Maharashtra Education Society's

ABASAHEB GARWARE COLLEGE (AUTONOMOUS)

KARVE ROAD, PUNE 411004

(Affiliated to Savitribai Phule Pune University)

Three year B. Sc. Degree Program in Statistics

(Faculty of Science and Technology)

Syllabus under Autonomy

S. Y. B. Sc. (Statistics)

Choice Based Credit System (C. B. C. S.) Syllabus

To be implemented from Academic Year 2023-2024

Title of the course: S. Y. B. Sc. (Statistics)

ELIGIBILITY:

1. Passed (with at least 22 credits) in F. Y. B. Sc. with Statistics as one of the subjects.
2. 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.

STRUCTURE OF THE COURSE: (Each theory and practical paper has 2 credit)

Semester	Course Type	Course Code	Course Title	Remark	No. of Lectures /Practicals to be conducted
III	Core Course	USST -231	Probability Distributions-I	Theory	36
		USST - 232	Probability Distributions-II	Theory	36
		USSTP -233	Practical Paper - I	Practical	11
	Ability Enhancement course	USLG-231	Language	Theory	36
		UEVS-231	Environmental Science	Theory	36
IV	Core Course	USST -241	Statistical Methods	Theory	36
		USST -242	Sampling Distributions and Exact Tests	Theory	36
		USSTP -243	Practical Paper - II	Practical	12
	Ability Enhancement course	USLG-241	Language	Theory	36
		UEVS-241	Environmental Science	Theory	36

GENERAL INSTRUCTIONS: **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 S.Y. B.Sc. Statistics students may be arranged and study report should be attached in the journal.

S.Y.B.Sc Syllabus in Statistics
SEMESTER – III

Paper Title: Probability Distributions – I

Paper Code: USST – 231 | **Number of credits – 2**
Number of Lectures - 36

Course outcomes:

At the end of this course, students are able to…

1) Understand about continuous univariate and bivariate random variables, their expectation, variance, higher order moments and their properties.
2) Get the knowledge of different standard continuous distributions, their M.G.F., C.G.F., skewness, kurtosis etc.
3) Find the distributions of transformed random variables using various techniques.
4) Identify various applications of these distributions in real life.
5) Simulate random samples from standard continuous distributions.

Units and Contents

Unit 1: Continuous Univariate Distribution **[10 Lectures]**

1.1 Continuous (Uncountably infinite) 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., finding probabilities.

1.2 Expectation of continuous random variable (r.v.), expectation of function of r.v., theorems on expectation: (i) expectation of constant a is the constant a itself (ii) effect of change of origin and scale. Variance and theorems on variance: (i) variance of constant is 0 (ii) effect of change of origin and scale. Mean, geometric mean, harmonic mean, moments about constant a, raw and central moments with their interrelation, skewness, kurtosis, mean deviation about mean.

1.3 Moment generating function (M.G.F.): Definition, properties (i) $M_X(0) = 1$ (ii) Uniqueness property: For every p.d.f. there is unique M.G.F. and vice versa (iii) $M_{cX}(t) = M_X(ct)$, where c is a constant. Deductions of raw and central moments using M.G.F.
Examples to find central moments using M.G.F. (M.G.F. of $X - \mu$ where μ be mean of the distribution), Cumulant generating function (C.G.F.): Definition, properties: (i) Additive property (ii) Effect of change of origin and scale on cumulants.

1.4 Mode, partition values: quartiles, deciles, percentiles, i^{th} quantile, quartile deviation.

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

1.6 Examples

Unit 2: Continuous Bivariate Distribution **[09 Lectures]**

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

2.2 Expectation of a function of bivariate r.v. i.e. E[g(X, Y)], joint raw and central moments, Cov (X,Y), Corr (X,Y), conditional mean, conditional variance

2.3 Concept of independence of random variables X and Y (probability density function approach and expectation function approach).

2.4 Results related expectation:

(i) E(X+Y) = E(X) + E(Y)

(ii) E(XY) = E(X) E(Y), if X and Y are independent.

(iii) E[E(X|Y = y)] = E(X)

(iv) V (X) = E[V (X|Y)] + V[E (X|Y)]

(v) E (aX + bY+c) and Var (aX + bY + c)

2.5 Moment generating function (M.G.F.): $M_{X,Y}(t_1, t_2)$, M.G.F. of marginal distribution of random variables, properties:

(i) $M_{X,Y}(t_1, t_2) = M_X(t_1)\, M_Y(t_2)$ if X and Y are independent r.v.s.,

(ii) $M_{X+Y}(t) = M_{X,Y}(t, t)$

(iii) $M_{X+Y}(t) = M_X(t)\, M_Y(t)$ if X and Y are independent r.v.s.

2.6 Probability distribution of transformation of bivariate r. v. (X, Y)

2.7 Examples

Unit 3: Continuous Uniform (Rectangular Distribution) **[03 Lectures]**

3.1 Probability density function (p.d.f.):

$$f(x) = \begin{cases} \frac{1}{b-a}\ , & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

Notation- $X \sim U(a, b)$. Plotting of p.d.f. curve for various parameter values, c. d. f., mean, variance, moments, skewness and kurtosis, M.G.F., C.G.F.

3.2 Distributions of (i) $\frac{X-a}{b-a}$, (ii) $\frac{b-X}{b-a}$, (iii) X + Y

3.3 Property- Let Y = F(X) where F(X) be the c.d.f. of continuous r. v. X, then Y follows Uniform Distribution (0,1).

3.4 Real life applications and examples.

Unit 4: Normal Distribution **[09 Lectures]**

4.1 Probability distribution function (p.d.f):

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; & -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0 \\ 0 ; & \text{otherwise} \end{cases}$$

Notation: $X \sim N(\mu, \sigma^2)$. Plotting of p.d.f. curve for various parameter values, identification of scale and location parameters, c.d.f.

4.2 Mean, variance, central moments of odd and even order, recurrence relation between even ordered moments, M.G.F., C.G.F., cumulants, points of inflexion of probability curve, mean deviation, additive property, area under the curve, distribution function ϕ(x), area between the ordinates $\mu - k\sigma$ and $\mu + k\sigma$ where k be the integer constant, mode, skewness, kurtosis, mode, quartiles, quartile deviation.

4.3 Probability distribution of: i) $Z = \frac{X-\mu}{\sigma}$, standard normal variable (S.N.V.) and its notation, distribution function ϕ(z) (If $\phi(z) > 0.5$ then z > 0, $\phi(z) < 0.5$ then z < 0 and $\phi(z) = 0$ then z = 0 where Z~N(0, 1)) ii) U = $aX + b$, iii) $V = aX + bY + c$, where X and Y are independent normal variates iv) $\bar{X}$, the mean of n independent and identically distributed (i.i.d.) $N(\mu, \sigma^2)$ r.v.s.

4.4 Computations of normal probabilities using normal probability integral tables. Examples for computation of c.d.f. for Z where Z ~ $N(0, 1)$.

4.5 Generation of random sample using Box-Muller transformation. Normal Probability Plot.

4.6 Real life applications and examples.

Unit 5: Lognormal Distribution **[05 Lectures]**

5.1 Probability distribution function (p.d.f):

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

Notation $X \sim LN\,(a, \mu, \sigma^2)$. Plotting of p.d.f. curve for various parameter values, moments (r^{th} moment of $X - a$), first four central moments of X, skewness, kurtosis, mode, quartiles, quartile deviation.

5.2 Relation with $N\,(\mu, \sigma^2)$ distribution

5.3 Distribution of $\prod X_i$, Xi's are independent lognormal variates.

5.4 Real life applications and examples.

Reference Books:

1. Goon A. M., Gupta, M. K. and Dasgupta, B. (1986), Fundamentals of Statistics, Vol. 1, World Press, Kolkata.
2. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002.
3. Hogg, R. V. and Craig, A. T., Mckean J. W. (2012), Introduction to Mathematical Statistics (Tenth Impression), Pearson Prentice Hall.
4. Meyer, P. L. (1970), Introductory Probability and Statistical Applications, Oxford and IBH Publishing Co. New Delhi.
5. 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.
6. Ross, S. (2003), A first course in probability (Sixth Edition), Pearson Education publishers, Delhi, India.
7. 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.
8. Weiss N. (2015), Introductory Statistics, Pearson education publishers.
9. Rohatgi, V. K. and Saleh, A. K. (2015). An Introduction to Probability and Statistics, Third Edition, John Wiley & Sons, Inc.

S.Y.B.Sc Syllabus in Statistics
Semester – III
Paper Title: Probability Distributions – II

Paper Code: USST – 232 | **Number of credits – 2**
Number of Lectures - 36

Course outcomes:

At the end of this course, students are able to…

1) Apply standard discrete probability distributions in real life applications.
2) Understand the lifetime probability distributions like exponential and gamma distributions.
3) Apply these lifetime distributions in real life applications.
4) Get the knowledge of truncated probability distributions and how to apply in real life situations.

Units and Contents

Unit 1: Some Standard Discrete Probability Distributions [16 Lectures]

Negative Binomial Distribution:

1.1 Probability mass function (p.m.f.):

$$P(X = x) = \begin{cases} \binom{x+k-1}{x} p^k q^x ; & x = 0, 1, 2, \dots, 0 < p < 1, q = 1 - p \text{ and } k > 0 \\ 0; & \text{otherwise} \end{cases}$$

Notation: $X \sim NB(k, p)$, where X is the number of failures before getting the k^{th} success. Plotting of p.d.f. curve for various parameter values, introduction of p.m.f. in terms of Y, where Y is the number of trials to get k^{th} success. Negative binomial distribution as a waiting time distribution.

1.2 Mean, variance, moments, skewness, kurtosis, M.G.F., C.G.F., additive property.

1.3 Relation between geometric distribution and negative binomial distribution.

1.4 Poisson approximation to negative binomial distribution.

1.5 Real life applications and examples. (Example based capture-recapture problem)

Multinomial Distribution:

1.6 Probability mass function (p.m.f.):

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!\, p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}}{x_1!\, x_2! \dots x_k!} ; x_i = 0, 1, 2, \dots, n;$$
$$i = 1, 2, \dots, , k$$
$$x_1 + x_2 + \dots + x_k = n;$$
$$0 < p_i < 1; i = 1, 2, \dots, k;$$
$$p_1 + p_2 + \dots + p_k = 1;$$
$$= 0 \quad ; \text{otherwise}$$

Notation: $(X_1, X_2, \dots, X_k) \sim MD(n, p_1, p_2, \dots, p_k)$ or $\underline{X} \sim MD\left(n, \underline{p}\right)$ where $\underline{X} = (X_1, X_2, \dots, X_k)$ and $\underline{p} = (p_1, p_2, \dots, p_k)$

1.7 Joint M.G.F. of $(X_1, X_2, \ldots., X_k)$, use of M.G.F. to obtain means, variances, covariances, total correlation coefficients, variance–covariance or dispersion matrix, rank of variance–covariance matrix and its interpretation.

1.8 Additive property of multinomial distribution, univariate marginal distribution, distribution of $X_i + X_j$, conditional distribution of X_i given $X_j = r$, conditional distribution of X_i given $X_i + X_j = r$.

1.9 Real life situations and applications. Examples.

Unit 2: Truncated Distributions **[08 Lectures]**

2.1 Truncated Distributions: Concept of truncated distribution, truncation to the right, left and on both sides. Binomial distribution left truncated at $X = 0$ (value zero is discarded), its p.m.f., mean and variance.

2.2 Poisson distribution left truncated at $X = 0$ (value zero is discarded), its p.m.f., mean and variance.

2.3 Normal distribution $N(\mu, \sigma^2)$ truncated (i) to the left below a (ii) to the right above b (iii) to the left below a, and to the right above b (a < b), its p.d.f., derivation of mean and variance.

2.4 Real life applications and examples.

Unit 3: Some Standard Continuous Probability Distributions **[12 Lectures]**

Exponential Distribution:

3.1 Probability density function (p. d. f.):

$$f(x) = \begin{cases} \alpha e^{-\alpha x}; & x \geq 0, \alpha > 0 \\ 0 \quad ; & \text{otherwise} \end{cases}$$

Notation: $X \sim Exp(\alpha)$. Plotting of p.d.f. curve for various parameter values, mean, variance, moments, skewness, kurtosis, c.d.f.,, survival function, M.G.F., C.G.F., quartiles, quartile deviation, mean deviation about mean.

3.2 Lack of memory property (with proof) its interpretation and illustrations.

3.3 Properties: (i) Distribution of sum of two i.i.d. exponential r.v.s. (ii) Distribution of min (X, Y) and max (X, Y) with X, Y $i.i.d.$ exponential r.v.s. (iii) If X follows U(0, 1) then Y=$-\frac{1}{\alpha}\log(1 - X)$ follows exp(α) distribution.

3.4 Applications: (i) interpretation of α as an interarrival rate of customers joining the queue and $\frac{1}{\alpha}$ as mean. (ii) as a lifetime distribution with constant hazard rate.

3.5 Real life applications and examples.

Gamma Distribution:

3.6 Probability density function (p. d. f.):

$$f(x) = \begin{cases} \dfrac{\alpha^{\lambda}}{\Gamma\lambda} x^{\lambda-1} e^{-\alpha x}; & x > 0, \quad \alpha, \lambda > 0 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Notation: $X \sim G(\alpha, \lambda)$ (α=scale parameter and λ= shape parameter). Plotting of p.d.f. curve for various parameter values, special cases: i) $\alpha = 1$, ii) $\lambda = 1$, mean, variance, moments, skewness, kurtosis, mode, M.G.F., C.G.F., additive property.

3.7 Properties of gamma distribution -

(i) Relation between distribution functions of Poisson and Gamma variables.

(ii) Relation between the normal and gamma distributions:

If X follows N(0, 1) then $X^{2 =}$ follows G(½, ½)

(iii) Distribution of sum of n i.i.d. exponential variables is a gamma variable.

(iv) If X and Y are independent G(α, λ_1) and G(α, λ_2) variables respectively then U=X + Y and V=X/(X+Y) are independent .

(v) If X follows G(α, λ_1) and Y follows G(α, λ_2) , where X and Y are independent then show that U=X+Y and V=X/Y are independent.

3.8 Real life applications and examples. (Example about gamma distribution as a lifetime distribution with increasing hazard rate)

Reference Books:

1. Gupta, S. C. and Kapoor, V. K. (2002). Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, New Delhi.
2. Gupta, S. P. (2002), Statistical Methods (Thirty First Edition), Sultan Chand and Sons, New Delhi.
3. Casella, G., & Berger, R. L. (2021). Statistical inference. Cengage Learning.
4. Ross S. (2002). A First Course in Probability, Sixth Edition, Pearson Education, Inc. & Dorling Kindersley Publishing, Inc.
5. Rohatgi, V. K. and Saleh, A. K. (2015). An Introduction to Probability and Statistics, Third Edition, John Wiley & Sons, Inc.
6. Johnson N. L. & Kotz. S. (1996): Distributions in Statistics Vol-I, II and III, John Wiley and Sons, New York.
7. Dasgupta, A. (2010) Fundamentals of Probability: A First Course, Springer.
8. Arora, S. and Bansi, Lal. (1989), New Mathematical Statistics. Satya Prakashan, New Delhi.

Paper Title - Practical Paper - I

Semester – III **Paper Code: USSTP – 233** **Number of credits – 2**

Course Outcomes:

At the end of this course, students are able to…

1) Learn to fit a suitable discrete and continuous probability distributions to the data.
2) Identify the suitable probability model for the population.
3) Generate random samples from the continuous probability distributions.
4) Learn the basic commands of R-software. Carry out data visualization, summary statistics, computation of probabilities for discrete distributions using R software.

Expt. No.	Title of experiment
1	Fitting of negative binomial distribution and computation of expected frequencies. Plot of expected frequencies vs. observed frequencies.
2	Fitting of normal distribution and computation of expected frequencies. Plot of expected frequencies vs. observed frequencies.
3	Applications of negative binomial and multinomial distributions.
4	Applications of exponential and normal distributions.
5	A) Generating random samples from exponential distribution using distribution function method. B) Generating random samples from normal and log normal distributions using i) distribution function ii) Box-Muller transformation.
6	Applications of truncated binomial, truncated Poisson and truncated normal distributions.
7	Use of basic R-software commands: c(), scan(), rep(), seq(), length(), min(), max(), sort(), extract(), data.frame(), matrix(), accessing resident data sets.
8	Diagrammatic (pie chart, bar diagram- simple, sub-divided, multiple) and graphical (histogram, stem and leaf plot, rod or spike plot, ogive curves, scatter plot) representation of data using R software.
9	Finding summary statistics using summary( ) and fivenum( ) functions. Calculate arithmetic mean (A.M.), geometric mean (G.M.), harmonic mean (H.M.), median, mode, quantiles, range, quartile deviation (Q.D.), variance, coefficient of variation (C.V.) (ungrouped data) using R software.
10	Computation of probabilities for binomial, hypergeometric, Poisson, geometric, negative binomial and multinomial distributions using R software.
11	Plot of p.m.f./p.d.f. and c.d.f. curve of standard discrete and continuous probability distributions using R software.

Total no. of experiments: 11

Each practical is of duration: 4 hours 20 minutes

S.Y.B.Sc Syllabus in Statistics
Semester – IV
Paper Title: Statistical Methods

Paper Code: USST – 241

Number of credits – 2
Number of lectures - 36

Course Outcomes:

At the end of this course, students are able to…

1) Understand the concepts of estimator and estimate. Also, learn the methods of estimation.
2) Identify the appropriate test of hypothesis to be used in a scenario at hand.
3) Aptly conclude whether a claim made about the population is valid or not through a corresponding test of hypothesis.
4) Infer about the validity of a hypothesis via various approaches.
5) Identify a situation where multiple linear regression can be used.
6) Select an appropriate response variable and corresponding regressors for fitting a regression plane.
7) Compute and interpret multiple and partial correlation coefficients.
8) Determine the strength and adequacy of fit of a multiple linear regression model.

Units and Contents

Unit 1: Methods of Estimation **[12 Lectures]**

1.1 Statistic and Parameter: Random sample $X_1, X_2, \dots, X_n$ from a distribution, concept of statistic, sampling distribution of a statistic, standard error of a statistic. Notion of parameter and parameter space. Concept of family of distributions.

1.2 Statistical Inference: Introduction to problem of estimation and testing of hypothesis. Estimator and estimate, difference between estimator and estimate. Point and interval estimation.

1.3 Method of maximum likelihood: Likelihood Function, Definition of likelihood as a function of unknown parameter for a random sample from a probability distribution.
Distinction between the likelihood function and p.d.f. or p.m.f.

1.4 Derivation of maximum likelihood estimator (M.L.E.) for parameters of standard distributions (case of only one parameter): Bernoulli, Binomial, Poisson, Geometric, Exponential, Normal (with known mean or known variance)

1.5 Method of moments: Introduction, derivation of moment estimator for standard distributions. Illustrations of situations where M.L.E. and moment estimators are distinct.

1.6 Method of moment estimator of the parameter for a given non-standard probability distribution. Examples.

Unit 2: Testing of Hypothesis **[14 Lectures]**

2.1 Statistical hypothesis, null and alternative hypotheses, simple and composite hypotheses, one-sided and two-sided alternative hypotheses, critical region, type-I and type-II error, notion of size and power of test, level of significance, p-value. Two-sided confidence interval.

2.2 Testing of hypotheses using (i) critical region approach, (ii) p-value approach and (iii) confidence interval approach.

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

(i) $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, $H_1 : \mu > \mu_0$, $H_1 : \mu < \mu_0$

(ii) $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, $H_1 : \mu_1 > \mu_2$, $H_1 : \mu_1 < \mu_2$

(iii) Construction and application of two sided confidence interval for:

a) μ b) $\mu_1 - \mu_2$

2.4 Tests for population proportions:

(i) $H_0 : P = P_0$ against $H_1 : P \neq P_0$, $H_1 : P > P_0$, $H_1 : P < P_0$

(ii) $H_0 : P_1 = P_2$ against $H_1 : P_1 \neq P_2$, $H_1 : P_1 > P_2$, $H_1 : P_1 < P_2$

(iii) Construction and application of two sided confidence interval for:

a) P b) $P_1 - P_2$

Unit 3: Multiple Linear Regression **[10 Lectures]**

3.1 Notion of multiple linear regression, extension of simple linear regression to multiple linear regression through illustrations, total correlation coefficients $r_{YX_1}, r_{YX_2}, r_{X_1X_2}$ matrix of correlation coefficients R.

3.2 Yule's notation for trivariate case taking Y as the response variable and X_1, X_2 as regressors. Fitting of regression plane of Y on X_1 and X_2 by principle of least squares, expressions of regression coefficients $b_{YX_1.X_2}$ and $b_{YX_2.X_1}$ in terms of total correlation coefficients, interpretation of regression coefficients.

3.3 Mapping of regression plane in Yule's notation to regression plane of Y on X_1 and X_2, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$.

3.4 Residuals: Definition, order of residuals, notations, properties of residuals, Variances of residuals (statements only).

3.5 Multiple Correlation Coefficient ($R_{Y.X_1X_2}$): Definition, meaning and interpretation, derivation of expression of $R_{Y.X_1X_2}$ in terms of total correlation coefficients.

3.6 Properties of multiple correlation coefficient (with proof):

(i) $0 \leq R_{Y.X_1X_2} \leq 1$ (ii) $R_{Y.X_1X_2} \geq \max\{r_{YX_1}, r_{YX_2}\}$

3.7 Coefficient of multiple determination (${R_{Y.X_1X_2}}^2$), Interpretation in terms of proportion of variation explained in linear regression. Interpretation of cases:

(i) ${R_{Y.X_1X_2}}^2 = 1$ (ii) ${R_{Y.X_1X_2}}^2 = 0$.

3.8 Notion and need of adjusted R-square.

3.9 Partial Correlation Coefficients ($r_{YX_1.X_2}$ and $r_{YX_2.X_1}$): Definition, meaning and interpretation, derivation of expressions of $r_{12\bullet3}$ and $r_{13\bullet2}$ in terms of total correlation coefficients.

Properties of partial regression coefficients (with proof)

(i) $-1 \leq r_{YX_1.X_2} \leq 1$ (ii) $-1 \leq r_{YX_2.X_1} \leq 1$

Expression of multiple correlation coefficient in terms of total and partial correlation coefficients.

3.10 Conditions of consistency of data in terms of: i) R matrix ii) ${R_{Y.X_1X_2}}^2$

Finding expressions for multiple and partial correlation coefficients if

$$(X_1, X_2, X_3) \sim MD(n, p_1, p_2, p_3)$$

Reference Books:

1. Buyan, K. C. (2010). Probability theory and Statistical inference,1stEdn., New Central Book Agency.
2. Goon A. M., Gupta, M. K. and Dasgupta, B. (1986), Fundamentals of Statistics, Vol. 2, World Press, Kolkata.
3. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002 .
4. Gupta, S. P. (2002), Statistical Methods (Thirty First Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi 110002.
5. Kulkarni, M. B., Ghatpande, S. B. and Gore, S. D. (1999), Common Statistical Tests, Satyajeet Prakashan, Pune 411029.
6. Mishra Amarendra (2020), Theory of Statistical Hypothesis Testing (First Edition), Notion Press.
7. Taff Arthur (2018), Hypothesis Testing: The Ultimate Beginner's Guide to Statistical Significance, CreateSpace Independent Publishing Platform.
8. Wilson J. H., Keating B. P., Beal-Hodges M. (2012), Regression Analysis, Business Expert Press.
9. Wayne W. Daniel (2006), Biostatistics: A Foundation for Analysis in Health Sciences, Seventh edition, Wiley India Pvt. Ltd.
10. Draper, N. R. and Smith, H. (1998), Applied Regression analysis, (John Wiley) Third Edition.
11. Montgomery, D.C., Peck, E. A. and Vining, G. G.(2003), Introduction to Linear Regression Analysis, Wiley.
12. Chatterjee S. and Hadi A.S. (2012), Regression Analysis by Example, Fifth Edition, Wiley.

S.Y.B.Sc Syllabus in Statistics
Semester – IV
Paper Title: Sampling Distributions and Exact Tests

Paper Code: USST – 242 **Number of credits – 2**
Number of lectures - 36

Course Outcomes:

At the end of this course, students are able to…

1) Distinguish between various sampling distributions.
2) Understand different properties of these sampling distributions.
3) Understand the interrelationships between these distributions.
4) Understand applications of hypothesis testing in real life based on sampling distributions.
5) Solve real life testing problems based on sampling distributions.

<u>Units and Contents</u>

Unit 1: Chi-square Distribution **[12 Lectures]**

1.1 Definition of Chi-square r.v. as a sum of squares of i.i.d. standard normal variables, derivation of the p.d.f. of Chi-square variable with n degrees of freedom (d.f.) using M.G.F. Notation: $X \sim \chi_n^2$. Plotting of p.d.f. curve for various parameter values.

1.2 Mean, variance, M.G.F., C.G.F., central moments skewness, kurtosis, mode, additive property. Use of chi-square tables for calculations of probabilities.

1.3 Normal approximation: $\frac{\chi_n^2 - n}{\sqrt{2n}}$ (statement only). Distribution of $\bar{X}$ and $\frac{ns^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2$ for a random sample from a normal distribution using orthogonal transformation, independence of $\bar{X}$ and S^2.

Unit 2: Student's – t distribution **[05 Lectures]**

2.1 Definition of t r.v. with n d.f. in the form of $t = \frac{U}{\sqrt{\frac{V}{n}}}$, where $U \sim N(0,1)$ and V is Chi-square with n d.f., where U and V are independent random variables. Notation: $t \sim t_n$.

2.2 Derivation of the p.d.f of t distribution. Plotting of p.d.f. curve for various parameter values, mean, variance, moments, mode. Use of t-tables for calculations of probabilities, statement of normal approximation.

2.3 Distinction between density curves of normal and t-distributions.

Unit 3: Snedecor's F – distribution **[06 Lectures]**

3.1 Definition of F r.v. with n_1 and n_2 d.f. as $F_{n_1,n_2} = \frac{X_1/n_1}{X_2/n_2}$ where X_1 and X_2 are independent Chi-square variables with n_1 and n_2 d.f., Notation: $F \sim F_{n_1,n_2}$.

3.2 Derivation of the p.d.f, plotting of p.d.f. curve for various parameter values, mean, variance, moments, mode.

3.3 Properties of F distribution.

3.4 Distribution of $\frac{1}{F_{n_1,n_2}}$,use of F - tables for calculation of probabilities. Interrelationship between Chi-square, t and F distributions.

Unit 4: Tests of Hypothesis based on sampling distributions [13 Lectures]

4.1 Tests based on chi-square distribution:

(a) Test for independence of two attributes arranged in 2×2 contingency table (with Yate's correction)
(b) Test for independence of two attributes arranged in $r \times s$ contingency table
(c) Mc Nemar's test
(d) Test for goodness of fit.
(e) Test for variance (H_0: $\sigma^2 = \sigma_0^2$) against one-sided and two-sided alternatives i) for known mean ii) for unknown mean.

4.2 Tests based on t - distribution:

a) Tests for population means:
   (i) Single sample with unknown variance and two sample for unknown equal variances (for one-sided and two-sided alternatives.)
   (ii) $100(1 - \alpha)\%$ two-sided confidence interval for population mean and difference of means of two independent normal populations.
b) Paired t-test for one-sided and two-sided alternatives.
c) Test for significance of population correlation coefficient.
d) Test for significance of linear regression coefficient for one regressor only.

4.3 Test based on F-distribution:

a) Test for H_0: $\sigma_1^2 = \sigma_2^2$: against one-sided and two-sided alternatives when i) means are known and ii) means are unknown.

Reference Books:

1. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002 .
2. Mishra Amarendra (2020), Theory of Statistical Hypothesis Testing (First Edition), Notion Press.
3. Taff Arthur (2018), Hypothesis Testing: The Ultimate Beginner's Guide to Statistical Significance, CreateSpace Independent Publishing Platform.
4. Buyan, K. C. (2010). Probability theory and Statistical inference,1stEdn., New Central Book Agency.
5. Goon A. M., Gupta, M. K. and Dasgupta, B. (1986), Fundamentals of Statistics, Vol. 2, World Press, Kolkata.
6. Wayne W. Daniel (2006), Biostatistics: A Foundation for Analysis in Health Sciences, 7th edition, Wiley India Pvt. Ltd.

Paper Title - Practical Paper – II

Semester – IV **Paper Code: USSTP – 243** **Number of credits - 2**

Course Outcomes:

At the end of this course students are able to…

1. Learn to test the significance of mean, proportions, attributes and variance for the sample.
2. Computation of probabilities of various continuous probability distributions using R-software.
3. Understanding of fitting of trivariate regression plane using R-software.
4. Analyzing and interpretation of the collected sample data.

Expt. No.	Title of Experiment
1	Fitting of lognormal distribution and computation of expected frequencies. Plot of expected vs. observed frequencies.
2	Test for proportions and construction of confidence interval. Verification of result using p-value approach.
3	Test for means and construction of confidence interval. Verification of result using p-value approach.
4	Test based on chi square distribution (i) Test for independence of attributes (2 x 2, r x s contingency table) (ii) Goodness of fit (iii) McNemar's test.
5	Tests for population variance (one sample and two sample).
6	Computation of probabilities of continuous uniform, exponential, normal, log normal, gamma, chi square, t, and F distributions using R software. Fitting of trivariate regression plane using R software.
7	Tests using R software for: (i) Proportions and means. (ii) Chi square, Student's-t and Snedecor's-F distributions. (iii) Significance of correlation coefficient and linear regression coefficient.
8	Model sampling from standard discrete and continuous distributions using R software.
9	**Project:** Project based on analysis of data collected by students in groups of maximum 6 students. (Project is equivalent to four practicals)

Total no. of experiments: 12

Each practical is of duration: 4 hours 20 minutes

Instructions for the completion of project:

1) Project activity should be carried out in a group consisting not more than 6 students.

2) A project group has to submit a single copy of project report to the department.
3) Different data sets from primary or secondary sources may be collected.
4) Students have to use statistical methods learnt in the respective semester/year and perform data analysis with the use of free statistical software.

Points to be included in the Project Report:

A). Title of the project, names of the students, name of the department and college. Acknowledgement, Data Sources, Description of the computing system/software(s), Programming language(s) used, etc. (if applicable)

B). Motivation for selecting the topic, abstract of the project, key-words of the project.

C). **Text of the project**: Broadly this should cover description of the selected problem using terminology in the field of application, conversion of the problem in statistical language, description of collected data, small illustrative data set, methodology for the analysis, interpretation of the results, validation of the results, conclusions in statistical as well as user's language, limitation of proposed solutions, directions for future work, references used, etc.

----------xxxxx-------xxxxx-------