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 N. E. P.

S. Y. B. Sc. (Statistics)

To be implemented from Academic Year 2024-2025

Title of the course: S. Y. B. Sc. (Statistics) (NEP 2023 pattern)

I. ELIGIBILITY:

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

II. STRUCTURE OF THE COURSE:

SEMESTER III

Course Type	Course Code	Course Title	Remark	Credit	No. of hours engaged
Major	STS-201-MJ	Theory of Continuous Probability Distributions	Theory	2	30
	STS-202-MJ	Discrete and Continuous Probability Distributions	Theory	2	30
	STS-203-MJ	Time Series Analysis	Theory	2	30
	STS-204-MJP	Statistics Practical – III	Practical	2	60
Minor	STS-241-MN	Fundamentals of Statistics	Theory	2	30
	STS-242-MNP	Practicals based on Fundamentals of Statistics and R - Software	Practical	2	60
OE	OE-201-STS	Survey Analysis	Theory	2	30
VSC	STS-221-VSC	Data Analysis with R-Software	Practical	2	60
FP	STS-231-FP	Field Project	-	2	60

SEMESTER IV

Course Type	Course Code	Course Title	Remark	Credit	No. of hours engaged
Major	STS-251-MJ	Statistical Methods	Theory	2	30
	STS-252-MJ	Sampling Distributions and Exact Tests	Theory	2	30
	STS-253-MJ	Statistical Process and Product Control	Theory	2	30
	STS-254-MJP	Statistics Practical – IV	Practical	2	60
Minor	STS-291-MN	Statistical Inference	Theory	2	30
	STS-292-MNP	Practicals based on Statistical Inference	Practical	2	60
OE	OE-251-STS	Basic Applied Statistics	Theory	2	30
SEC	SEC-251-STS	Introduction to Python	Practical	2	60
CEP	STS-281-CEP	Community Engagement Practices	-	2	60

III. 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 / Field visit for S.Y.B.Sc. Statistics students may be arranged and a study tour report must be submitted to the concerned teacher.

IV. Syllabus:

SEMESTER – III

MAJOR

STS-201-MJ – Theory of Continuous Probability Distributions

Semester – III **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

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 Hours]

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.

1.3 Mean, geometric mean, harmonic mean, moments about constant a, raw and central moments with their interrelation, skewness, kurtosis, mean deviation about mean.

1.4 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.

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

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

1.7 Examples

Unit 2: Continuous Bivariate Distribution **[10 Hours]**

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.

2.3 Results related expectation:

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

(ii) E(XY) = E(X) E(Y), if X and Y are independent, generalization to k variables.

(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) **[02 Hours]**

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 Y = F(X) where F(X) is 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 **[08 Hours]**

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

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

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), (also $\phi(z) > 0.5$ then z > 0, $\phi(z) < 0.5$ then z < 0 and $\phi(z) = 0$ then z = 0 where Z~N(0, 1)), 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, 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 $N(0, 1)$.

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

4.6 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.

STS-202-MJ - Discrete and Continuous Probability Distributions

Semester – III **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course outcomes:

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

1) Understand about some standard discrete and continuous probability distributions and their real life applications.
2) Understand about some lifetime distributions like exponential and gamma distributions. Their real life applications and examples.
3) Understand the concept of truncated probability distributions.
4) Apply truncated probability distributions in real life situations.

Units and Contents

Unit 1: Some Standard Discrete Probability Distributions [14 Hours]

1.1 Negative Binomial Distribution:

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

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

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.

1.1(b) Mean, variance, moments, skewness, kurtosis (recurrence relation between moments is not expected), M.G.F., C.G.F., additive property.

1.1(b) Relation between geometric distribution and negative binomial distribution.

1.1(c) Poisson approximation to negative binomial distribution.

1.1(d) Real life applications and examples. (Example based on capture-recapture problem)

1.2 Multinomial Distribution:

1.2(a) Probability mass function (p.m.f.):

$$P(X_1=x_1, X_2=x_2,\ldots,X_k=x_k)=\frac{n!\,p_1^{x_1}p_2^{x_2}\ldots p_k^{x_k}}{x_1!\,x_2!\ldots x_k!};\ x_i=0,1,2,\ldots,n;\ i=1,2,\ldots,,k$$

$$x_1+x_2+\cdots+x_k=n;$$
$$0<p_i<1;\ i=1,2,\ldots,k;$$
$$p_1+p_2+\cdots+p_k=1;$$
$$=0 \qquad ;\text{otherwise}$$

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

1.2(b) 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.

1.2(c) 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 $X_i + X_j = r$.

1.2(d) Real life situations and applications. Examples.

Unit 2: Truncated Distributions **[06 Hours]**

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 **[10 Hours]**

3.1 Exponential Distribution:

3.1(a) 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, hazard rate, M.G.F., C.G.F., quartiles, quartile deviation, mean deviation about mean.

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

3.1(c) 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.

3.1(d) 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.

3.1(e) Real life applications and examples.

3.2 Gamma Distribution:

3.2(a) Probability density function (p. d. f.):

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

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, c.d.f, survival function, M.G.F., C.G.F., additive property.

3.2(b) Properties of gamma distribution -

(i) Relation between the normal and gamma distributions:
If X follows N(0, 1) then X^{2} = follows G(½, ½)

(ii) 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 .

(iii) 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.

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

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

3.3(c) 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.

STS-203-MJ - Time Series Analysis

Semester – III **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course outcomes:

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

1) Understand the concept of time series and its applications.
2) Identify various components of time series and deal with real life situations.
3) Forecast using exponential smoothing.
4) Understand the concept of stationarity and Box-Jenkins technique for fitting ARMA models.
5) Analyze time series data.

Units and Contents

Unit 1: Introduction [14 Hours]

1.1 Examples of time series, Objectives of time series, Meaning and utility of time series. Simple time series models viz. (i) Some Zero-Mean models (ii) Models with trend and seasonality.

1.2 Components of time series: trend, seasonal variations, cyclical variations, irregular (error) fluctuations or noise.

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

1.4 Methods of trend estimation: (i) moving average, (ii) linear, parabolic, exponential, Pareto curve fitting by least squares principle (iii) exponential smoothing.

1.5 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.

1.6 De- trending and de- seasonalizing a series, Analysis of residuals

1.7 Analysis of Irregular components to examine whether the effect of trend, seasonality is removed

1.8 Case studies of real-life Time Series: Price index series, share price index series, economic time series: temperature and rainfall time series, wind speed time series, pollution levels.

Unit 2: Forecasting techniques using Exponential smoothing [5 Hours]

2.1 Forecasting based on exponential smoothing. Estimating mean square error (M.S.E.) of forecasting. Double exponential smoothing i.e, Holt-Winters method.

2.2 Choosing constants for smoothing and forecasting.

2.3 Forecasts based on AR(p) and MA(q) models, MSE of forecasts [under the assumption that parameters are estimated].

Unit 3: Stationary Processes **[5 Hours]**

3.1 Introduction to weak and strong stationarity of time series and their properties.

3.2 Introduction to ARMA Processes, properties, examples, properties of sample mean and autocorrelation function.

3.3 General approach to time series modelling. Autocorrelation, Autocovariance Function and Stationary Models.

Unit 4: Introduction to Box Jenkins Techniques **[6 Hours]**

4.1 Study of weak stationarity through plots involving differencing and seasonal differencing concepts, consequences of over differencing

4.2 Transformation of data: Transformation and differencing, AR(p) and MA(q) models.

4.3 Fitting of autoregressive model AR(p) model, p=1,2.

Reference Books:

1. Brockwell, P.J. and Davis, R.A. (2002). Introduction to time series and forecasting. New York, NY: Springer New York.
2. Christopher Chatfield (1975): The Analysis of Time Series, 6th edition, CRC Press.
3. Farmum, N.R. and Stantorr, L.W. (1989): Quantitative Forecasting Methods, PWS Kent Publishing Company, Boston.
4. Montgomery, D.C. and Johnson L.A. (1976): Forecasting and Time Series Analysis, McGraw Hill.
5. Mukhopadhyay, P (2011): Applied Statistics, 2nd edition revised reprint, Books and Allied (P) Ltd.

STS-204-MJP – Statistics Practical – III

Semester – III **Course type: Practical**

No. of Credits: 2 **No. of Contact Hours: 60**

Course Outcomes:

At the end of this course, students will be able to:

1) 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) Apply suitable truncated probability distribution in real life situations.
5) Estimate trend and seasonal variations components of time series.
6) Fit AR(1) model to the data.

List of practicals

1. Fitting of negative binomial distribution and computation of expected frequencies.
2. Fitting of normal distribution and computation of expected frequencies.
3. Generating random samples from exponential distribution using distribution function and from gamma(α, λ) distribution using interrelation with exponential distribution.
4. Generating random samples from normal distribution using (i) distribution function (ii) Box-Muller transformation.
5. Applications of normal distribution.
6. Applications of negative binomial and multinomial distributions.
7. Applications of an exponential distributions.
8. Applications of truncated normal distribution.
9. Applications of truncated binomial and truncated Poisson distributions
10. Time series: Estimation of trend by moving averages and regression approach, plotting of residuals.
11. Time series: Estimation of seasonal variations by ratio to trend and ratio to moving average.
12. Time series: Fitting of AR(1) model, exponential smoothing and double exponential smoothing.

MINOR

STS-241-MN – Fundamentals of Statistics

Semester – III **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course Outcomes:

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

1. Identify the appropriate discrete probability distribution for a given scenario.
2. Compute probabilities of events corresponding to standard discrete and continuous probability distributions.
3. Understand about continuous univariate random variables.
4. Gain knowledge of some standard continuous distributions and their properties.

Units and Contents

Unit 1: Standard Discrete Probability Distributions [10 hours]

1.1 Poisson distribution-p.m.f., mean, variance, mode, real life situations where it is applicable, statement of additivity property, statement of limiting case of B(n, p).

1.2 Geometric distribution-p.m.f., mean, variance, mode, real life situations where it is applicable. Statement of lack of memory property.

1.3 Negative binomial distribution-p.m.f., mean, variance, real life situations where it is applicable. Relation with geometric distribution, concept of waiting time distribution.

1.4 Numerical problems.

Unit 2: Continuous Univariate Distributions [5 hours]

2.1 Definition of ccontinuous random variable, probability density function (p.d.f.).

2.2 Cumulative distribution function (c.d.f.) and its properties.

2.3 Mathematical expectation, variance, quartiles and quartile deviation.

2.4 Numerical problems.

Unit 3: Standard Continuous Probability Distributions [10 hours]

3.1 Uniform distribution- statement of p.d.f., c.d.f., mean, variance, nature of p.d.f. curve

3.2 Normal distribution- statement of p.d.f., c.d.f., mean, variance, median, mode, nature of p.d.f. curve, statement of additive property, skewness. Standard normal distribution, statement of Central Limit Theorem (CLT).

3.3 Exponential distribution- statement of p.d.f., c.d.f., mean, variance, nature of p.d.f.

curve, statement of lack of memory property.

3.4 Numerical problems.

Unit 4: Correlation and Regression **[5 hours]**

4.1 Bivariate data, Scatter diagram. Concept of correlation between two variables. Types of correlation.

4.2 Karl Pearson's coefficient of correlation (r), limits of r, computation for ungrouped data with interpretation.

4.3 Meaning of regression, connection between correlation and regression.

4.4 Linear regression, fitting of straight line using least square method, concept of coefficient of determination.

4.5 Non-linear regression models: second degree curve.

4.6 Numerical problems.

Reference books:

1. Brase C. H., Brace C. P (2016). Understandable Statistics, Concepts and methods, 12th Edition, Cengage Learning.
2. Freedman D., Pisani R., Purves R. (2007). Statistics, 4th Edition, W. W. Norton and Company.
3. Gupta, S. C. and Kapoor, V. K. (2002). Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, New Delhi.
4. Gupta, S. P. (2002), Statistical Methods (Thirty First Edition), Sultan Chand and Sons, New Delhi.
5. 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.
6. Weiss N. (2015), Introductory Statistics, Pearson education publishers.

STS-242-MNP – Practicals based on Fundamentals of Statistics and R - Software

Semester – III **Course type: Practicals**

No. of Credits: 2 **No. of Contact Hours: 60**

Course Outcomes:

At the end of this course, students will be able to:

1. Compute summary statistics.
2. Apply standard discrete and continuous probability distributions to solve real life problems.
3. Use R – software to find the summary statistics.
4. Compute probabilities of the distributions using R – software.
5. Represent statistical data by diagrammatically and graphically using R-software

List of Practicals:

Following practicals are to be performed using scientific calculator:

1. Summary statistics (ungrouped data).
2. Summary statistics (grouped data).
3. Applications of binomial and Poisson distribution.
4. Applications of geometric and negative binomial distributions.
5. Applications of normal and exponential distribution.
6. Scatter diagram, computation of correlation coefficient (ungrouped data). Fitting of line of regression.

Following practicals are to be performed using R – software.

7. Introduction to R – software: Installation, Use of basic commands, accessing resident data sets.
8. Diagrammatic representation of data (pie chart, simple, sub-divided).
9. Graphical representation of data (histogram, boxplot, spike, scatter).
10. Finding summary statistics using summary( ). Calculate arithmetic mean (A.M.), median, variance (ungrouped data). Correlation and line of regression.
11. Computation of probabilities for binomial, Poisson, geometric, negative binomial, normal and exponential distributions.

GENERIC / OPEN ELECTIVE (GE / OE)

OE-201-STS – Survey Analysis

Semester – III **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course Outcomes:

At the end of this course, students will be able:

1. To identify the appropriate data type suitable for an intended survey.
2. To identify the most suited sampling method for an intended survey.
3. To anticipate difficulties/problems in data collection and take proactive measures to resolve them.
4. To create survey forms using various available platforms.
5. To compute coefficients used to assess internal consistency of collected data.
6. To effectively represent collected data through diagrams and/or graphs.

Units and Contents

Unit 1: Types of Data: (3 hours)

1.1: Types of characteristics: Attributes and Variables, Measurement scale for attributes: Nominal and Ordinal scale, Measurement scale for variables: Ratio and Interval scale, Likert scale. (all with relevant illustrations).

1.2: Types of data: Primary data, Secondary data, Cross sectional data, Time series data. (all with relevant illustrations).

Unit 2: Methods of Sampling (5 hours)

2.1 Statistical population, Finite population, Infinite population, homogeneous population, heterogeneous population, Concept of sample.

2.2 Random sampling: Simple random sampling with and without replacement, Stratified sampling, Systematic sampling, Two-stage sampling, Non-random sampling: Purposive sampling, Snowball sampling, Convenience sampling, Crowdsourcing sampling. (only descriptions and illustrations of all methods).

Unit 3: Questionnaires, Errors and Consistency: (14 hours)

3.1 Characteristics of a good questionnaire, Problems faced in data collection: problem of non-response, sampling errors, non-sampling errors, validity (internal consistency) using Cronbach's alpha and Kuder Richardson's coefficient (KR-20).

3.2 Designing questionnaires for surveys (i.e. survey forms) (with various sections if required) using various online platforms such as Google Forms, Survey Monkey, Form Façade (an add-on to customize Google Forms), etc., Aesthetic presentation of survey forms, Flow of questions.

3.3: Processing Survey Data: Downloading responses as MS-Excel sheets, computation of

summated scores (for Likert scale based questions), coding data, computation of Cronbach's alpha and Kuder Richardson's coefficient (KR-20).

Unit 4: Basic Exploratory Analysis: (8 hours)

4.1 Selection of appropriate diagram/graph type

4.2 Construction of diagrams (simple bar, subdivided bar, multiple bar, pie), graphs (histogram, boxplot, etc.)

4.3 Use of MS-Excel to carry out basic exploratory analysis of data collected via survey forms.

Reference books:

1. Groves R. M. and et.al. (2009), Survey Methodology, Second Edition, Wiley Series in Survey Methodology.
2. Mukhopadhyay P. (2008), Theory and Methods of Survey Sampling, Prentice-Hall India, New Delhi.
3. Murthy M. N., (1977), Sampling Theory and Methods, Statistical Publishing Society, Kolkata.
4. Singh, D. and Chaudhary, F.S., (1986). Theory and analysis of sample survey designs.
5. Sukhatme P. V., Sukhatme B. V. (1984), Sampling Theory of Surveys with Applications, Indian Society of Agricultural Statistics, New Delhi.

VOCATIONAL SKILL COURSE

STS-221-VSC – Data Analysis with R – Software

Semester – III **Course type: Practical**

No. of Credits: 2 **No. of Contact Hours: 60**

Course Outcomes:

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

1. Install and Use R-software for basic Statistical calculations.
2. Represent the data diagrammatically or graphically in R software.
3. Calculate the central tendency and dispersion, skewness and kurtosis for grouped and ungrouped data.
4. Calculate the probabilities and plot the p.m.f./p.d.f and c.d.f. curves of standard discrete and continuous distributions.
5. Fit the regression and time series models to the data using R software.

List of Practicals

1) Use of basic R-software commands.
2) Diagrammatic representation of data.
3) Graphical representation of data.
4) Measures of central tendency and dispersion, skewness and kurtosis (Ungrouped Data).
5) Measures of central tendency and dispersion, skewness and kurtosis (Grouped Data).
6) Computation of probabilities for Standard discrete and continuous distributions.
7) Plot of p.m.f./ p.d.f and c.d.f. curve of standard discrete and continuous distributions.
8) Model sampling from standard discrete and continuous distributions.
9) Scatter plot and Correlation. Simple linear regression.
10) Fitting of trivariate regression.
11) Time Series Analysis -I.
12) Time Series Analysis -II.

FIELD PROJECT

STS-231-FP – Field Project

Semester – III

No. of Credits: 2 **No. of Contact Hours: 60**

Guidelines: Field Project

1. For field project maximum 5 students are allowed in a group.
2. **The collection of a primary data through the field work is mandatory.**
3. Copy of the field project report to be made available to the examiner on the first day of practical examination.
4. The following points should be included in the Report/ Dissertation:
   a) Title of the field project, name(s) of the student(s), 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 should cover description of the selected problem using terminology in the field of application, conversion of the problem in statistical language, literature survey, 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.
5. The division of 50 marks for the field project evaluation is,

A) Internal evaluation:	**25 Marks** **(Based on project report and viva)**
B) End semester evaluation	
i) Project report:	**15 Marks**
ii) Viva based on power point presentation:	**10 Marks**
Total marks (A+B):	**50 Marks**

I) Aspects of Assessment and marks assigned (The following are the guidelines; some modification can be done as and when required by the examiners)

Scheme of the marking of project report for end semester examination:

Sr. No.	Aspects in Project Report	Marks
i)	Dissertation in the proper format which includes (Title, abstract, Key words, methodology, conclusion, references, limitations and source of data etc.)	4
ii)	Conversion of the real life problem in statistical language	2
iii)	Appropriateness of tools (preferably learned in syllabus) used for analysis, testing the assumptions needed for analysis, methodology, program coding (if any) and numerical computations.	6
iv)	Conclusions in relevant language	3
Total		**15**

II) Viva of each project group consisting maximum 5 students is conducted using power point presentation only (Time duration: 10 Minutes).

Division of marks for viva as:

Presentation (through slides)	**Marks**
Understating the Project problem	**4**
Understanding of statistical techniques used to solve the problem	**4**
Overall presentation	**2**
Total	**10**

SEMESTER – IV

MAJOR

STS-251-MJ – Statistical Methods

Semester – IV **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course Outcomes:

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

1) Understand various methods of estimation.
2) Identify the appropriate test of hypothesis to be used in a scenario at hand.
3) Infer about the validity of a hypothesis via various approaches.
4) Identify a situation where multiple linear regression can be used.
5) Compute and interpret multiple and partial correlation coefficients.
6) Determine the strength and adequacy of fit of a multiple linear regression model.

Units and Contents

Unit 1: Methods of Estimation **[10 Hours]**

1.1 Statistic and Parameter: Random sample $X_1, X_2, \ldots, X_n$ 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 **[10 Hours]**

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 Hours]

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.

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 Property of multiple correlation coefficient (with proof), $0 \leq R_{Y.X_1X_2} \leq 1$.

3.7 Coefficient of multiple determination $\left(R^2_{Y.X_1X_2}\right)$, Interpretation in terms of proportion of variation explained in linear regression.

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 $r_{YX_1.X_2}$ and $r_{YX_2.X_1}$ in terms of total correlation coefficients. Properties of partial regression coefficients (with proof)

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

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

Reference Books:

1. Buyan, K. C. (2010). Probability theory and Statistical inference,1st Edn., New Central Book Agency.
2. Chatterjee S. and Hadi A.S. (2012), Regression Analysis by Example, Fifth Edition,

Wiley.

3. Daniel, W.W. (2000) Applied Nonparametric Statistics, Duxbury Press Boston.
4. Draper, N. R. and Smith, H. (1998), Applied Regression analysis, (John Wiley) Third Edition.
5. Gibbons J.D. and Chakraborti, S. (2010). Non parametric Statistical Inference, CRC Press, London
6. Goon A. M., Gupta, M. K. and Dasgupta, B. (1986), Fundamentals of Statistics, Vol. 2, World Press, Kolkata.
7. Gross D., Shortle J. F., Thompson J. M., Harris C. M. (2012), Fundamentals of Queuing Theory (Fourth Edition), Wiley Series in Probability and Statistics.
8. Gupta, S. C. and Kapoor V. K. (2007), Fundamentals of Applied Statistics (Fourth Edition), Sultan Chand and Sons, New Delhi.
9. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002.
10. Gupta, S. P. (2002), Statistical Methods (Thirty First Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi 110002.
11. Kulkarni, M. B., Ghatpande, S. B. and Gore, S. D. (1999), Common Statistical Tests, Satyajeet Prakashan, Pune 411029.
12. Mishra Amarendra (2020), Theory of Statistical Hypothesis Testing (First Edition), Notion Press.
13. Montgomery, D.C., Peck, E. A. and Vining, G.G.(2003), Introduction to Linear Regression Analysis, Wiley.
14. Taff Arthur (2018), Hypothesis Testing: The Ultimate Beginner's Guide to Statistical Significance, CreateSpace Independent Publishing Platform.
15. Wayne W. Daniel (2006), Biostatistics: A Foundation for Analysis in Health Sciences, Seventh edition, Wiley India Pvt. Ltd.
16. Wilson J. H., Keating B. P., Beal-Hodges M. (2012), Regression Analysis, Business Expert Press.

STS-252-MJ – Sampling Distributions and Exact Tests

Semester – IV **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course Outcomes:

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

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

Units and Contents

Unit 1: Chi-square Distribution [6 Hours]

1.1 Definition of Chi-square r.v. as a sum of squares of i.i.d. standard normal variates, derivation of the p.d.f. of Chi-square variable with n degrees of freedom (d.f.) using M.G.F. Notation: $X \sim \chi^2_n$. 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 Numerical problems.

Unit 2: Student's – t distribution [5 Hours]

2.1 Definition of $t = \frac{U}{\sqrt{\frac{V}{n}}}$ r.v. with 'n' d.f. 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.

2.4 Numerical problems.

Unit 3: Snedecor's F – distribution [5 Hours]

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 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.

3.4 Numerical problems.

Unit 4: Sampling distributions **[4 Hours]**

4.1 Random sample from a distribution as i.i.d. r.v.s

4.2 Notion of a statistic as function of $X_1, X_2, \ldots, X_n$ with illustrations

4.3 Sampling distribution of a statistic, distribution of sample mean from normal, exponential and gamma distribution, notion of standard error of a statistic.

4.4 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 5: Tests of Hypothesis based on sampling distributions **[10 Hours]**

5.1 Tests based on chi-square distribution:

(a) Test for independence of two attributes arranged in $r \times s$ contingency table (with Yate's correction)

(b) Mc Nemar's test

(c) Test for goodness of fit.

(d) 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.

5.2 Tests based on t - distribution:

(e) 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.

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

(g) Test for significance of population correlation coefficient.

5.3 Test based on F-distribution:

(h) 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. 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. (2007), Fundamentals of Applied Statistics (Fourth Edition), Sultan Chand and Sons, New Delhi.
4. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002 .
5. Mishra Amarendra (2020), Theory of Statistical Hypothesis Testing (First Edition), Notion Press.
6. Taff Arthur (2018), Hypothesis Testing: The Ultimate Beginner's Guide to Statistical Significance, CreateSpace Independent Publishing Platform.
7. Wayne W. Daniel (2006), Biostatistics: A Foundation for Analysis in Health Sciences, 7^{th} edition, Wiley India Pvt. Ltd.

STS-253-MJ – Statistical Process and Product Control

Semester – IV **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course Outcomes:

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

1) Understand various tools of Statistical quality control and its purpose.
2) Construct control charts for range, sample mean, p-chart and c-chart.
3) Various terms related to control charts.
4) Develop single and double sampling plans.

Units and Contents

Unit 1: Introduction to Statistical Quality Control (SQC) (5 Hours)

1.1 Meaning of quality, total quality management (TQM), six-sigma, ISI, ISO,

1.2 Meaning and purpose of Statistical Quality Control (SQC), Seven Process Control

1.3Tools of Statistical Quality Control (SQC) (i) Check Sheet, (ii) Cause and effect diagram (CED), (iii) Pareto Diagram, (iv) Histogram, (v) Control chart, (vi)Scatter Diagram,(vii) Design of Experiments (DOE). (Only introduction of 7 PC tools is expected).

Unit 2: Control Charts (4 Hours)

2.1 Chance causes and assignable causes of variation, statistical basis of control charts, exact probability limits, k-sigma limits, justification for the use of 3- sigma limits for normal distribution and using Chebychev's inequality for non-normal distributions.

2.2 Criteria for detecting lack of control situations:

2.3 At least one point outside the control limits

2.4 A run of seven or more points above or below central line.

2.5 Presence of a non random pattern eg. cycle or linear trends etc.

2.6 Control chart technique as hypotheses testing problem.

2.7 Construction of control charts for (i) standards given, (ii) standards not given.

Unit 3: Control charts for variables and attributes (12 Hours)

3.1 Introduction of R chart and $\bar{X}$ chart

a) Construction of R chart when the process standard deviation is specified: control limits, drawing of control chart, plotting of sample ranges, drawing conclusion-determination of state of control process, corrective action if the process is out of statistical control.

Construction of $\bar{X}$ chart when the process average is specified: control limits, drawing of control chart, plotting of sample means. Drawing conclusion - determination of state of control of process, corrective action if the process is out of statistical control.

b) Construction of R chart when the process standard deviation (σ) is not given:

control limits, drawing of control chart, plotting sample range values, revision of control limits if necessary, estimate of σ for future use.

3.2 Construction of $\bar{X}$ chart when the process average (μ) is not given : control limits based $\hat{\mu} = \bar{X}, \hat{\sigma} = \bar{R}d_2$ drawing of control chart, plotting sample means, revision of control

limits of $\bar{X}$ chart, if necessary.

3.3 p- chart: a) Construction and working of p-chart when subgroup sizes are same and value of the process fraction defective P is specified: control limits, drawing of control chart, plotting of sample fraction defectives. Determination of state of control of the process.
b) p-chart when subgroups sizes are different and value of the process fraction defective P is not specified with i) separate control limits ii) average sample size iii) standardized control chart. Drawing of control chart, plotting sample fraction defectives, determination of state of control of the process. Interpretation of high and low spots. Identification of real life situations. Probability of catching a shift.

3.4 C - Chart: a) Construction of c-chart when standard is given; control limits justification of 3 sigma limits, drawing of control chart, plotting number of defects per unit. (b) Construction of c chart when standard is not given; control limits, explanation for the use of 3-sigma limits, drawing of control chart. Plotting number of defects per unit. Determination of state of control, interpretation of high and low spots in above cases. Identification of real life situations.

Unit 4: Sampling Plans. (9 Hours)

4.1 Acceptance Sampling for Attributes: Introduction, Concept of sampling inspection plan, Comparison between 100% inspection and sampling inspection. Procedures of acceptance sampling with rectification, Single sampling plan and double sampling plan. Probabilities of acceptance and rejection

4.2 Explanation of the terms: Producer's risk and Consumer's risk, Operating characteristic (OC) curve, Acceptable Quality Level (AQL), Lot Tolerance Fraction Defective (LTFD) and Lot Tolerance Percent Defective (LTPD), Average Outgoing Quality (AOQ) and Average Outgoing Quality Limit (AOQL), AOQ curve, Average Sample Number (ASN), Average Total Inspection (ATI)

4.3 Single Sampling Plan: Computation of probability of acceptance using Poisson approximation, Derivation of AOQ and ATI. Graphical determination of AOQL, Determination of a single sampling plan by: a) lot quality approach b) average quality approach.

4.4 Double Sampling Plan: Evaluation of probability of acceptance using Poisson distribution, Structure of OC Curve, Derivation of AOQ, ASN and ATI (with complete inspection of second sample), Graphical determination of AOQL, Comparison of single sampling plan and double sampling plan.

Reference Books:

1. Besterfield, D.H. and Michna , C.B. et al. (2009). Total Quality Management, 3rd edition,Pearson Education, Delhi.34
2. Dodge, H.F. and Roming, H.G. Sampling Inspection tables, John Wiley and Sons, Inc. New York
3. Duncan A.J. (1974). Quality Control and Industrial Statistics, fourth edition D.B. TaraporewalaSons and Co. Pvt. Ltd., Mumbai.

4. Grant, E. L. and Leavenworth (1980). Statistical Quality Control, fifth edition, Mc-Graw Hill,New Delhi.
5. Johnson, N.L. and Kotz, S. (1993). Capability Studies, Chapman and Hall Publishers.
6. Kamji and Asher (1996). 100 Methods of TQM, Sage Publishers, Delhi
7. Montgomery, D. C. (2008). Statistical Quality Control, 6thEdn., John Wiley, New York.
8. SP20 : Handbook of SQC, Bureau of Indian Standards.

STS-254-MJP – Statistics Practical – IV

Semester – IV **Course type: Practical**

No. of Credits: 2 **No. of Contact Hours: 60**

Course Outcomes:

At the end of this course students are expected to be able to:

1) Test the significance of mean, proportions, attributes and variance for the sample.
2) Test significance of correlation using t-test.
3) Construct various control charts for variables and attributes and take the decision about the process.
4) Determine single and double sampling plans for attributes.

List of practicals:

1. Test for proportions (one sample problem and two sample problem) and construction of confidence interval. Verification of result using p-value.
2. Test for means (one sample problem and two sample problem) and construction of confidence interval (large sample test). Verification of result using p-value.
3. Test for means (one sample problem) and construction of confidence interval (small sample test). (for one-sided and two-sided alternatives).
4. Test for means (two sample problem) and construction of confidence interval (small sample test) (for one-sided and two-sided alternatives).
5. Paired t-test (for one-sided and two-sided alternatives) and significance of correlation using t-test.
6. Test for independence of attributes (2 x 2 and r x s contingency table) and McNemar's test.
7. Test for goodness of fit using chi-square distribution.
8. Tests for population variance (one sample and two sample problem) (for one-sided and two-sided alternatives).
9. Construction of R and $\bar{X}$ control chart. Probability of catching shift.
10. p - chart for (i) fixed sample size (ii) variable sample size based on individual control limits, probability of detecting shift.
11. Single sampling plan for attributes (OC Curve, Producer's and Consumer's risks, AOQ, AOQL, ATI).
12. Double sampling plan for attributes (OC curve, AOQ, AOQL, ATI, ASN using Poisson distribution).

MINOR

STS-291-MN – Statistical Inference

Semester – IV **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course Outcomes:

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

1) Understand the concept of testing of hypothesis.
2) Test the hypothesis for population means and population proportions for large sample.
3) Understand the concept of sampling distributions and its various applications in real life.
4) Test the independence of two attributes using chi-square test.
5) Test the population mean and population variance.

Units and Contents

Unit 1: Testing of Hypothesis **[12 Hours]**

1.1 Notion of statistic, notion of a standard error of a statistic, parameter, parameter space, statistical hypothesis, null and alternative hypotheses, one-sided and two-sided alternative hypotheses, critical region, type-I and type-II error, level of significance, p-value, Two-sided confidence interval.

1.2 Testing of hypotheses using critical region and p-value approach.

1.3 Tests for population proportions- one sample and two sample problem, confidence interval for population mean.

1.4 Tests for population means (large sample) one sample and two sample confidence interval for population mean.

1.5 Numerical problems.

Unit 2: Sampling Distributions **[6 Hours]**

2.1 Chi-square distribution – statement of p.d.f., mean and variance, applications of the distribution, use of chi-square tables for calculations of probabilities.

2.2. Student's t distribution - statement of p.d.f, mean and variance. applications of the distribution. use of t-tables for calculations of probabilities.

2.3 Snedecor's F distribution - statement of p.d.f,, mean and variance. applications of F distribution, use of F - tables for calculation of probabilities.

Unit 3: Tests of Hypothesis based on sampling distributions [12 Hours]

3.1 Test for independence of two attributes arranged in $r \times s$ contingency table.

3.2 Test for goodness of fit.

3.3 Tests for population means: a) one sample and two sample b) two-sided confidence interval for population mean.

3.4 Paired t-test. Significance of correlation test.

3.5 Test for equalities of variance when i) means are known and ii) means are unknown.

3.6 Examples.

Reference Books:

1. Buyan, K. C. (2010). Probability theory and Statistical inference,1stEdn., New Central Book Agency.
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. Kulkarni, M. B., Ghatpande, S. B. and Gore, S. D. (1999), Common Statistical Tests, Satyajeet Prakashan, Pune 411029.
4. Mishra Amarendra (2020), Theory of Statistical Hypothesis Testing (First Edition), Notion Press.
5. Taff Arthur (2018), Hypothesis Testing: The Ultimate Beginner's Guide to Statistical Significance, CreateSpace Independent Publishing Platform.

STS-292-MNP – Practicals based on Statistical Inference

Semester – IV **Course type: Practical**

No. of Credits: 2 **No. of Contact Hours: 60**

Course Outcomes:

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

1) Test the hypothesis for population means and population proportions for large sample.
2) Test the independence of two attributes using chi-square test.
3) Test the population mean and population variance for small sample problem.
4) Use of R - software to test the hypothesis for large sample and small sample problem.

List of practicals

Following practicals to be performed using scientific calculator:

1. Test for proportions (one sample problem and two sample problem)
2. Test for means (large sample test).
3. Test for means (one sample problem) (small sample test).
4. Test for means (two sample problem) (small sample test).
5. Paired t-test.
6. Test for independence of attributes and goodness of fit.
7. Tests for population variance (two sample problem only).

Following practicals to be performed using R-software:

8. Test for proportions.
9. Test for means (small sample test).
10. Paired t-test and significance of correlation
11. Test for independence of attributes.
12. Tests for population variance.

GENERIC / OPEN ELECTIVE (GE / OE)

OE-251-STS – Basic Applied Statistics

Semester – IV **Course type: Theory**

No. of Credits: 2 **No. of Contact Hours: 30**

Course Outcomes:

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

1) Understand meaning and scope of index numbers.
2) Construct index number by considering various considerations.
3) Compute mortality, fertility and reproduction rates for the population.
4) Interpret and use different rates in daily life.
5) Assess relationship between two attributes.

Units and Contents

Unit 1: Index Number **[10 Hours]**

1.1 Introduction, definition, meaning and scope of Index Numbers.

1.2 Problems/ Considerations in the construction of Index Numbers.

1.3 Simple and weighted Price Index Numbers based on price relatives

1.4 Laspeyre's, Paasche's and Fisher's Index Numbers.

1.5 Cost of living index number

1.6 Shifting of base and splicing

1.7 Description of the index numbers: BSE and SENSEX

1.8 Numerical Problems

Unit 2: Demography **[13 Hours]**

2.1 Vital events, vital statistics, methods of obtaining vital statistics, rates of vital events, sex ratios, dependency ratio.

2.2 Death/Mortality rates: Crude death rates, specific (age, sex etc.) death rate, standardized death rate (direct and indirect), infant mortality rate.

2.3 Birth/ Fertility rate: Crude birth rates, general fertility rate, specific (age, sex etc.) fertility rates, total fertility rates.

2.4 Growth/Reproduction rates: Gross reproduction rate, net reproduction rate.

2.5 Interpretations of different rates, uses and applications.

2.6 Numerical Problems

Unit 3: Theory of Attributes [7 Hours]

3.1 Likert Scale, Notion of classification - dichotomous and manifold, class frequencies and their categories

3.2 Method of dot operator for relationships between class frequencies (two attributes),

3.3 Consistency of data (up to two attributes)

3.4 Concept of independence and association of two attributes

3.5 Yule's coefficient of association (Q) and its interpretation, range of Q.

Reference Books:

1) Freedman D., Pisani R., Purves R. (2007). Statistics, 4th Edition, W. W. Norton and Company.
2) Freund J. E. (1977). Modern Elementary Statistics. 4th Edition, Prentice Hall of India Private Limited, New Delhi.
3) Goon A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1. 6th Revised Edition, The World Press Pvt. Ltd., Calcutta.
4) Gupta S. C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics. 8th Edition, Sultan Chand and Sons Publishers, New Delhi.
5) Gupta S. C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, 3rd Edition, Sultan Chand and Sons Publishers, New Delhi.
6) Utts J. M., Heckard R. F. (2010). Mind On Statistics, 4th Edition, Richard Stratton Publisher.

SKILL ENHANCEMENT COURSE

SEC-251-STS – INTRODUCTION TO PYTHON

Semester – IV **Course type: Practical**

No. of Credits: 2 **No. of Contact Hours: 60**

Course Outcomes:

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

1) Work comfortably with Python. Execute basic actions in Python.
2) Use appropriate built-in functions for required actions.
3) Import libraries and use them in a Python code.
4) Define and use functions that are user-defined, suited to the problem at hand.

List of Practicals:

1) Basics
2) Conditions and Loops
3) Strings.
4) Data Visualization - I.
5) Data Visualization - II.
6) Python for Descriptive Statistics - I.
7) Python for Descriptive Statistics - II
8) Python for Probability Distributions-I.
9) Python for Probability Distributions-II.
10) User defined Functions.
11) Modules – I
12) Modules - II

STS-281-CEP : COMMUNITY ENGAGEMENT PRACTICE

Semester – IV **Course type: Practical**

No. of Credits: 2 **No. of Contact Hours: 60**

GUIDELINES –

1. Community engagement activity shall be carried out group wise (minimum 5 students in 1 group).
2. Every student should devote minimum 4 hrs. per week for the community engagement activities.
3. Fortnightly report of community engagement activities should be submitted to the concerned teacher. Report should be validated by the concerned authority.
4. CIA will be based on fortnightly reports submitted.
5. ESE will be based on final report submission and presentation.