Syllabus
M.Sc. Statistics

Students with B.Sc. Statistics / Mathematics or studied at least one paper of Statistics at graduation level of B.Sc. or B.Com. with advance statistics are applicable for this course.

<table>
<thead>
<tr>
<th>Paper no.</th>
<th>Title of Paper</th>
<th>No. of Hrs. Per Week</th>
<th>Weightage For Internal Examination</th>
<th>Weightage For Semester end Examination</th>
<th>Total Marks</th>
<th>Duration of Semester end Exam in Hrs.</th>
<th>Course Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS-101</td>
<td>Basics of Statistical Methods</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-102</td>
<td>Statistical Computing and Numerical Methods</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-103</td>
<td>Statistical Inference and Non – Parametric Test</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-104</td>
<td>Probability and Distribution Theory</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-105</td>
<td>Mathematical Statistics</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-106</td>
<td>Practical</td>
<td>12</td>
<td>-</td>
<td>100</td>
<td>100</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>32</td>
<td>150</td>
<td>450</td>
<td>600</td>
<td>20</td>
<td>24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paper no.</th>
<th>Title of Paper</th>
<th>No. of Hrs. Per Week</th>
<th>Weightage For Internal Examination</th>
<th>Weightage For Semester end Examination</th>
<th>Total Marks</th>
<th>Duration of Semester end Exam in Hrs.</th>
<th>Course Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS-201</td>
<td>Data Warehousing and Data Mining</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-202</td>
<td>Planning and Analysis of Industrial Experiments</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-203</td>
<td>Applied Multivariate Analysis</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-204</td>
<td>Sampling Techniques</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-205</td>
<td>Base SAS Programming</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-206</td>
<td>Practical</td>
<td>12</td>
<td>-</td>
<td>100</td>
<td>100</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>32</td>
<td>150</td>
<td>450</td>
<td>600</td>
<td>20</td>
<td>24</td>
</tr>
</tbody>
</table>
### Semester - 3

<table>
<thead>
<tr>
<th>Paper no.</th>
<th>Title of Paper</th>
<th>No. of Hrs. Per Week</th>
<th>Weightage For Internal Examination</th>
<th>Weightage For Semester end Examination</th>
<th>Total Marks</th>
<th>Duration of Semester end Exam in Hrs.</th>
<th>Course Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS-301</td>
<td>Survival analysis and Clinical Trials</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-302</td>
<td>Industrial Statistics</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-303</td>
<td>Optimizing Techniques</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-304</td>
<td>Stochastic Process</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-305</td>
<td>Applied Econometrics</td>
<td>4</td>
<td>30</td>
<td>70</td>
<td>100</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MS-306</td>
<td>Practical</td>
<td>12</td>
<td>-</td>
<td>100</td>
<td>100</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>32</td>
<td>150</td>
<td>450</td>
<td>600</td>
<td>20</td>
<td>24</td>
</tr>
</tbody>
</table>

### Semester – 4

<table>
<thead>
<tr>
<th>Paper no.</th>
<th>Title of Paper</th>
<th>No. of Hrs. Per Week</th>
<th>Weightage For Internal Examination</th>
<th>Weightage For Semester end Examination</th>
<th>Total Marks</th>
<th>Duration of Semester end Exam in Hrs.</th>
<th>Course Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS-401</td>
<td>Project &amp; Viva</td>
<td>32</td>
<td>-</td>
<td>300</td>
<td>300</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>32</td>
<td>-</td>
<td>300</td>
<td>300</td>
<td>20</td>
<td>24</td>
</tr>
</tbody>
</table>
SEMESTER -1

MS – 101: BASICS OF STATISTICAL METHODS
1. Exploratory data analysis and Descriptive Statistics: Random Variables, Types of Variable and Data Types, Graphical Displays of Sample Data, Histograms, Box plot, Scatter plot, Bar chart, Measures of Centre Tendency, Measures of Dispersion, Moments, Skewness and Kurtosis.
3. Classical Probability Distributions:
   Discrete Distributions: Bernoulli, Binomial, Poisson, Negative Binomial, Geometric, Hypergeometric,
   Continuous Distributions: Normal, Uniform, Gamma, Beta distribution of first kind, Beta distribution of second kind, Exponential, Weibul, Cauchy, Central Limit Theorem.
4. Sampling Distributions: Chi-Square Distribution, t-distribution, F- distribution.
5. Statistical Inference and Hypothesis Testing: One sample tests, two sample tests, several sample tests. Applications: Case-Control Studies, Test of Association.
6. Correlation and Regression: Karl Pearson’s Coefficient of Correlation, Spearman’s rank correlation coefficient, Linear Regression.
7. SPSS: Introduction to SPSS, Statistical analysis using SPSS.

References
3. B. L. Agarwal: Basic Statistics.

MS – 102: STATISTICAL COMPUTING AND NUMERICAL METHODS
2. Generation of continuous random variables covering Exponential, Normal, Gamma, Chi-square, Bivariate exponential, Bivariate Normal distributions, and mixture of distributions.
3. Excel Introduction to MSEXCEL and exercises on using EXCEL for Statistical analysis covering frequency distribution, histograms, t-test, and test for Independence in 2x2 contingency tables.
5. Solution to a nonlinear equation: Bisection method, Newton-Raphson method.
6. Iterative methods: Jacobi, Gauss-Seidel methods with convergence analysis.
References

MS – 103: STATISTICAL INFERENCE AND NON PARAMETRIC TESTS
3. Test of hypothesis: simple and composite hypothesis, two types of errors, critical regions, randomized tests, power function, and most powerful and most powerful tests. Neyman Pearson lemma, generalized Neyman Pearson lemma.
4. Unbiased tests: uniformly most powerful unbiased test, similar test, relation between UMP unbiased test and a UMP similar test, application to one parameter exponential family. Tests with Neyman structure. Inference on scale and location parameters: estimation and tests.
5. Rank test, locally most powerful rank test, linear rank statistics and their distributional properties under null hypothesis.

Reference:

MS – 104: PROBABILITY AND DISTRIBUTION THEORY
2. Law of large numbers: Weak law of large numbers, Strong law of large numbers for i.i.d. sequence, strong law of large numbers, Kolmogorov’s strong law of large numbers. Borel 0-1 law, Borel-Cantelli lemma, Kolmogorov 0 – 1 law.
3. Convergence in Probability, in distribution and in mean. Central limit theorem for a sequence of independent random variables under Lindberg’s condition, Liapounov’s CLT (only statement).
4. Brief review of basic distribution theory. Joint, Marginal and Conditional probability mass function (pmf) and probability density function (pdf). Discrete and continuous distributions.
5. Function of random variables and their distributions using Jacobian of transformation and their tools, probability distribution of a random variables, properties of distribution functions, Characteristics functions and its properties, Inversion theorem, uniqueness theorem and Convolutions.
6. Power series distribution: its mean, variance, mgf, cgf, and recurrence relations. Various discrete distributions as its particular cases.
7. Sampling distributions: Non central chi square, t and f - distributions and their properties. Distributions of quadratic form under normality.
8. Order statistics: their distribution and properties, joint and marginal distributions of order statistics. Extreme values and their asymptotic distributions (Statement only) and its applications.

References:
MS – 105: MATHEMATICAL STATISTICS

1. Set of real numbers, countable and uncountable sets, countability of rationals and uncountability of the interval (0,1) Supremum and Infimum of bounded sets, limit point of a set, open, closed, dense and compact sets. Bolzano-Weierstrass and Heine-Borel Theorems (Statements only). Applications of these theorems.
2. Sequence of real numbers, convergence, divergence. Cauchy sequence. Convergence of bounded monotone sequence. Limit inferior and limit superior of the sequences.
3. Series of numbers, tests for convergence (without proof) test for absolute convergence, convergence of sequences of non-negative terms.
4. Vector space, subspace, linear dependence and independence, basis, dimension of a vector space, example of vector spaces.

References:

MS-106 Practical
MAS-106 (a) Practical Based on MAS-101 & MAS-103
MAS-106(b) Practical Based on MAS-102
**SEMESTER -2**

**MS-201: Data Warehousing and Data Mining**

1. Introduction of Data Warehouse: Operational and Informational systems, OLTP and DSS systems, Characteristics of Data Warehouse, Data Warehouse software and hardware architecture, Basic steps to develop data warehouse architecture, Architectural components of data warehouse, Data warehouse system architecture (Two-Tiered and Three-Tiered).

2. Data Marts: Data Mart structure, Usage of Data Mart, Security in Data Mart, Data warehouse and Data Mart

3. Online Analytical Transactional Process: OLTP and OLAP systems, Types of OLAP (MOLAP, ROLAP and HOLAP) with advantages and disadvantages.

4. ETL: Extraction of Data, Transformation of Data, Loading of Data, Comparison and contradiction of various ETL tools, Practical study of popular ETL tools.

5. Data Mining: Foundation of Data Mining, Data Mining Process, Data Understanding, Data Preparation, Creating database for data mining, Exploring database, creating for data mining model, building a data mining model, evaluating a data mining model, deployment of data mining model.


9. Case Study: Insurance, Financial services, Healthcare and medicine, Telecommunications, Retail Marketing, Government, Education,

**Reference Books:**
(1) Data mining Explained A manager’s guide to customer centric business intelligence by Rhonda Delmater, Monte Hancock, Digital Press
(2) Data mining by Pieter Adriaans, DolfZantinge
(4) Data warehousing in the real world A practical guide for business DSS by Sam Analory, Dennis Murray

**MS - 202: PLANNING AND ANALYSIS OF INDUSTRIAL EXPERIMENTS**

1. Introduction to designed experiments. General block design and its information matrix C. Properties of block design: Connectedness, balance and orthogonality.

2. Balanced incomplete block design, its properties, parametric relations, intra block analysis of BIB design. Finite group and finite field geometry projective and Euclidean. Mutually orthogonal lattice square design. Construction of (1) MOLS and (2) BIB designs using MOLS, PG (N, S), EG(N, S) and other methods.

4. General factorial experiment, main effects and interaction effects. $2^n$ and $3^n$ factorial experiment. Analysis of $2^n$ and $3^n$ factorial experiments in randomized block. Confounding experiments: complete partial and balanced confounding and its ANOVA table.

5. Two associated PBIB design association scheme and intra block analysis. Group divisible designs, dual and linked block designs, resolvable and affined resolvable designs.

1. Diallel Crosses: complete diallel crosses, its analysis and efficiency factor, optimal diallel crosses plane. Robustness of designs. Robustness of diallel crosses plan against the lost of $1<=s<=k$ observation in a block.

References:

4. Das,M.N. : Design of Experiment
5. Alok, Dey : Incomplete Block Design

MS - 203: APPLIED MULTIVARIATE ANALYSIS

1. Multivariate normal distribution, two definitions and their equivalence, singular and nonsingular normal distribution, characteristic function, moments, marginal and conditional distributions.
2. Maximum likelihood estimators of the parameters of the multivariate normal distribution and their sampling distributions.
5. Discrimination and classification. Fisher’s discriminant function and likelihood ratio procedure, minimum ECM rule, Rao’s U statistics and its use in tests associated with discriminant function, classification with three populations.
6. Principal components. Dimension reductions, Canonical Correlation and canonical variables.

Reference:


MS - 204: SAMPLING TECHNIQUES
1. Concept of population and sample, Need for Sampling, census & sample surveys, basic concepts in sampling and designing of large-scale surveys design, sampling scheme and sampling strategy. Basic methods of sample selection: SRSWR, SRSWOR.
3. Systematic sampling: The sample mean and its variance, comparison of systematic with random sampling, comparison of systematic sampling with stratified sampling, comparison of systematic with simple and stratified random sampling for certain specified population. Estimation of variance, Two stage sample: Equal first stage units, Two stage sample: Unequal first stage units; systematic sampling of second stage units.
4. PPSWR methods: Cumulative total method, Lahiri’s method related estimation Problems and PPSWOR methods and related estimation of a finite population mean (Horwitz- Thompson and Des Raj estimators for a general sample size and Murthy’s estimator for a sample of size 2), Midzuno sampling.
5. Use of supplementary information for estimation: ratio and regression estimators and Their properties. Unbiased and almost unbiased ratio type estimators, Double sampling.
6. Cluster sampling. Two – stage sampling with equal number of second stage units.

References:

MS - 205 Base SAS Programming
1. Introduction to SAS programs, running SAS Programs, diagnosing and correcting syntax errors. Producing List Reports using PRINT procedure; sequencing and grouping observations, using special WHERE statement operators; customizing report appearance – formatting data values, creating HTML reports. Programming with the DATA step – reading SAS data sets and creating variables, executing statement conditionally, dropping and keeping variables. Assigning and Changing variable attributes, Combining merging and SAS Data Sets Producing Summary Reports using REPORT procedure.

3. Using SQL with SAS: Understanding the purpose, design, uses, and terminology of SQL; Basic Queries, using SQL procedure, summarizing data with column and row functions, grouping data, performing analyses on groups of data, subquerying, and remerging, ordering data, customizing query output. Combining Tables – querying multiple tables using joins, using union, intersect, and other set operators to combine tables. Creating and Modifying Tables and Views, using views to simplify queries and access changing data, creating and using indexes; maintaining tables, views, and indexes.

Note: The examination of this paper is based on short answer questions of one mark each.

Reference:

MS - 206 PRACTICAL
MS-206 (a) Practical based on MS-201, MS-202
MS-206 (b) Practical based on MS-205
SEMESTER-3

MS - 301 SURVIVAL ANALYSIS AND CLINICAL TRIALS
1. Life time distributions, survival functions, hazard rate, cumulative hazard function, residual life time, survival function of residual life time, mean residual life time, one correspondence of these functions. Computation of these function for Common life time distributions: exponential Weibull, Gamma, Pareto, Rayleigh, log-normal etc: computation of survival and failure rate function proportional hazard models and proportional hazard model.
2. Concept of censoring, various types of censoring, Estimation and Testing of parameters of exponential distribution under various types of censoring.
5. Introduction to clinical trials and other types of clinical research, bias and random error in clinical studies, overview of Phase I-IV trials.
6. Design of Phase 1-3 clinical trials: parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, review of factorial designs, objectives and endpoints of clinical trials, formulation of appropriate hypotheses (equivalence, non-inferiority, etc.); sample size calculation.
7. Analysis of Phase 1-3 trials: Use of generalized linear models; analysis of categorical outcomes.

References:

MS - 302: INDUSTRIAL STATISTICS
1. Basic concept of quality control, process control and product control, seven SPC tools Flowchart. Histogram, Check sheet. Ishikawa diagram, Pareto chart, Defect
concentration diagram, control chart. Deming’s PDCA cycle for continuous improvements and its applications.

2. Control charts for measurements and attributes $x$, $R$, $S$, $p$, np. Charts with subgrouping, CUSUM chart, tabular form and V-mask use of these charts for process control. Moving average and exponentially weighted moving average charts.

3. Sampling Inspection plans: for attribute inspection: Single, double & sequential sampling plans and their properties. Dodge & Roming characterization by OC curve and ARL-Inspection by variables for one or two sided specifications.


5. Simulation of X-bar and R control charts, estimation of ARL and process capability indices.

References:

MS - 303: OPTIMIZATION TECHNIQUES
1. Linear programming problem (LPP): Theorems related to the development of Simplex algorithm, Proof of the theorems related to a basic feasible solution (b.f.s); Reduction of a feasible solution to a basic feasible solution, Improvement of a basic feasible solution, Existence of unbounded solution, Optimality conditions. For other related theorems, statements only.
2. Artificial variable technique; two phase and Big M method, the case of redundancy. Revised simplex method.
3. Concept of Duality, theorems related to duality, complementary slackness property and development of dual simplex algorithm.
4. Theory of games: two person zero sum games, minimax and maximin principles, Saddle point, mixed strategies; rules of dominance, solution of 2 x 2 game by algebraic method, 13 Graphical method, Reduction of the game problem as LPP, Minimax and maximin theorem (without proof).
7. Replacement Problem, Project Management: PERT and CPM, Inventory Control Models, Queuing theory.

References:

**MS - 304: STOCHASTIC PROCESSES.**

1. Introduction to stochastic process (sp’s): classification of sp’s according to state space and time domain. Countable state markov chain (mc’s), stationary process, classification of states, transition probability, Chapman kolmogorow questions, calculations of n steps transition probability (higher transition probability) and its limit.

2. Random walk and gambler’s ruin problem: effect of changing bet, duration of game, probability of gamblers ruin in exactly n games, one and two dimensional random walk.

3. Poisson process: introduction, probability mass function, probability generating function and property of Poisson process.


5. Branching process: Galton Watson branching process, its mean and variance, probability of ultimate extinction.

**References**

5. Ross, S.M. stochastic process, new age international, New Delhi.

**MS - 305 APPLIED ECONOMETRICS**


5. Estimation on simultaneously equation model, 2 SLS estimators, limited information estimator, 3 SLS estimation, full information maximum likelihood method. Monte Carlo studies and simulation.
Reference:

MS - 306 PRACTICAL
MS-306 (a) Practical based on MAS-301 & MAS-302
MS-306 (b) Practical based on MAS-305

SEMESTER-4

MS – 401: Project and Viva