

Department of Statistics
St. Xavier's College, Kolkata
Autonomous
(Under Calcutta University)

Description of Courses
(Statistics Honours)

MODULES	COURSES	SEMESTER
DS 101 PT 101 RA 101 EI 100	Descriptive Statistics I Probability Theory I Real Analysis I Economic Statistics and Indian Statistical System	Sem I
DS 202 PT 202 RA 202 PS 200	Descriptive Statistics II Probability Theory II Real Analysis II Population Statistics	Sem II
PT 303A PT 303B LA 301 TS 300	Probability Theory IIIA Probability Theory IIIB Linear Algebra I Time Series Analysis	Sem III
SD 400 SI 400 LA 402 SQ 400	Sampling Distribution Elements of Statistical Inference Linear Algebra II Statistical Quality Control	Sem IV
ML 501 ET 501 SS 501 AD 501	Multivariate Analysis Theory of Estimation and Testing of Hypotheses I Sample Survey Methods I ANOVA Regression and ANCOVA	Sem V
ML 602 ET 602 SS 602 AD 602	Large Sample Theory Theory of Estimation and Testing of Hypotheses II Sample Survey Methods II Design of Experiments	Sem VI

Description of Courses (Statistics General)

MODULES	COURSES	SEMESTER
DS 111 PT 111	Descriptive Statistics I Probability Theory I	SEM I
DS 212 PT 212	Descriptive Statistics II Probability Theory II	SEM II
SI 311 SE 311	Sampling Distributions and Statistical Inference I Sample Survey	SEM III
SI 412 SE 412	Sampling Distributions and Statistical Inference II Design of Experiments	SEM IV
EI 510 PS 510	Economic Statistics and Indian Statistical System Population Statistics	SEM V
TS 610 SQ 610	Time Series Analysis Statistical Quality Control	SEM VI

STATISTICS (HONOURS) SYLLABUS

THEORY

Descriptive Statistics- I (DS101)

Introduction: Nature of Statistics, Uses of Statistics, Statistics in relation to other disciplines.
(3L)

Types of Data: Concepts of population & sample, Quantitative & Qualitative data, Cross-sectional & Time-series data, Discrete & Continuous data, Different types of scales. (7L)

Collection & Scrutiny of Data: Primary data, Secondary data. Complete enumeration, Controlled experiments, Observational studies & Sample Surveys. Scrutiny of data for internal consistency & Detection of errors in recording. (7L)

Presentation of Data: Construction of tables with one or more factors of classification. Diagrammatic representations. Frequency distributions, Cumulative Frequency Distributions & their graphical representations, Steam-Leaf displays. (10L)

Analysis of Quantified Data:

Univariate Data- Different measures of location, dispersion, skewness & kurtosis. Moments & Sheppard's correction for grouped data (without proof). Quantiles & measures based on them, Box plot.

Comparison of Data-sets- Percentile ranks, Z-scores & relative dispersion. (23L)

Probability Theory –I (PT 101)

Random Experiments: Trials, Sample points, Sample space, Events, Classes of events. (5L)

Definition of Probability: Long-run relative frequency, Classical definition, Kolmogorov's Axiomatic definition (detailed discussion on discrete space only). Limitations of classical definition. Probability of union and intersection of events. Probability of occurrence of exactly m & at least m events out of n events. Conditional probability & Independence of events. Bayes' theorem & its applications. Examples based on classical approach. Repeated trials. (20L)

Real Analysis – I (RA101)

{concepts, statements (without proof) of major results and their applications}

Sets and Functions: Countable & Uncountable sets. Sequence & Series of real numbers. Convergence, Limits, Absolute convergence. Concepts of o_h & O_h . Comparison, Ratio & Root tests. (15L)

Continuity & Differentiability of real-valued functions. Riemann integration. Fundamental Theorem of integration. Integration by parts. Differentiation under integration, Improper integrals and their convergence. Gamma and Beta integrals. (10L)

Economic Statistics & Indian Statistical System (ES100)

Economic Statistics

Index Numbers: Price, Quantity & Value indices. (2L)

Price Index Numbers: Construction, Uses, Limitations, Tests for index numbers, Various formulae & their comparisons, Chain Index Number. (10L)

Some Important Indices: Consumer Price Index, Wholesale Price Index & Index of Industrial Production – methods of construction & uses. (5L)

National Accounts: Estimation of national income & its components. (8L)

Measures of Inequality: Gini's coefficient, Lorenz curves. (5L)

Indian Statistical System

The Statistical System in India: Central & State Government organizations, Functions of Central Statistical Organization (CSO), National Sample Survey Organization (NSSO) & West Bengal Bureau of Applied Economics and Statistics. (10L)

Sources of Official Statistics: Official Statistics in India & West Bengal relating to population, agriculture, industry, trade, price, finance & employment (10L)

Descriptive Statistics- II (DS202)

Analysis of Quantified Data:

Bivariate Data- Scatter plots, Correlation coefficient & its properties. Correlation Ratio. Correlation Index. Intraclass Correlation with equal & unequal group sizes. Principles of least squares, Regression curves & related results, Fitting of polynomial & exponential curves, Residual plots. Rank correlation- Spearman's & Kendall's measures. (30L)

Analysis of Categorical Data: Consistency of data, Independence & association of attributes. Measures of association- Odds ratio, α & δ measures, Yule's & Cramer's V^2 , Pearson's, Tschuprow's measures, Kendall's τ & τ_b , Goodman-Kruskal's γ , Somer's d . Fitting of Logistic Regression equations. (20L)

Probability Theory- II (PT202)

Random Variables: Definition of discrete & continuous random variables, Cumulative distribution function (cdf) & its properties (with proof), Probability mass function (pmf) & Probability density function (pdf), Expectation & moments, Theorems on sum & product of expectations of random variables, Dispersion, Skewness, Kurtosis, Quantiles. Probability generating function (pgf) & Moment generating function (mgf) in the univariate case. (25L)

Real Analysis-II (RA202)

Sequence & Series of functions. Pointwise & Uniform convergence. Simple tests, Power Series, Taylor's series expansion, Error, Differentiation & Integration of series. Double integration. Evaluation of double integrals- repeated integrals & change of variables. (25L)

Population Statistics (PS200)

Introduction: Sources of population data – Census data, Registration data & errors. Rates & Ratios of Vital events. (3L)

Measurements of Mortality: Crude Death Rate, Specific Death Rate, Standardized Death Rate, Cause of death rate, Infant Mortality Rate. Survival probability, Forces of mortality, Makeham & Gompertz laws, Fitting by selected points. (14L)

Measurements of Morbidity: Morbidity Incidence & Morbidity Prevalence Rates. (3L)

Life Tables: Description of complete & abridged life tables, Uses. Cohort vs current life tables, Stable population & Stationary population, Construction of complete life table from census & registration data. (10L)

Measurements of Fertility: Crude Birth Rate, General Fertility Rate, Age Specific Fertility Rate, Total Fertility Rate. (4L)

Measurement of Population Growth: Crude Rate of Natural Increase & Vital Index, Gross & Net Reproduction Rates. (4L)

Population Estimation, Projection & Forecasting: Use of A.P. and G.P. methods of population estimates, Use of Component method for population projection, Logistic curve for population forecasting & fitting by Rhode's method. (12L)

Probability Theory-III A (PT303A)

Random Variables: The cdf, pmf & pdf in the Bivariate case. Marginal & Conditional distributions, Independence, Conditional Expectation & Variance, Correlation & Regression. Bivariate pgf & mgf. (20L)

Probability Inequalities: Chebyshev's Lemma, Markov & Chebyshev inequalities. (10L)

Limit Theorems: Convergence in Probability, Weak Law of Large Numbers and its applications, Convergence in Distribution, De-Moivre-Laplace Limit Theorem, Normal approximation to the Poisson distribution, Statement of Central Limit Theorem (iid case) & its applications. (15L)

Probability Theory-III B (PT303B)

Univariate Discrete Distributions: Uniform, Binomial, Hypergeometric, Poisson, Truncated Poisson, Geometric, Negative Binomial distributions & their properties. (10L)

Univariate Continuous Distributions: Rectangular, Normal, Truncated Normal, Cauchy, Gamma, Exponential, Beta, Log-normal distributions & their properties. Use of Normal distribution in scaling of data. (10L)

Bivariate Continuous Distribution: Bivariate Normal distribution & its properties. (5L)

Linear Algebra-I (LA301)

Linear Equations: Systems of Linear Equations, Gauss method of successive elimination, Rectangular arrays of coefficients. Matrices (definition & types) & Determinants (definition & statement of properties), Cramer's rule. (6L)

Vector Algebra: n-dimensional Euclidean space. Vector spaces, Linear independence of vectors, Basis & Dimension of a vector space. Orthogonal vectors, Gram-Schmidt orthogonalization. (19L)

Time Series Analysis (TS300)

Introduction: Observations recorded in order of time as a sequence of random variables. Examples from various fields. Different types of variation and their modelling- Seasonal, Trend, Cyclical & Stationary. (5L)

Analysis of Variations in Time Series: Seasonality Analysis- Method of average; Trend Analysis- Method of simple & weighted Moving Averages & Method of Regression (fitting of polynomials & curves reducible to polynomials). Analysis of composite time series. Analysis of Cyclical Variation- Method of Regression (fitting of harmonic curves). (20L)

Analysis of Stationary Time Series: Weak stationarity, Autocorrelation function & Correlogram. MA, AR & ARMA processes. Estimation of the parameters in AR(1)& AR(2). Yule-Walker equations. (20L)

Forecasting: Fitting of asymptotic growth curves, Exponential smoothing. (5L)

Sampling Distribution (SD400)

Introduction: Concepts of Random Sampling, Statistics and Sampling Distributions of Statistics. Uses of Distribution Function, Moment Generating function & Transformation of Variables. Multiple Integration, Orthogonal & Polar Transformations. (15L)

Some Standard Sampling Distributions: χ^2 distribution. Distributions of Mean and Variance of a random sample from a Normal population, t & F distributions; Distributions of mean, variance & correlation coefficient (null case) of a random sample from a Bivariate Normal population, Distribution of simple Regression Coefficient (for both stochastic & non stochastic independent variable cases). Distribution of Order Statistics (Basic results and simple problems). (25L)

Elements of Statistical Inference (SI400)

Elements of Point Estimation: Concepts of Point & Interval Estimation, Requirements of a good estimator- notions of Mean Square Error, Unbiasedness, Minimum Variance, Best Linear Unbiasedness. Methods of Estimation- Method of Moments & Least Square method. (10L)

Elements of Hypothesis Testing and Interval Estimation: Null & Alternative hypotheses, Simple & Composite hypotheses, Critical Region, Type I & Type II Errors, Level of Significance & Size, p-value, Power. Confidence Interval and Confidence Coefficient. (10L)

Applications: Estimation, Tests of Significance related to: single Binomial proportion & Poisson parameter, Two Binomial proportions & Poisson parameters. The mean & variance of an univariate Normal distribution, The difference of means & ratio of variances of two independent Normal distributions & associated confidence intervals. Difference of means, Ratio of variances & the regression & correlation coefficients of a Bivariate Normal distribution. (20L)

Linear Algebra-II (LA402)

Matrix Algebra: Matrix operations, Partitioned matrices, Matrix inverse, Orthogonal matrix. Row space & column space of a matrix, Rank of a matrix, Elementary matrices & their uses, Null space & nullity. Consistency of a system of linear equations. Linear transformation & matrices (ideas only). (19L)

Quadratic Forms: Classification & related results, Characteristic roots & vectors (definitions only), Canonical reduction. (6L)

Statistical Quality Control (SQ400)

Introduction: Concepts of Quality & Quality Control, Process Control & Product Control. (5L)

Process Control: Rational Subgroups, 3-sigma limits & Probability limits. Control charts & their uses. Choice of Subgroup sizes. Construction of \bar{X} , R, SD, np, p & c charts (fixed & variable sample sizes). Interpretation of non-random patterns of points. Modified control charts. (20L)

Product Control: Description of Rectifying Sampling Inspection Plans. Producer's Risk, Consumer's Risk, Single & Double sampling plans by attributes, OC, ASN (ATI), AQL, LTPD & AOQL. Single sampling plan for inspection by variables (one sided specification, known & unknown sigma cases), Selection of plans using IS2500 plans & tables. (25L)

Multivariate Analysis & Large Sample Theory-I (ML501)

Multivariate Analysis

Multivariate Data: Multiple correlation, Partial correlation & their properties. Multiple regression & related results. (14L)

Random Vector: Probability mass & density functions, Distribution function, Mean vector & Dispersion matrix, Marginal & Conditional distributions, Ellipsoid of Concentration, Multiple correlation, Partial correlation. (16L)

Multivariate Distributions: Multinomial & Multivariate Normal distributions and their properties. (12 L)

Theory of Estimation & Testing of Hypothesis-I (ET501)

Point Estimation: Sufficiency, Completeness, Factorization theorem (proof for discrete cases only), Exponential family of distribution, Properties of Minimum Variance Unbiased Estimators, Consistent estimators & Asymptotic efficiency. Cramer-Rao lower bound (derivation for continuous cases only), Rao-Blackwell theorem, Lehmann-Scheffe theorem, Maximum Likelihood estimators & their properties (excluding proofs of large sample properties). Method of Minimum χ^2 (38L)

Theory of Hypothesis Testing: Most Powerful (MP), Uniformly Most Powerful (UMP) & Uniformly Most Powerful Unbiased (UMPU) tests. Randomized & non-randomized tests. Fundamental Neyman-Pearson Lemma (proof for the sufficiency part only) & its use in the construction of MP & UMP tests (single parameter with range independent of the parameter), Monotone Likelihood Ratio. (32L)

Sample Survey Methods-I (SS501)

Introduction: Concepts of a finite population & a sample, Need for sampling, Complete enumeration & sample surveys, Judgment & Probability sampling. (5L)

Sampling & Estimation Procedures: Simple Random Sampling with & without replacement. Random Number Tables & their uses. Stratified Random Sampling. Cluster Sampling, Two-Stage (with equal sized first stage units) Sampling with equal selection probabilities at each stage; Associated unbiased estimators of population total, mean & proportion, Associated variances & unbiased variance estimators. Determination of sample size in simple random sampling. Allocation problem in stratified random sampling, optimum choice of sampling & sub-sampling fractions in two-stage sampling. (35L)

ANOVA REGRESSION AND ANCOVA (AD501)

Linear Model: Statement of Gauss-Markov theorem (without proof), Estimability & BLUE, Orthogonal splitting of total variation & statement of Fisher-Cochran theorem (without proof), Linear hypothesis. (20L)

ANOVA: One-way classified data, Two-way classified data with equal number of observations per cell for fixed & random effects model only.

Analysis of Regression: Testing simple regression coefficients, Correlation ratio, Linearity of simple regression, Multiple correlation & Partial correlation coefficients. (20L)

ANOCOVA: (10L)

Large Sample Theory (ML602)

Variance Stabilization: Transformation of Statistics, Derivation and use of \sin^{-1} , square root, logarithmic & Fisher's Z- transformations. (8L)

Pearsonian χ^2 : Large Sample distribution of Pearsonian χ^2 statistic & its uses (goodness of fit, independence, homogeneity). Yates' correction in a 2x2 contingency table. (10L)

Standard Errors of Statistics: Derivation of large sample standard error of sample moments, Standard deviation, Coefficient of Variation, b_1 & b_2 measures, Correlation coefficient & their uses in large sample tests. Asymptotic distribution of sample quantiles. (17L)

Large Sample Tests: Binomial proportions, Poisson means (single & two independent sample cases). (3L)

Theory of Estimation & Testing of Hypothesis-II (ET602)

Theory of Hypothesis Tests: Likelihood Ratio Tests & its applications to test for the equality of means & variances of several Normal populations. Combination of Probabilities in tests of significance. Wald's Sequential Probability Ratio Test for simple hypothesis against simple alternatives, Approximate expressions for OC & ASN functions (without proof), Tests regarding Binomial proportion & Normal mean. (25L)

Interval Estimation: Confidence Intervals (Exact & Asymptotic), Shortest expected length. (5L)

Non-Parametric Methods: Sign test, Median test, Wilcoxon Signed-rank test, Run test, Mann-Whitney U test, Confidence limits for quantiles, Tolerance limits. (20L)

Sample Survey Methods-II (SS602)

Sampling & Estimation Procedures: Linear & Circular Systematic sampling, Interpenetrating Sub-Sampling technique for unbiased variance estimation in systematic sampling. Ratio & Regression methods of estimation in simple random sampling. Double Sampling for ratio & regression estimators. (35L)

General Ideas: Planning & Execution of sample surveys, Designing questionnaires & schedules, Types of non-sampling errors. (5L)

Design of Experiments (AD602)

Principles of Experimental Design: Randomization, Replication & Local control. Uniformity trials, Shapes & sizes of plots & blocks. (5L)

Standard Designs & Analysis: Completely Randomized Design (CRD), Randomized Block Design (RBD), Latin Square Design (LSD), Comparison of efficiencies, Missing plot techniques (one missing observations). (25L)

Groups of Experiments: Groups of experiments using RBD & LSD. (10L)

Factorial Experiments: 2ⁿ experiments- Uses & Advantages, Total & Partial Confounding, Analysis. (20 L)

PRACTICAL

SEMESTERS 1-4:

Data Analysis using MINITAB
Elements of C-Programming

SEMESTER 5

R-Module (25 marks)
Data Analysis using MINITAB (75 marks)

SEMESTER 6

Project (25 marks)
Data Analysis using MINITAB (75 marks)

Continued...

STATISTICS (GENERAL) SYLLABUS

Descriptive Statistics- I (DS111)

Types of Statistical data. Population & Sample. Compilation & Classification of data. Tabulation & Diagrammatic representation of data. Frequency distribution, Cumulative frequency distribution & their graphical representation- Histogram, Frequency polygon, Frequency curve & Ogive. (10L)

Analysis of Univariate Quantitative Data- Concepts of central tendency, dispersion, relative dispersion, skewness & kurtosis. Measures based on quantiles & moments. (10L)

Analysis of Categorical Data- Independence & Association of attributes, Measures of association- Odds Ratio, Yule's & Cramer's measures. Pearson's measures for $k \times l$ contingency table. (5L)

Probability Theory- I (PT111)

Random experiments, Events, Statistical regularity & meaning of probability, Classical & Axiomatic definitions of probability (discrete sample space only). Conditional probability, Independence of events, Principal theorems including union & intersection of events. Bayes' theorem. Examples based on classical approach. Repeated trials. (15L)

Random variable & its probability distribution, Cumulative distribution function (cdf), Probability mass function (pmf), Probability density function (pdf), Mathematical expectation, variance & moments. (10L)

Descriptive Statistics- II (DS212)

Analysis of bivariate quantitative data- Scatter plot, Product-moment correlation coefficient & Coefficient of determination, Regression analysis, Fitting of linear & quadratic curves by the principles of least squares. Spearman's rank correlation coefficients. (12L)

Analysis of multivariate quantitative data- Multiple regression, Multiple correlation & Partial Correlation in three variables- Measures & related results. (10L)

Fitting of Binomial, Poisson & Normal distributions (3L)

Probability Theory- II (PT212)

Joint distribution of two random variables, Marginal & Conditional distributions, Covariance & Correlation, Simple theorems on expectation & variance of a sum of random variables, Expectation of product of random variables. (7L)

Standard univariate discrete distributions & their properties- Uniform, Binomial, Hypergeometric, Geometric & Poisson distributions. (6L)

Standard univariate continuous distributions & their properties- Uniform, Normal, Exponential, Gamma and Beta (5L)

Bivariate Normal distribution & statement of its general properties. (2L)

Chebychev's inequality, Weak law of large numbers, Bernoulli's theorem, Statement of Central Limit theorem (iid case) & its uses. (5L)

Sampling Distribution and Statistical Inference-1 (SI 311)

Sampling Distributions

Population and sample, Random Sampling and Sampling Distributions of Statistics, Sampling Distributions of sum of independent Binomial and Poisson variables, Transformation of continuous variables (2 – variable case only) and simple problems. χ^2 , t and F distributions (derivations excluded), sampling distributions of mean and variance of independent Normal variables (derivations excluded). (15L)

Point Estimation

Point Estimation of population parameter, Bias and Standard Error of an estimator, concepts of Unbiasedness, Minimum Variance, Consistency and Asymptotic Efficiency of an estimator. Method of moments and Maximum Likelihood method, Point estimators of the parameters of Binomial, Poisson and Univariate Normal distributions. (10L)

Sample Survey (SE 311)

Concepts of a finite population and a sample, Need for sampling. Complete enumeration and Sample Surveys. Judgment and Probability sampling. Sampling Error. Simple random sampling with and without replacement – associated unbiased estimators of population mean, total, proportion & ratio, their variances and variance estimators. Random Number Tables and their uses. (15L)

Stratified random sampling, Cluster sampling (equal sizes) – associated unbiased estimators of population mean, total and proportion, their variances and variance estimators. Notions of Two Stage Sampling. (10L)

General Ideas: Planning and execution of sample surveys. Questionnaires and Schedules. Non-Sampling Errors. (10L)

Sampling Distribution and Statistical Inference-2 (SI 412)

Testing of Hypothesis and Interval Estimation

Testing of Hypotheses – Null and Alternative hypotheses, Types of Errors, Critical Region, Level of Significance, Power and p-value, Exact tests of hypotheses under Normal set-up for a single mean, the equality of two means, a single variance and equality of two variances, Test of significance of sample correlation coefficient (null case) and tests of hypotheses for the equality of means and equality of variances of a bivariate Normal distribution. (12L)

Interval estimation – Confidence Interval and Confidence Coefficient, Exact confidence interval under Normal set-up for a single mean, single variance, the difference of two means and the ratio of two variances. (5L)

Large Sample Tests and related Interval Estimates of a single mean and a single proportion and difference of two means and two proportions. Pearsonian χ^2 tests for Goodness of fit and Independence in a contingency table. (8L)

Anova & Design of Experiments

Analysis of Variance in one-way classified data and two-way classified data with equal number of observations in each cell (fixed effects models only).

Basic principles of design of experiments: Randomization, Replication and Local Control. Completely Randomized Design and Randomized Block Design: layouts and analyses including paired comparisons. Factorial Experiments – main effects and interaction effects in 2^2 experiments. (10L)

Economic Statistics and Indian Statistical System (EI 510)

Economic Statistics

Index Numbers: Price, Quantity and Value indices. Price Index Numbers: Construction, Uses, Limitations, Tests of Index Numbers, Various formulae and their comparisons. Chain Index Numbers. Some important indices: Consumer Price Index, Wholesale Price Index, Index of Industrial Production – methods of construction and uses. Measures of Inequality: Gini's coefficient, Lorenz curves (20L)

Population Statistics (PS 510)

Sources of population data: census data and registration data, Rates and Ratios of Vital Events, Measurements of Mortality: Crude Death Rate, Specific Death Rate, Standardized Death Rate, (6L)

Measurement of Fertility and Reproduction: Crude Birth Rate, General Fertility Rate, Age Specific Fertility Rate and Total Fertility Rate, Crude Rate of Natural Increase, Vital Index, Gross and Net Reproduction Rates. (10L)

Time Series Analysis (TS 610)

Observations recorded in order of time as a sequence of random variables. Examples from various fields. Different types of variation and their modelling: Seasonal variation, Trend, Cyclical and Stationary. (5L)

Analysis of Seasonal Variation: Method of Averages, Analysis of Trend: Method of Simple Moving Averages and Method of Regression (fitting of linear, log-linear and quadratic curves). Analysis of Composite time series. (7L)

Statistical Quality Control (SQ 610)

Concepts of Quality and Quality Control, Process and Product Control. (3L)

Process Control: Rational Sub-groups, 3-sigma limits. Control Charts and their uses, choice of Sub-group sizes, Construction of \bar{X} , R, S, np charts, p and C charts (fixed sample sizes), interpretation of non-random pattern of points. (8L)

Single Sampling Plans by attributes (4L)

PRACTICAL

Practical involves solution of numerical problems from topics covered in the syllabus using a scientific non-programmable calculator.