

INDIAN STATISTICAL SERVICE EXAM

STANDARD AND SYLLABI

The standard of papers in General English and General Studies will be such as may be expected of a graduate of an Indian University.

The standard of papers in the other subjects will be that of the Master's degree examination of an Indian University in the relevant disciplines. The candidates will be expected to illustrate theory by facts, and to analyse problems with the help of theory. They will be expected to be particularly conversant with Indian problems in the field(s) of Statistics.

GENERAL ENGLISH

Candidates will be required to write an essay in English. Other questions will be designed to test their understanding of English and workman like use of words. Passages will usually be set for summary or precis.

GENERAL STUDIES

General knowledge including knowledge of current events and of such matters of everyday observation and experience in their scientific aspects as may be expected of an educated person who has not made a special study of any scientific subject. The paper will also include questions on Indian Polity including the political system and the Constitution of India, History of India and Geography of a nature which a candidate should be able to answer without special study.

STATISTICS-I (OBJECTIVE TYPE)

(i) Probability:

Classical and axiomatic definitions of Probability and consequences. Law of total probability, Conditional probability, Bayes' theorem and applications. Discrete and continuous random variables. Distribution functions and their properties.

Standard discrete and continuous probability distributions - Bernoulli, Uniform, Binomial, Poisson, Geometric, Rectangular, Exponential, Normal, Cauchy, Hyper geometric, Multinomial, Laplace, Negative binomial, Beta, Gamma, Lognormal. Random vectors, Joint and marginal distributions, conditional distributions, Distributions of functions of random variables. Modes of convergences of sequences of random variables - in distribution, in probability, with probability one and in mean square. Mathematical expectation and conditional expectation. Characteristic function, moment and probability generating functions, Inversion, uniqueness and continuity theorems. Borel 0-1 law, Kolmogorov's 0-1 law. Tchebycheff's and Kolmogorov's inequalities. Laws of large numbers and central limit theorems for independent variables.

(ii) Statistical Methods:

Collection, compilation and presentation of data, charts, diagrams and histogram. Frequency distribution. Measures of location, dispersion, skewness and kurtosis. Bivariate and multivariate data. Association and contingency. Curve fitting and orthogonal polynomials. Bivariate normal distribution. Regression-linear, polynomial. Distribution of the correlation coefficient, Partial and multiple correlation, Intraclass correlation, Correlation ratio.

Standard errors and large sample test. Sampling distributions of sample mean, sample variance, t, chi-square and F; tests of significance based on them, Small sample tests. Non-parametric tests-Goodness of fit, sign, median, run, Wilcoxon, Mann-Whitney, Wald-Wolfowitz and Kolmogorov-Smirnov. Order statistics-minimum, maximum, range and median. Concept of Asymptotic relative efficiency.

(iii) 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-eighth 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.

(iv) Computer application and Data Processing:

Basics of Computer: Operations of a computer, Different units of a computer system like central processing unit, memory unit, arithmetic and logical unit, input unit, output unit etc., Hardware including different types of input, output and peripheral devices, Software, system and application software, number systems, Operating systems, packages and utilities, Low and High level languages, Compiler, Assembler, Memory – RAM, ROM, unit of computer memory (bits, bytes etc.), Network – LAN, WAN, internet, intranet, basics of computer security, virus, antivirus, firewall, spyware, malware etc.

Basics of Programming: Algorithm, Flowchart, Data, Information, Database, overview of different programming languages, frontend and backend of a project, variables, control structures, arrays and their usages, functions, modules, loops, conditional statements, exceptions, debugging and related concepts.

STATISTICS- II (OBJECTIVE TYPE)

(i) 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.

One way and two-way classifications, fixed, random and mixed effects models. Analysis of variance (two-way classification only), multiple comparison tests due to Tukey, Scheffe and Student-Newmann-Keul-Duncan.

(ii) Statistical Inference and Hypothesis Testing:

Characteristics of good estimator. Estimation methods of maximum likelihood, minimum chi-square, moments and least squares. Optimal properties of maximum likelihood estimators. Minimum variance unbiased estimators. Minimum variance bound estimators. Cramer-Rao inequality.

Bhattacharya bounds. Sufficient estimator. factorization theorem. Complete statistics. Rao-Blackwell theorem. Confidence interval estimation. Optimum confidence bounds. Resampling, Bootstrap and Jackknife.

Hypothesis testing: Simple and composite hypotheses. Two kinds of error. Critical region. Different types of critical regions and similar regions. Power function. Most powerful and uniformly most powerful tests. Neyman-Pearson fundamental lemma. Unbiased test. Randomized test. Likelihood ratio test. Wald's SPRT, OC and ASN functions. Elements of decision theory.

(iii) Official Statistics:

National and International official statistical system

Official Statistics: (a) Need, Uses, Users, Reliability, Relevance, Limitations, Transparency, its visibility (b) Compilation, Collection, Processing, Analysis and Dissemination, Agencies Involved, Methods

National Statistical Organization: Vision and Mission, NSSO and CSO; roles and responsibilities; Important activities, Publications etc.

National Statistical Commission: Need, Constitution, its role, functions etc; Legal Acts/ Provisions/ Support for Official Statistics; Important Acts

Index Numbers: Different Types, Need, Data Collection Mechanism, Periodicity, Agencies Involved, Uses

Sector Wise Statistics: Agriculture, Health, Education, Women and Child etc. Important Surveys & Census, Indicators, Agencies and Usages etc.

National Accounts: Definition, Basic Concepts; issues; the Strategy, Collection of Data and Release.

Population Census: Need, Data Collected, Periodicity, Methods of data collection, dissemination, Agencies involved.

Misc: Socio Economic Indicators, Gender Awareness/Statistics, Important Surveys and Censuses.

STATISTICS- III (DESCRIPTIVE TYPE)

(i) Sampling Techniques:

Concept of population and sample, need for sampling, complete enumeration versus sampling, basic concepts in sampling, sampling and Non-sampling error, Methodologies in sample surveys (questionnaires, sampling design and methods followed in field investigation) by NSSO.

Subjective or purposive sampling, probability sampling or random sampling, simple random sampling with and without replacement, estimation of population mean, population proportions and their standard errors. Stratified random sampling, proportional and optimum allocation, comparison with simple random sampling for fixed sample size.

Covariance and Variance Function.

Ratio, product and regression methods of estimation, estimation of population mean, evaluation of Bias and Variance to the first order of approximation, comparison with simple random sampling.

Systematic sampling (when population size (N) is an integer multiple of sampling size (n)). Estimation of population mean and standard error of this estimate, comparison with simple random sampling.

Sampling with probability proportional to size (with and without replacement method), Des Raj and Das estimators for $n=2$, Horvitz-Thomson's estimator

Equal size cluster sampling: estimators of population mean and total and their standard errors, comparison of cluster sampling with SRS in terms of intra-class correlation coefficient.

Concept of multistage sampling and its application, two-stage sampling with equal number of second stage units, estimation of population mean and total. Double sampling in ratio and regression methods of estimation.

Concept of Interpenetrating sub-sampling.

(ii) 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 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.

(iii) 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, Income distribution-Pareto and Engel curves, Concentration curve, Methods of estimating national income, Inter-sectoral flows, Interindustry table, Role of CSO. Demand Analysis

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, computations based on Fourier transform.

STATISTICS-IV (DESCRIPTIVE TYPE)

(Equal number of questions i.e. 50% weightage from all the subsections below and candidates have to choose any two subsections and answer)

(i) Operations Research and Reliability:

Definition and Scope of Operations Research: phases in Operation Research, models and their solutions, decision-making under uncertainty and risk, use of different criteria, sensitivity analysis.

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 2×2 , $2 \times m$ and $m \times n$ games.

Analytical structure of inventory problems, EOQ formula of Harris, its sensitivity analysis and extensions allowing quantity discounts and shortages. Multi-item inventory subject to constraints. Models with random demand, the static risk model. P and Q- systems with constant and random lead times.

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. M/G/1 queue and Pollazcek-Khinchine result.

Sequencing and scheduling problems. 2-machine n-job and 3-machine n-job problems with identical machine sequence for all jobs

Branch and Bound method for solving travelling salesman problem.

Replacement problems – Block and age replacement policies.

PERT and CPM – basic concepts. Probability of project completion.

Reliability concepts and measures, components and systems, coherent systems, reliability of coherent systems.

Life-distributions, reliability function, hazard rate, common univariate life distributions – exponential, weibull, gamma, etc. Bivariate exponential distributions. Estimation of parameters and tests in these models.

Notions of aging – IFR, IFRA, NBU, DMRL and NBUE classes and their duals. Loss of memory property of the exponential distribution.

Reliability estimation based on failure times in variously censored life-tests and in tests with replacement of failed items. Stress-strength reliability and its estimation.

(ii) 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. National life tables. UN model life tables. 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. Projection method including logistic curve fitting. Decennial population census in India.

(iii) Survival Analysis and Clinical Trial:

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.

Introduction to clinical trials: the need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I – IV trials, multicenter trials.

Data management: data definitions, case report forms, database design, data collection systems for good clinical practice.

Design of clinical trials: parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, review of factorial designs, objectives and endpoints 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,

Reporting and analysis: analysis of categorical outcomes from Phase I – III trials, analysis of survival data from clinical trials.

(iv) 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) , (\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; control by gauging; moving average and exponentially weighted moving average charts; Cu-Sum charts using V-masks and decision intervals; Economic design of \bar{X} -bar chart.

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 specification.

(v) Multivariate Analysis:

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 T^2 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.

(vi) Design and Analysis of Experiments:

Analysis of variance for one way and two way classifications, Need for design of experiments, basic principle of experimental design (randomization, replication and local control), complete analysis and layout of completely randomized design, randomized block design and Latin square design, Missing plot technique. Split Plot Design and Strip Plot Design.

Factorial experiments and confounding in 2^n and 3^n experiments. Analysis of covariance.

Analysis of non-orthogonal data. Analysis of missing data.

(vii) Computing with C and R :

Basics of C: Components of C language, structure of a C program, Data type, basic data types, Enumerated data types, Derived data types, variable declaration, Local, Global, Parametric variables, Assignment of Variables, Numeric, Character, Real and String constants, Arithmetic, Relation and Logical operators, Assignment operators, Increment and decrement operators, conditional operators, Bitwise operators, Type modifiers and expressions, writing and interpreting expressions, using expressions in statements. Basic input/output.

Control statements: conditional statements, if - else, nesting of if - else, else if ladder, switch statements, loops in c, for, while, do - while loops, break, continue, exit (), goto and label declarations, One dimensional two dimensional and multidimensional arrays. Storage classes: Automatic variables, External variables, Static variables, Scope and lifetime of declarations.

Functions: classification of functions, functions definition and declaration, assessing a function, return statement, parameter passing in functions. Pointers (concept only).

Structure: Definition and declaration; structure (initialization) comparison of structure variable; Array of structures : array within structures, structures within structures, passing structures to functions; Unions accessing a union member, union of structure, initialization of

a union variable, uses of union. Introduction to linked list, linear linked list, insertion of a node in list, removal of a node from list.

Files in C: Defining and opening a file, input – output operation on a file, creating a file, reading a file.

Statistics Methods and techniques in R.