SYLLABUS

STATISTICS PAPER – 2

Unit 1: Multivariate Analysis and Linear Models

Multivariate normal distribution and its properties. Random sampling from multivariate normal distribution. Maximum likelihood estimators of parameters, distribution of sample mean vector. Wishart matrix – its distribution and properties, distribution of sample generalized variance, null and non-null distribution of multiple correlation coefficients. Hotelling's T2 and its sampling distribution, application in test on mean vector for one and more multivariate normal population and also on equality of components of a mean vector in multivariate normal population. Classification problem: Standards of good classification, procedure of classification based on multivariate normal distributions. Principal components, dimension reduction, canonical variates and canonical correlation — definition, use, estimation and computation. Linear Models: Theory of linear estimation, Gauss-Markov linear models, estimable functions, error and estimation space, normal equations and least square estimators, estimation of error variance, estimation with correlated observations, properties of least square estimators, generalized inverse of a matrix and solution of normal equations, variances and covariances of least square estimators.

Unit 2: Demography and Vital Statistics

Sources of demographic data, census, registration, ad-hoc surveys, Hospital records, Demographic profiles of the Indian Census. Complete life table and its main features, Uses of life table. Makehams and Gompertz curves. Abridged life tables. Stable and stationary populations. Measurement of Fertility: Crude birth rate, General fertility rate, Age specific birth rate, Total fertility rate, Gross reproduction rate, Net reproduction rate. Measurement of Mortality: Crude death rate, Standardized death rates, Age specific death rates, Infant Mortality rate, Death rate by cause. Internal migration and its measurement, migration models, concept of international migration. Net migration. International and postcensal estimates. Decennial population census in India.

Unit 3: Econometrics

Nature of econometrics, the general linear model (GLM) and its extensions, ordinary least squares (OLS) estimation and prediction, generalized least squares (GLS) estimation and prediction, heteroscedastic disturbances, pure and mixed estimation. Auto correlation, its consequences and tests. Theil BLUS procedure, estimation and prediction, multi-collinearity problem, its implications and tools for handling the 2 problem, ridge regression. Linear regression and stochastic regression, instrumental variable estimation, errors in variables, autoregressive linear regression, lagged variables, distributed lag models, estimation of lags by OLS method, Koyck's geometric lag model. Simultaneous linear equations model and its generalization, identification problem, restrictions on structural parameters, rank and order conditions. Estimation in simultaneous equations model, recursive systems, 2 SLS estimators, limited information estimators, k-class estimators, 3 SLS estimator, full information maximum likelihood method, prediction and simultaneous confidence intervals.

Unit 4: Applied Statistics

Index Numbers: Price relatives and quantity or volume relatives, Link and chain relatives composition of index numbers; Laspeyre's, Paasches', Marshal Edgeworth and Fisher index numbers; chain base index number, tests for index number, Construction of index numbers of wholesale and consumer prices. Demand analysis, Income Distribution-Pareto and Engel curves. Time Series Analysis: Economic time series, different components, illustration, additive and multiplicative models, determination of trend, seasonal and cyclical fluctuations. Time-series as discrete parameter stochastic process, auto covariance and autocorrelation functions and their properties. Exploratory time Series analysis, tests for trend and seasonality, exponential and moving average smoothing. Holt and Winters smoothing, forecasting based on smoothing. Detailed study of the stationary processes: (1) moving average (MA), (2) auto regressive (AR), (3) ARMA and (4) AR integrated MA (ARIMA) models. Box-Jenkins models, choice of AR and MA periods. Discussion (without proof) of estimation of mean, auto covariance and autocorrelation functions under large sample theory, estimation of ARIMA model parameters. Spectral analysis of weakly stationary process, periodogram and correlogram analyses.

Unit 5: Statistical Quality Control Operations Research

Quality Control: Statistical process and product control: Quality of a product, need for quality control, basic concept of process control, process capability and product control, general theory of control charts, causes of variation in quality, control limits, sub grouping summary of out of control criteria, charts for attributes p chart, np chart, c chart, V chart, charts for − variables: $\bar{X}$,R,($\bar{X}$,R) and ($\bar{X}$,σ) charts. Basic concepts of process monitoring and control; process capability and process optimization. General theory and review of control charts for attribute and variable data; O.C. and A.R.L. of control charts; Acceptance sampling plans for attributes inspection; single and double sampling plans and their properties; plans for inspection by variables for one-sided and two sided specifications. Operations Research: Transportation and assignment problems. Bellman's principle of optimality, general formulation, computational methods and application of dynamic programming to LPP. Decision-making in the face of competition, two-person games, pure and mixed strategies, existence of solution and uniqueness of value in zero-sum games, finding solutions in 2x2, 2xm and mxn games. Analytical structure of inventory 3 problems, EOQ formula of Harris, its sensitivity analysis and extensions allowing quantity discounts and shortages. Queuing models – specification and effectiveness measures. Steady-state solutions of M/M/1 and M/M/c models with associated distributions of queue-length and waiting time. PERT and CPM – basic concepts. Probability of project completion.

Unit 6: Survival Analysis and Numerical Analysis

Concept of time, order and random censoring, likelihood in the distributions – exponential, gamma, Weibull, lognormal, Pareto, Linear failure rate, inference for these distribution. Life tables, failure rate, mean residual life and their elementary classes and their properties. Estimation of survival function – actuarial estimator, Kaplan – Meier estimator, estimation under the assumption of IFR/DFR, tests of exponentiality against non-parametric classes, total time on test. Two sample problem – Gehan test, log rank test. Semi-parametric regression for failure rate – Cox's proportional hazards model with one and several covariates, rank test for the regression coefficient. Competing risk model, parametric and non-parametric inference for this model. Numerical Analysis: Finite differences of different orders: Δ, E and D operators, factorial representation of a polynomial, separation of symbols, sub-division of intervals, differences of zero. Concept of interpolation and extrapolation: Newton Gregory's forward and backward interpolation formulae for equal intervals, divided differences and their properties, Newton's formula for divided difference, Lagrange's formula for unequal intervals, central difference formula due to Gauss, Sterling and Bessel, concept of error terms in interpolation formula. Inverse interpolation: Different methods of inverse interpolation. Numerical differentiation: Trapezoidal, Simpson's one-third and three-eight rule and Waddles rule. Summation of Series: Whose general term (i) is the first difference of a function (ii) is in geometric progression. Numerical solutions of differential equations: Euler's Method, Milne's Method, Picard's Method and Runge-Kutta Method.