Hands-On Guide, Axelrod Schnall Publishers; 2nd edition

SYLLABUS FOR SEMESTER-VI

Subject: Statistics (General-III)
ST-33876: Operations Research (CC 1F)

Objectives:

The main objective of this course is to impart the knowledge in concepts and tools of Operations Research. After successfully completion of this course students are expected,

- To find optimal solution for given linear function or real life problems.
- To solve transportation and assignment problems using operation research techniques.
- To use CPM and PERT.
- To apply optimization techniques constructively to make effective decisions.
- To analyze the data using appropriate software and interpret results.

Outcomes: At the end of course,

- Students will familiar with basic concepts of operation research, Linear Programming, CPM, PERT, etc.
- Students will familiar with mathematical models used in operation research.
- Students will get idea regarding a use of optimization techniques in real life situations.
- Students will be able to analyze data with suitable software and interpret results.

Contents:

1. Operation Research (6L)

1.1 History (Sec. 1.2)

1.2 Definition – (Sec. 1.3 – i. Daellenbach & George, ii. Theierauf & Klekamp, H. A. Taha)

1.3 Features of Operations Research Approach – Interdisciplinary Approach, Scientific Approach, Holistic Approach, Objective-Oriented Approach (Sec. 1.4).

1.4 Operations Research Approach to problem solving – Judgment, Research, Action (Sec. 1.5).

1.5 Classification of models based on Static models, Dynamic models, Deterministic models, Analytical model, Simulation model (Sec. 1.6.3 & 1.6.4,1.6.5).

1.6 Methods for solving models – Analytical method, Numerical method, Monte Carlo Method (Sec. 1.8).

1.7 Advantages, Opportunities and disadvantages of Operations research (Sec. 1.10, Sec. 1.11).

1.8 Applications of Operation Research – Finance and Accounting, Marketing, Purchase, Procurement and Exploration, Production Management (Sec. 1.13).

2. Linear Programming (20L)

2.1 Statement of the linear Programming Problem (LPP), Formulation of problem as L.P. problem. L.P. Problem in (i) Canonical form (ii) standard form. Definition 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.

2.2 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, criteria for deciding whether obtained solution is optimal criteria for unbounded solution, no solution,and more than one solution. 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.

2.3 Duality Theory: i) Writing dual of a primal problem, ii) Solution of a L.P.P. by using its dual problem.

3. Transportation and Assignment Problem (12L)

3.1 Transportation problem (T.P.), statement of T.P., balanced and unbalanced T.P.

3.2 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).

3.3 U-V method of obtaining Optimal solution of T.P., degenerate solution, uniqueness and non-uniqueness of optimal solutions.

3.4 Assignment problems: i) Statement of an assignment, balanced and unbalanced problem, ii) relation with T.P. iii) optimal solution of an assignment problem using Hungarian method.

3.5 Examples and numerical problems.

4. CPM (6L)

4.1 Introduction (Sec. 13.1)

4.2 Basic differences between PERT and CPM. (Sec. 13.2).

4.3 Definition of (i) Events with Merge and Burst Events, (ii) Node, (iii) Activities – Predecessor, Successor, Dummy, (iv) Critical Activity, (v) Project Duration. (Sec. 13.4).

4.4 CPM: 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 (Sec. 13.5).

5. PERT (4L)

5.1 Construction of network; (i) pessimistic time estimate, (ii) optimistic time estimate (iii) most likely time estimates, Determination of critical path, determination of mean and standard deviation of project duration, computations of probability of completing the project in a specified duration. (Sec. 13.6).

Note: ***All sections are from the book "Operations Research-Theory and Application" by J. K. Sharma.***

References

1. Gass, S.L. (1997). Linear programming methods and applications, Narosa Publishing House, New Delhi.
2. Gupta, P.K. and Hira, D.S.(2008). Operation Research, 3rd edition S. Chand and company Ltd., New Delhi.
3. Kapoor, V. K.(2006). Operations Research, S. Chand and Sons. New Delhi.
4. Phillips, D.T and Solberg, R.A.(1976). Operation Research principles and practice, John Willey and sons Inc.
5. Saceini, M., Yaspan,A.. and Friedman, L.(2013). Operation Research methods and problems, Willey International Edition.
6. Sharma, J.K. (1989). Mathematical Models in Operation Research, Tata McGraw Hill Publishing Company Ltd., New Delhi.
7. Shrinath.L.S (1975). Linear Programming, Affiliated East-West Pvt. Ltd, New Delhi.
8. Taha, H.A. (2007). Operation research: An Introduction, eighth edition, Prentice Hal of India, New Delhi.

Subject: STATISTICS (SPECIAL- III)
ST-33886: Distribution Theory–II (DSE 1D)

Objectives: The main objective of this course is to get acquainted with different univariate and bivariate distributions. Differentiate the various distributions as per their properties such as mean,variance, symmetry and study their relations under certain assumptions. Student should be able to study these distributions and apply in their real life situations.

Outcomes: At the end of the course, the students are able to

- correlate Weibull distribution with other distributions such as Exponential and Gamma distribution.
- develop the relation of Laplace's distributionwith Exponential distribution under certain conditions.
- prove non-existence of moments of Cauchy's distribution.
- develop the relation between Normal and Lognormal Distribution.
- use of bivariate Normal distribution and its applications and relation with Cauchy's distribution under assumptions.

Contents:

1. Weibull Distribution **(6L)**

1.1 p.d.f .

$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\alpha}\right)^{\beta}\right\}$, $x \geq 0$, $\alpha, \beta > 0$

$= 0$, elsewhere

Notation: $X \sim W(\alpha, \beta)$.

1.2 Probability curve, location parameter, shape parameter, scale parameter. Derivation of distribution function, quartiles, mean and variance, coefficient of variation relationship with gamma and exponential distribution.

2. Laplace (Double Exponential) Distribution **(8L)**

2.1 p.d.f. $f(x) = \frac{\lambda}{2}\exp(-\lambda|x-\mu|)$; $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\lambda > 0$,

$= 0$; elsewhere

Notation: $X \sim L(\mu, \lambda)$

2.2 Nature of the probability curve.

2.3 Derivation of distribution function, quartiles.

2.4 MGF, CGF, Moments and cumulants, skewness and kurtosis

2.5 Derivation of Laplace distribution as the distribution of the difference of two i.i.d. exponential random variables with mean $\frac{1}{\lambda}$.

3. Cauchy Distribution **(10L)**

3.1 p.d.f. $f(x) = \frac{\lambda}{\pi}\frac{1}{\lambda^2+(x-\mu)^2}$; $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\lambda > 0$.

Notation: $X \sim C(\mu, \lambda)$

3.2 Nature of the probability curve for various choice of parameters, comparison with tails of normal distribution.

3.3 Derivation of distribution function, quartiles .Non – existence of moments, Statement of distribution of $aX + b$. Derivation of distribution of i) $\frac{1}{X}$ ii) X^2 where $X \sim C(0,1)$, Problems based on these results.

3.4 Statement of additive property for two Independent Cauchy variates, statement of distribution of the sample mean, comment on limiting distribution of X.

3.5 Statement of relationship with uniform, student's t and normal distributions.

4. Lognormal Distribution **(10L)**

4.1 p.d.f $f(x) = \frac{1}{(x-a)\sigma\sqrt{2\pi}} \exp\left\{\frac{-1}{2\sigma}[\log_e (x-a)-\mu]^2\right\}$, a<x, -∞< μ <∞, σ >0,

= 0 ; elsewhere

Notation: X~ LN (a ,μ , σ^2)

4.2 Derivation of relation with N (μ, σ^2) distribution

4.3 Nature of the probability curve.

4.4 Derivation of moments (r-th moment of X-a), mean, variance, quartile, mode, Karl Pearson's and Bowley's coefficient of skewness and kurtosis, derivation of quartiles and mode.

4.5 Distribution of (Π Xi), when Xi's independent lognormal random variables

5. Bivariate Normal Distribution **(14L)**

5.1 Bivariate normal distribution p.d.f.

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho 2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]\right\}$$

-∞< x, y <∞,

-∞<μ_1 , μ_2<∞ ,

σ_1, σ_2 >0,-1<ρ<+1

Notation: (X, Y) ~ BN($\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$).

5.2 Nature of surface of p. d. f., marginal and conditional distributions, identification of parameters, regression of Y on X, independence and uncorrelatedness, derivation of MGF and moments. Statement of distribution of aX+ bY+ c , X/Y.

5.3 Applications and real life situations.

6. Research project/Field work/etc. **(1 credit)**

References:

1. H. Cramer: (1962) Mathematical Method of Statistics, Asia Publishing House , Mumbai
2. Mood. A.M. ,Graybill , F. Bose ,D.C.: (1974) Introduction to theory of Statistics. (IIIrd Edition) Mc- Graw Hill Series.
3. B.W. Lindgren: (1976) Statistical Theory (III rd Edition) Collier Macmillan international Edition, Macmillan Publishing Co. Inc. New York.
4. Hogg, R.V.and Craig A.T.(1970). Introduction Mathematical Statistics (IIIrd Edition) Macmillan Publishing Company Inc New York
5. Sanjay Arora and Bansi Lal (1989). Mathematical Statistics (Ist Edition), Satya Prakashan 16/17698, New Delhi.
6. S.C Gupta and V.K.Kapoor: (2006). Fundamental Mathematical Statistics, Sultan Chand and Sons, 88 Daryaganj, New Delhi.
7. Rohatgi , V.K. (1975) An Introduction to probability Theory and Mathematical Statistic ,Wiley Eastern Ltd .New Delhi.
8. Mukhopdhyay, P (1996). Mathematical Statistics, New Central Book Agency.
9. Adke, S.R., Manjunath, S.M. (1984). An introduction to finite Markov processes, Wiley Eastern.
10. Bhat, B.R. (2000). stochastic models: Analysis and applications, New Age International.
11. Ross, S. (2000). Introduction to probability models, 7thedn, Academic Press

Subject: STATISTICS (SPECIAL-IV)
ST-33896: Statistics Practical (DSE 2D)

Objective: Prepare the student to apply various statistical tools in real life problems.

Outcomes: At the end of the course, the student able to,

- Draw/Simulate the sample observations from Cauchy and Laplace distribution.
- Construct various parametric and non-parametric test various population parameters.
- Formulate and obtain the optimal solution to Linear Programming Problem.
- To solve problems related to Transportation, Assignment, CPM and PERT.

Sr.No	Title of the Experiment	No.of practicals
1.	Model sampling from Cauchy and Laplace distributions	1
2.	Fitting of lognormal distribution	1
3.	Applications of Weibull, Cauchy, Laplace, Lognormal and Bivariate normal distribution	2
4.	Constriction of MP and UMP test. Plotting of power function of a test.	2
5.	Non-parametric Tests-I: Run test, sign test, Wilcoxon's signed rank test.	1
6.	Non-parametric tests-II: Mann-Whitney U test, median test and Kolmogorov–Smirnov test.	2
7.	SPRT for Bernoulli, Poisson distributions. (Graphical representation also)	2
8.	SPRT for normal, exponential distribution (Graphical representation also)	1
9.	Linear programming problem I (Simplex Method)	1
10.	Linear Programming Problem II (simplex Method)	1
11.	Transportation Problem	1
12.	Assignment Problem	1
13.	CPM	1
14.	PERT	1
	Total Practicals	18
15.	**Research project/Field work/etc. based on Semester VI for individual student**	**1 credit**

NOTE: Mathematical Statistics can be offered only as a general level subject. A student of three year B.A. Degree Course opting Mathematical Statistics will not be allowed to opt Applied Statistics in any of the three years of the course.

Subject: Mathematical Statistics (General-IV)
ST-33276: Testing of Hypothesis (CC 2F)

Objectives: The students shall get

- various parametric and non-parametric tests
- Applications of parametric tests on various standard probability distributions
- Comparison between parametric and non-parametric tests

Outcomes: At the end of the course the students are expected to apply the testing of hypothesis on real life data set.

Contents:

1. Parametric Tests (15L)

1.1Statistical hypothesis, problem of testing of hypotheses. Definition and illustrations of (1) simple hypothesis, (2) composite hypothesis, (3) test of hypothesis, (4) critical region, (5) type I and type II errors. probabilities of type I error and type II error. Problem of calculating the probabilities of errors of two kinds.

1.2 Definition and illustrations of (i) level of significance, (ii) observed level of significance (p-value), (iii) size of a test, (iv) power of a test.

1.3 Definition of most powerful (M.P.) level α test of simple null hypothesis against 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.

1.4 Power function of a test, power curve, definition of uniformly most powerful (UMP) level α test for one sided alternative. Illustrations

2. Likelihood ratio tests (9L)

2.1 Notion of likelihood ratio test (LRT), $\Lambda(x) = \text{Sup } L(\theta_0|x) / \text{Sup } L(\theta_1|x)$

2.2 Construction of LRT for $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$ for the mean of normal distribution for i) known σ^2 ii) unknown σ^2 (one sided and two sided alternatives).

2.3 LRT for variance of normal distribution for i) known μ ii) unknown μ (one sided and two sided alternative hypotheses).

2.4 LRT for parameters of binomial and exponential distribution for two sided alternatives only.

2.5 LRT as a function of sufficient statistics, statement of asymptotic distribution of $-2\log_e\lambda(x)$.

3. Sequential Tests (9L)

3.1 Sequential test procedure for simple null hypothesis against simple alternative hypothesis and its comparison with fixed sample size N-P test procedure.

3.2 Definition of Wald's SPRT of strength (α, β). Illustration for standard distributions like Bernoulli, Poisson, Normal and Exponential. SPRT as a function of sufficient statistics.

3.3 Graphical representation of SPRT.

4. Non-parametric Tests **(15L)**

4.1 Concept of non- parametric tests. Distinction between a parametric and a non-parametric tests. Concept of distribution free statistic. One tailed and two tailed test procedure of i) Sign test, ii) Wilcoxon signed rank test, iii) Mann Whitney U test, iv) Run test, one sample and two samples problems.

4.2 Empirical distribution function Sn(x). Properties of Sn(x) as estimator of F(.). Kolmogorov – Smirnov test for completely specified univariate distribution (one Sample problem only) for two sided alternative hypotheses. Comparison with chi-square test.

References:

1. Agarwal, B.L. (2003). Programmed Statistics, second edition, New Age International Publications, Delhi
2. Arora, S. and Bansi Lal. (1989): New Mathematical Statistics, first edition, Satya Prakashan, New Delhi.
3. Daniel, W.W. (2000) Applied Nonparametric Statistics, Duxbury Press Boston.
4. Dudeweitz, E.J. and Mishra, S.N. (1988). Modern Mathematical Statistics, John Wiley and Sons, Inc.
5. Gibbons J.D. (1971). Non parametric Statistical Inference, McGraw Hill Book Company, New York.
6. Hoel, P.G.,Port, S. and Stone, C. (1971). Introduction to Statistical Theory, Houghton Mifflin Company (International)
7. Hogg, R.V. and Craig, R.G. (1989). Introduction to Mathematical Statistics (fourth edition,. Collier Macmillan International Edition, Macmillan Publishing Co. Inc., New York.
8. Kale, B.K. and Muralidharan, K. (2015). Parametric Inference: An Introduction. Narosa, New Delhi
9. Kendall, M. and Stuart, A. (1943) The advanced Theory of Statistics, Vol 1, Charles and Company Ltd., London.
10. Lindgren, B.W.(1976). Statistical Theory (third edition), Collier Macmillan International Edition, Macmillan publishing Co., Inc. New York.
11. Kunte, S ., Purohit, S.G. and Wanjale, S.K. : Lecture Notes On Nonparametric Tests
12. Mood, A.M., Graybill, F. and Bose, D. C .(1974). Introduction to the theory of Statistics (third edition) International Student Edition, McGraw Hill.
13. Rohatgi, V.K. (1976). An introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd., New Delhi.
14. Siegel, S. (1956). Nonparametric methods for the behavioral sciences, International Student Edition, McGraw Hill, New York.

Subject: Skill Enhancement Course (SEC)
SEC 2D:DATA ANALYTICS

Objectives: Students shall get aquatinted with field of data analytics

Outcomes: At the end of course, students are able

- To apply data cleaning tools and data mining process.
- To apply various types of classification techniques.
- To do market basket analysis.

Contents:

1. Introduction to Data Mining (8L)

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 (8L)

2.1 Problem of classification.

2.2 Classification techniques: k-nearest neighbor, decision tree, Naïve Bayesian, classification based on logistic regression

3. Model Evaluation, Selection and Classification Accuracy (10L)

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 Ada Boost, Improving Classification Accuracy of Class Imbalanced Data.

4. ANN, SVM and Clustering (10L)

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, A priori Algorithm, data attributes, applications to electronic commerce.

References

1. Berson andSmithS.J.(1997):Datawarehousing,DataMining,andOLAP,McGraw-Hill.
2. Breiman J.H Friedman, R.A. Olshen and Stone C.J.(1984):Classification and Regression Trees, Wadsworth and Brooks/Cole.
3. Han,J.and Kamber, M.and Pei,J.(2012):Data Mining: Concepts and Techniques. Morgan Gaufmann.3rd Edition.
4. Mitchell T.M.(1997):Machine Learning, McGraw-Hill.
5. RipleyB.D.(1996):PatternRecognitionandNeuralNetworks.CambridgeUniversityPress.
6. Mehrika, K.,Mohan,C.,and Ranka(1997) Elements of Artificial neural networks. Penram international.
7. HastieT,Tibshirani R, Friedmant J,(2009):The elements of statistical Learning, Springer.
8. Chattamvelli, R.(2015).Data mining methods. Alpha Science International.

Subject: GENERIC ELECTIVE COURSE(GE)
GE2: OPERATIONS MANAGEMENT

Objectives: To explain the various concept related to Operations Research.

Outcomes: At the end of course, the students are able

- To solve the problems of replacement theory.
- To solve various problems related to inventory, decision theory, game theory and sequencing.

Contents:

1. **Replacement Theory** **(5L)**
   - 1.1 Introduction(Sec.17.1)
   - 1.2 Types of failures–Gradual failure and sudden failure(Sec.17.2)
   - 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 (Sec. 17.3 – Model I, Model II).
2. **Inventory Theory** **(9L)**
   - 2.1 Introduction, meaning, functional role and reasons for carrying inventory (Sec.14.1,14.2,14.3,14.4).
   - 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(Sec.14.13).
   - 2.3 Description of generalized inventory model (Sec14.6).
   - 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),
     - iii) The economic lot size model with uniform demand, instantaneous replenishment with shortage (without derivation),
     - iv) The economic lot size model with uniform rate of demand, finite replenishment rate with shortage (without derivation) (Sec.14.7).
3. **Decision Theory** **(9L)**
   - 3.1 Introduction, Types of Decision making Environment (Sec.11.1,11.3)
   - 3.2 Decision making under uncertainty-(i) Laplace criterion, (ii) Minimax/Maximax criterion, (iii)Savage minimax regret criterion, (iv)Hurwitz criterion (Sec.11.4).
   - 3.3 Decision making under risk: (i) Expected Monetary Value, (ii) Expected Opportunity Loss, (iii) Expected Value of Perfect Information (Sec. 11.5).
   - 3.4 Decision Trees (Sec.11.7)
4. **Game Theory** **(8L)**
   - 4.1 Introduction, Definitions–(i)Number of players, (ii)sum of gains or losses, (iii) Strategy and its type (Sec.12.1).
   - 4.2 Two Person Zero-Sum Game(Sec.12.2)
   - 4.3 Pure & Mixed Strategies (Se.12.3,12.4)
   - 4.4 Principles of Dominance(Sec.12.5)
   - 4.5 Solution Methods for Games without Saddle Point – Algebraic, Arithmetic, Matrix and Graphical Method.

5. Sequencing **(5L)**

5.1 Introduction, notations, terminology, assumptions (Sec.20.1,20.2)

5.2 Processing n-jobs through two machines (Sec.20.3)

5.3 Processing n jobs through three machines, (reducible to two machines and n-jobs) .(Sec.20.4)

5.4 Examples and problems.

Note: All sections are from the book "Operations Research-Theory and Application" by J.K. Sharma

Books Recommended

1. Gass, E.: Linear programming method and applications, Narosa Publishing House, New Delhi.
2. Taha,R.A.: Operation research, An Introduction, fifth edition, Prentice Hall of India, New Delhi.
3. Saceini, Yaspan, Friedman: Operation Research methods and problems, Willey International Edition.
4. Shrinath.L. S: Linear Programming, Affiliated East-West Pvt. Ltd, New Delhi.
5. Phillips, D.T, Ravindra, A, Solberg,I.: Operation Research principles and practice, John Willey and sons Inc.
6. Sharma, J.K.: Mathematical Models in Operation Research, Tata McGraw Hill Publishing Company Ltd., New Delhi.
7. Kapoor,V.K.: Operations Research, Sultan Chand and Sons. New Delhi.
8. Gupta,P.K. and Hira,D.S.: Operation Research, S.Chand and company Ltd., New Delhi.

XXXXXXXXXXXXXXXXX