



Haribhai V. Desai College of Arts, Science & Commerce, Pune-02



(Affiliated to Savitribai Phule Pune University)

Four Year B.Sc. / B. A Degree Program in Statistics
(Faculty of Science & Technology)

F. Y. B. Sc./ F. Y. B. A. Statistics Syllabi in NEP
(Level 4.5)

National Education Policy Syllabus
To be implemented from Academic Year 2024-2025

Title of the program: F. Y. B. Sc. / F. Y. B. A. Statistics**Preamble of the syllabus:**

The word *Statistics* is used in different ways in different contexts. To a cricket fan, Statistics is the information about runs scored or wickets taken by a player. To the manager of a manufacturing unit, Statistics may be the information about the process control. To a medical researcher investigating the effects of a new drug, Statistics are evidence of research efforts. To a college student, Statistics are the grades or marks scored in a course. Thus, in all these illustrations Statistics word refers to quantitative data in the area under study. Statistics as a subject is an important branch of knowledge and is devoted to various techniques of collection, presentation, analysis, and interpretation of data. It is a science of learning from data.

Statistics provides tools for making decisions when conditions of uncertainty prevail. Hence these tools and techniques are used in almost all fields. Statistics is indispensable for people working in fields like agriculture, business, management, economics, finance, insurance, education, biotechnology, and medical science etc. Since last two decade, with the help of computers large amount of data can be handled and more sophisticated statistical techniques can be used in an effective manner. Knowledge of different aspects of Statistics has become crucial. There is a continuous demand for statisticians in every field – education, industry, software, insurance, clinical trials data and research. The syllabus of the three-Year B. Sc. degree course in Statistics is framed in such a way that the students at the end of the course can apply judiciously the statistical tools to a variety of data sets to arrive at some conclusions.

Statistics can be divided into two broad categories, (1) exploratory statistics or descriptive statistics, which is concerned with summarizing data and describing these data, and (2) confirmatory statistics or inferential statistics, which is concerned with making decisions about the population based on the sample.

Up to higher secondary school, students are mostly exposed to descriptive statistics. These techniques are briefly reviewed but the emphasis in degree course is on inferential statistics. At the end of the degree course a student is expected to apply statistical tools to real life data and analyze it.

Introduction: National Education Policy (NEP) 2023.

B. Sc. in Statistics program is of four years duration, with semester pattern for all the four years. At **first year of under-graduation**, students will be given the basic information that includes – methods of data representation and summarization. Correlation and regression are the forecasting

tools that are frequently used in statistical analysis. These topics are studied in one of the papers each semester. Further they are introduced to probability and different discrete probability distributions along with applications in the other paper. Relevant experiments on these topics will be included in the practical course. Further the students are expected to start using statistical software MS-Excel and verify the computations during practical. It is a skill-oriented part of the course. Also, students are expected to study skills Counting Principles, Permutations and Combinations, Scope of Statistics and Statistical Organizations in India: binomial theorem and its applications, preparation of questionnaire statistical population and sampling methods:

At second year of under-graduation, we offer statistics as minor subject at our college and students are expected to study various probability distributions and its applications to real life situations. It is a foundation for further theory. An important branch of Statistics, viz. testing of hypotheses related to mean, variance, proportion, correlation etc. will be introduced. Some topics related to applications of Statistics will also be introduced. Further the students are expected to start using statistical software *R* and verify the computations during practical. It is a skill-oriented part of the course.

Structure of Evaluation: (For each paper)

1. Continuous Internal Assessment (CIA): 15 Marks (Internal and Assignment)
2. End Semester Examination (ESE): 35 Marks
3. Total: 50 Marks

Structure of Four Year-degree Program

Structure of the course for four years, the pattern of examination and question papers are as per Savitribai Phule Pune University:

Statistics as Major (Core) Subject and any other subject as Minor (each theory / practical paper has 2 credits).

Year / Level	Sem	Code Number	Title of the paper (Theory / Practical)	Credits allotted	Lecture/Practical hours per week
I 4.5/100	I	STS-101-TH	Univariate and bivariate data analysis	02	02
		STS-102-PR	Statistics Practical-I	02	04
	II	STS-151-TH	Theory of Probability and Discrete Probability Distributions	02	02
		STS-152-PR	Statistics Practical-II	02	04

List of Vocational Skill Courses (VSC):

Year / Level	Sem	Code Number	Title of the paper (Theory / Practical)	Credits allotted	Lecture/Practical hours per week
II 5.0/200	III	STS-221-VSC-PR	Introduction to MS-EXCEL -I (Practical Course)	02	04
	IV	STS-271-VSC-PR	Introduction to MS-EXCEL -II (Practical Course)	02	04

Statistics as Minor Subject and any other subject as Major (each theory / practical paper has 2 credits)

Year / Level	Sem	Code Number	Title of the paper (Theory / Practical)	Credits allotted	Lecture/Practical hours per week
II 5.0/200	III	STS-241-MN-TH	Probability Distributions and Demography	02	02
		STS-242-MN-TH	Statistics Practical on Probability Distributions and Demography	02	04
	IV	STS-291-MN-TH	Sampling Distributions and Statistical Inference	02	02
		STS-292-MN-PR	Practical on Sampling Distributions and Statistical Inference	02	04
III 5.5/300	V	STS-341-MN-TH	Statistical Methods	02	02

List of Generic / Open Electives (OE):

Year / Level	Sem	Code Number	Title of the paper (Theory / Practical)	Credits allotted	Lecture/Practical hours per week
I 4.5/100	I	OE-101-STS-TH	Elementary Statistics for Social Sciences-I OR	02	02
		OE-102-STS-TH	Elementary Commercial Statistics-I	02	02
	II	OE-151-STS-TH	Elementary Statistics for Social Sciences -II OR	02	02
		OE-152-STS-TH	Elementary Commercial Statistics-II	02	02
II 5.0/200	III	OE-201-STS-TH	Business Statistics-I OR	02	02
		OE-202-STS-TH	Applied Statistics-I	02	02
	IV	OE-251-STS-TH	Business Statistics-II OR	02	02
		OE-252-STS-TH	Applied Statistics-II	02	02

List of Skill Enhancement Courses (SEC):

Year / Level	Sem	Code Number	Title of the paper (Theory / Practical)	Credits allotted	Lecture/Practical hours per week
I 4.5/100	I	SEC-101- STS-PR	MS-EXCEL for Data Analysis (Practical Course)	02	04
	II	SEC-151- STS-PR	Computational Statistics using MS- EXCEL (Practical Course)	02	04
II 5.0/200	IV	SEC-251- STS-PR	Descriptive Statistics using R-software (Practical Course)	02	04

Indian Knowledge System (IKS):

Year / Level	Sem	Code Number	Title of the paper (Theory / Practical)	Credits allotted	Lecture/Practical hours per week
I 4.5/100	I	IKS -101-TH	Generic	02	02
II 5.0/200	III	IKS -201-TH	Evolution / Development of Statistics in India	02	02

SEMESTER I

Paper Code and Title: STS-101-TH: Univariate and bivariate data analysis**Course type- Theory****No. of credits – 2****No. of contact hours - 30****Course Outcomes:**

By the end of the course, students should be able to:

1. Identify the appropriate scale of measurement for a particular characteristic under study.
2. Calculate and describe data through measures of central tendency and dispersion.
3. Interpret the utilization of measures of central tendency and dispersion to compare groups.
4. Calculate and interpret coefficients of skewness and kurtosis.
5. Fit linear and non-linear regression models.

Unit No.	Content	No. of Hours
1	Population and Sample	04
1.1	Types of characteristics: a) Attributes: Nominal scale, ordinal scale, b) Variables: Interval scale, ratio scale, discrete and continuous variables, difference between linear scale and circular scale. Real life illustrations.	
1.2	Types of data: Primary data, Secondary data, Cross-sectional data, time series data, directional data.	
1.3	Notion of a statistical population: Finite population, infinite population, homogeneous population, and heterogeneous population. Notion of a sample and a random sample. Methods of sampling (Description only): Simple random sampling with and without replacement (SRSWR and SRSWOR), stratified random sampling, systematic sampling, cluster sampling and two-stage sampling.	
2	Summary Statistics:	12
2.1	Classification of data: Raw data and its classification, ungrouped frequency distribution, Sturges' rule, grouped frequency distribution, cumulative frequency distribution, Inclusive and exclusive methods of classification, Open end classes, relative frequency distribution and frequency density. Numerical problems.	

2.2 Measures of Central Tendency:

Concept of central tendency of statistical data, Statistical averages, characteristics of a good statistical average.

Arithmetic Mean (A.M.): Definition, effect of change of origin and scale, combined mean of several groups, merits and demerits, trimmed arithmetic mean.

Geometric Mean (G.M.): Definition, formula, merits, and demerits.

Harmonic Mean (H.M.): Definition. Formula, merits, and demerits.

Order relation between arithmetic mean, geometric mean, harmonic mean.

Mode and Median: Definition, formulae (for ungrouped and grouped data), merits and demerits. Empirical relation between mean, median and mode.

Partition Values: Quartiles, Deciles and Percentiles (for ungrouped and grouped data).

Weighted Mean: weighted A.M., G.M. and H.M.

Situations where one kind of average is preferable to others. Numerical problems.

2.3 Measures of Dispersion:

Concept of dispersion, characteristics of good measure of dispersion.

Absolute and relative measure of dispersion. Range and coefficient of range,

Semi-interquartile range (Quartile deviation): Definition, merits and demerits

and coefficient of Quartile deviation, Mean deviation: Definition, merits, and demerits, minimality property (without proof) and coefficient of Mean

deviation, Mean square deviation: Definition, minimality property of mean

squared deviation (with proof), coefficient of mean square deviation. Variance

and standard deviation: Definition, merits and demerits, effect of change of origin and scale, combined variance for n groups. Coefficient of variance.

Numerical problems.

3 Moments, Skewness and Kurtosis

04

3.1 Raw moments for ungrouped data and frequency distribution.

Central moments for ungrouped data and frequency distribution, Effect of change of origin and scale (without proof). Relations between central moments and raw moments, up to 4th order (without proof).

3.2 Concept of skewness for frequency distribution, positive skewness, negative skewness, symmetric frequency distribution. Assessment of skewness from Box plot. Bowley's coefficient of skewness. Bowley's coefficient of skewness lies between -1 to 1 (with proof). Karl Pearson's coefficient of skewness. Measures of skewness based on moments (Pearsonian's coefficient).

3.3 Concepts of kurtosis, Types of kurtosis: leptokurtic, mesokurtic, and platykurtic frequency distributions. Measures of kurtosis based on moments (Pearsonian's coefficient).

4 Correlation and Regression

10

4.1 Bivariate data, Scatter diagram and its interpretation.

Concept of correlation between two variables. Types of correlation (positive correlation, negative correlation, no correlation). Covariance between two variables. Karl Pearson's coefficient of correlation (r): Definition, computation for ungrouped data with interpretation. Its properties.

Spearman's rank correlation coefficient: Definition, computation (without ties).

4.2 Meaning of regression, relation between correlation and regression.

Simple linear regression model: $Y = a + bX + \epsilon$, where ϵ is a continuous random variable with $E(\epsilon) = 0$, $V(\epsilon) = \sigma^2$. Assumptions of regression model, Estimation of parameters of the model by method of least squares. Interpretation of parameters.

Concept of error in regression, mean residual sum of squares, Explained and unexplained variation, coefficient of determination. Concept of residual, plot of residual against X . Standard error of an estimate of line of regression.

4.3 Non-linear regression, Necessity and importance of drawing second degree curve. Fitting of second-degree curve ($Y = a + bX + cX^2$), Fitting of exponential curves of the type $Y = a b^X$ and $Y = a X^b$. In all these curves constants or parameters are estimated by the method of least squares.

(Justification via determinant of matrix of second derivative/second derivative test). Mean residual sum of squares as a criterion to decide the best fit of the curve.

References:

1. Agarwal, B. L. (2003). Programmed Statistics, Second Edition, New Age International Publishers, New Delhi.
2. Brase C. H., Brace C. P (2016). Understandable Statistics, Concepts and Methods, 12th Edition, Cengage Learning.
3. Freedman D., Pisani R., Purves R. (2007). Statistics, 4th Edition, W. W. Norton, and Company.
4. Freund J. E. (1977). Modern Elementary Statistics. 4th Edition, Prentice Hall of India Private Limited, New Delhi.
5. Ghosh, J. K. and Mitra, S. K., Parthsarthy, K. R. (1993). Glimpses of India's Statistics Heritage, Wiley publishing Co.
6. Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, The World Press Pvt. Ltd., Calcutta.
7. Gupta, S. C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, Eighth Edition, Sultan Chand and Sons Publishers, New Delhi.
8. Gupta, S. C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, Third Edition, Sultan Chand and Sons Publishers, New Delhi.
9. Heumann C., Schomaker, M., Shalabh (2016). Introduction to Statistics and Data Analysis. 1st Edition, Springer, Germany.
10. Moore D. S., Notz W. I., Fligner M. A. (2013). The Basic Practice of Statistics, 6th Edition, Ruth Baruth.
11. Utts J. M., Heckard R. F. (2010). Mind On Statistics, 4th Edition, Richard Stratton Publisher.
12. Zealure C. H. (1998). Fundamentals of Descriptive Statistics. 1st Edition, Routledge, U. K. (Taylor and Francis Group).
13. Neil A. Weiss, (2016). Introductory Statistics, Tenth Edition, Pearson.
14. Snedecor G. W. and Cochran W. G. (1989). Statistical Methods, Eighth Ed. East- West Press.

Paper Code and Title: STS-102-PR: Statistics Practical -I**Course type- Practical****No. of credits – 02 No. of contact hours –60(48+12)****Course Outcomes:**

By the end of the course, students should be able to:

1. have knowledge of preparing the frequency distribution and presentation using tabular form.
2. use various diagrammatic and graphical techniques to represent statistical data and interpret.
3. analyze data pertaining to discrete variables and to interpret the results.
4. compute various measures of central tendency, dispersion, skewness, and kurtosis.
5. compute correlation coefficient for given data and interpret results

Sr. No.**Title of the experiment**

1. Preparation of frequency distribution for discrete and continuous variables for raw data.
2. Tabulation
3. Diagrammatic representation of statistical data I: simple bar diagram, subdivided bar diagram, multiple bar diagram, percentage bar diagram.
4. Diagrammatic representation of statistical data-II: pie diagram, spike plot for Likert scale. Data interpretation from diagrams.
5. Graphical representation of statistical data I: Histogram and frequency curve. Determination of mode graphically.
6. Graphical representation of statistical data-II: ogive curves and Pareto chart. Determination of median graphically. Data interpretation from graphs.
7. Use of random number tables to draw SRSWOR, SRSWR, stratified sample and systematic sample.
8. Summary statistics – I: Computation of measures of central tendency and dispersion (ungrouped data). Use of an appropriate measure and interpretation of results and computation of partition values.
9. Summary statistics – II: Computation of measures of central tendency and dispersion (grouped data). Use of an appropriate measure and interpretation of results and computation of partition values.

- 10 Computations of central moments
- 11 Measures of skewness and kurtosis. Box plot.
- 12 Scatter diagram and computation of Karl Pearson's correlation coefficient

Note: Every practical is equivalent to four theory lectures per batch per week.

Generic / Open Elective (OE)

Paper code and title – OE-101-STS-TH: Elementary Statistics for Social Sciences-I

Course type: Theory

No. of Credits: 02

No. of Contact Hours: 30

Course Outcomes:

At the end of this course, students will be able:

1. To identify the appropriate data type suitable for an intended survey.
2. To identify the most suited sampling method for an intended survey.
3. To anticipate difficulties/problems in data collection and take proactive measures to resolve them.
4. To create structured and organized survey forms.
5. To compute coefficients used to assess internal consistency of collected data.
6. To effectively represent collected data through diagrams and/or graphs.

Unit No.	Content	No. of Hours
1	Types of Data	03
	1.1 Types of characteristics: Attributes and Variables, Measurement scale for attributes: Nominal and Ordinal scale, Measurement scale for variables: Ratio and Interval scale, Likert scale. (all with relevant illustrations). 1.2 Types of data: Primary data, Secondary data, Cross sectional data, Time series data. (all with relevant illustrations).	
2	Methods of Sampling	07
	Statistical population, Finite population, Infinite population, homogeneous population, heterogeneous population, Random sampling: Simple random sampling with and without replacement, Stratified sampling, Systematic sampling, Cluster Sampling, Two-stage sampling, Non-random sampling: Purposive sampling, Snowball sampling, Convenience sampling, Crowdsourcing sampling. (only descriptions and illustrations of all methods).	
3	Questionnaires, Errors and Consistency	14
	3.1 Characteristics of a good questionnaire, Problems faced in data collection: problem of non- response, sampling errors, non-sampling errors, validity	

(internal consistency) using Cronbach's alpha and Kuder Richardson's coefficient (KR-20).

3.2 Designing questionnaires for surveys (i.e. survey forms) (with various sections if required) using various online platforms such as Google Forms, Survey Monkey, Form Façade (an add-on to customize Google Forms), etc., Aesthetic presentation of survey forms, Flow of questions.

3.3 Processing Survey Data: Downloading responses as MS-Excel sheets, computation of summated scores (for Likert scale-based questions), coding data, computation of Cronbach's alpha and Kuder Richardson's coefficient (KR-20).

4 Basic Exploratory Analysis

06

Selection of appropriate diagram/graph type, Construction of diagrams (simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram).

References:

1. R. M. and et.al. (2009), Survey Methodology, Second Edition, Wiley Series in Survey Groves Methodology.
2. Mukhopadhyay P. (2008), Theory and Methods of Survey Sampling, Prentice-Hall India, New Delhi.
3. Murthy M. N., (1977), Sampling Theory and Methods, Statistical Publishing Society, Kolkata.
4. Sukhatme P. V., Sukhatme B. V. (1984), Sampling Theory of Surveys with Applications, Indian Society of Agricultural Statistics, New Delhi.

Generic / Open Elective (OE)**Paper code and title – OE-102-STSTH: Elementary Commercial Statistics-I****Course type: Theory****No. of Credits: 02****No. of Contact Hours: 30****Course Outcomes:**

At the end of the course, students should be able to:

1. Students will understand the key concepts of the theory of attributes and their applications in data classification.
2. Students will analyse relationships between two attributes using different.
3. Students will create and interpret tables for both qualitative and quantitative data, recognizing their advantages and limitations.
4. Students will construct various graphical representations, such as histograms and ogives, to effectively visualize data distributions.
5. Students will classify and summarize data using ungrouped and grouped frequency distributions.
6. Students will compute and evaluate measures of central tendency and understand their respective merits and demerits.

Unit No.	Content	No. of Hours
	Classification and Tabulation	07
1.1	Raw data, class frequency, Discrete frequency distribution, Continuous frequency distribution, Inclusive and Exclusive methods of classifications, class limits, class boundaries, mid values, class width, open end class, Cumulative frequencies of less than and more than type. relative frequencies, frequency density.	
1.2	Definition and concept of tabulation, objectives of tabulation, Components of table, Advantages and limitation of tabulation.	
1.3	Tabulation for quantitative data.	
1.4	Definition and concept of tabulation, objectives of tabulation, Components of table, Advantages and limitation of tabulation.	
1.6	Numerical problems related to real life situations.	

- 2 Graphical Representation 06**
- 2.1** Diagrammatic representation of statistical data: simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram.
- 2.2** Graphical representation of statistical data: Line graph, histogram (with equal and unequal class width) frequency curve, ogive curves.
- 2.3** Numerical Examples
- 3 Univariate Data Analysis 10**
- 3.1** Concept of univariate data. Concept of central tendency, Measures of central tendency. Arithmetic mean: Definition. Computation of arithmetic mean for ungrouped and grouped frequency distribution, Merits and Demerits.
- 3.2** Median: Definition. Computation of median for ungrouped and grouped frequency distribution, Merits and Demerits. Mode: Definition. Computation of mode for ungrouped and grouped frequency distribution, Merits and Demerits.
- 3.3** Concept of dispersion and measures of dispersion, absolute and relative measures of dispersion. Range and Quartile Deviation: definition for ungrouped and grouped data and their coefficients, merits and demerits.
- 3.4** Variance and Standard deviation: definition for ungrouped and grouped data, coefficient of variation.
- 3.5** Numerical problems related to real life situations.
- 4 Bivariate Data Analysis 07**
- 4.1** Bivariate data, Scatter diagram and interpretation. Concept of correlation between two variables, positive correlation, negative correlation, no correlation.
- 4.2** Covariance between two variables: Definition Karl Pearson's coefficient of correlation (r): Definition, computation for ungrouped data and interpretation. Range of correlation coefficient. *i.e.* $-1 \leq r \leq 1$ (without proof).
- 4.3** Concept of dependent and independent variables in regression
- 4.4** Meaning of regression, difference between correlation and regression, Connection between correlation and regression.

- 4.5 Fitting of line $Y = a + bX$, a and b are estimated using least square principle method (without proof). Interpretation of regression coefficient.
- 4.6 Numerical Problems related to real life situations

References:

1. Agarwal B. L. (2003). Programmed Statistics, 2nd edition, New Age International Publishers, New Delhi.
2. Deborah J. Rumsey, Statistics for Dummies.
3. David S. Moore, George P. McCabe, and Bruce A. Craig, Introduction to the Practice of Statistics.
4. Gupta S.C. and Kapoor, V. K. (1983). Fundamentals of applied Statistics, 3rd Edition, Sultan Chand and Sons Publishers, New Delhi.

Skill Enhancement Course (SEC)

Paper code and title SEC-101-STS-PR— MS-EXCEL for Data Analysis

Course type: Practical

No. of Credits: 02

No. of Contact Hours: 60(48+12)

Course Outcomes:

At the end of this course, students will be able:

1. Students will be familiar with MS-Excel.
2. Students will be able to create spreadsheets, enter data, and maintain data.
3. Students will be able to handle data using existing MS-Excel functions.
4. Students will be able to draw appropriate diagrams or graphs to the given data.

**Unit
No.**

Content

1 Getting Acquainted with MS-Excel

1.1 Introduction to MS-Excel

1.2 The Excel Environment: Cells, Rows, and Columns, Title Bar, Ribbon, Scroll Bars, Quick Access Toolbar, Formula Bar, Workbook View Buttons, Zoom Slider, Mini Toolbar, Keyboard Shortcuts, Formulas, Sheet Tabs, Page Margins, Page Orientation, Page Breaks and Printing.

1.3 Worksheets and Workbooks: Definition of Worksheets and Workbooks, creating and saving new worksheet, Naming of Worksheets, Adding and Deleting Worksheets, Hiding/ Unhiding Worksheets, Hiding Columns and Rows, Saving Workbooks, Saving an Existing File, Headers and Footers, Inserting, Deleting, copy and Renaming of Worksheets. Conditional Formatting and cell styles

2 Entering and Editing Information

2.1 Import external data, Entering Data, create a table, Labels and Values, Copying Cells, Rows and Columns, Pasting Cells, Rows, and Columns, Paste an Item from the Clipboard

2.2 Inserting and Deleting Rows and Columns, Filling and Editing Cell Data, Find and Replace, Go To Cell Data, Locking Rows and Columns, Spell Check, AutoCorrect.

3 Formatting & Adding Elements to a Worksheet

3.1 Change Font Styles and Sizes, Adding Borders and Colours to Cells, change Column Width, change Row Height, Merge Cells, Applying Number Formats

3.2 Creating Custom Number Formats, Align Cell Contents, Cell Styles, Conditional Formatting, Freeze and Unfreeze Rows and Columns, Adding and Modifying Images

3.3 Removing A Background, Cropping and Rotating an image, compressing a Picture, Inserting AutoShapes, Adding WordArt, Clip Art, and a Hyperlink.

4 Inbuild MS-Excel Functions

4.1 Logical: IF, AND, NOT, OR, LET, LAMBDA, TRUE, FALSE, SWITCH, etc.

4.2 Mathematical: ABS, EXP, CEILING, FLOOR, INT, EVEN, ODD, COMBIN, COMBINA, FACT, FACTDOUBLE, GCD, LCM, LN, LOG, LOG10, MOD, MULTINOMIAL, POWER, PRODUCT, RAND, RANDARRAY, RANDBETWEEN, ROUND, SIGN, SQRT, etc.

4.3 Lookup: LOOKUP, HLOOKUP, VLOOKUP, XLOOKUP, etc.

4.4 Other functions: Date and Time Functions, Text functions, sort, duplicate, Pivot table and Pivot chart

5 Graphical Representation

5.1 Titles, legend, data labels, creating a New Chart, Formatting the Chart, Types of charts, Using Chart Templates.

5.2 Simple bar diagram, subdivided bar diagram, multiple bar diagram, percentage bar diagram, pie diagram, rod or spike plot, histogram, frequency curve and ogive curves, Pareto chart.

List of practical:

Sr. No.	Title of the Experiment	No. of Experiments
1	Introduction to MS-Excel	3
2	Basic Mathematical Functions	2
3	Conditional Logic Functions	1
4	Lookup Functions	1
5	Diagrammatic or graphical representation using MS-Excel	3
6	Pivot Table	1
7	Pivot Chart	1

Note: Every practical is equivalent to four theory hours per batch per week.

References:

1. Frag Curtis (2013). Step by Step Microsoft Excel 2013, MS Press.
2. Frye Curtis D. (2007). Step by step Microsoft Office Excel 2007, Microsoft Press.
3. John Walkenbach (2013). 101 Excel 2013 Tips, Tricks and Time savers, Wiley.
4. Kumar Bittu (2013). Microsoft Office 2010, V&S Publishers.
5. Salkind Neil J. and Frey Bruce B. (2021). Statistics for people who (Think They) Hate Statistics, Using MS- Excel, Sage Publications.
6. Sanjay Saxen (2007). MS Office 2000 for everyone, Vikas Publishing House.

SEMESTER II

**Paper Code and Title: STS-151-TH: Theory of Probability and Discrete Probability
Distributions**

Course type- Theory

No. of credits – 02

No. of contact hours - 30

Course Outcomes:

By the end of the course, students should be able to:

1. familiarized with basic concepts of random experiments, random variable, probability, etc.
2. get ideas regarding the use of probability in real life situations. Also, can find out the probability of real-life events.
3. identify the nature of data using moments, skewness, kurtosis, etc.
4. see the dependency of events using probability.
5. apply standard discrete probability distribution to different real situations.

Unit No.	Content	No. of Hours
1	Basics of Probability	06
	1.1 Experiments/Models, Ideas of deterministic and non-deterministic models. Random Experiment, concept of statistical regularity.	
	1.2 Definitions: i) Sample space, ii) Discrete sample space: finite and countably infinite, iii) Event, iv) Elementary event, v) Complement of an event. vi) Certain event, vii) Impossible event. Concept of occurrence of an event. Algebra of events and its representation in set theory notation. Occurrence of following events: i) at least one of the given events, ii) none of the given events, iii) all the given events, iv) mutually exclusive events, v) mutually exhaustive events, vi) exactly one event out of the given events.	
	1.3 Classical definition of probability and its limitations. Probability model, probability of an event, equiprobable and non-equiprobable sample space,	
	1.4 Axiomatic definition of probability. Theorems and results on probability with proofs based on axiomatic definition such as $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Generalization to $P(A \cup B \cup C)$, $0 \leq P(A) \leq 1$, $P(A) + P(A^c) = 1$, $P(\Phi) = 0$, $P(A) \leq P(B)$ when $A \subset B$. Boole's inequality.	
	1.5 Numerical Examples	

- 2 Conditional Probability and Independence 02**
- 2.1** Definition of conditional probability of an event. Results on conditional probability. Definition of independence of two events $P(A \cap B) = P(A) \cdot P(B)$. Independence of A and B' , A' and B , and A' and B' . Pairwise independence and mutual independence for three events. Multiplication theorem $P(A \cap B) = P(A) \cdot P(B|A)$. Generalization to $P(A \cap B \cap C)$.
- 2.2** Numerical Examples
- 3 Univariate Probability Distributions (Defined on Discrete Sample Space) 02**
- Concept and definition of a discrete random variable. Probability mass function (p.m.f.) and cumulative distribution function (c.d.f.), $F(\cdot)$ of discrete random variable, properties of c.d.f.. Mode and median of a univariate discrete probability distribution
- 4 Mathematical Expectation (Univariate Random Variable) 05**
- 4.1** Definition of expectation (Mean) of a random variable, expectation of a function of a random variable, properties of expectation of r.v., 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** Numerical Examples.
- 5 Standard Discrete Probability Distributions-I 08**
- 5.1 Degenerate distribution** (one point distribution): $P(X=c) = 1$, mean and variance.
- 5.2 Uniform discrete distribution on integers 1 to n:** p.m.f., c.d.f., mean, variance, real life situations, comments on mode and median.
- 5.3 Bernoulli Distribution:** p.m.f., mean, variance.
- 5.4 Binomial Distribution:** p.m.f.

$$P(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & x = 0, 1, 2, \dots, n \\ 0, & \text{otherwise} \end{cases}$$

Notation: $X \sim B(n, p)$. Recurrence relation for successive probabilities, computation of probabilities of different events, mode of the distribution, mean, variance, m. g. f. and c. g. f. Situations where this distribution is applicable.

6 Standard Discrete Probability Distributions-II

07

6.1 Poisson Distribution:

p.m.f. of the distribution

$$p(x) = \frac{e^{-m} m^x}{x!} \quad x = 0, 1, 2, \dots \dots \dots m > 0$$

$$= 0, \text{ otherwise}$$

Notation: $X \sim P(m)$.

m.g.f. and c.g.f. Moments, mean, variance, skewness and kurtosis.

Situations where this distribution is applicable.

6.2 Geometric Distribution:

Notation: $X \sim G(p)$,

Geometric distribution on support $(0, 1, 2, \dots)$ with p.m.f. $p(x) = pq^x$

Geometric distribution on support $(1, 2, \dots)$ with p.m.f. $p(x) = pq^{x-1}$

$0 < p < 1, q = 1 - p$

Mean, variance, m.g.f. and c.g.f.

Situations where this distribution is applicable.

6.3 Numerical Examples.

References:

1. Agarwal B. L. (2003). Programmed Statistics, 2nd edition, New Age International Publishers, New Delhi.
2. Brase C. H., Brace C. P (2016). Understandable Statistics, Concepts and Methods, 12th Edition, Cengage Learning.
3. Freedman D., Pisani R., Purves R. (2007). Statistics, 4th Edition, W. W. Norton and Company.
4. Gupta S.C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, 8th Edition, Sultan Chand and Sons Publishers, New Delhi.
5. Hoel P. G. (1971). Introduction to Mathematical Statistics, John Wiley and Sons, New York.

6. Hogg R.V. and Craig R.G. (1989). Introduction to Mathematical Statistics, Ed. MacMillan Publishing Co., New York.
7. Mayer P. (1972). Introductory Probability and Statistical Applications, Addison Wesley Publishing Co. London.
8. Mood A. M. and Graybill, F. A. and Boes D.C. (1974). Introduction to the Theory of Statistics, 3rd Edition, McGraw Hill Book Company.
9. Moore D. S., Notz W.I., Fligner M.A. (2013). The Basic Practice of Statistics, 6th Edition, Ruth Baruth.
10. Rohatgi V.K. and Saleh, A. K. (2015). An Introduction to Probability and Statistics, 3rd Edition, John Wiley & Sons, Inc.
11. Ross S.(2002). A First Course in Probability, Sixth Edition, Pearson Education, Inc.& Dorling Kindersley Publishing, Inc.
12. Utts J. M., Heckard R. F. (2010). Mind On Statistics, 4th Edition, Richard Stratton Publisher.

Paper Code and Title: STS-152-PR: Statistics Practical -II**Course type- Practical****No. of credits – 02 No. of contact hours –60(48+12)****Objectives:**

At the end of this course students are expected to be able:

1. To fit line of regression, second-degree curve, and exponential curve.
2. To construct bivariate probability distributions, marginal probability distribution, and conditional probability distribution
3. To fit binomial distributions
4. To compute probabilities of bivariate probability distributions.
5. To draw random samples from binomial distributions

Sr. No.	Title of the experiment	No. of Experiments
1	Computation of Spearman's rank correlation coefficient.	01
2	Fitting of line of regression $Y = a + bX$	01
3	Fitting of second-degree curve $Y = a + bX + cX^2$	01
4	Fitting of exponential curve of type $Y = ab^X, Y = aX^b$.	01
5	Fitting of Binomial distribution and computation of expected frequencies.	01
6	Model sampling from Binomial distribution...	01
7	Applications of Binomial	01
8	Fitting of Poisson Distribution and computation of expected frequencies.	01
9	Applications of Poisson distribution and Geometric distribution.	01
10	Model sampling from Poisson distribution.	01
11	Project	02

Note: Every practical is equivalent to four theory lectures per batch per week

Generic / Open Elective (OE)**Paper code and title – OE-151-STTS-TH: Elementary Statistics for Social Sciences-II****Course type: Theory****No. of Credits: 02****No. of Contact Hours: 30****Course Outcomes:**

At the end of the course, students should be able to:

1. Students will understand the key concepts of the theory of attributes and their applications in data classification.
2. Students will analyse relationships between two attributes using different.
3. Students will create and interpret tables for both qualitative and quantitative data, recognizing their advantages and limitations.
4. Students will construct various graphical representations, such as histograms and ogives, to effectively visualize data distributions.
5. Students will classify and summarize data using ungrouped and grouped frequency distributions.
6. Students will compute and evaluate measures of central tendency and understand their respective merits and demerits.

Unit No.	Content	No. of Hours
	Classification and Tabulation	09
	1.1 Raw data, class frequency, Discrete frequency distribution, Continuous frequency distribution, Inclusive and Exclusive methods of classifications, class limits, class boundaries, mid values, class width, open end class, Cumulative frequencies. relative frequencies, frequency density.	
	1.2 Numerical Examples.	
	1.3 Definition and concept of tabulation, objectives of tabulation, Components of table, Advantages and limitation of tabulation.	
	1.4 Tabulation for quantitative data.	
2	Graphical Representation	07
	2.1 Histogram, frequency curve, frequency Polygon, less than O'give curve, more than O'give Curve, Stem and leaf chart.	
	2.2 Numerical Examples.	

3	Theory of Attributes	04
3.1	Manifold classification, Dichotomy, order of class, Positive class frequency, Negative class frequency, Ultimate class frequency.	
3.2	Consistency of data up to two attributes.	
3.3	Concepts of independence and association of two attributes	
3.4	Yule's coefficient of association.	
5	Measures of Central Tendency	10
5.1	Concept of central tendency, Requisites of good measure of central tendency, Measures of central tendency.	
5.2	Arithmetic mean: Definition. Computation of arithmetic mean for ungrouped and grouped frequency distribution, Merits and Demerits.	
5.3	Median: Definition. Computation of median for ungrouped and grouped frequency distribution, Merits and Demerits	
5.4	Mode: Definition. Computation of mode for ungrouped and grouped frequency distribution, Merits and Demerits.	

References:

1. Agarwal B. L. (2003). Programmed Statistics, 2nd edition, New Age International Publishers, New Delhi.
2. Deborah J. Rumsey, Statistics for Dummies.
3. David S. Moore, George P. McCabe, and Bruce A. Craig, Introduction to the Practice of Statistics .
4. Gupta S.C. and Kapoor, V. K. (1983). Fundamentals of applied Statistics, 3rd Edition, Sultan Chand and Sons Publishers, New Delhi.

Generic / Open Elective (OE)**Paper Code and Title: OE-152-STS-TH: Elementary Commercial Statistics-II**

Course type- Theory

No. of credits – 2

No. of contact hours –30

Course Outcomes:

By the end of the course, students should be able to:

1. Identify the appropriate scale of measurement for a particular characteristic under study.
2. Represent data using appropriate diagram/graph.
3. Calculate and describe data through measures of central tendency and dispersion.
4. To compute the correlation coefficient for bivariate data and interpret it.
5. To fit linear regression.
6. Interpret the utilization of measures of central tendency and dispersion to compare more group results.

Unit No.	Content	No. of Hours
1	Types of data and Basic Exploratory Analysis:	06
1.1	Primary data, Secondary data, Time Series data, Cross-sectional data, Directional data, Survival data.	
1.2	Frequency classification: Raw data, ungrouped frequency distribution, grouped frequency distribution, cumulative frequency distribution, inclusive and exclusive methods of classification, open end classes.	
1.3	Diagrammatic representation of statistical data: simple bar diagram, subdivided bar diagram, multiple bar diagram, percentage bar diagram, pie diagram.	
1.4	Graphical representation of statistical data: Line graph, histogram (with equal and unequal class width) frequency curve, ogive curves, Stem and leaf chart.	
1.5	Numerical problems related to real life situations.	

- 2 Measures of central tendency** **06**
- 2.1 Concept of central tendency, requisites of good measures of central tendency.
- 2.2 Arithmetic mean: Definition, computation for ungrouped and grouped data, combined mean, weighted mean, merits and demerits
- 2.3 Geometric mean and harmonic mean.
- 2.4 Median and Mode: Definition, formula for computation for ungrouped and grouped data, graphical method, merits, and demerits. Empirical relation between mean, median and mode. (without proof).
- 2.5 Numerical problems related to real life situations.
- 3 Measures of Dispersion** **06**
- 3.1 Concept of dispersion and measures of dispersion, requisites of good measures of dispersion, absolute and relative measures of dispersion.
- 3.2 Range and Quartile Deviation: definition for ungrouped and grouped data and their coefficients, merits and demerits.
- 3.3 Variance and Standard deviation: definition for ungrouped and grouped data, coefficient of variation.
- 3.4 Numerical problems related to real life situations.
- 4 Correlation and Regression Analysis** **12**
- 4.1 Bivariate data, Scatter diagram and interpretation
- 4.2 Concept of correlation between two variables, positive correlation, negative correlation, no correlation.
- 4.3 Covariance between two variables: Definition, computation, effect of change of origin and scale (without proof).
- 4.4 Karl Pearson's coefficient of correlation (r): Definition, computation for ungrouped data and interpretation. Properties: i) $-1 \leq r \leq 1$ (without proof), ii) Effect of change of origin and scale (without proof).

- 4.5 Spearman's rank correlation coefficient: Definition, computation, and interpretation (without ties). In the case of ties, compute Karl Pearson's correlation coefficient between ranks. (Spearman's rank correlation coefficient formula with correction for ties not expected.)
- 4.6 Concept of dependent and independent variables in regression.
- 4.7 Identification of response and predictor variables and relation between them.
- 4.8 Meaning of regression, difference between correlation and regression, Connection between correlation and regression.
- 4.9 Fitting of line $Y = a + bX$, a and b are estimated using least square principle method (without proof). Interpretation of regression coefficient.
- 4.10 Numerical Problems related to real life situations

References:

1. Brase C. H., Brace C. P (2016). Understandable Statistics, Concepts and Methods, 12th Edition, Cengage Learning.
2. Freedman D., Pisani R., Purves R. (2007). Statistics, 4th Edition, W. W. Norton and Company.
3. Freund J. E. (1977). Modern Elementary Statistics. 4th Edition, Prentice Hall of India Private Limited, New Delhi.
4. G.W. Snedecor, W.G. Cochran (1989), Statistical Methods, John Wiley & sons.
5. Goon A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1. 6th Revised Edition, The World Press Pvt. Ltd., Calcutta.
6. Gupta S. C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics. 8th Edition, Sultan Chand and Sons Publishers, New Delhi.
7. Kennedy, W.J. and Gentle, J.E., (2021). Statistical computing. Routledge.
8. Moore D. S., Notz W. I., Fligner M. A. (2013). The Basic Practice of Statistics, 6th Edition, Ruth Baruth.
9. Utts J. M., Heckard R. F. (2010). Mind On Statistics, 4th Edition, Richard Stratton Publisher.
10. Zealure C. H. (1998). Fundamentals of Descriptive Statistics. 1st Edition, Routledge, U.K. (Taylor and Francis Group).

Skill Enhancement Course (SEC)

Paper code and title – SEC-151-STS-PR: Computational Statistics using MS-EXCEL

Course type: Practical

No. of Credits: 02

No. of Contact Hours: 60(48+12)

Objectives:

The main objective of this course is to introduce basics of statistics to the students using MS-Excel.

Course Outcomes:

At the end of the course, students are expected to be able,

1. To analyze data pertaining to discrete variables and to interpret the results.
2. To compute various measures of central tendency, dispersion, skewness, and kurtosis using MS-Excel and interpret the values of summary statistics.
3. To summarize and analyze the data using MS-Excel.
4. To apply standard discrete probability distribution to different real situations.
5. To compute Correlation and fit regression to the real data using MS-Excel.

Unit No.

Content

- | | |
|----------|---|
| 1 | Summary Statistics (Grouped and Ungrouped Data): |
| 1.1 | Measures of central tendency: A.M., G.M., H.M., Trimmed arithmetic mean, Mode
Partition values: Median, Quartiles, Deciles, percentiles. |
| 1.2 | Measure of Dispersion: Range and coefficient of range, Quartile Deviation and
coefficient of Quartile Deviation, Mean deviation about: mean, median and mode,
Mean square deviation, variance and standard deviation, Coefficient of Variation
(C.V.). |
| 1.3 | Measures of skewness and kurtosis: raw moments, central moments, Karl Pearson's
coefficient of skewness, Bowley's coefficient of skewness, Pearson coefficient of
skewness and kurtosis, boxplot. |
| 2 | Computation of probabilities: |
| 2.1 | Degenerate, Discrete Uniform |
| 2.2 | Bernoulli, Binomial, |

- 2.3 Hypergeometric
- 2.4 Poisson, Geometric, Negative Binomial, (p. m. f. and c. d. f. plot for all above distributions.)
- 3 Model Sampling, correlation, and regression**
- 3.1 Random number generation, scatter plot, Covariance, Correlation.
- 3.2 Karl Pearson's coefficient of correlation.
- 3.3 Regression Models: linear, exponential, second degree, power.
- 3.4 coefficient of determination.

List of Practicals:

Sr. No.	Title of the Experiment	No. of Experiments
1	Measures of Central Tendency (Ungrouped Data)	1
2	Measures of Central Tendency (Grouped Data)	1
3	Measures of Dispersion (Ungrouped Data)	1
4	Measures of Dispersion (Grouped Data)	1
5	Measures of Moments, Skewness, Kurtosis (Ungrouped Data)	1
6	Measures of Moments, Skewness, Kurtosis (Grouped Data)	1
7	Computation of probabilities from Binomial, Poisson, Geometric, Negative Binomial, Hypergeometric distribution.	2
8	Model Sampling.	1
9	Correlation	1
10	Fitting of linear, exponential, second degree, power regression models.	2

Note: Every practical is equivalent to four theory hours per batch per week.



Haribhai V. Desai College of Arts, Science & Commerce, Pune-02

(Affiliated to Savitribai Phule Pune University)

Faculty of Science & Technology

F.Y.B.Sc. (Computer Science) Statistics Syllabi in NEP

(Level 4.5)

National Education Policy Syllabus

To be implemented from Academic Year 2024-2025

Title of the program: B.Sc.(Computer Science) Statistics

Preamble of the syllabus:

Statistics is a branch of science that can be applied practically in every walk of life. Statistics deals with any decision-making activity in which there is certain degree of uncertainty and Statistics helps in taking decisions in an objective and rational way. The student of Statistics can study it purely theoretically which is usually done in research activity, or it can be studied as asystematic collection of tools and techniques to be applied in solving a problem inreal life.

In last 15 to 20 years, computers have been playing very crucial role in society. Theuse of computers has horizontally spread and penetrated vertically in the society. It has become a part and parcel of common man. Thus, there is a hugedemand for computer education.

The Savitribai Phule Pune University of Pune had done a pioneering work in this area and Three-year degree course B. Sc. (Computer Science) of University of Pune is very popular among the student community and I. T. Industry. This course covers various subjects which are required directly or indirectly forbecoming computer professional. Statistics is one such important subject which is required and is extensively used in a vast spectrum of computer-based applications. **Data Mining and Warehousing, Big Data Analytics, Theoretical Computer Science, Reliability of a computer Program or Software, Machine Learning, Artificial Intelligence, Pattern Recognition, Digital Image Processing, Embedded Systems**are just few applications to name where Statistics can be extensively used.

Introduction: National Education Policy (NEP) 2020:**To be implemented from 2024-25:**

The syllabus of Statistics for **First Year** of this course covers development of statistical required for I.T. industry and Data Science. The teachers teaching this syllabus and students should give emphasis on understanding the concepts and ability to apply statistical tools and techniques and not on the theoretical discussion. It isexpected that at the end of the course, a student should be well equipped tolearn and apply acquired techniques in computer -based applications.

At **second year of under-graduation**, students are expected to perform statistical analysis using software *R* and apply the tools in the field of technology. Students are expected to study various probability distributions and their applications to real life situations. An important branch of Statistics, viz. testing of hypotheses related to mean, variance, proportion, correlation etc. will be

introduced. Some topics related to applications of Statistics will also be introduced. It is a skill-oriented part of the course.

Structure of Evaluation: (For each paper)

1. Continuous Internal Assessment (CIA): 15 Marks
2. End Semester Examination (ESE): 35 Marks
3. Total: 50 Marks

Structure of Statistics in B.Sc. (Computer Science) degree Program

Structure of the course, the pattern of examination and question papers are as per Savitribai Phule Pune University:

List of Skill Enhancement Courses (SEC):

Year / Level	Sem	Code Number	Title of the paper (Practical)	Credits allotted	Lecture/Practical hours per week
I 4.5/100	I	SEC-101-CS-PR	Statistical Methods for Computer Science-I (Practical Course)	02	04
	II	SEC-151-CS-PR	Statistical Methods for Computer Science-II (Practical Course)	02	04
II 5.0/200	IV	SEC-251-CS-PR	Statistical Analysis using R-software (Practical Course)	02	04

SEMESTER -I

Paper Code and Title: SEC-101-CS-PR: Statistical Methods for Computer Science-I

Course type- Practical

No. of credits – 02

No. of contact hours –60(48+12)

Course Outcomes:

By the end of the course, students should be able to:

1. Present the complex data in tabular format.
2. Use various diagrammatic and graphical techniques to represent statistical data and interpret the data.
3. Compute various measures of central tendency, dispersion, skewness, and kurtosis using MS-Excel and interpret the results
4. Establish relation between variables and estimate response for given bivariate data using software and interpret the results

Sr. No.	Title of the experiment	No. of Experiments
1.	Tabulation and construction of frequency distribution. (Use of at least two data sets more than 50 observations- each for constructing frequency distribution)	1
2.	Diagrammatic representation of statistical data using MS-EXCEL and data interpretation. (simple bar diagram, subdivided bar diagram, multiple bar diagram, percentage bar diagram, pie diagram) Data interpretation from diagrams.	1
3.	Graphical representation of statistical data using MS-EXCEL: Histogram and frequency curve, Ogive curves. Determination of median graphically. Data interpretation from graphs.	1
4.	Summary statistics – I: Computation of measures of central tendency for ungrouped data (AM, Median and Mode) using MS-Excel by regular formula method and using direct command. (Discuss use of an appropriate measure).	1

5.	Summary statistics – II: Computation of measures of central tendency for grouped data (AM, Median and Mode) using MS-Excel by regular formula method. (Discuss use of an appropriate measure).	01
6.	Summary statistics – III: Computation of measures of dispersion for ungrouped data (Range, quartiles, variance, standard deviation, coefficient of variation) using MS-Excel by regular formula method and by direct command. Use of an appropriate measure and interpretation of results.	01
7	Summary statistics – IV: Computation of measures of dispersion for grouped data (Range, quartiles, variance, standard deviation, coefficient of variation) using MS-Excel by regular formula method. Use of an appropriate measure of dispersion and interpretation of results.	01
8.	Computations of raw and central moments (not using the relation formula), measures of skewness and kurtosis (using Karl Pearson's formula and moments) using MS-Excel by regular formula method and by direct command. Interpretation of results.	01
9.	Measures of Attributes	01
10	Scatter diagram, computation of covariance and Karl Pearson's correlation coefficient using MS-EXCEL by regular formula method and by direct command. Interpretation of results.	01
11	Fitting of line of regression $Y = a + bX$, using MS-Excel by regular formula method and by direct command (use scatter plot for explaining the linear relationship).	01
12	Data Collection, its condensation and representation using MS-Excel.	01

Note:

1. Every practical is equivalent to four theory lectures per batch per week.
2. One hour is reserved for theory explanation of corresponding practical.
3. For project, a group of maximum 8 students be made. All the students in a group are given equal marks for project. Different data sets from primary or secondary sources may be collected.

SEMESTER II

Paper Code and Title: SEC-151-CS-PR: Statistical Methods for Computer Science-II

Course type- Practical

No. of credits – 02

No. of contact hours –60(48+12)

Objectives:

At the end of this course students are expected to be able:

1. To fit second-degree curve, and exponential curves.
2. To estimate trend by using time series data.
3. To understand concept of probability.
4. To estimate probabilities of standard probability distributions
5. To perform tests based on normal, Chi-Square, t and F distributions.

Sr. No.	Title of the experiment	No. of Experiments
1	Fitting of second-degree curve $Y = a + bX + cX^2$ using MS-Excel. (Use of scatter plot for explaining the nonlinear relationship between two variables)	01
2	Fitting of exponential curve of type $Y = ab^x$, $Y = aX^b$ using MS-Excel. (Use of scatter plot for explaining the nonlinear relationship between two variables)	01
3	Time Series- Estimation of trend by using the method of moving averages using MS-Excel by using regular formula method and by direct command.	01
4	Time Series- Estimation of trend by using exponential smoothing using MS-Excel by regular formula and by direct command.	01
5	Problems on probability theory ($P(A \cup B)$, $P(A' \cup B')$, $P(A' \cap B)$, $P(A' \cap B')$). Use Venn diagram whenever possible.	01
6	Computation of probability values for Normal distribution and Chi-square distribution using MS Excel command.	01
7	Computation of probability values for t distribution and F distribution using MS Excel command.	01
8	<p>Test for means:</p> <p>i) $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$, $H_1: \mu > \mu_0$, $H_1: \mu < \mu_0$ when σ is known.</p> <p>ii) $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$, $H_1: \mu_1 > \mu_2$, $H_1: \mu_1 < \mu_2$ when σ is known</p> <p>Test for proportion:</p>	01

- i) $H_0: P = P_0$ against $H_1: P \neq P_0, H_1: P > P_0, H_1: P < P_0$.
- ii) $H_0: P_1 = P_2$ against $H_1: P_1 \neq P_2, H_1: P_1 > P_2, H_1: P_1 < P_2$
- 9 **Test based on students t-distribution:** 01
- i) $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0, H_1: \mu > \mu_0, H_1: \mu < \mu_0$ when σ is unknown.
- ii) $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2, H_1: \mu_1 > \mu_2, H_1: \mu_1 < \mu_2$ when σ is unknown
- iii) Paired t-test
- 10 **Test based on χ^2 distribution:** 01
- i) Goodness of fit
- ii) Independence of attributes (2 x 2).
- iii) Independence of attributes (2 x 3 or 3 x 2 or 3 x 3)
- 11 Tests based on F-distribution: 01
- i) $H_0: \sigma_1^2 = \sigma_2^2$ against $H_1: \sigma_1^2 \neq \sigma_2^2, H_1: \sigma_1^2 > \sigma_2^2, H_1: \sigma_1^2 < \sigma_2^2$, known means
- ii) $H_0: \sigma_1^2 = \sigma_2^2$ against $H_1: \sigma_1^2 \neq \sigma_2^2, H_1: \sigma_1^2 > \sigma_2^2, H_1: \sigma_1^2 < \sigma_2^2$, unknown means
- 12 Project (Part-II) - Analysis of data collected in semester-I 01

Note:

1. Every practical is equivalent to four theory lectures per batch per week
2. One hour is reserved for theory explanation of corresponding practical.
3. For project, a group of maximum 8 students be made. All the students in a group are given equal marks for project. Different data sets from primary or secondary sources may be collected.
