Progressive Education Society's

Modern College of Arts, Science andCommerce Pune 16

(An Autonomous College Affiliated to Savitribai Phule Pune University)

Three Year B.Sc. Degree Program in Statistics

(Faculty of Science & Technology)

S. Y. B. Sc. Statistics

Choice Based Credit System Syllabus

To be implemented from Academic Year 2023-2024

Notes:

1. A student of the three year B. Sc. Degree course will not be allowed to offer Statistics and Statistical Techniques simultaneously in any of the three years of the course.
2. Students offering Statistics at the first year of the three year B. Sc. course may be allowed to offer Statistical Techniques as one of their subjects in the second year of the three year $B.Sc.$ Course in the place of Statistics.
3. Students offering Statistical Techniques at the first of the three year B. Sc. course may be allowed to offer Statistics as one of their subjects in the second year of the three year B. Sc. Course in place of Statistical Techniques provided they satisfy other requirements regarding subject combinations, if any.
4. Students must complete all the practical to the satisfaction of the teacher concerned. At the time of practical examination, a student must produce the laboratory journal along with the completion certificate signed by the Head of the Department.
5. **Study Tour**: In order to acquaint the students with applications of Statistical methods in various fields such as industries, agricultural sectors , Government Institutes etc. study tour may be arranged and the report be attached in practical journal.

Structure of the S. Y. B. Sc. Statistics Course

Sem. No.	Paper No.	Paper code and paper title	Credits	Lectures per Week	Marks	
					Internal	External
III	1	23-ST-231 Discrete Probability Distributions and Time Series	02	03	15	35
	2	23-ST-232 Continuous Probability Distributions				
	3	23-ST-233 StatisticsPractical I				
IV	1	23-ST-241 Tests of Significance And Statistical Methods	02	03	15	35
	2	23-ST-242 Continuous Probability Distributions And Exact tests				
	3	23-ST-243 StatisticsPractical II				

*** (each lecture of 50 minutes)

Objectives:

1. **To fit various discrete and continuous probability distributions and to study various** real life situations.
2. To identify the appropriate probability model that can be used.
3. To use forecasting and data analysis techniques in case of uni-variate and multivariate datasets.
4. To use statistical software packages.
5. To test various hypotheses of significance like means, proportions, independence ofattributes, variance etc. included in theory (using calculators, software).
6. To compute probabilities of discrete and continuous probability distributions usingMS-Excel and/or R software (whichever is applicable).
7. To study applications of statistics in the field of demography etc.

Structure of evaluation of practical paper at *S. Y.* B. Sc.

A) Continuous Internal Assessment for practical CIA)

Section	Marks
1) Journal	05
2) Viva-voce	05
3) Project	05
Total of A	**15**

B) End of Semester Examination (ESE)

Section	Nature	Marks	Time
I	Note: Question No.1 is compulsory.	5	10 Min
II	Using R-software: Note : Any two questions out of 4 Questions)	26 (13 Marks per Question)	1 Hr. 30 Min
III	Viva -Voce	4	10 Min
Total of B		35	1 Hr. 50 Min
Total of A+ B			

Preparation by Internal Examiner:

1. Keep at least 15 **computers** with latest configuration ready with battery backup and necessary software, printers, at the examination laboratory.
2. Any other types of data required as per slip also be entered in computer spreadsheet.

Instructions to Examiners:

1. Students are not expected to fill data items at the time of examination. They are expected to use MS Excel and R −commands (whichever is applicable) to operate on the data set which are already fed.
2. The questions on section I (On line examination Using $Ms - EXCEL$ / R−commands (whichever applicable)) are compulsory and there is no internal option.
3. The slips made available for Section I shall be allotted to the candidate's at random.

Instruction for theory Examination:

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. Medium of Instruction: English.
4. Pattern of examination: Semester wise.
5. Standard of passing : As per norms of University

SEMESTER – III
PAPER - I
23-ST – 231: DISCRETE PROBABILITY DISTRIBUTIONS AND TIME SERIES

Course Outcomes:-

CO1) Students will learn new distributions like Negative binomial, multivariate, truncated distribution.
CO2) Students learn new concept of time series and their components.
CO3) Students learn exploratory data analysis in time series.
CO4) Students learn the application of the techniques like fitting of curve and their statistical analysis for time series.

1. Time Series: (08L)

1.1 Meaning and utility of time series, components of time series: trend, seasonal variations, cyclical variations, irregular (error) fluctuations or noise.

1.2 Exploratory data analysis: Time series plot to (i) check any trend and seasonality in the timeseries (ii) identify the nature of trend .

1.3 Methods of trend estimation and smoothing: (i) moving average, (ii) linear parabolic, exponential, curve fitting by least squares principle (iii) exponential smoothing, Choosing parameters for smoothing and forecasting. Forecasting based on exponential smoothing.

1.4 Measurement of seasonal variations: i) simple average method, ii) ratio to moving average method, iii) ratio to trend where linear trend is calculated by method of least squares. (Numericalexamples with heavy computations are to be asked preferably in practicals).

1.5 Fitting of autoregressive model $AR\ (p)$, where $p = 1,2$.

2. Negative Binomial Distribution: **(08L)**

2.1 Probability mass function (p. m. f.), Notation: X ~ N B (k, p).

2.2 Graphical nature of p. m. f., negative binomial distribution as a waiting time distribution,

2.3 Momentgenerating function (MGF), cumulant generating function (CGF), mean, variance, skew-ness, kurtosis (recurrence relation between moments is not expected),

2.4 Additive property of NB (k, p).

2.5 Relation between geometric distribution and negative binomial distribution.

2.6 Poisson approximation to negative binomial distribution. Real life situations.

3. Multinomial Distribution: (12L)

3.1 Probability mass function (p. m. f.))

Notation: $\underline{X} = (X_1, X_2 \ldots X_k), p = (p_1, p_2 \ldots p_k)$, $\underline{X} \sim (n,)$,

3.2Joint MGF of $(X_1, X_2, \ldots, X_k)$, use of MGF to obtain means, variances, covariance

3.3 Total correlation coefficients, variance – covariance matrix, rank of variance – covariance matrix and its interpretation,

3.4 Additive property of multinomial distribution, uni-variate marginal distribution, distribution of Xi+Xj, conditional distribution of $Xi \mid Xj = r$, conditional distributionof Xi given Xi+ Xj = r, real life situations and applications.

4. Truncated distributions (08L)

4.1 Concept of truncated distribution, truncation to the right, left and on both sides.

1. Binomial distribution left truncated at X = 0 (value zero is discarded), its p. m. f., mean and variance.
2. Poisson distribution left truncated at X = 0 (value zero is discarded), its p. m. f., mean and variance. Real life situations and applications.

SEMESTER – III

PAPER - II

23-ST 232: CONTINUOUS PROBABILITY DISTRIBUTIONS

Course Outcomes:

CO1) Students will understand the concept of continuous random variable and its probability distribution.

CO2) Students will able to describe and study the different kinds of continuous probability distributions such as Uniform distribution, Normal distribution and Exponential distribution.

CO3) They can find relations among aforesaid continuous random variables.

CO4) Students can implement these probability distributions in handling the real life Data.

1. Continuous Uni-variate Distributions: (08L)

1.1 Continuous sample space: Definition, illustrations.

Continuous random variable: Definition, probability density function (p. d. f.), cumulative distribution function (c. d. f.), properties of c. d. f. (without proof), and probabilities of events relatedto random variable.

1.2 Expectation of continuous r. v., expectation of function of r. v. $[g(X)]$, mean,

variance, geometric mean, harmonic mean, raw and central moments, skewness, kurtosis, mean deviation about mean.

1.3 Moment generating function (MGF): Definition, properties. Cumulant generating function (CGF): Definition.

1.4 Mode, partition values: quartiles, deciles, percentiles.

1.5 Probability distribution of function of r. v. : $Y = g(X)$ using i) Jacobean of transformationfor $g(.)$ monotonic function and one-to-one, on to functions, ii) Distribution function for $Y = X^2, Y = |X|$ etc., iii) M.G.F. of $g(X)$.

2. Continuous Bivariate Distributions: (07L)

2.1 Continuous bivariate random vector or variable (X, Y): Joint p. d. f., joint c. d. f, properties (without proof), probabilities of events related to random variables (events in terms ofregions bounded by regular curves, circles, straight lines). Marginal and conditional distributions.

2.2 Expectation of r. v.(X,Y), expectation of function of r. v. $E[g(X, Y)]$, joint moments, $Cov(X, Y)$, $Corr\ (X, Y)$, conditional mean, conditional variance, $E[(X|Y = y)] = E(X)$ & $E[(Y/X = x) = (Y)$, regression as a conditional expectation.Theorems on expectation: i) $E(X + Y) = E(X) + (Y)$, (ii) $(XY) = E(X)\ (Y)$, if X $and\ Y$ are independent, generalization to k variables. $(aX + bY + c)$, $Var\ (aX + bY + c)$ (statement only proofnot expected).

2.3 Independence of random variables X $and\ Y$ and also its extension to k random variables.

2.4 Moment generating function (MGF):$M_{X,Y}\ (t_1, t_2)$, properties, MGF of marginal distribution ofrandom variables, properties i) $M_{X,Y}\ (t_1, t_2) = M_X\ (t_1, 0)\ M_Y\ (0, t_2)$ if X and Y are independent random variables

2.5 Probability distribution of transformation of bivariate r. v. $U = \phi_1\ (X, Y)$, $V = \phi_2(X, Y)$.

3. Standard Uni-variate Continuous Distributions:

3.1 Uniform or Rectangular Distribution: (03L)

3.1.1Probability density function (p. d. f.)
Notation: X ~ U (a, b), p. d. f., sketch of p. d. f., c. d. f.,

3.1.2 mean, variance, symmetry, MGF.

3.1.3 Distributions of i)$\frac{X-a}{b-a}$, ii)$\frac{b-X}{b-a}$, iii) Y= F(x) where F(x) is the c. d. f. of continuous r. v.

3.1.4 Application of the result to model sampling. (Distributions of X+Y, X–Y, X/Y and XY are not expected)

3.2 Normal Distribution: (10L)

3.2.1 Probability density function (p. d. f.)
Notation: X~ N (μ, σ2)

3.2.2 P. d. f. curve, identification of scale and location parameters, nature of probability curve, Points of inflexion of probability curve, computations of normal probabilities using normal probability integral tables.

3.2.3 mean, variance, mode, quartiles (Q1, Q2, Q3), mean deviation, ,

3.2.4 MGF, CGF, additive property, central moments, cumulants, skewness, kurtosis,

3.2.4 Probability distribution of: i), standard normal variable (S.N.V.), ii) aX+b iii) aX+bY+ c, where X and Y are independent normal variates. Probability distribution of the mean of n. r. v s.,

3.2.5 Central limit theorem (CLT) for r. v. s with finite positive variance (statement only), its illustration for Poisson and Binomial distributions.

***(Box-Muller transformation and normal probability plot to be covered in practical)

3.3 **Gamma Distribution:** **(05L)**

3.3.1 Notation: X~ G (α, λ)

3.3.2 Nature of probability curve,

3.3.3 MGF, CGF, moments, cumulants, skewness, kurtosis, additive property.

3.3.4 Relation between distribution function of Poisson and Gamma variates.

3.4 Exponential Distribution **(03L)**

3.4.1 Exponential Distribution as a special case of Gamma distribution

3.4.2 Distribution function, quartiles (Q_1, Q_2, Q_3), Lack of memory property,

3.4.3, mean deviation about mean,

3.4.4 M.g.f., distribution of sum of two i.i.d exponential random variables.

3.4.5. Distribution of $min(X, Y)$ and $max(X, Y)$ with X, Y as i. i. d. standard exponential random variables.

****

SEMESTER – III
PAPER - III
23-ST-233: PRACTICAL

Sr. No.	Title of the Practical	No. of Practicals
1	Introduction to R: 1. c function, scan 2. data Frame, edit () 3. Matrix form 4.importing data file, accessing the data from R library 5.seq() , rep() functions 6.subset and transform 7.Basic commands (summary(), fivenum(), length())	1
2	Representation of data using R commands: 1. Diagrams (Simple, multiple, subdivided bar diagram) 2. Graphs (Histogram, ogive curves, boxplot diagram)	1
3.	Calculations of measures of 1.Central tendency 2. Dispersion 3. Skewness 4. Kurtosis for raw data.	1
4.	Calculations of using R 1. Probabilities 2. Quantiles Model sampling from probability distribution using R Understanding location, scale parameter using R (For Normal and Gamma distribution)	1
5	Fitting of negative binomial distribution and computation of expected frequencies and testing the goodness of fit graphically using R	1
6	Fitting of normal distribution and computation of expected frequencies. and testing the goodness of fit graphically using R	1
7	Fitting of Exponential distribution and computation of expected frequencies and testing the goodness of fit graphically using R	1
Practicals based on MS-EXCEL		
8	Computation of Moving average	1
9	Computation of seasonal indices	1
10	Project: Project based on analysis of data collected by students in groups of Maximum 6 students.	3
**** Study tour Report (if any)		

****

SEMESTER – III
PAPER - I
23-ST – 241: TESTS OF SIGNIFICANCE AND STATISTICAL METHODS.

Course Outcomes:

CO1) Students will learn new techniques like testing of hypotheses.
CO2) Students will learn multiple regression which is the extension of simple linear regression.
CO3) Students learn Demography and the various rates of vital statistics.
CO4) Students will learn new application which is queuing model.

1. Tests of Significance: (14L)

1.1 Random sample from a distribution as i. i. d. r.vs. $X_1, X_2, X_3, \ldots\ldots, X_n$.

1.2 Statistic and Parameter. Sampling distribution of a statistic, standard error of a statistic with illustrations. **Statistical Inference**: Introduction to problem of Estimation and testing of hypothesis. Estimator and estimate. Unbiased estimator (definition and simple illustrations only). Point and interval estimation. Statistical hypothesis, null and alternative hypothesis,simple and composite hypothesis, one sided and two sided alternative hypothesis, critical region, $type - I$ and $type - II$ error, level of significance, $p - value$. Two sided confidence interval. Tests of hypotheses using i) critical region approach, ii) $p - value$ approach and iii) confidence interval approach.

1.3 Tests for population means (large sample / approximate tests):

i) single population(two sided, one sided test, variance known)
ii) two populations(two sided, one sided test, variance known)
iii) Construction of two sided confidence interval for μ and $\mu_1 - \mu_2$

1.4 Tests for population proportions:

i) single population(two sided, one sided test)
ii) two population(two sided, one sided test).
iii) Construction of two sided confidence interval for P and $P_1 - P_2$.

2. Multiple Linear Regression Model: (08L)

2.1 Notion of multiple linear regression. Yule's notation (trivariate case) (statement only). Fitting of regression plane of Y on X_1 and X_2, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + s$ by the method of least squares; obtaining normal equations, solution of normal equations. Definition and interpretation of partial regression coefficients β_1 and β_2. (relations between partial regressioncoefficients and multiple correlations on are not expected). Residual: Definition, order, derivation of variance, properties. Finding multiple and partialcorrelation coefficients if (X_1, X_2, X_3) ~> MD(n, P_1, P_2 ,P_3)

2.2 Definition of multiple correlation coefficient. Derivation of the expression formultiple correlation coefficient. Properties of multiple correlation coefficient.
i)range ii) relation between multiple and total correlation coefficient
2.3 Coefficient of determination. Interpretation of coefficient of multiple determination
2.4 Partial correlation coefficient: Definition and derivation of partial correlation coefficient Property of partial correlation coefficient
i)range ii) relation between multiple and partial correlation coefficient

3. Demography: (08L)

3.1 Vital events, vital statistics, methods of obtaining vital statistics, rates of vital events,sex ratios, dependency ratio.
3.2 Death/Mortality rates: Crude death rate, specific (age, sex etc.) death rate, standardizeddeath rate (direct and indirect), infant mortality rate.
3.3 Fertility/Birth rate: Crude birth rate, general fertility rate, specific (age, sex etc.)fertility rates, total fertility rate.
3.4 Growth/Reproduction rates: Gross reproduction rate, net reproduction rate. (Numerical examples with heavy computations are to be asked preferably in practicals).
3.5 Interpretations of different rates, uses and applications.
3.6 Trends in vital rates as revealed in the latest census.

4. Queuing Model: (06L)

4.1 Introduction to queueing model. as an application of exponential distribution, Poissondistribution and geometric distribution.
4.2 Kendall's notation $M/M/1: FIFO$.
4.3 Inter arrival rate, service rate, traffic intensity, queue disciplines.
4.4 Probability distribution of number of customers in queue, average queue length, average waitingtime
In i) queue, ii) system. (Without derivations) statement of Little's formula / relations.

SEMESTER – III
PAPER - II

23-ST-242: CONTINUOUS DISTRIBUTIONS AND EXACT TESTS.

Course Outcomes:

CO1) Students will understand the concept of testing of hypothesis.
CO2) Students will able to describe and study the different kinds of continuous probability distributions such as Chi square distribution, t and F distributions.
CO3) They can find relations among aforesaid continuous random variables.
CO4) Students can implement the tests based on sampling distributions in the real life situations

1. Chi-square Distribution: (06L)

1.1 Definition of chi-square r.v. as a sum of squares of i.i.d. standard normal variables, Notation.

1.2 Derivation of the p.d.f. of Chi-square variable with n degrees of freedom (d.f.).

1.3 Chi square distribution as a special case of Gamma distribution.
mean, variance, MGF, CGF, central moments skewness, kurtosis, mode, additive property.

1.4 Use of chi-square tables for calculations of probabilities. Normal approximation of Chi square random variable to standard normal variate (statement only)

1.5 Distribution of $\bar{X}$ and S^2 for a random sample taken from normal distribution using orthogonal transformation, independence of $\bar{X}$ and S^2.

2. Student's t −distribution: (05L)

2.1 Definition of t r.v. with n d.f., Notation: $t \sim t_n$

2.2 Derivation of the p.d.f of t distribution, nature of probability curve, mean, variance, moments, Mode

2.3 Use of t-tables for calculations of probabilities, statement of normal approximation.

3. Snedecore's F −distribution: (05L)

3.1 Definition of F random variable as a ratio of two independent standard normal variates.
Notation: $F \sim F_{n_1,n_2}$

3.2 Derivation of the p.d.f, nature of probability curve,

3.3 mean, variance, moments, mode.

3.4 Distribution of $\frac{1}{F_{n1,n2}}$,

3.5 use of tables for calculation of probabilities

3.6 Interrelationship between Chi-square, t and F distributions.

4. Test of Hypothesis: (12L)

4.1Tests based on chi-square distribution:

a) Test for independence of two attributes arranged in 2 × 2 contingency table(with Yate'scorrection)

b) Test for independence of two attributes arranged in $r \times s$ contingency table, Mc Nemar'stest

c) Test for goodness of fit.

d)Test for variance (Ho: $\sigma^2 = \sigma^2$)against one-sided and two-sided alternatives i) for known mean, ii) for unknown mean.

4.2Tests based on distribution:

a) Tests for population means:

(i)Single sample with unknown variance and two sample for unknown equal variances testsfor one-sided and two-sided alternatives.

(ii)$100(1 - \alpha)$ % two sided confidence interval for population mean and difference of meansof two independent normal populations.

b) Paired t-test for one-sided and two-sided alternatives.

4.3Test based on F −distribution:

Test for equality of two population variances with

i) population means are known ii) Population means are unknown.

5. Bivariate normal distribution: (08L)

5.1 P. d. f. of a bivariate normal distribution

Notation :(, Y) ~ BN (μ_1, μ_2, ${\sigma_1}^2$, ${\sigma_2}^2$, ρ), $X \sim Np$ (μ, Σ)

5.2 Nature of surface of p. d. f., marginal and conditional distributions, identification of parameters,

5.3 Regression of Y on X and of X on Y, independence and un-correlatedness,

5.4 MGF and moments.

5.5 Distribution of $aX + bY + c$, X/Y.

5.6 Applications and real life situations

SEMESTER – III
PAPER - III

23-ST-243: PRACTICALS

Pre-requisites: Knowledge of the topics in theory.

Objectives:

1. To conduct various tests of significance like averages, population proportions, independence of attributes, variance etc. included in theory (using calculators, R software).
2. To compute probabilities of discrete and continuous probability distributions using R software.

Sr. No.	Title of the Practical	No. of Practical
1	Computations of GRR and NRR using R	1
2	Large sample tests using R 1. testing population proportion (single population, two populations) 2. Testing population mean (single population, two populations, variance known, unknown)	2
3	Small sample tests using R 1. Testing population mean (single population, two populations, variance known, unknown) 2. Paired t-test 3. Testing population variance (single population, two populations, mean known, unknown.)	2
4.	Tests based on chi-square distribution 1. Goodness of fit 2. Independence of attributes	1
6	Fitting of regression model using R 1. Simple regression model 2. Multiple regression model 3.Quadratic regression model	3
7.	**Writing the comment using if… then statement** **Finding frequency distribution using for() statement**	1
10	**Project**: Project based on analysis of data collected by students in groups of maximum **6** students. (Project is equivalent to three practical's)	3
	Study tour report (if any)	-

.....

Books Recommended:

1. Brockwell P.J.and Davis R.A. (2003), Introduction to Time Series and Forecasting (Second Edition),Springer Texts in Statistics.
2. Chatfield C. (2001), The Analysis of Time Series An Introduction, Chapman and Hall / CRC, Texts inStatistical Science.
3. Goon A. M., Gupta, M. K. and Dasgupta, B. (1986), Fundamentals of Statistics, Vol. 2, World Press,Kolkata.
4. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), SultanChand and Sons, 23, Daryaganj, New Delhi, 110002.
5. Gupta, S. C. and Kapoor V. K. (2007), Fundamentals of Applied Statistics (Fourth Edition), Sultan Chandand Sons, New Delhi.
6. Gupta, S. P. (2002), Statistical Methods (Thirty First Edition), Sultan Chand and Sons, 23, Daryaganj, NewDelhi 110002.
7. Hogg, R. V. and Craig, A. T., Mckean J. W. (2012), Introduction to Mathematical Statistics (TenthImpression), Pearson Prentice Hall.
8. Kulkarni, M. B., Ghatpande, S. B. and Gore, S. D. (1999), Common Statistical Tests, Satyajeet Prakashan,Pune 411029
9. Medhi, J., Statistical Methods, Wiley Eastern Ltd., 4835/24, Ansari Road, Daryaganj, New Delhi – 110002.
10. Meyer, P. L., Introductory Probability and Statistical Applications, Oxford and IBH Publishing Co. NewDelhi.
11. Mood, A. M., Graybill F. A. and Bose, F. A. (1974), Introduction to Theory of Statistics (Third Edition,Chapters II, IV, V, VI), McGraw - Hill Series G A 276
12. Mukhopadhya Parimal (1999), Applied Statistics, New Central Book Agency, Pvt. Ltd. Kolkata
13. Purohit S. G., Gore S. D. and Deshmukh S. R. (2008), Statistics using R, Narosa Publishing House, NewDelhi.
14. Ross, S. (2003), A first course in probability (Sixth Edition), Pearson Education publishers, Delhi, India.
15. Walpole R. E., Myers R. H. and Myers S. L. (1985), Probability and Statistics for Engineers and Scientists (Third Edition, Chapters 4, 5, 6, 8, 10), Macmillan Publishing Co. Inc. 866, Third Avenue, New York 10022.
16. Weiss N., Introductory Statistics, Pearson education publishers.

****