Ahmednagar Jilha Maratha Vidya Prasarak Samaj's

New Arts, Commerce and Science College, Ahmednagar (Autonomous)
(Affiliated to Savitribai Phule Pune University, Pune)

Choice Based Credit System (CBCS)

F.Y.B. Sc. Detailed Syllabus

Bachelor of Science (B. Sc.) Statistics (F.Y.B.Sc.)

Implemented from

Academic year 2021 -22

F.Y. B. Sc.(Statistics)

Semester - I

Course Type	Course Code	Course Title	Credits
DSCC A - 1	BSC-ST 101 T	Descriptive Statistics- I Using MS-Excel	02
DSCC A - 2	BSC-ST 102 T	Introduction to Probability	02
DSCC A -3 Practical	BSC-ST 103 P	Practical (Based on MS-Excel)-I	1.5

1. Prologue/ Introduction of the programme:

It is known that in economic activities are of three types, agriculture, industrial and service. In the same way the subject Statistics is a SERVICE SCIENCE having potential to address the problems in these three fields. In research application of Statistics is mandatory. In the present days, apart from traditional field of career, Data Science, Data Analytics, Data Mining, Data Visualization are the upcoming field of career for Statistics students. In these field student must have mathematical ability, statistical thinking, computer (Software and programming) knowledge and communication (Verbal and written). These points are taken into consideration to design the syllabus and examination pattern of Statistics. In addition to academics, the department takes care to arrange a series of lectures on interview skills, preparation of CV, improve communication skill and overall personality development. The students are given the task of event management so that they can practice the principles of management such as leadership, creativity, communication, time management, group activity, team work, etc. In general, through curricular, co-curricular and extra-curricular activities student in three years is developed as thought provoker, problem solver, technologically sound, with command on communication, strong self-confidence.

B. Sc. in Statistics program is of three years' duration, with semester pattern for all the three years. The important feature of the syllabus is that, all practical's form first year to third year will be conducted on computer using MS-EXCEL/ R Suit, Python programming and Tableau. The another feature is at the in all courses, end of every chapter, self-learning activities are listed. These self-learning activities will play important role in creating interest in the subject and also boost their confidence. Further group activities will give the chance to explore their creativity and ideas. In addition, the verbal and written communication will be improved. These self-learning activities are expected to motivate students to participate in various student related academic events organized by home college or by other colleges too.

The course on Tableau will give an opportunity to learn thousands of various data presentation types and to present the complex data by easy way. The practical examinations of all courses will be on computer.

In T.Y.B.Sc. examination of one theory course at each semester will be on computer. In short, maximum exposure is given to students to work on computer and evaluate them on computer.

The syllabus is framed with appropriate weightage of theory, applied and skill enhancement courses. After receiving B.Sc. degree, student is expected to have minimum knowledge of various courses and student will have ability to analyse the data with relevant interpretation of results.

2. Programme outcomes (PoS)

Students enrolled in the program complete a curriculum that exposes and trains students in a full range of essential skills and abilities. They will have the opportunity to master the following objectives.

1. Student will achieve the skill of understanding the data.
2. Student will be able to develop the data collection instrument.
3. Student will have skill to write a story using data visualization.
4. Student will understand the interdisciplinary approach to correlate the statistical concepts with concepts in other subjects.
5. Student will be made aware of history of Statistics and hence of its past, present and future role as part of our culture.
6. Students will demonstrate conceptual domain knowledge of the Statistics in an integrated manner.
7. Student will play the key role in management for effective functioning of MIS.

Semester -I	Paper -I
Course Code: BSC-ST 101 T	Title of the Course: Descriptive Statistics-I Using MS Excel
Credits: 02	Total Hours: 30

Course Outcomes (CoS):

1. The newly introduced concept, "Self-Learning Activity" is a unique feature of this course. With due participation in this activity student will learn all concepts taught in this course by self-practice.
2. Along with regular teaching and learning, student will groom to design posters and power point presentations.
3. Teaching-learning and Self-Learning Activity will improve the process of logical thinking.
4. Students will be aware of the variety of fields in which Statistics is used widely.
5. Student will also gain the knowledge of computational tools.

Detailed Syllabus:

Unit I: (8L)

1) Introduction of Statistics
   - 1.1 Meaning of Statistics
   - 1.2 Importance of Statistics
   - 1.3 Scope of Statistics (Field of Industry, Medical Science, Economics, Social Science, Biological Science, Agriculture, and Psychology, Clinical Trial, Decision Theory)

Self-Learning Activity:

1. Prepare a poster of applications of Statistics
2. Prepare a poster on biography of Statistician
3. Power Point Presentation on definitions of Statistics

2) Statistics in Data Science:
   - 2.1 Concepts of big data, properties of big data- velocity, volume, varsity, verity
   - 2.2 Applications: Fraud detection in Banking, Costumer churn

Self-Learning Activity:

1. Power Point Presentation on Big data
2. Power Point Presentation on applications of statistics in data science
3. Write a note any one big data
4. Collect the list of organization where decisions are based on big data

Types of data:

3.1 Nature of Data:

Attributes (qualitative): Nominal and Ordinal Scale, Likert scale

Variables (Quantitative): Interval scale, ratio scale, and discrete scale and continuous scale

3.2 Primary data and Secondary data.

3.3 Data collection methods: register, questionnaire, interview method

3.4 Categorical data, directional data, Binary data, time series data, Panel data and

3.5 Cross sectional data.

3.6 Image, Voice, Audio, Animated images, Text, Video data

Self-learning activity:

1. Identify attribute and variables according to their nature in one of the big data
2. Conduct a small survey
3. Prepare a questionnaire,
4. Prepare an interview Schedule
5. Prepare a note on real life examples of Categorical data, directional data, Binary data, time series data, Panel data and Cross sectional data.
6. Prepare a note on real life examples of Image, Voice, Audio, Animated images, Text, Video data

Unit II: (7L)

1) Sampling methods

1.1 Definition of population: Finite population, Infinite population, Homogenous population, Heterogeneous population

1.2 Definition of Sample: Types of sampling method (Probability and Non probability sampling and their types (only description)

Probability sampling - SRS, SRSWR, SRSWOR, Stratified, Systematic, Cluster sampling.

Non probability sampling- Judgment, Quota, Convenience, snowball sampling

1.3 Using MS Excel function draw different types of random samples

Self-learning activity:

1. Visit National Sample Survey Organization website and write a note
2. Power Point Presentation on sampling methods
3. Conduct an experiment illustrating different sampling methods.

Collect different sampling frames (ex; voter list, blood donors)

2) Data Presentation:

2.1 Frequency classification: Raw data, data, Tally mark, frequency distribution (Tabulation) by Sturge's rule, grouped and ungrouped data, inclusive, Exclusive, open end classes, cumulative frequency, relative frequency.

2.2 Diagrammatic representation: Pictograms, Cartograms, Bar Diagrams (simple bar diagrams, multiple bar diagrams and sub-divided bar, diagrams) and Pie Diagrams

2.3 Graphical Representation: (Box and Whisker Plot, Time series plot [Line Graph, Bar graph, Histogram, Stem and leaf chart, Frequency polygon, Frequency curve, ogive curve] (to be covered in practical)]

2.4 Using MS Excel for diagrammatic representation for data visualization in topic 2.2.

2.5 Using MS Excel for graphical representation for data visualization in topic 2.3.

Self-Learning Activity:

1. Conduct an experiment to demonstrate discrete frequency distribution
2. Conduct an experiment to demonstrate Continues frequency distribution using Sturge's rule
3. Present data in 1. and 2 using appropriate diagram or graph or chart
4. Give real life examples where open end classes can be used

Unit III: (8L)

1)Measures of Central Tendency:

1.1 Concept and Definition of Central Tendency

1.2 Characteristics of good measures of Central Tendency.

1.3 Types of central Tendency;

Arithmetic Mean (A.M): Definition of Mean, formulae for ungrouped and grouped data (without proof) Properties of A.M, Trimmed AM, Weighted AM

Median: Definition of Median, Formulae for ungrouped and grouped data, Graphical representation

Partition values: Quartiles, Deciles, Percentiles, Quantiles, and their interrelationship

Mode: Definition of Mode, formulae for ungrouped and grouped data. Graphical Representation. Empirical relation between mean, median and mode

Geometric mean: Definition of G.M, formulae,

Harmonic Mean: Definition of H.M, formulae

merits and demerits of AM, Median, Mode, HM, GM, Relation between A.M, G.M and H.M

1.4 Using MS Excel functions find Mean (AVARAGE), median (MEDIAN), Quartiles (QUARTILE), Percentiles (PERTCENTILES) mode(MODE), GM(GEOMEAN) and Harmonic Mean(HERMEAN) for un grouped data

1.5 Using MS Excel functions find Mean, median mode, GM and Harmonic Mean for grouped data

1.5 Using MS Excel chart draw Median and Mode graphically

1.6 Use of Data Analysis tool pack for various concepts of topic 1 of Unit III

Self-Learning Activity:

1. PPT to demonstrate mean, median, mode on selected data
2. Power Point Presentation to demonstrate Partition values on selected data
3. Power Point Presentation to demonstrate GM, HM on selected data
4. Summarize the data on height and weight of students in your class
5. Summarize the data on electricity units consumed and bill from certain locality

2). Measures of Dispersion:

2.1 Concept and Definition of dispersion

2.2 Characteristics of good measures of Dispersion.

2.3 Types of Dispersion:

Absolute and relative measures of dispersions

Range: Definition, formula of range, for ungrouped and grouped data, merits and demerits of range Coefficient of range

Mean deviation: definition, formula. for ungrouped and grouped data Merits and demerit. Coefficient of mean deviation, minimal property of MD.

Variance and Standard deviation: definition, formula. for ungrouped and grouped data. Merits and demerit, combined variance. Minimal property of variance (Mean square Deviation, coefficient of quartile deviation and coefficient of mean deviation, coefficient of variation (C.V)

2.4 Using MS Excel functions find variance (VAR) and Standard deviation(STDEV) for un grouped data

2.5 Using MS Excel functions find range, mean deviation, variance and Standard deviation for combined variance, grouped data.

2.6 Use Data Analysis tool pack for various concepts in topic 2 of Unit III

Self-Learning Activity:

1. Study the variations in data on height and weight of students in your class
2. Prepare Power Point Presentation on different absolute and relative measures of dispersion using suitable data
3. Prepare Power Point Presentation on study of variability and consistency of data using CV. Prepare Power Point Presentation to demonstrate the application on combined mean and combined variance

Unit IV: (7L)

1) Moments, Measures of Skewness and kurtosis

1.1 Raw moments for grouped and ungrouped data

1.2 Central moments for grouped and ungrouped, effect of change of origin and scale. Relation between central moments and raw moments, up to 4^{th} order (without prof)

1.3 Concept of skewness of frequency distribution, and their types, Karl person's Coefficient of Skewness, Bowley's coefficient of Skewness and them interpretation. Skewness based on moments. Importance of skewness

1.4 Concept of Kurtosis of frequency distribution, and their types, kurtosis based on moments. Importance of kurtosis.

1.5 Using MS Excel functions find Skewness (SKEW) and Kurtosis (KURT)

1.6 Use of MS Excel for computation of Bowleys coefficient of skewness, moments based on moments.

Self -Learning Activity:

1. Collect different data sets and Prepare Power Point Presentation to demonstrate their skewness
2. Collect different data sets and Prepare Power Point Presentation to demonstrate their Kurtosis.

References:

1. Agarwal, B. L. (2003). Programmed Statistics, Second Edition, New Age International Publishers, New Delhi.
2. Gupta, S. C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, Eleventh Edition, Sultan Chand and Sons Publishers, New Delhi.
3. Mood, A.M. Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw-Hill Pub. Co. Ltd.
4. Sarma, K. V. S. (2001). Statistics Made It Simple: Do it yourself on PC. Prentice Hall of India, New Delhi.
5. Snedecor G. W. and Cochran W. G. (1989). Statistical Methods, Eighth Ed. East-West Press. 6. Gupta, S. C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, 3rd Edition, Sultan Chand and Sons Publishers, NewDelhi.
7. Mukhopadhyay P. (2015). Applied Statistics, Publisher: Books & Allied (P) Ltd.
8. Agarwal, B. L. (2003). Programmed Statistics, 2nd Edition, New Age International Publishers, NewDelhi.
10. Purohit, S. G., Gore S. D., Deshmukh S. R. (2008). Statistics Using R, Narosa Publishing House, NewDelhi.

Semester -I	Paper -II
Course Code: BSC-ST 102 T	**Title of the Course** : Introduction to Probability
Credits: 02	**Total Hours: 30**

Course Outcomes (CoS):

1. The course will give the overall idea about the uncertain situations that are expressed in probabilistic form.
2. Statistical thinking will help one's success in life and career by quantifying uncertainty using probability.
3. Student will learn the use of probability for better decisions.
4. Self-Learning activity will give them an opportunity to collect the data related to uncertain situation and interpret the probabilities.

Detail Syllabus:

Unit I: Introduction to Probability (8L)

1. Basics of Probability

1.1 Counting Principles: Additive principle, multiplicative principle.
Counting Rules: Permutations and combinations. Rules and relationship between Permutations and combinations (without proof).

1.2 Concept of deterministic and non-deterministic models (Random experiments)

1.3 Definitions of sample space and types of sample space:
(i) Sample space, (ii) Types of sample space: finite, countably infinite and uncountable.
Real life examples.
Definitions of Event and types of event: Event and concept of occurrence of an event
(i) Elementary event, (ii) complement of an event, (iii) certain event, (iv) impossible event, (v) Relative complement event
Mutually exclusive events or Disjoint events (for two and three events), mutually Exhaustive events (for two and three events), mutually exclusive and exhaustive events, Partition of sample space. Algebra of events including De Morgan's rules and its representation in set theory notation.

Occurrence of following events (with the help of listing and Venn diagram)

1) Complement of an event,
2) At least one of the two given events,
3) At least one of the three given event
4) None of the given two events,
5) None of the given three events,
6) Simultaneous occurrence of the two events,
7) Simultaneous occurrence of the three events
8) Mutually exclusive events (for two and Three events)
9) Mutually exhaustive events (for two and three events)
10) mutually exclusive and exhaustive events (for two and three events)

11) Partition of sample space
12) Exactly one event out of the two events,
13) Exactly one event out of the three events,
14) Verification of De Morgan's rules.

1.4 Classical definition of probability and its limitations.
Equiprobable and non-equiprobable sample space, classical definition of probability, Addition theorem on probability, limitations of classical definition. Situations where classical definition of probability is applicable.

1. 5 Axiomatic approach of probability.
Axioms of Probability, Situations where axiomatic approach of probability is applicable. Addition theorem on probability and its generalization. Various results on Probability Boole's inequality. Numerical examples and problems.

Self -Learning Activity:

1. Perform experiment to demonstrate permutation and combination.
2. Perform experiment to clarify the distinction between permutation and combination.
3. Perform some simple non-deterministic experiments.
4. Study the sample space corresponding to different non-deterministic experiments.
5. Study different events corresponding to non-deterministic experiments.
6. Prepare poster to show non-deterministic experiments and its sample space.
7. Prepare poster to show different events corresponding to non-deterministic experiments.
8. Perform experiment to demonstrate partition of sample space.
9. Perform experiment to demonstrate mutually exclusive events.
10. Perform experiment to demonstrate exhaustive events.

Unit II. Conditional Probability and Bayes' Theorem: (7L)

2.1 Definition of conditional probability of an event. Results on conditional probability.
Definition of independence of two events $P(A \cap B) = P(A) * P(B)$
Pairwise independence and mutual independence for three events, Multiplication theorem $P(A \cap B) = P(B)*P(A|B)$. Generalization to $P(A \cap B \cap C)$.

2.2 Prior and posterior probabilities.
Bayes' theorem. Applications of Bayes' theorem in real life. Concept of True positive (TP, Sensitivity), False positive (FP), True negative (TN, Specificity), False negative (FN). Numerical examples and problems.

Self -Learning Activity:

1. Perform experiment to demonstrate conditional probability.
2. Perform experiment to understand pairwise independence.
3. Perform experiment to mutual independence. 4. Perform experiment to understand Bayes' theorem.
5. Perform experiment to understand prior and posterior probabilities.

6. Perform experiment and compute True positive (TP, Sensitivity), false positive (FP), True negative (TN, Specificity), False negative (FN).
7. Prepare poster to illustrate True positive (TP, Sensitivity), false positive (FP), True negative (TN, Specificity), False negative (FN).
8. Power Point Presentation on conditional probability and independence. 9. Power Point Presentation on Bayes' theorem.
10. Prepare a poster on multiplication theorem.
11. Study the Bayes' theorem using urn of balls and or pebbles.

Unit III. Univariate Probability Distributions (Defined on Discrete Sample Space): (8L)

3.1: Univariate probability mass function (p.m.f.):

Concept and definition of a random variable. Types of random variable.

Concept and definition of a discrete random variable.

3.2: Probability mass function (p.m.f) and cumulative distribution function (c.d.f), F(.) of discrete random variable, properties of c.d.f., graphical representation of p.m.f. and c.d.f.

3.3: Mode and median of discrete probability distribution, Numerical examples and problems.

Self-Learning Activity:

1. Perform some non-deterministic experiments to understand random variable.
2. Perform an experiment to construct a p.m.f. and c.d.f.
3. Perform an experiment and compute mode, median.
4. Perform an experiment and graphically present p.m.f and c.d.f.
5. Perform a dice throwing experiment 100 times and prepare ppt to show observed probability, theoretical probability and plot p.m.f. and c.d.f. of both.
6. Perform an experiment of tossing a coin 100 times prepare ppt to show observed probability, theoretical probability and plot p.m.f. and c.d.f. of both.
7. Perform an experiment of tossing two coins simultaneously 100 times prepare Power Point Presentation to show observed probability, theoretical probability and plot p.m.f. and c.d.f. Of both.
8. Perform an experiment of tossing three coins simultaneously 100 times prepare Power Point Presentation to show observed probability, theoretical probability and plot p.m.f. and c.d.f. of both.
9. Perform an experiment of throwing two dice simultaneously 100 times and obtain the actual probability distribution of sum of two faces.
10. Record mobile handsets used by 100 people. Obtain modal handset.
11. Record the service provider of communication of 100 individuals and obtain modal service provider.

Unit IV. Mathematical Expectation (Univariate Random Variable): (7L)

4.1 Definition of expectation (mean) of a random variable, expectation of a function of a random variable, m.g.f. and c.g.f. properties of m.g.f. and c.g.f.

4.2 Definitions of variance, standard deviation(s.d.) and coefficient of variation(C.V.) of univariate probability distribution, effect of change of origin and scale on mean, variance and s.d.

4.3 Definition of raw, central and factorial raw moments of univariate probability distribution and their interrelations (with proof).

4.4 Coefficients of skewness and kurtosis based on moments.

Numerical examples and problems.

Self -Learning activity

1) Perform an experiment of green peas to obtain p.m.f. and mathematical expectation of number of peanuts observed in 100 peas. (If possible record video)
2) Conduct a survey on financial apps such as Phone Pe, Google Pay, BHIM etc. and obtain probabilities of using different apps.
3) Conduct a survey on online shopping.
4) Perform an experiment of 250 gm. Lemon and find the variation in number of lemons.
5) Conduct an experiment to count the number of lemons and number of apple bares of 250 gm. Weight and compare the variation.
6) Collect a data of 100 people (50 males, 50 female) having weekly fast including Special fast. Study it from probabilistic point of view.

References:

1. Agarwal B. L. (2003). Programmed Statistics, second edition, New Age International Publishers, New Delhi.
2. Devore/ Peck: Statistics (The Exploration and Analysis of Data), Duxbury.
3. Gupta, S.C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, Eighth Edition, Sultan Chand and Sons Publishers, New Delhi.
4. Hoel P. G. (1971). Introduction to Mathematical Statistics, John Wiley and Sons, New York.
5. Hogg, R. V. and Craig R. G. (1989). Introduction to Mathematical Statistics, Ed. 4. MacMillan Publishing Co., New York.
6. Mayer, P. (1972). Introductory Probability and Statistical Applications, Addison Wesley Publishing Co., London.
7. Statistical Methods: Welling, Khandeparkar, Pawar, Naralkar Manan Publications. First edition.
8. Theory and Problems of Statistics: Spiegel M.R. – Schaums Publishing Series, Tata Mcgraw - Hill. First edition
9. Rohatgi V. K. and Saleh A. K. Md. E. (2002): An Introduction to probability and statistics. John wiley & Sons (Asia)
10. Gupta V.K. & Kapoor S.C. Fundamentals of Mathematical Statistics. - Sultan & Chand
11. Mukhopadhyay P. (2006): Probability. Books and Allied (P) Ltd

Semester -I	Paper -III
Course Code: BSC-ST 103 P	**Title of the Course: Practical (Based on MSEXCEL) –I**
Credits: 1.5	**Total Hours : 45**

Course Outcome (CoS):

1. On completion of all the practical's student will have sufficient computational skill through software and programming.
2. Student will understand the difference in data visualization by MS-Excel and R.
3. Student will understand the difference in output of analysis by MS-Excel and R.
4. The overall logical thinking as a base of data science will be improved.
5. Student will have skill of result interpretation.

List of Practical's

1) Introduction to R
2) Introduction to MS EXCEL
3) Identification of different data types in given big data.
4) Draw SRSWR and SRSWOR, Stratified and systematic samples from a given population using R Software
5) Data Visualization using MS EXCEL
6) Measures of central tendencies using MS EXCEL
7) Measures of dispersion using MS EXCEL
8) Measures of Skewness and Kurtosis for different data types using R.
9) Verification of different rules of counting principles using R
10) Drawing of different events using vein- diagram using R Software
11) Computation of probabilities, Conditional probabilities and Bayes theorem using MS EXCEL
12) Plot of probability mass function and cumulative distribution function using R
13) Survey report based on self-developed questionnaire (Equivalent to three Practical)

F.Y.B.Sc. Semester – II

Course Type	Course Code	Course Title	Credits
DSCC A-4	BSC-ST 201 T	Descriptive Statistics-II Using R	02
DSCC A-5	BSC-ST 202 T	Discrete Probability Distributions	02
DSCC A-6 Practical	BSC-ST 203 P	Practical (Based on R) - II	1.5

Semester -II	Paper -I
Course Code: BSC-ST 201 T	**Title of the Course:** Descriptive Statistics -II Using R
Credits:2	**Total Hours: 30**

Course Outcome(CoS):

1) Student will learn basics of explanatory data analysis.
2) Visualization of relationship between data features.
3) Student will learn basics of Statistical modelling
4) Student will gain the knowledge of computational tool R software. 5) Student will be exposed to categorical analysis.

Unit I: Correlation and Regression: (15 L)

1.1 Concept of Bivariate data, examples of bivariate data, concept of correlation, types of correlation with illustration, scatter diagram, interpretation of scatter diagram according to pattern of plotted points, merits and demerits of scatter diagram, definition of Covariance for bivariate raw data and bivariate frequency distribution, proof of the following properties of Covariance

1) $Cov\ (X, X) = Var(X)$
2) Effect of Change of Origin and Scale
3) If X, Y, Z are three random variables, then $Cov\ (X + Y, Z) = Cov\ (X, Z) + Cov\ (Y, Z)$

Draw scatter diagram and compute covariance using R function, verify properties of covariance using R command.

1.2 Karl Pearson's coefficient of Correlation: Definition, formula in terms of covariance, Variance of Linear Combination

Proof of the following properties:

1) Effect of change of Origin and scale
2) $Corr\ (X, X) = 1$
3) Correlation coefficient always lies between -1 and 1.

Merits and Demerits of Karl Pearson's coefficient, Ranking, Rank, tie, Spearman Rank Correlation, derivation of formula for Spearman rank correlation, spearman rank correlation lies between -1 and 1, rank correlation with ties.

Using R Command to compute correlation coefficient and verify properties of correlation coefficient, Using R Command compute rank correlation coefficient.

1.3 Linear Regression: Concept of dependent (response) and independent (predictor) variables, Identification of response and predictor variables and relation between them, Meaning of regression, difference between correlation and regression, assumptions of regression, fitting of line $Y = a + bX + \epsilon$ and X= $a_1 + b_1X + \epsilon_1$, a, a_1 b and b_1 are estimated using method of least squares, Regression coefficient, interpretation of regression coefficient Proof of following Properties of Regression Coefficient:

1) Correlation and regression coefficient have same algebraic sign
2) Correlation coefficient is a geometric mean of the regression coefficients

3) Both the regression coefficients cannot exceed unity simultaneously.
4) Regression coefficient is invariant to change of origin but not the change of scale.
5) If r = $\pm$ 1 then regression coefficients are reciprocal of each other.

Explained and unexplained variation, coefficient of determination, standard error of an estimate of line of regression, analysis of residuals, model adequacy using charts.

1.4 Use R Software to fit linear regression equation, find residual and coefficient of determination, by using R Software check model adequacy by graphical method.

Unit II: Non- Linear Regression: (5L)

2.1 Necessity and importance of drawing second degree curve, Fitting of second-degree curve Y = a + b X + cX^2, Fitting of exponential curves of the type $Y = a\ b^X$ and Y =ae^b, unknowns a, b, c are estimated by using the method of least squares.

2.3 Use R Software to fit non-linear regression models in topic 2.1 also find residual. By using R Software check model adequacy by graphical method.

Unit III Theory of Attributes: (10L)

2.1 Concept of attributes, dichotomous and manifold classification of attributes, Likert's Scale, Analysis of response using Likert's scale, class, class frequency, Positive and negative attributes, Positive and negative class, ultimate class frequencies, total number of class frequencies, relation among the class frequencies.

2.2 Relation between the frequencies with three attributes, methods of dot operator, Consistency of data, independence of attributes, Association and Dissociation, Yule's coefficient of association, coefficient of colligation.

2.3 Use R Software to compute Yule's coefficient of association, coefficient of Colligation and to check the consistency of the data.

References:

1. Agarwal, B. L. (2003). Programmed Statistics, Second Edition, New Age International Publishers, New Delhi.
2. Gupta, S. C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, Eleventh Edition, Sultan Chand and Sons Publishers, New Delhi.
3. Mood, A.M. Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw-Hill Pub. Co. Ltd.
4. Snedecor G. W. and Cochran W. G. (1989). Statistical Methods, Eighth Ed. East-West Press.
5. Gupta, S. C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, 3rd Edition, Sultan Chand and Sons Publishers, NewDelhi.
6. Mukhopadhyay P. (2015). Applied Statistics,Publisher: Books & Allied (P) Ltd.
7. Agarwal, B. L. (2003). Programmed Statistics, 2nd Edition, New Age International Publishers, NewDelhi.

Semester -I	**Paper -II**
Course Code: BSC-ST 202T	**Title of the Course:** Discrete Probability Distributions
Credits: 2	**Total Hours: 30**

Course Outcome(CoS):

1) Understand the applicability of probability model in real life
2) Learn interrelation among the different probability distributions with real life problems
3) Understanding of different concepts of probability using MS EXCEL e
4) Understand the appropriateness of probability distribution in particular real life application
5) Understand various features of output of MS EXCEL.

.

Unit 1: Discrete Probability Distributions on Countably Finite Sample Space: (15 L)

1.1 Concept of Probability Model, need of the probability model, patterns of probability model, independent and identical trials (Random variables) one-point distribution (Singular Distribution): Situations where One-point distribution is used, probability mass function (p.m.f.), mean and variance, application of one-point distribution in probability theory.

1.2 **Discrete uniform (DU) distribution:** Real life situations, definition of DU distribution, plot of p.m.f., cumulative distribution function (CDF), plot of CDF, first four raw and central moments, coefficient of skewness and kurtosis, moment generating function (MGF), distribution of sum of two independent discrete uniform random variables.
Use of MS EXCEL functions to compute probabilities of DU, plot of pmf and CDF of DU, study of skewness and kurtosis for different parameters.

1.3 **Bernoulli Distribution**: Concept of Bernoulli trials, genesis of p.m.f. of Bernoulli distribution, definition of Bernoulli distribution with parameter p, Notation, real life situations, plot of probability mass function, cumulative distribution function (CDF), plot of CDF, raw and central moments of Bernoulli distribution, condition under which Bernoulli distribution is symmetric, MGF, deduction of raw moments from MGF., distribution of sum of independent and identical Bernoulli random variables, distribution of product of n independent Bernoulli random variable with parameter p.
Use of MS EXCEL functions to compute probabilities of Bernoulli distribution plot of pmf and CDF of Bernoulli distribution, MS EXCEL functions to understand Skewness and Kurtosis of Bernoulli distribution for different parameters of the distribution.

1.4 **Binomial Distribution**: Definition of binomial distribution with parameters n and p, Notations, distribution of number of successes in n independent Bernoulli trial as a Binomial distribution with parameters n and p, conditions for the applications of binomial distribution, an illustration of use of binomial distribution in SRSWR, real

life situations, raw and central moments, coefficient of skewness and kurtosis (different cases such as $p > 0.5$, $p < 0.5$ and $p = 0.5$), MGF, deduction of raw moments by using MGF, cumulant generating function (CGF), distribution of $n - X$ if X has B (n, p), recurrence relation between probabilities of binomial distribution, mode of the binomial distribution (case when $(n + 1)\ p$ is integer and not integer), recurrence relation between raw moments , recurrence relation between central moments, additive property of binomial distribution, conditional distribution of X given $X + Y = n$

Use of MS EXCEL functions to compute probabilities of Binomial distribution plot of pmf and CDF of Binomial distribution, MS EXCEL functions to understand Skewness and Kurtosis of Binomial distribution, Verifications of additive property of Binomial Distribution using MS EXCEL. MS Excel to verify the relation between the distribution of X and that of (n-X).

1.5 **Hypergeometric Distribution**: Failure of assumptions of binomial distribution in SRSWR, genesis of p.m,f. of hypergeometric distribution with parameters $N, M\ and\ n$, difference between hypergeometric and binomial distribution, conditions for the applications of hypergeometric distribution, real life situations, mean and variance, r^{th} factorial moment, binomial approximation to Hypergeometric distribution.

Use of MS EXCEL functions to compute probabilities of Hyper geometric distribution plot of pmf and CDF of Hyper geometric distribution, MS EXCEL functions to understand Skewness and Kurtosis of Hyper geometric distribution, Verifications of Binomial approximation to Hyper geometric distribution using MS EXCEL.

Unit 2: **Discrete Probability Distributions on Countably Finite Sample Space: (15L)**

2.1 **Poisson Distribution**: Poisson distribution as a model for the situations where chances of occurrence of an event in a short time interval is with high probability, real life situations, definition of Poisson distribution, Notation, mean and variance, MGF, deduction of raw and central moments from MGF, CGF, all the cumulants are equal for the Poisson distribution, central moments using CGF, coefficient of skewness and kurtosis, interpretation from these coefficients, nature of distribution $as\ m \to \omega$, additive property, generalization of additive property, conditional distribution of X_1 given $X_1 + X_2 = n$, Recurrence relation for the probabilities of Poisson distribution, Poison distribution as a limiting form of binomial distribution, fitting of Poisson distribution, mode of Poisson distribution.

Use of MS EXCEL functions to compute probabilities of Poisson distribution plot of pmf and CDF of Poisson distribution, MS EXCEL functions to understand Skewness and Kurtosis of Poisson distribution, Verifications of Poison distribution as a limiting form of binomial distribution, Fitting of Poisson distribution using MS EXCEL

2.2 **Geometric distribution**: Genesis of p.m.f. of geometric distribution (for both forms), definition of Geometric distribution, Notation, Geometric distribution as waiting time distribution, mean, variance, mode, relation between mean and variance, moment generating function, deduction of mean and variance from MGF, CGF, deduction of first four central moments from CGF, recurrence relation between probabilities,

distribution function, Lack of memory property and its interpretation, real life applications.

Use of MS EXCEL functions to compute probabilities of Geometric distribution plot of pmf and CDF of Geometric distribution, MS EXCEL functions to understand Skewness and Kurtosis of Geometric distribution, Verifications of lack of memory property using MS-EXCEL

2.3 **Negative Binomial Distribution (NBD)**: Sum of two (or more) geometric random variables as a NBD, derivation of p.m.f. of negative binomial distribution, mean, variance, relation between mean and variance, factorial moments, MGF, CGF, deduction of mean, variance and third central moment from CGF, coefficient of skewness, additive property of NBD.

Use of MS EXCEL functions to compute probabilities of Negative Binomial Distribution plot of pmf and CDF of Negative Binomial Distribution, MS EXCEL functions to understand Skewness and Kurtosis of Negative Binomial Distribution, Verifications of additive property using MS EXCEL

References:

1) Gupta, S.C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, Eighth Edition, Sultan Chand and Sons Publishers, New Delhi.
2) Agarwal B. L. (2003). Programmed Statistics, second edition, New Age International Publishers, New Delhi.
3) Hoel P. G. (1971). Introduction to Mathematical Statistics, John Wiley and Sons, New York.
4) Hogg R.V. and Craig R.G. (1989). Introduction to Mathematical Statistics, MacMillan Publishing Co., New York.
5) Mood, A. M. and Graybill, F. A. and Boes D.C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill Book Company.
6) Rao, VLS Prakash (2008). First Course in Probability and Statistics, New Age International Publishers, New Delhi.
7) Ross S. (2002). A First Course in Probability, Sixth Edition, Pearson Education, Inc. & Dorling Kindersley Publishing, INC.

Semester -I	**Paper -III**
Course Code: BSC-ST 203 P	**Title of the Course: Practical (Based on R) - II**
Credits: 1.5	**Total Hours : 45**

Course Outcome (CoS):

1. On completion of all the practical's student will have sufficient computational skill through software and programming.
2. Student will understand the difference in data visualization by R and MS-Excel
3. Student will understand the difference in output of analysis by R and MS-Excel
4. The overall logical thinking as a base of data science will be improved.
5. Student will gain skill of interpreting the outputs

List of Practical's

1) Correlation and Regression for ungrouped data using R Software
2) Fitting of Nonlinear Regression using R software
3) Fitting of Binomial distribution and computation of expected frequencies Using Excel
4) Fitting of Poisson distribution and computation of expected frequencies using Excel
5) Fitting of Geometric distribution and computation of expected frequencies using Excel
6) Fitting of Negative Binomial Distribution and computation of expected frequencies using Excel
7) Applications of Binomial, hypergeometric using R Software
8) Application of Poisson Distributions using R Software
9) Applications of geometric and Negative Binomial Distribution using R Software
10) Model sampling from Poisson and Binomial distributions. using R
11) Model Sampling from Negative Binomial Distribution using R
12) Coefficient of association in attribute using R software
13) Short project on any one topic (equivalent to three practical's)