Deccan Education Society's
FERGUSSON COLLEGE, PUNE
(AUTONOMOUS)

SYLLABUS UNDER AUTONOMY

THIRD YEAR B.Sc. (Statistics)

SEMESTER –V

SYLLABUS FOR T.Y. B.Sc.
Academic Year 2018-2019

T.Y. B.Sc. (Statistics) Semester V
Statistics Paper -1 (STS3501)
Distribution Theory

[Credits-3]

The main objective of this course is to acquaint students with the following.

1.Continuous distributions.
2. Fitting of various continuous probability distributions and to study various real life situations.
3. Identification of the appropriate probability model that can be used.

Unit I		Beta distribution	(9L)
	1.1	**Beta distribution of first kind** $f(x) = \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1}, 0 \le x \le 1, \quad m, n > 0$ $= 0 \qquad$ otherwise Notation: X~ β_1 (m,n) Nature of probability curve, Derivation of mean, variance, r^{th} raw moment, harmonic mean, mode. Symmetry of the distribution. Relation with U (0, 1). Probability distributions of $\frac{1}{X}, X+Y, X-Y, XY, \frac{X}{Y},$ where X and Y are iid β_1 (1,1). Statement of relation between distribution function of β_1(m, n) and binomial distribution. Application of these models for real life data.	
	1.2	**Beta distribution of second kind** $f(x) = \frac{1}{\beta(m,n)} \frac{(x)^{m-1}}{(1+x)^{m+n}}, x \ge 0, \quad m, n > 0$ $= 0 \qquad$ otherwise Notation: X~ β_2(m,n). Nature of probability curve , Derivation of mean, variance, r^{th} raw moment, harmonic mean, mode . Derivation of interrelation between β_1(m, n) and β_2(m,n). Derivation of distribution of $\frac{X}{Y}, \frac{X}{(X+Y)}$ when X and Y are independent gamma variates. Application of this models for real life data.	

Unit II		**Weibull Distribution**	**(5 L)**
	2.1	$f(x)=\frac{\beta}{\alpha}\left(\frac{x-\gamma}{\alpha}\right)^{(\beta-1)}\exp\left\{-\left(\frac{x-\gamma}{\alpha}\right)^{\beta}\right\}$, $x\geq\gamma,\quad \alpha,\beta>0$ $=0$ otherwise. Notation: X~W(γ, α, β).	
	2.2	Probability curve, location parameter, shape parameter, scale parameter. Derivation of distribution function, quartiles, mean and variance, coefficient of variation, relationship with gamma and exponential distribution. Application of this model for real life data.	
Unit III		**Cauchy Distribution**	**(6L)**
	3.1	$f(x)=\frac{\lambda}{\pi}\frac{1}{\lambda^2+(x-\mu)^2}$, $-\infty<x<\infty, -\infty<\mu<\infty, \lambda>0.$ Notation: X~C(μ,λ).	
	3.2	Nature of the probability curve, comparison with tails of normal distribution.	
	3.3	Derivation of distribution function, quartiles. Non – existence of moments, Statement of distribution of aX+b. Derivation of distribution of i) $\frac{1}{X}$ ii) X^2 ,where X ~ C (0,1), Problems based on these results.	
	3.4	Statement of additive property for two Independent Cauchy variates, statement of distribution of the sample mean, comment on limiting distribution of $\overline{X}$.	
	3.5	Statement of relationship with uniform, student's t and normal distributions. Application of this model for real life data.	

Unit IV		**Laplace (Double Exponential) Distribution**	**(6 L)**
	4.1	$f(x) = \frac{\lambda}{2} \exp(-\lambda \mid x - \mu \mid), \quad -\infty < x < \infty, -\infty < \mu < \infty, \lambda > 0.$ Notation: $X \sim L(\mu, \lambda)$ Nature of probability curve.	
	4.2	Derivation of distribution function, quartiles.	
	4.3	MGF, CGF, Moments and cumulants, skewness and kurtosis.	
	4.4	Derivation of Laplace distribution as the distribution of the difference of two i.i.d. exponential random variables with mean $\frac{1}{\lambda}$ Application of this models for real life data.	
Unit V		**Lognormal Distribution**	**(6 L)**
	5.1	$f(x) = \frac{1}{(x-a)\sigma\sqrt{2\pi}} \exp\{\frac{-1}{2\sigma^2}[\log_e(x-a) - \mu]^2, x > a,$ $, -\infty < \mu < \infty, \sigma > 0.$ $= 0$ otherwise.. Notation : $X \sim LN(a, \mu, \sigma^2)$	
	5.2	Derivation of relation with $N(\mu, \sigma^2)$ distribution.	
	5.3	Nature of the probability curve.	
	5.4	Derivation of moments (r^{th} moment of X), mean, variance, Karl Pearson's and Bowley's coefficient of skewness. Coefficient of kurtosis, derivation of quartiles and mode.	
	5.5	Distribution of $\underset{i=1}{\overset{n}{\pi}} X_i$ when Xi's are independent lognormal random variables. Application of this model for real life data.	
Unit VI		**Bivariate Normal Distribution.**	**(10L)**
	6.1	$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{\frac{-1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2 + (\frac{y-\mu_2}{\sigma_2})^2 - 2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_1})\}$	

		$-\infty < x, y < \infty, -\infty < \mu_1, \mu_2 < \infty, \quad \sigma_1, \sigma_2 > 0, -1 < \rho < 1.$ $= 0$ otherwise. Notation : $(X , Y) \sim BN(\mu_1 , \mu_2, \sigma_1^2 , \sigma_2^2, \rho)$.	
	6.2	Nature of surface of p. d. f., marginal and conditional distributions , identification of parameters , regression of Y on X, independence and un-correlatedness, Derivation of MGF and moments. Statement of distribution of aX+ bY+ c , $\frac{X}{Y}$. Application of this models for real life data.	
Unit VII		**Order statistics**	**(6 L)**
		Order Statistics for a random sample of size n from a continuous distribution, definition, derivation of distribution function and density function of the i^{th} order statistics $X_{(i)}$, particular cases for i = 1 and i = n. Joint distribution of $(X_{(i)}, X_{(j)})$ for a random sample from uniform and exponential distribution. Definition of p^{th} sample quantile $X_{[(np]+1)}$. Distribution of sample median for a random sample from uniform distribution.	
		References: 1. Hogg, R.V. and Craig A.T.(1970). Introduction to Mathematical Stastistics, III rd Editior, Macmillan Publishing Company .Inc. New York. 2. Lindgren B.W.: (1976) Statistical Theory III rd Edition Collier Macmillan international Edition, Macmillan Publishing Co. Inc. New York. 3. Mood. A.M. , Graybill , F.Bose ,D.C.: (1974) Introduction to theory of Statistics. III rd Edition Mc- Graw Hill Series. 4. Mukhopdhyay, P (1996). Mathematical Statistics, New Central Book Agency. 5. Rohatgi , V.K. (1975) An Introduction to probability Theory and Mathematical Statistics Wiley Eastern Ltd .New Delhi. 6. Casella G. and Berger Robert L. (2002) Statistical Inference. 2nd Edition, Duxbury Advanced series. 7. Dasgupta A. (2010) Fundamentals of Probability : A first course, , Springer, New York.	

****

T.Y. B.Sc. (Statistics) Semester V
Statistics Paper - 2 (STS3502)
Theory of Estimation

[Credits-3]

The main objective of this course is to introduce students to

1. Notion of parameter and estimator.
2. Learn and apply various methods of estimation.
3. Understand concept of point estimation and interval estimation.

Unit I		**Point Estimation**	**(4L)**
	1.1	Notion of a parameter, parameter space, sample space as a set of all possible values of $(X_1, X_2, \ldots, X_n)$, general problem of estimating an unknown parameter by point and interval estimation.	
	1.2	Point Estimation: Definition of an estimator, distinction between estimator and estimate, illustrative examples.	
	1.3	Mean Square Error (MSE) of an estimator.	
Unit II		**Methods of Estimation**	**(10L)**
	2.1	Method of moments: Derivation of moment estimators for standard distributions.	
	2.2	Definition of likelihood as a function of unknown parameter, for a random sample from i) discrete distributions ii) continuous distributions, distinction between likelihood function and p.d.f./ p.m.f.	
	2.3	Method of maximum likelihood: Derivation of maximum likelihood estimator (M.L.E.) for parameters of only standard distributions (case of two unknown parameters only for normal distribution). Use of iterative procedure to derive M.L.E. of location parameter μ of Cauchy distribution. Invariance property of M.L.E.	
	2.4	M.L.E. of θ in uniform distribution over i) $(0, \theta)$ ii) $(-\theta, \theta)$ iii) $(m\theta, n\theta)(m<n)$ M.L.E. of θ in $f(x ; \theta)= \exp\{-(x-\theta)\}, x > \theta$. M.L.E. of location parameter in Laplace distribution. Illustrations of situations where M.L.E. and moment	

		estimators are distinct and their comparison using mean square error.	
Unit III		**Properties of estimators.**	
	3.1	Unbiasedness	**(4L)**
		Definition of an unbiased estimator, Proofs of the following results regarding unbiased estimators: a) Two distinct unbiased estimators of θ give rise to infinitely many estimators. b) If T is an unbiased estimator of θ , then ø (T) is unbiased estimator of ø (θ) provided ø (.) is a linear function.	
	3.2	Variance of the estimator	**(4L)**
		Notion of the Best Linear Unbiased Estimator and uniformly minimum variance unbiased estimator (UMVUE), uniqueness of UMVUE whenever it exists. Illustrations.	
	3.3	Sufficiency	**(7L)**
		Concept and definition of sufficiency, statement of the Fisher-Neyman factorization theorem with proof for discrete probability distribution. Pitmann – Koopman form and sufficient statistic; Exponential family of probability distributions and sufficient statistic. Proofs of the following properties of sufficient statistic: i) If T is sufficient for θ, then ø (T) is also sufficient for θ provided ø is a one to one and onto function. ii) If T is sufficient for θ then T is also sufficient for ø (θ). iii) M.L.E. is a function of sufficient statistic.	
	3.4	Efficiency	**(7L)**
		Fisher information function**:** Amount of information contained in statistic $T=T(X_1, X_2, \ldots, X_n)$. Statement regarding information in sample and in a sufficient statistic T. Cramer- Rao Inequality**:** Statement and proof of Cramer - Rao inequality, Cramer – Rao Lower Bound (CRLB),definition of minimum variance bound unbiased estimator (MVBUE) of ø(θ).Proofs of following results: a) If MVBUE exists for θ then MVBUE exists for ø (θ) where ø is a linear function. b) If T is MVBUE for θ then T is sufficient for θ. Comparison of variance with CRLB, relative efficiency of T_1	

		w.r.t. T_2 for (i) unbiased (ii) biased estimators. Efficiency of unbiased estimator T w.r.t. CRLB.	
	3.5	**Asymptotic Behaviour of an Estimator**	**(6L)**
		Consistency: Definition, proof of the following theorems: a. An estimator is consistent if its bias and variance both tend to zero as the sample size tends to infinity. b. If T is consistent estimator of θ and ø (.) is a continuous function, then ø (T) is a consistent estimator of ø (θ)	
Unit IV		**Interval Estimation**	**(6L)**
		Notion of interval estimation, definition of confidence interval (C.I),length of C.I., Confidence bounds, confidence coefficient. Definition of pivotal quantity and its use in obtaining confidence intervals. Interval estimation for the following cases: i) Mean (μ) of normal distribution (σ^2known and σ^2 unknown). ii) Variance (σ^2) of normal distribution (μ known and μ unknown). iii) Median, quartiles using order statistics.	
		References: 1. Dudewicz, E.J. and Mishra, S.N. (1988). Modern Mathematical Statistics, John Wiley, New York. 2. Hoel, P.G. Port, S. and Stone, C.(1972). Introduction to Statistical Theory, Houghton Mifflin Company (International) Dolphin Edition. 3. Hogg, R.V. and Craig, A.T. (1978). Introduction to Mathematical Statistics (fourth edition), Collier Macmillan International Edition, Macmillan Publishing Co. Inc.,New York. 4. Mood, A.M., Graybill, F. and Bose, D.C. (1974). Introduction to the theory of Statistics (third edition) International Student Edition, McGraw Hill. 5. Rohatagi , V.K. (2001). An introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd., New Delhi. 6. Kale B.K. and Murlidharan K. (2015) Introduction to Parametric Inference , Narosa Publication House, New Delhi.	

****

T.Y. B.Sc. (Statistics) Semester V
Statistics Paper – 3 (STS3503)
Introduction to Regression Analysis

[Credits-3]

The main objective of this course is to introduce students to

1. Fitting of simple, Multiple and Logistic Regression models.
2. Model building , residual diagnostics, corrective measures and polynomial regression model.
3. Tests of hypothesis of model parameters, AIC and BIC criteria.
4. Interpretation of output produced by glm command in R.

Unit I		**Simple linear regression model**	**(14L)**
	1.1	Review of simple linear regression model: $Y = \beta_0 + \beta_1 X + \varepsilon$ where ε is a continuous random variable with $E(\varepsilon) =0$, $V(\varepsilon) = \sigma^2$ Estimation of β_0 and β_1, by the method of least squares.	
	1.2	Properties of estimators of β_0, and β_1.	
	1.3	Estimation of σ^2.	
	1.4	Assumption of normality of ε .Tests of hypothesis of β_1.	
	1.5	Interval estimation in simple linear regression model.	
	1.6	Coefficient of determination.	
	1.7	Residual analysis: Standardized residuals, Studentized residuals, residual plots.	
	1.8	Detection and treatment of outliers.	
	1.9	Interpretation of four plots produced by lm command in R.	
Unit II		**Multiple linear regression model**	**(20L)**
	2.1	Review of multiple linear regression model $Y = \beta_0 + \beta_1 X_1 + . . + \beta_p X_p + \varepsilon$, where ε is a continuous random variable with $E(\varepsilon) =0$, $V(\varepsilon) = \sigma^2$. Estimation of regression parameters β_0, β_1, . . . and β_p by the method of least squares, obtaining normal equations, solutions of normal equations.	
	2.2	Estimation of σ^2.	
	2.3	Assumption of normality of ε .Tests of hypothesis of	

		regression parameters.	
	2.4	Interval estimation in multiple linear regression model.	
	2.5	Variable selection and model building.	
	2.6	Residual diagnostics and corrective measures such as transformation of response variable, weighted least squares method.	
	2.7	Introduction of multicolinearity, computation of VIF and brief introduction to ridge regression.	
Unit III		**Logistic regression model**	**(14L)**
	3.1	Binary response variable, Logit transform, estimation of parameters, interpretation of parameters.	
	3.2	Tests of hypotheses of model parameters, model deviance, LR test.	
	3.3	AIC and BIC criteria for model selection.	
	3.4	Interpretation of output produced by glm command in R.	
	3.5	Multiple logistic regression.	
		References: 1. Draper, N. R. and Smith, H. (1998). Applied Regression Analysis Third Edition, John Wiley. 2. Hosmer, D. W. and Lemeshow, S. (1989). Applied Logistic Regression, Wiley. 3. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). Introduction to Linear Regression Analysis Wiley. 4. Neter, J., W., Kutner, M. H. , Nachtsheim, C.J. and Wasserman, W.(1996) : Applied Linear Statistical Models, fourth edition, Irwin USA. 5. Chatterjee S. and Hadi A.S.(2012) : Regression Analysis by Example, 5th Edition,Wiley. 6. Kleinbaum G. and Klein M. (2011) : Logistic Regression, IIIrd Edition A Self learning text, Springer.	

****

T.Y. B.Sc. (Statistics) Semester V
Statistics Paper - 4 (STS3504)
Design of Experiments

[Credits-3]

The main objective of this course is to introduce students to

1. Identifying relationships between cause and effect.
2. Providing an understanding of interactions among causative factors.
3. Determining the levels at which to set the controllable factors in order to optimize reliability by minimizing experimental error.
4. Improving robustness of the design or process variation.

Unit I		**Design of Experiments**	**(24L)**
	1.1	Analysis of variance (ANOVA): concept and technique.	
	1.2	Basic terms of design of experiments: Experimental unit, treatment, layout of an experiment.	
	1.3	Basic principles of design of experiments : replication, randomization and local control, choice of size and shape of a plot for uniformity trials, the empirical formula for the variance per unit area of plots.	
	1.4	Completely Randomized Design (CRD) : Application of the principles of design of experiment in CRD, Layout, Model: $X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ i= 1,2, …….t. j = 1,2,……., $n_{i,}$ assumptions and interpretations. Testing normality graphically. Breakup of total sum of squares into components. Estimation of parameters, expected values of mean sums of squares, components of variance, preparation of (ANOVA) table, testing equality of treatment effects, hypothesis to be tested $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$. Comparison of treatment means using box plot techniques. Statement of Cochran's theorem. F test for testing Ho with justification (independence of chi-square is to be assumed), test for equality of two specified treatment effects using critical difference (C.D).	
	1.5	Randomized Block Design (RBD) : Application of the principles of design of experiments in RBD, Layout , Model: $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ i= 1,2, …….t. j = 1,2,…… b , assumptions and interpretations, Breakup of total sum of squares into	

		components. Estimation of parameters, expected values of mean sums of squares, components of variance, preparation of analysis of variance table, Hypotheses to be tested H_{01} : $\alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_t = 0$ H_{02} : $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_b = 0$ F test for testing H_{01} and H_{02} with justification (independence of chi-squares is to be assumed), test for equality of two specified treatment effects using critical difference (CD).	
	1.6	Latin Square Design (LSD): Application of the principles of design of experiments in LSD , layout Model : $X_{ij(k)} = \mu + \alpha_i + \beta_j + Y_k + \varepsilon_{ij(k)}$, i = 1.2 … m, j = 1, 2 …, m, k = 1, 2 …., m. Assumptions and interpretations. Break up of total sum of squares into components . Estimation of parameters , expected values of mean sum of squares, components of variance , preparation of ANOVA table , hypothesis to be tested H_{01}: $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ H_{02} : $\beta_1 = \beta_2 \cdots = \beta_m = 0$ H_{03} : $\gamma_1 = \gamma_2 = \cdots = \gamma_m = 0$ and their interpretations . Justification of F test for H_{01}, H_{02} and H_{03} (independence of chi- square is to be assumed). Preparation of ANOVA table and F test for H_{01}, H_{02} and H_{03} testing for equality of two specified treatment effects, comparison of treatment effects using critical difference, linear treatment contrast and testing its significance.	
	1.7	Linear treatment contrasts, orthogonal contrasts. Scheffe's method for comparing contrasts, Tukey's procedure for comparing pairs of treatment means(applicable to C.R.D., R.B.D. and L.S.D.).	
	1.8	Identification of real life situations where the above designs are used.	
	1.9	Analysis of non- normal data using. i) Square root transformation for counts. ii) Sin^{-1}(.) transformation for proportions. iii) Kruskal Wallis test.	
Unit II		**Efficiency of Design**	**(5L)**
	2.1	Concept and definition of efficiency of a design.	

	2.2	Efficiency of RBD over CRD.	
	2.3	Efficiency of LSD over (i) CRD (ii) RBD.	
Unit III		**Analysis of Covariance (ANOCOVA) with One Concomitant Variable**	**(7L)**
	3.1	Situations where analysis of covariance is applicable.	
	3.2	Model for covariance in CRD, RBD. Estimation of parameters (derivations are not expected).	
	3.3	Preparation of analysis of variance – covariance table, test for $\beta=0$, test for equality of treatment effects (computational technique only).	
Unit IV		**Factorial Experiments**	**(12L)**
	4.1	General description of m^n factorial experiment, 2^2 and 2^3 factorial experiments arranged in RBD.	
	4.2	Definitions of main effects and interaction effects in 2^2 and 2^3 factorial experiments.	
	4.3	Yate's procedure, preparation of ANOVA table, test for main effects and interaction effects.	
	4.4	General idea of confounding in factorial experiments.	
	4.5	Construction of layouts in total confounding and partial confounding in 2^2 and 2^3 factorial experiments.	
	4.6	Total confounding (confounding only one interaction) ANOVA table, testing main effects and interaction effects.	
	4.7	Partial confounding (confounding only one interaction per replicate); ANOVA table, testing main effects and interaction effects.	
		References: 1. Cochran W.G. and Cox, C.M. (1968) Experimental Design, John Wiley and Sons, Inc., New York. 2. Dass, M.N. and Giri,N.C. (1986) Design and Analysis of Experiments, II Edition Wiley Eastern Ltd., New Delhi.	

		3. Federerer W.T. (1967) Experimental Design : Oxford and IBH Publishing Co., New Delhi. 4. Goon, A.M., Gupta,M.K. and Dasgupta, B. (1998). Fundamentals of Statistics, Vol.II, The world Press Pvt. Ltd. Kolkata. 5. Johnson, R.A., Miller, I. and Freund, J.(2010). Probability and Statistics for engineers, Prentice Hall, India. 6. Montgomery, D.C. (2001). Design and Analysis of Experiments, John Wiley and sons Inc., New Delhi. 7. Snedecor, G.W. and Cochran, W.G. (1994). Statistical Methods, 8th edition, Affiliated East – West Press, New Delhi.	

****

T.Y. B.Sc. (Statistics) Semester V
Statistics Paper - 5 (STS3505)
Actuarial Statistics

[Credits-3]

The main objective of this course is to introduce students to

1. Identify and analyze consequences of events involving risk and uncertainty.
2. To study survival distribution and construction of life tables.
3. Apply appropriate modeling techniques for lifetime random variables involved in the field of Insurance.
4. Study of annuities and benefit premium in case of n-year term life insurance, whole life insurance and endowment insurance.

Unit I		**Insurance Business**	**(3L)**
	1.1	Insurance companies as business organizations.	
	1.2	Role of insurance business in Economy.	
	1.3	Concept of risk, types of risk, characteristics of insurable risk.	
	1.4	Working of insurance business, introduction of terms such as premium, policy, policyholder and benefit.	
	1.5	Role of Statistics in insurance.	
	1.6	Insurance business in India.	
Unit II		**Feasibility of Insurance Business**	**(4L)**
	2.1	Measurement of adverse financial impact, expected value principle.	
	2.2	Concept of utility function.	
	2.3	Feasibility of insurance business.	
	2.4	Illustrative examples.	
Unit III		**Survival Distribution and Life Tables**	**(12L)**
	3.1	Time- until death random variable, its d.f. and survival function in actuarial notation.	
	3.2	Force of mortality.	
	3.3	Interrelations among d.f., survival function, force of mortality and	

		p.d.f.	
	3.4	Curtate future life random variable, its p.m.f. and survival function in actuarial notation.	
	3.5	Construction of life table using random survivorship approach.	
Unit IV		**Models for Life Insurance**	**(11L)**
	4.1	Theory of compound interest, effective rate of interest, discount factor.	
	4.2	Insurance payable at the end of the year of death, present value random variable, actuarial present value.	
	4.3	Derivation of actuarial present value for n-year term life insurance, whole life insurance and endowment insurance.	
Unit V		**Annuities**	**(10L)**
	5.1	Annuities – certain, annuity due, annuity immediate.	
	5.2	Discrete life annuities: n-year temporary life annuity due and a whole life annuity due, present value random variables of the payment, and their actuarial present values.	
Unit VI		**Benefit Premiums**	**(8L)**
	6.1	Concept of a loss random variable.	
	6.2	Equivalence principle.	
	6.3	Computation of fully discrete premium for n-year term life insurance, whole life insurance and endowment insurance.	
	6.4	Variance of loss random variable.	
		References: 1. Bowers N.L. Jr., H.S.Gerber, J.C. Hickan, D.A.Jones, C.J.Nesbitt, (1997). Actuarial Mathematics, Society of Actuaries, U.S. 2. Deshmukh, S. R. (2009). Actuarial Statistics, Universities Press, Hyderabad, India.	

****

T.Y. B.Sc. (Statistics) Semester V Statistics Paper - 6 (STS3506) Demography and Indian Official Statistics [Credits-3]
The main objective of this course is to introduce students to 1. Applications of statistical methods to demography and population projection. 2. Understanding mortality, fertility and reproduction rates, their computation and interpretation. 3. Indian official statistics pertaining to agriculture, industry and concept of national income and methods of computation.

Unit I		**Demography**	**(12L)**
	1.1	Measures of mortality: Crude death rate, specific death rates-age, sex, occupation, standardized death rates(based on age specific death rates) direct and indirect methods, comparative study of these measures, numerical problems on these measures, infant mortality rate, neo-neonatal mortality rate, maternal mortality rate, old age mortality rate.	
	1.2	Measures of fertility : Crude birth rate, general fertility rate, age specific fertility rate, total fertility rate, comparative study of these measures, numerical problems on these measures.	
	1.3	Reproduction rates : GRR, NRR, comparison, interpretation, numerical problems.	
	1.4	Population Projection : Component method (description only), mathematical method – derivation of logistic law of population growth, interpretation of constants in the equation of the logistic curve.	
Unit II		**Life Tables**	**(5L)**
		Construction and uses of a life table. Numerical problems.	
Unit III		**Census**	**(6L)**
		General principles of census, utility of census methods, de-facto and de-jure methods, Indian Census 2011 : Houselist, establishment schedule, individual slip, population record, findings of 2011 census, estimation of intercensal estimates	

		using standard methods.	
Unit IV		**Indian Official Statistics**	**(18L)**
	4.1	Agricultural Statistics in India (i) Statistics of land utilization (ii) Statistics of crop output (iii) miscellaneous of crop output (not to be studied in detail) (iv) Indices of agricultural production. Defects of Indian agricultural statistics.	
	4.2	Price statistics : usefulness of price statistics, wholesale price statistics, index number of wholesale prices. Retail price statistics- labour bureau index, number of retail prices for urban and rural areas, consumer price index for industrial workers, non manual employees and agricultural labourers, limitations of price statistics.	
	4.3	Industrial Statistics : primary sources of industrial statistics, statistics collected (description in brief), limitations of industrial statistics, index number of industrial production. Method of compilation. Index number of industrial profits revised series.	
	4.4	Educational Statistics : Description of different statistics relating to education, compiled and published by the ministry of education of the India Govt. Number of educational institutions, education of scheduled castes, tribes and backwards classes. Number of scholars, number of teachers, examination result. Sources of publications, critical study of educational statistics in India.	
Unit V		**National Income**	**(4L)**
	5.1	Definition (three approaches: product, income and expenditure). Methods of estimating national income : product method, income method, expenditure method and social accounting method.	
Unit VI		**Pareto and Lognormal laws of income distribution**	**(3L)**
		References: 1. Spiegelaman M. (1955) Introduction to Demography, Society of Actuaries. 2. Wolfendon H. H. (1954) Population Statistics and their compilation, Society of Actuaries. 3. Goon Gupta and Dasgupta (1968) Fundamental of	

		Statistics Vol. II., World Press. 4. Asthana B. N. and Srivastava S. S. (1969) Applied Statistics of India. Chaitanya Publishing House, Allahabad. 5. Gupta, S.C and Kapoor, V.K. (2008): Fundamentals of Applied Statistics,Sultan Chand & Sons Pvt. Ltd. New Delhi. 6. RamKumar, R. (1986): Technical Demography, Wiley Eastern Limited. 7. Pathak, K.B. and Ram, F. (1992): Techniques of Demographic Analysis, Himalaya publishing House censusindia.gov.in and www.rchiips.org. 8. Cox. P. (1959): Demography, Cambridge University Press. 9 . Shrivastava, O.S. (1995): Demography and Population Studies, Vikas Publishing society.	

****

T.Y. B.Sc. (Statistics) Semester V
Statistics Paper - 7 (STS3507)
Operations Research

[Credits-3]

The main objective of this course is to acquaint students with the following concepts.

1. Convert given practical situations to the format of linear programming problem.
2. Understand concept of optimal solution.
3. Learn special cases of LPP viz. transportation problem, assignment problem.
4. Understand concepts related to project management technique.

Unit I		**Linear Programming**	**(16L)**
	1.1	Statement of the linear Programming Problem (LPP) with minimization or maximization of objective function. Formulation of problem as L.P. Problem. Definition of (i) A slack variable, (ii) A surplus Variable. L.P. Problem in (i) Canonical form (ii) standard form. Definition of (i) a solution (ii)basic and	

		non basic variables (iii) a feasible solution (iv) a basic feasible solution, (v) a degenerate and non–degenerate solution (vi) an optimal solution.	
	1.2	Solution of L.P.P by Simplex Method: Obtaining Initial Basic Feasible Solution (IBFS), criteria for deciding whether obtained solution is optimal ,criteria for unbounded solution, no solution, more than one solutions, introduction of artificial variable, Big-M method.	
	1.3	Duality Theory: Writing dual of a primal problem, solution of a L.P.P. by using its dual problem.	
Unit II		**Transportation Problem**	**(12L)**
	2.1	Transportation problem (T.P.), statement of T.P., balanced and unbalanced T.P. Minimization and maximization problem.	
	2.2	Obtaining basic feasible solution of T.P. by (i) Least cost method (ii) Vogel’s approximation method (VAM).	
	2.3	u-v (MODI) method of obtaining Optimal solution of T.P., uniqueness and non- uniqueness of optimal solutions, degenerate solution. Removing degeneracy and obtaining optimal solution.	
	2.4	Assignment Problem : Statement of an assignment problem , Minimization and maximization problem , balanced and unbalanced problem ,relation with transportation problem , optimal solution using Hungarian method , maximization case.	
Unit III		**Critical Path Method (CPM) and Project Evaluation and Review Techniques (PERT)**	**(15L)**
	3.1	Definition of (i) Event,(ii) Node,(iii)Activity,(iv)Critical Activity,(v)Project Duration.	
	3.2	CPM: Construction of network, Definitions (i) earliest start time. (ii) earliest finish time. (iii) latest start time. (iv) latest finish time for an activity. Critical Path, Types of float, total floats, free float, independent float and their significance. Determination of critical path.	

	3.3	PERT: Construction of network; (i) pessimistic time estimate, (ii) Optimistic time estimate (iii) most likely time estimates, Determination of critical path, determination of mean and standard deviation of project duration computations of probability of completing the project in a specified duration.	
	3.4	Cost Benefit Analysis Definition of normal time, crash time, normal cost, crash cost, cost slope, direct cost, indirect cost, project cost, Determination of project duration and its associated cost when (i) Normal time are considered. (ii) Crash time are considered. Determination of optimal network.	
Unit IV		**Decision Theory**	**(5 L)**
	4.1	Decision under risk : Expected value criterion.	
	4.2	Decision Trees.	
	4.3	Decision under uncertainty. (i) Laplace criterion (ii) Mini max criterion. (iii) Savage minimax regret criterion. (iv) Hurwitz criterion.	
		References: 1. Gass, S.L. (1997). Linear programming methods and applications, Narosa Publishing House, New Delhi. 2. Gupta, P.K. and Hira, D.S.(2008). Operation Research, 3rd edition S. Chand and company Ltd., New Delhi. 3. Kapoor, V. K.(2006). Operations Research, S. Chand and Sons. New Delhi. 4. Saceini, M., Yaspan,A. and Friedman, L.(2013). Operation Research methods and problems, Willey International Edition. 5. Sharma, J.K. (1989). Mathematical Models in Operation Research, Tata McGraw Hill Publishing Company Ltd., New Delhi. 6. Shrinath.L.S (1975). Linear Programming, Affiliated East-West Pvt. Ltd, New Delhi.	

		7. Taha, H.A. (2010). Operation research: An Introduction, 9th edition, Prentice Hall of India, New Delhi.	

****

T.Y. B.Sc. (Statistics) Semester V Statistics Paper – 8 (STS3508) Biostatistics [Credits-3]			
The main objective of this course is to introduce students to 1. Concepts of health, illness, disease and socially defined concept of sickness. 2. Understanding of the principles of epidemiology and discussion of the concept of natural history of disease particularly with respect to possible clinical interventions. 3. Understanding of the purposes for conducting clinical trials. 4. The phases of clinical trials and the types of study designs typically used in clinical trials, use of randomization and blinding.			
Unit I		**Epidemiology**	**(16L)**
	1.1	Introduction to Epidemiology	
	1.2	Odds, odds ratio, relative risk.	
	1.3	Estimation of odds ratio (OR), Confidence interval for OR. Relation	

		with parameter in a logit model.	
	1.4	Symmetry in square contingency tables, collapsing tables and Simpson's paradox.	
Unit II		**Clinical trials**	**(12 L)**
	2.1	General information on history of drug discovery including Louis Pasteur (rabies and small pox), Ronald Ross and malaria, Alexander Fleming and penicillin, Jonas Salk and polio, cholera, asthma, diabetes, blood pressure, heart attack, arthritis.	
	2.2	Phases of clinical trial, purpose, duration, cost, drug regulatory bodies, ICH, statistical analysis plan, clinical study report.	
Unit III		**Design of clinical trials**	**(15L)**
	3.1	Design of clinical trials: parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, objectives and endopoints of clinical trials, design of phase I trials, design of single – stage and multi-stage phase II trials. Design and monitoring of phase III trials with sequential stopping, design of bio-equivalence trials. Inference for 2x2 crossover design: Classical methods of interval testing for bioequivalence, nonparametric methods.	
	3.2	2 treatments, 2 periods cross over design.	
Unit IV		**Bioequivalence and bio-availability**	**(5L)**
	4.1	Bioequivalence and bio-availability, non-inferiority trial, non parametric methods.	
	4.2	Practice based medical research, evidence based medicine.	
		References: 1. A.p .Gore and S. A, Paranjape ,(2000) Course on mathematical and statistical Ecology, Kluwer, publishing Holland. 2. M.B. Kulkarni, V.R. Prayag, (2004) “ Introduction to Statistical Ecology, SIPF Academy, Nasik 41. 3. Agresti A. (1996) Categorical Data Analysis. Wiley, New York. 4. J.N.S. Matthews (2006) Introduction to Randomized Controlled clinical Trials, Chapman and Hall.	

		5. Stephen Sann (2000) Statistical Issues in drug Development, John Wiley. 6. Steven Diantadosi (2000) Clinical Trials – A methodological Perspective, John Wiley. 7. L.M. Friedmon, C.D. Forbes, D.L. Demats (2000) Fundamentals of Clinics Trials, Spinner. 8. Steve selvin (2004)Epidemiologic Analysis, Oxford Press. 9. M.M. Shoukni, C.A. Pavse (1999) Statistical Methods for Health Sciences, CPC Pree. 10. Steve Salvin, (1999) Statistical Analysis of Epidomiologic Data, Oxford.	

****

T.Y. B.Sc. (Statistics) Semester V
Statistics Practical - I (STS3511)

[Credits-2]

The main objective of this course is to introduce students to

1. Programming in C
2. Programming skills to solve statistical problems.

Sr. No.	Title of the experiment
1.	To carry out arithmetic calculations. To check whether given number is odd or even. To check whether given number m is divisible by n or not. To find maximum of 2 numbers or 3 numbers. To find area of triangle and circle. To find roots of quadratic equation. To check whether integer is prime or not.
2.	To find sum of digits of a number. To solve simultaneous linear equations.(two equations in two variables) To evaluate simple and compound interest To prepare multiplication table.

3.	Programs using string function To test palindrome string using string function. To sort a string using string function. To search string using string function. To combine given two strings using string function.
4.	To evaluate exp(x), sin (x), log (x) etc. using Taylor series expansion
5.	To solve transcendental equations using Newton Raphson method.
6.	Program in C to prepare a discrete frequency distribution with given class interval from raw data and hence fit a discrete probability distribution
7.	To find mean, variance ,coefficient of variation of n numbers
8.	To obtain correlation coefficient for given bivariate data and fit line of regression
9.	To arrange the observations in ascending order of magnitude and find median. Hence obtain m.l.e. of a specific probability distribution
10.	Fitting of lognormal distribution, Model sampling from Cauchy and Laplace distribution

****

T.Y. B.Sc. (Statistics) Semester V
Statistics Practical - II(STS3512)

[Credits-2]

The main objective of this course is to introduce students to

1. Principles of Design of Experiments (DOE),and planning of experiment
2. Analysis of various designs.
3. Confidence interval estimation.

Sr. No.	Title of the experiment
1.	Analysis of CRD (equal and unequal replications parirwise comparison of treatments, using critical difference (C.D.) check normality using normal probability plot.
2.	Analysis of R.B.D. pairwise comparison of treatments using i) C. D ii) Tukey and Scheffe's procedure. Efficiency of RBD w.r.t. CRD.
3.	Analysis of LSD, pairwise comparison of treatments using C.D. and box plot, efficiency of LSD w.r.t. i) CRD ii) RBD.

4.	Kruskal Wallis test.
5.	Analysis of covariance in CRD, testing $\beta = 0$.
6.	Analysis of covariance in RBD, testing $\beta = 0$.
7.	Analysis of 2^2 and 2^3 factorial experiments in RBD.
8.	Analysis of 2^3 factorial experiments in RBD (partial confounding).
9.	Analysis of 2^3 factorial experiments in RBD (total confounding).
10.	Construction of confidence interval for population median and quartiles, based on order statistics.

****

T.Y. B.Sc. (Statistics) Semester V

Statistics Practical – III (STS3513)

[Credits-2]

The main objective of this course is to acquaint students with the following concepts.

1. The use of basic as well as advanced R-commands.
2. Drawing sample from standard discrete and continuous distributions
3. Analysis of regression models
4. Understand certain sampling techniques through R

Sr. No.	Title of the experiment
1.	Introduction to R- Software
2.	Model sampling from standard discrete distributions.
3.	Model sampling from standard continuous distributions.
4.	Fitting of Lognormal distribution.
5.	Computation of probabilities of type $P(a<X<b)$, $P(X<b)$ etc. for standard

	discrete and continuous distributions.
6.	Simple linear regression analysis.
7.	Multiple Linear regression analysis.
8.	Logistic regression analysis.
9.	Testing normality of number of samples.
10.	Computation of benefit premium for n – year term insurance, whole life insurance and endowment insurance.

****

Deccan Education Society's

FERGUSSON COLLEGE, PUNE

(AUTONOMOUS)

SYLLABUS UNDER AUTONOMY

THIRD YEAR B.Sc. (Statistics)

SEMESTER –VI

SYLLABUS FOR T.Y. B.Sc.

Academic Year 2018-2019

T.Y. B.Sc. (Statistics) Semester VI
Statistics Paper – 1 (STS3601)
Introduction to Stochastic Processes

[Credits-3]

The main objective of this course is to introduce students to

1. Concept of stochastic processes and Markov chains.
2. The situation where Markov chains can be applied.
3. Poisson process as a particular case of stochastic processes.

Unit I		Definition of a Stochastic process, state space, parameter space, types of stochastic processes, Markov chains (MC) $\{X_n, n \geq 0\}$, finite MC, time homogeneous M.C. one step transition probabilities, and transition probability matrix (t.p.m.),stochastic matrix, Chapman Kolmogorov equation, n-step transition probability matrix , initial distribution, joint distribution function of $\{X_0, X_1, \ldots, X_n\}$, partial sum of independent and identically distributed random variables as Markov chain, illustrations such as random walk, Gambler's ruin problem, Ehrenfest chain.	**(18L)**
Unit II		Classification of states: Communicating states, first return probability, probability of ever return Classification of states, as persistent and transient states. Decomposition of state space, closed set of states, irreducible set of states, irreducible MC, periodicity of M.C. aperiodic M.C. ergodic M. C.	**(12L)**
Unit III		Stationary distribution for a irreducible ergodic finite long run behavior of a M.C.	**(6L)**
Unit IV		**Homogeneous Poisson process**: Postulates and properties of Poisson process, probability distribution of N(t), the number of occurrences of the event in (o, t], Poisson process and probability distribution of inter-arrival time,mean, variance and covariance functions Definition of compound Poisson process mean and variance functions and its applications.	**(12L)**
		References: 1. Bhat, B.R. (2000): Stochastic models: Analysis and applications, New Age International.	

		2. Hoel , P. G., Port, S.C. and Stone, C.J. (1972) : Introduction to stochastic processes, Wiley Eastern. 3. Medhi J. (2009): Stochastic processes, New age International. 4. Ross, S. (2014): Introduction to probability models, 11^{th}edn, Elsevier. 6. Ross, S. (1996): Stochastic processes, 2^{nd} edn. John Wiley. 7. Taylor, H N and Karlin, S. (2010) : An introduction to stochastic modelling 4^{th} edn. Academic Press. 8. Vidyadhar Kulkarni (2016) : Modelling and Analysis Stochastic Systems.3^{rd} edn., CRC press.	

****

T.Y. B.Sc. (Statistics) Semester VI
Statistics Paper - 2 (STS3602)
Testing of Hypotheses

[Credits-3]

The main objective of this course is to introduce students to

1. Understand concept of statistical hypothesis and use of its in real life situations.
2. Understand M P and UMP test procedure and difference between then.
3. Understand Parametric and Nonparametric tests.

Unit I		**Most Powerful tests**	**(15L)**
	1.1	(a) Statistical hypothesis, problem of testing of hypotheses. Definition and illustrations of (i) simple hypothesis, (ii) composite hypothesis, (iii) test of hypothesis, (iv) critical region, (v) type I and type II errors. probabilities of type I error and type II error. Problem of controlling the probabilities of errors of two kinds. (b) Definition and illustrations of (i) level of significance, (ii) observed level of significance (p-value), (iii) size of a test, (iv) power of a test.	
	1.2	Definition of most powerful (M.P.) level α test of simple null hypothesis against simple alternative. Statement of Neyman - Pearson (N-P) lemma for constructing the most powerful level α test of simple null hypothesis against simple alternative hypothesis. Illustrations.	
	1.3	Power function of a test, power curve, definition of uniformly most powerful (UMP) level α test for one sided alternative. Illustrations.	
Unit II		**Likelihood ratio tests**	**(9L)**
		Notion of likelihood ratio test (LRT), $\lambda(x) = \sup_{\theta} £(\theta \mid x) / \sup_{\theta} £(\theta \mid x)$ Construction of LRT for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ for the mean of normal distribution for i) known σ^2 ii) unknown σ^2 (one sided and two sided alternatives). LRT for variance of normal distribution for i) known μ ii) unknown μ (one sided and two sided alternatives hypotheses). LRT for parameters of binomial and exponential distribution for two sided alternatives only. LRT as a function of sufficient statistics, statement of asymptotic distribution of $-2 \log_e \lambda(x)$ for testing $H_o : \theta \in Ho$ again $H_1 : \theta \in H_1$.	
Unit		**Sequential Tests**	**(9L)**

III			
		Wald's Sequential probability test procedure of strength (α,β) for simple null hypothesis of against simple alternative hypothesis and its comparison with fixed sample size N-P test procedure. Definition of Wald's SPRT of strength (α, β). Illustration for standard distributions like Bernoulli, Poisson, Normal and Exponential. SPRT test statistic as a function of sufficient statistics. Graphical representation of SPRT.	
Unit IV		**Non-parametric Tests**	**(15L)**
	4.1	Concept of non- parametric tests. Distinction between parametric and nonparametric tests.. Concept of distribution free statistic. One tailed and two tailed test procedure of (i) Sign test, ii) Wilcoxon signed rank test (iii) Mann-Whitney U test, (iv) Run test, one sample and two samples problems	
	4.2	Empirical distribution function S_n (x). Properties of S_n (x) as estimator of F (.). Kolmogorov – Smirnov test for completely specified univariate distribution (one Sample problem only) for two sided alternative hypotheses. Comparison with chi-square test.	
		References: 1. Daniel, W.W. (2000): Applied Nonparametric Statistics, 2nd edn. Duxbury Press Boston. 2. Dudeweitz, E.J. and Mishra, S.N. (1988): Modern Mathematical Statistics, John Wiley and Sons, Inc. 3. Gibbons J.D.(2010): Non parametric Statistical Inference, 5th edn. CRC Press Book. 4. Hogg, R.V. and Craig, R.G. (1989): Introduction to Mathematical Statistics, fourth edition,.Collier Macmillan International Edition, Macmillan Publishing Co. Inc., New York. 5. Kale, B.K. and Muralidharan, K. (2015): Parametric Inference: An Introduction. Narosa, New Delhi. 6. Mood, A.M., Graybill, F. and Bose, D. C .(1974) : Introduction to the theory of Statistics, third edition International Student Edition, McGraw Hill. 7. Rohatgi, V.K. (2008) : An introduction to Probability Theory Statistics, Wiley Eastern Ltd., New Delhi.	

T.Y. B.Sc. (Statistics) Semester VI
Statistics Paper – 3 (STS3603)
Reliability and Survival Analysis

[Credits-3]

The main objective of this course is to acquaint students with the following concepts :

1. Structural properties of coherent system.
2. Reliability of coherent system.
3. Ageing properties.
4. Censoring and Non-parametric estimation of survival function.

Unit I		**Structural Properties of coherent system**	**(13L)**
	1.1	Binary system of independent components, order of the system, different types of systems, concept of the structure function, structure function of series system, parallel system, k- out of- n system, (essentially parallel and series system), reliability block diagram, guidelines for construction of reliability block diagram.	
	1.2	Coherent structure function (maximum 4 components), relevant component , increasing structure function, pivotal decomposition of structure function, dual of a structure function (proof of dual of series system of order n is parallel system of order n, dual of the parallel system of order n is a series system of order n, dual of k-out -of –n system is (n-k+1)-out of –n system). , path sets, cut sets, minimal path and cut set, representation of coherent system in terms of minimal path sets and cut sets, dual coherent structure function, relative importance of components, module of the coherent system, modular decomposition of coherent system	
Unit II		**Reliability of coherent system**	**(8L)**
	2.1	Reliability of system of independent components, Basic properties of system reliability (such as reliability function is increasing function, system and component redundancy etc.), computation of reliability of coherent system by using minimal path and cut set representation, upper and lower bound on system reliability by using exact system reliability, relative importance of a component	
Unit III		**Ageing Properties**	**(9+9L)**

	3.1	Survival function, probability density function, hazard function, cumulative hazard rate, mean residual life function, equilibrium residual life function , interrelation between all these function, no ageing, proof of following properties of no ageing 1) Cauchy functional equation 2) Constant failure rate 3) Constant mean residual life 4) Exponential life distribution	
	3.2	**Positive and negative ageing**: IFR, DFR ,IFRA, DFRA, bathtub failure rate, Classification of following parametric families of life distribution according to aging**:** Weibull, Gamma, lognormal, linear failure rate, Makeham, Pareto, Lehman.	
Unit IV		**Censoring and Nonparametric estimation of survival function**	**(9L)**
	4.1	Concept of censoring, order censoring, time censoring, right random censoring, left random censoring, undersigned censoring, Nonparametric estimation of survival function, confidence band on survival function, actuarial estimator of survival function, Greenwood's formula, Kaplan Meier estimator of survival function in the presence of censored observations.	
		References: 1 .Barlow, R. E. and Proschan F. (1975) : Statistical theory of Reliability and Life testing: Probability Models Holt, Rinehart and Winston Inc. 2. Barlow, R. E. and Proschan F. (1996) : Mathematical Theory of Reliability. John Wiley. 3. Cox, D.R. and Oakes, D. (1984) : Analysis of Survival Data, Chapman and Hall 4. Deshpande, J.V. and Purohit S.G. (2005) : Life Time Data: Statistical Models and Methods, Word Scientific. 5. Tobias, P.A. and Trindane, D. C. (1995) : Applied Reliability. Second edition. CRC Press 6. J.Gertsbakh (2005) : Reliabity theory with applications to prevective maintenance, springer 7. Jerald F. Lawless (2002):Statistical methods and methods for lifetime data 2nd edition. Wiley 8. William Q. *Meeker* and Luis A. Escobar (1998) : Statistical methods for reliability data, John Wiley & Sons.	

****

T.Y. B.Sc. (Statistics) Semester VI
Statistics Paper - 4 (STS3604)
Sampling Methods

[Credits-3]

The main objective of this course is to acquaint students with the following concepts :

1. Designing and executing sample surveys.
2. Estimating population parameters using sampling distribution of estimator and obtaining estimators of standard error in estimation under various sampling procedures.
3. Determining adequate sample size for various sampling procedures.

Unit I		**Sampling**	**(8L)**
	1.1	Concept of distinguishable elementary units, sampling units, sampling frame, random sample, requisites of a good sample. Simple random sampling from finite population of size (N) (i) with replacement (SRSWR) ii) without replacement (SRSWOR) definitions, population mean and population total as parameters, inclusion probabilities.	
	1.2	(a) Sample mean $\overline{y}$ as an estimator of population mean, derivation of expectation and standard error of $\overline{y}$, confidence interval for population mean, population total . (b) N $\overline{y}$ as an estimator of population total, derivation of expectation and standard error of N $\overline{y}$ (c)Estimator of above standard errors, both in case of SRSWR and SRSWOR.	
	1.3	Sampling for proportion as an application of a simple random sampling with x_i as zero or one. (a) sample proportion as an estimator of population proportion of units possessing a certain attribute, derivation of expectation and standard error of (p). (b) Np as an estimator of total number of units in the population possessing a certain attribute, derivation of expectation and standard error of Np (c) Estimator of above standard error both in case of SRSWR and SRSWOR.	

Unit II		**Determination of Sample Size**	**(8L)**
	2.1	Determination of the sample size for the given: i) Margin of error and confidence coefficient. ii) Coefficient of variation of the estimator and confidence coefficient.	
Unit III		**Stratified Random Sample Size**	**(12L)**
	3.1	Stratification, basis of stratification, real life situation where stratification can be used.	
	3.2	Stratified random sampling as a sample drawn from individual strata using SRSWOR in each stratum.	
	3.3	(a) $\overline{y_{st}} = \frac{\sum N_i \overline{y_i}}{N}$ as an estimator of population mean ($\overline{Y}$) Derivation of expectation and standard error of $\overline{y_{st}}$. (b) N $\overline{y_{st}}$. as an estimator of population total, derivation of expectation and standard error of N $\overline{y_{st}}$. (c) Estimator of above standard errors.	
	3.4	Problem of allocation, proportional allocation, Neyman's allocation, derivation of the expressions for the standard errors of the above estimators when these allocations are used.	
	3.5	Gain in precision due to stratification, comparison amongst SRSWOR, stratification with proportional allocation and stratification with Neyman's allocation.	
	3.6	Cost and variance analysis in stratified random sampling, minimization of variance for fixed cost, minimization of cost for fixed variance optimum allocation Neyman's allocation as a particular case of optimum allocation in cost and variance analysis.	
Unit IV		**Ratio and Regression Methods of Estimation for SRSWOR**	**(6L)**
	4.1	Rationale behind using auxiliary variates in estimation	
	4.2	Situations where (i) ratio method is appropriate, (ii) regression method is appropriate	

	4.3	Ratio and regression estimators of the population mean and population total	
	4.4	Comments regarding bias statement of standard errors of ratio and regression estimators, relative efficiency of these estimators with respect to SRSWOR	
Unit V		**Systematic Sampling (Population size divisible by sample size)**	**(6L)**
	5.1	Real life situations where systematic sampling is appropriate. Techniques of drawing a sample using systematic sampling.	
	5.2	Estimation of the population mean and population total, standard error of these estimators.	
	5.3	Comparison of systematic sampling with SRSWOR	
	5.6	Comparison of systematic sampling with SRSWOR and stratified sampling in the presence of linear trend.	
Unit VI		**Role of Sample Surveys in Research Methodology**	**(8L)**
	6.1	Objectives of a sample survey	
	6.2	Designing a questionnaire, characteristics of a good questionnaire (Questions with codes & scores are to be discussed). Reliability and validity testing by using (i) Test – Retest method (ii) Internal Consistency: (A) Kuder Rechardson Coefficient (KR-20) (B) Cronbach's Coefficient Alpha	
	6.3	Planning, execution and analysis of a sample survey, practical problems at each of these stages	
	6.4	Sampling and non-sampling errors with illustrations	
	6.5	Study of some surveys illustrating the above ideas, rounds conducted by National Sample Surveys organization	
		References: 1. Arnab R. (2017) : Survey Sampling : Theory and Applications . Academic Press 2. M.M. *Desu* and D. Raghavarao (1990) : *Sample Size* Methodology, Academic Press, New York 3. Mukhopadhyay P (2008): Sampling theory and	

		methods of survey sampling. Prentice-Hall of India, New Delhi. 4. Sarjinder Singh (2003) : Advanced Sampling theory with applications, Kluwer Academic publishers. 5. Singh, D. and Chaudhary, F. S. (1986): Theory and Analysis of Sample Survey Designs, 1st edn. New Age International. 6. Sukhatme, P.V., Sukhatme, B. V. and Ashok (1984) : Sampling theory of Surveys with Applications, Indian Society of Agricultural Statistics, New Delhi.	

****

T.Y. B.Sc. (Statistics) Semester VI
Statistics Paper - 5 (STS3605)
Time Series Analysis

[Credits-3]

The main objective of this course is to introduce students to

1. The analysis of time series data.
2. To differentiate between various time series models.
3. To forecast future observations of the time series.

Unit I		**Features of Time Series Data**	**(8 L)**
	1.1	Introduction to Time Series, components of Time Series.	
	1.2	Trend, seasonality, changing variability.	
	1.3	Test for randomness of a series against trend and seasonality: Run test, Chi-sq test.	
Unit II		**Estimation of Trend and Seasonal Indices**	**(18 L)**
	2.1	Moving average and exponential smoothing.	
	2.2	Forecasting based on smoothing.	
	2.3	Double exponential smoothing, Halt Winter method with seasonality.	
	2.4	Choosing parameters for smoothing and forecasting.	
	2.5	Estimating mean square error of forecasting, Prediction, Intervals based on normality assumption.	
	2.6	Estimation of seasonal Indices using ratio to trend, ratio to moving average and link relative method.	
Unit III		**Time Series Analysis Through Regression Approach**	**(10L)**
	3.1	Regression models for trend and seasonality.	
	3.2	De- trending and de- seasonalizing a series.	
	3.3	Analysis of Irregular components to examine whether the effect of trend , seasonality is removed.	
Unit		**Introduction to Box Jenkins Techniques**	**(12L)**

IV			
	4.1	Study of weak stationarity through plots involving differencing and seasonal differencing concepts, consequences of over differencing when its variance increases.	
	4.2	Transformation of data : Transformation and differencing.	
	4.3	AR(I) and MA (I) models, Autocorrelatation and Autocovariance , forecasts based on AR(I) and MA (I) models, MSE of forecasts. [under the assumption that parameters are estimated.	
		References: 1. Montgomery , D.C. and Johnson L.A. (1976): Forecasting and Time Series Analysis, McGraw Hill. 2. Farmum, N.R. and Stantorr, L.W. (1989): Quantitative Forecasting Methods, PWSKent Publishing Company, Boston. 3. Christopher Chatfield (1975): The Analysis of Time Series, 6th edition, CRC Press. 4. Mukhopadhyay, P (2011): Applied Statistics, 2nd edition revised reprint, Books and Allied(P) Ltd.	

****

T.Y. B.Sc. (Statistics) Semester VI Statistics Paper -6 (STS3606) Statistical Ecology [Credits-3]
The main objective of this course is to acquaint students with the following concepts. 1. To understand the concepts of Ecology. 2. To understand the concepts of Biodiversity. 3. To study different models to analyze Statistical Ecology.

Unit I		**Population Dynamics**	**(16L)**
	1.1	Introduction: Ecology, Statistical Ecology.	
	1.2	Linear Growth $dN_t/dt = C$.	
	1.3	Exponential Model: Solving $\frac{dN_t}{dt} = kN_t$.	
	1.4	Logistic growth model: Density dependence, solving differential equation.	
	1.5	$\frac{dN_t}{dt} = a.N_t\ (k - N_t)$ Properties ,carrying capacity, Interpretation, Scope and limitation.	
	1.6	Geompertz Curve: Solving Differential equation $\frac{dN_t}{dt} = a \log\left(\frac{k}{N_t}\right)$ Asymptotically stable equilibrium, Properties, Interpretation, Scope and Limitation. Fitting the above growth models to data by linearization and regression.	
	1.7	**Life tables:** Force of mortality, stable population and stationary population. Cohort, columns of life table, interrelation between columns, interpretation, construction of life table, uses and applications.	
	1.8	Leslie matrix Models: fecundity and survival matrix, $n_t = n_o M_t$, future projections, stable age distribution, interpretation of largest eigen value of M.	
Unit II		**Smoothing Procedures**	**(16L)**
	2.1	Poisson forest, Aggregated, Regular spatial point pattern, estimation of population density by quadrate sampling,	

		nearest neighbor distances (Point to individual, individual to individual), i^{th} order nearest neighbor distance. $\lambda = \frac{n}{\pi} x2$ M.L.E. for Poisson forest, Bias and S.E. of λ estimate.	
	2.2	Line transect method: Drawing random line transect, exponential detection function, mle of population density, other detection functions.	
	2.3	Capture–recapture models: Closed population, Open population, Peterson estimator for single recapture, Multiple captures, iterative method to find mle of N, Population size.	
	2.4	Removal method: Zippin's estimator for closed population.	
Unit III		**Diversity Indices**	**(8L)**
	3.1	Concept of Biodiversity, need to protect it.	
	3.2	Richness indices, Simpson's index, Shannon's index.	
	3.3	Rare fraction Curves, Real life examples for computing these indices.	
Unit IV		**Distribution Models**	**(8L)**
	4.1	Use of geometric distribution, lognormal distribution in ecology.	
		References: 1. Pielou,E.C.(1977):An Introduction to Mathematical Ecology, Wiley. 2. Seber,G.A.F.(1982): The estimation of animal abundance and related parameters, C. Griffin. 3. Ludwig,J.A. and Regnold J.F (1988): Statistical Ecology, A primer on methods and computing, John Wiley and Sons. 4. Gore A.P. and Paranjape S.A.(2000): A Course on Mathematical and Statistical Ecology, Kluwer Academic Publishers. 5. Kot M. (2001): Elements of Mathematical Ecology, Cambridge University Press. 6. Maynard Smith (1974): Models in Ecology, Cambridge University press.	

****

T.Y. B.Sc. (Statistics) Semester VI
Statistics Paper - 7 (STS3607)
Elements of Statistical computing and Data mining

[Credits-3]

The main objective of this course is to acquaint students with the following concepts.

1.Random number generators, simulation
2. Monte Carlo integration
3. Data mining techniques

Unit I		**Random number generators**	**(12L)**
	1.1	Concept of randomness , revision of random sample , generating randomness , practical difficulties in generating randomness Mid square algorithm and its limitations Requisites of good random number generator	
	1.2	Test for randomness : Run test, poker's test Concept of entropy as measure of randomness Significance of Kendall's τ / spearman's rank correlation empirical test viz. Kolmogorov Smirnov test	
	1.3	Specific random number generators : Linear Congruential Generator(L.C.G.), Mixed L.C.G., Multiplicative generator.	
Unit II		**Generating random variates**	**(12 L)**
	2.1	Methods to generate random variates. (a) Inverse transformation method. (b) Method of composition (c) Method of convolution (d) Acceptance Rejection method	
	2.2	Generating discrete random variates. (Using recurrence relation and inverse transformation method) Generating samples from Bernoulli distribution Binomial distribution (Also using composition method.) Poisson distribution. Geometric distribution (also using relation with exponential distribution.)	

	2.3	Generating continuous random variates. Generating samples from Uniform distribution using inverse transformation Exponential distribution Normal distribution using (i) Box muller method. (ii) Atkinson pearce method. (iii) Acceptance rejection method. (i)using Laplace distribution as majorising function (ii)using exponential distribution as majorising function (iii)using Cauchy as majorising function	
Unit III		**Applications of random number generation and random variate generations.**	**(12 L)**
		Monte carlo integration Computation of definite integral , Error in computation.(confidence interval for the value of integral) Computation of some special integrals empirical MGF empirical Laplace transform Estimating moments.	
Unit VI		**Introduction to data mining**	**(12 L)**
	6.1	Concept of database.	
	6.2	Extraction of data from database wrt research objective.	
	6.3	Concept of receiving operating characteristics (ROC) of algorithm / method.	
	6.4	Concept of specificity, sensitivity, accuracy.	
	6.5	Mathew’s correlation coefficient , applications of ROC to association of some simple statistical procedures.	
		References: 1. Averill M. Law (2000): Simulation Modelling and Analysis, M.C. Graw Hill 2. Debasis Kundu, Ayanendranath Basu (2004) : Statistical Computing, Narosa Publishing House 3. Do Le *Minh* (2001) : Applied Probability Models, Duxbury.	
		**	

T.Y. B.Sc. (Statistics) Semester VI
Statistics Paper - 8 (STS3608)
Statistical Quality Control

[Credits-3]

The main objective of this course is to acquaint students with the following concepts.

1. To construct control charts for variables and attributes.
2. To understand process capability.
3. To study acceptance sampling plans.

Unit I		**Introduction**	**(4L)**
	1.1	Meaning and purpose of Statistical Quality Control (SPC), on line process control methods (control charts) and offline process control methods (Sampling plans). **Seven Process Control (PC) Tools of SPC** (i) Check Sheet, (ii) Cause and effect diagram (CED),(iii) Pareto Diagram, (iv) Histogram, (v) Control chart, (vi)Scatter Diagram,(vii) Design of Experiments (DOE). (Only introduction of 7 PC tools is expected)	
Unit II		**Control charts**	**(24L)**
	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. Criteria for detecting lack of control situations: (i) At least one point outside the control limits (ii) A run of seven or more points above or below central line. (iii)Presence of a non random pattern eg. cycle or linear trends etc. Control chart technique as hypotheses testing problem. Construction of control charts for (i) standards given, (ii) standards not given.	
	2.2	**Control charts for variables**	
		(I) R chart and $\overline{X}$ chart Purpose of R and $\overline{X}$ chart, normal probability plot for checking	

		normality assumption, 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 $\overline{X}$ chart when the process average is specified: control limits, drawing of control chart, plotting of sample means. Drawing conclusion - determination of state of control of process, corrective action if the process is out of statistical control. (II) 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. Construction of $\overline{X}$ chart when the process average μis not given : control limits based on $\hat{\mu} = \overline{\overline{X}}$, $\hat{\sigma} = \frac{\overline{R}}{d_2}$, drawing of control chart, plotting sample means, revision of control limits of chart, if necessary. Probability of catching a shift. **Note**: To find revised control limits of any control chart delete the sample points above UCL and points below LCL (assuming a search for assignable causes at those points), in case of R and $\overline{X}$ charts, first of all, revisions of control limits of R is to be completed and then by using the observations for which R chart shows the process is under control, the control limits for $\overline{X}$ chart should be determined. Revision of control limits of $\overline{X}$ chart be continued without revising the value of $\overline{R}$ or $\hat{\sigma}$ Estimate of μ and σ for further use. Determination of state of control of the process. Identification of real life situations where this technique can be used. Limitations of $\overline{X}$, R charts.	
	2.3	**Control charts for Attributes**	
		(I) 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 separate control limits, 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. **(II) 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 III		**Capability Studies**	**(6L)**
	3.1	Specification limits, natural tolerance limits and their comparisons, decisions based on these comparisons, estimate of percent defectives.	
	3.2	Capability ratio and capability indices *(Cp)*, capability performance indices *Cpk* with respect to machine and process, interpretation, relationship between (i) *Cp* and *Cpk* (ii)defective parts per million and *Cp.*	
Unit IV		**Offline Methods (Lot Control)** **Acceptance Sampling for Attributes**	**(14L)**
	4.1	**Introduction:** Concept of sampling inspection plan, comparison between 100% inspection and sampling inspection. Procedure of acceptance sampling with rectification.	
	4.2	**Single Sampling Plan**: Working of SSP, Evaluation of probability of acceptance using Poisson distribution. Producer's risk. Consumer's risk, Acceptable Quality Level (AQL). Lot Tolerance Fraction Defective (LTFD), LTPD, Average Outgoing Quality (AOQ), Average Outgoing Quality Limit (AOQL), Average Sample Number (ASN), Average Total Inspection (ATI),	

		Operating characteristic (OC) curve, AOQ curve.	
	4.3	**Double Sampling Plan**: Working of DSP, Evaluation of probability of acceptance using Poisson distribution. Producer's risk. Consumer's risk, O.C.curve, Average Outgoing Quality (AOQ), AOQ curve, Average Outgoing Quality Limit (AOQL), Average Sample Number (ASN), Average Total Inspection (ATI) (with complete inspection of second sample).	
		References: 1. Grant, E. L. and Leavenworth (1980): Statistical Quality Control, fifth edition, Mc-Graw Hill, New Delhi. 2. Johnson, N.L. and Kotz, S. (1993): Capability Studies, Chapman and Hall Publishers. 3. Montgomery, R.C. (1985): Introduction to Statistical Quality Control, Fourth edition, Wiley. 4. Mittag, H.J. & Rinne, H. (1993): Statistical Methods for Quality Assurance, Chapman & Hall. 5. Schilling, E.G. (1982): Acceptance Sampling in Quality Control, Marcel Dekker. 6. Amitava Mitra (2001): Fundamentals of Quality Control and Improvement – Pearson Education Asia. 7. Duncan, A.J. (1986) : Quality control and Industrial Statistics. McGraw-Hill 8. Chin-Knei Cho (1987) : Quality Programming, John Wiley.	

****

T.Y. B.Sc. (Statistics) Semester VI
Statistics Practical – IV (STS3611)

[Credits-2]

The main objective of this course is to acquaint students with the following concepts.

1. To develop MP test and UMP tests.
2. To develop SPRT for discrete and continuous distribution.
3. To study Markov Chain and types of stochastic processes

Sr. No.	Title of the experiment
1.	Construction of confidence interval for population median and quartiles, based on order statistics.
2.	Testing of hypotheses (Probability of type I and II errors, power of a test etc).
3.	Construction of most powerful (MP) test.
4.	Construction of uniformally most powerful (UMP) test, plotting of power function of a test.
5.	Non-parametric tests: Sign test, Wilcoxon's signed rank test. Mann – whitney U test.
6.	Non-parametric tests: Run test, median test. Kolmogorov - Smirnov test.
7.	SPRT for Bernoulli, Binomial, Poisson, Hypergeometric distributions. (graphical representation also).
8.	SPRT for normal, exponential distribution. (graphical representation also).
9.	Stochastic Process-I
10.	Stochastic Process-II

****

T.Y. B.Sc. (Statistics) Semester VI

Statistics Practical – V (STS3612)

[Credits-2]

The main objective of this course is to acquaint students with the following concepts.

1. To learn estimation and confidence interval estimation for population parameters using different sampling techniques
2. To analyse different time series models
3. To study reliability of coherent system

Sr. No.	Title of the experiment
1.	Simple random sampling for population mean, population total , Simple random sampling for proportions : (i)with replacement , (ii) without replacement
2.	Stratified random sampling : Proportional and Neyman allocation, comparison with SRSWOR
3.	Stratified random sampling : cost and variance analysis
4.	Ratio and Regression methods of estimation, comparison with SRSWOR
5.	Determination of minimal i) cut sets ii) path sets for given coherent system.
6.	Finding reliability h (p) of coherent system with i.i.d. components each with reliability p. Graph of h (p) against p. S shaped ness property.
7.	Analysis of Time Series _I (Estimation of trend in Time series by fitting of AR (1) model , exponential smoothing, moving averages)
8.	Analysis of Time Series _II (Estimation of seasonal indices by ratio to trend method and method of link relatives)
9.	Analysis of Time Series _III
10.	Analysis of Time Series _IV

****

T.Y. B.Sc. (Statistics) Semester VI
Statistics Practical – VI(STS3613)

[Credits-2]

The main objective of this course is to acquaint students with the following concepts.

1. To apply different methods of collection of data if any
2. To analyze the data using statistical tools
3. To prepare the report of the project
4. To make the presentation of the project work

The students are supposed to carry out project work during this semester. This project work should be done in a group of not more than 6 students.
The project will be evaluated on the following parameters

(1) CE - 50 marks
(2) ESE -50 marks at the end of the semester

Concurrent Evaluation (CE)

(i) Initial presentation of the problem -10 marks

(ii) Data collection/ Methods to be used / layout of project ---10 marks

(iii) Submission of the project & final presentation in scheduled time----20 marks

(iv) Attendance & active participation for project work ----10 marks

End Semester Examination (ESE)

Following are the details for evaluation of the project:

Details about submission marks		
1.	Technical problem and motivation	2
2.	Abstract and key words	2
3.	Conversion of the technical problem into statistical language	4
4.	Representation of raw data / questionnaire	2
5.	Exploratory data analysis	5
6.	Statistical analysis	5
7.	Conclusions in relevant language , scope / limitations	5
Total		25

Note: Viva would be conducted individually. Viva will be taken for 15 minutes for each candidate.

- The division of 25 marks for viva is as follows:

1.	Understanding of technical problem.	10
2.	Understanding of Statistical Techniques used to solve the problem	10
3.	Overall impression	5
Total		25

****