



**Progressive Education Society's
Modern College of Arts, Science and Commerce
Pune 16**
(An Autonomous College Affiliated to Savitribai Phule Pune University)

Faculty of Science & Technology

F.Y.B.Sc. (Computer Science) Statistics

**Choice Based Credit System Syllabus
To be implemented from Academic Year 2022-2023**

Title of the Course: B. Sc. (Computer Science) STATISTICS**Preamble of the Syllabus:**

Statistics is an applied Mathematics 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 a systematic collection of tools and techniques to be applied in solving a problem in real life.

In last 15 to 20 years, computers are playing very crucial role in the society. The use of computers has horizontally spread and also penetrated vertically in the society. It has become a part and parcel of common man. Thus there is a huge demand for computer education.

The 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 for becoming 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: The syllabus of Statistics for First Year of this course covers basic concepts and terminology in Statistics and covers basic tools and methods required for data analysis. 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 is expected that at the end of the course, a student should be well equipped to learn and apply acquired techniques in computer based applications.

Structure of the Subject

Structure of the subject and the pattern of examination and question papers are as specified below.

Structure of F. Y. B. Sc. (Computer Science) Statistics

Semester	Paper code	Paper	Paper title	Credits	Marks		
					CIA	ESE	Total
1	22-CSST-111	I	Descriptive Statistics I	2	15	35	50
	22-CSST-112	II	Mathematical Statistics	2	15	35	50
	22-CSST-113	III	Statistics Practical Paper I	1.5	15	35	50
2	22-CSST-121	I	Methods of Applied Statistics	2	15	35	50
	22-CSST-122	II	Continuous Probability Distributions and Testing of Hypothesis	2	15	35	50
	22-CSST-123	III	Statistics Practical Paper II	1.5	15	35	50

Semester I
Paper-I
22-CSST-111: Descriptive Statistics

No. of Credits: 2

No. of lectures: 40

Course Outcomes:

The students will be able to:

- CO1) recall the concept of Statistical data. They will understand how to collect and condense data using various statistical methods and how to classify and represent that data graphically.
- CO2) remember through various statistical measures such as measures of central tendency, dispersion.
- CO3) understand the concept of comprehensive introduction to descriptive statistics which are required for becoming computer professional.
- CO4) describe the moments skewness and kurtosis.
- CO5) apply the concept of Attributes

TOPICS/CONTENTS:**UNIT1: Data Description and Presentation of Data (9L)**

- Definition, importance, scope and limitations of statistics.
- Types of data (Primary and secondary), Attributes and variables, discrete and Continuous variables.
- Graphical Representation: Histogram, Ogive Curves, Steam and leaf chart, Mosaic plot
[Note: Theory paper will contain only procedures. Problems to be included in practical]
- Numerical problems related to real life situations.

UNIT 2: Descriptive Statistics (14L)

- Measures of central tendency: Concept of central tendency, requisites of good measures of central tendency.
- Arithmetic mean: Definition, computation for ungrouped and grouped data, properties of arithmetic mean (without proof) combined mean, weighted mean (Index number), merits and demerits.
- Median and Mode: Definition, formula for computation for ungrouped and grouped data, location of median and mode using appropriate graph, merits and demerits. Empirical relation between mean, median and mode
- Partition Values: Quartiles, Box Plot and interpretation of box plot.
- Concept of dispersion, requisites of good measures of dispersion, absolute and relative measures of dispersion.
- Measures of dispersion : Range, Semi-range and Quartile Deviation for ungrouped and

- grouped data and their coefficients, merits and demerits,
- Variance and Standard deviation for ungrouped and grouped data, coefficient of variation, combined variance (for two groups only), merits and demerits.
- Numerical problems related to real life situations.

UNIT 3: Moments, Skewness and Kurtosis**(10L)**

- Concept of Raw and Central moments: Formulae for ungrouped and grouped data (only first four moments), relation between central and raw moments upto fourth order. (without proof)
- Measures of Skewness: Concept of symmetry, Types of skewness, Pearson's and Bowley's coefficient of skewness, Measure of skewness based on moments.
- Measure of Kurtosis: Types of kurtosis, Measure of kurtosis based on moments and based on partition values.
- Numerical problems related to real life situations

UNIT4: Theory of Attributes**(7L)**

- Attributes: Concept, classification, notion of manifold classification, dichotomy, Concept of a Likert scale, class- frequency, order of a class, positive class frequency, negative class frequency, ultimate class frequency, relationship among different class frequencies (up to two attributes) using dot operator
- Consistency of data upto two attributes.
- Concepts of independence and association of two attributes using Mosaic plot
- Yule's coefficient of association (Q), $-1 \leq Q \leq 1$, interpretation.

References:

Sr. No.	Title of the book	Name of Author	Publications
1	Statistical Methods	George W. Snedecor, William G. Cochran	Wiley & sons
2	Programmed Statistics	B.L. Agarwal	New Age International Publishers
3	Modern Elementary Statistics	Freund J.E	Pearson Publication 2005
4	Fundamentals of Applied Statistics	Gupta and Kapoor	S.Chand and Sons, New Delhi, 1987.
5	An Introductory Statistics	Kennedy and Gentle	
6	Fundamentals of Statistics, Vol. 1, Sixth Revised Edition,	Goon, A. M., Gupta, M. K. and Dasgupta, B.	The World Press Pvt. Ltd., Calcutta 1983

Semester I
Paper-II
22-CSST-112: Mathematical Statistics

No. of Credits : 2

No. of lectures: 40

Course Outcomes:

The students will be able to:

CO1) recall the concept of Probability. They will understand how to determine deterministic and non-deterministic models, events, random experiment and how to calculate numerical problems using real life data.

CO2) learn conditional probability and Bayes theorem which is useful for calculating posterior probabilities.

CO3) understand the concept random variables and types of random variables.

CO4) obtain the probability distributions of random variables.

CO5) learn the concept of discrete random variables and will be able to apply the standard discrete probability distributions like Binomial, Poisson, Geometric to different real life situations

TOPICS/CONTENTS:

UNIT 1: Theory of Probability

(10L)

- Counting Principles, Permutation, and Combination.
- Pigeonhole Principle
- Deterministic and non- deterministic models.
- Random Experiment, Sample Spaces (Discrete)
- Events: Types of events, Operations on events.
- Probability - classical approach, probability models approach, axiomatic approach.
- Theorems of probability (without proof)
 - i) $0 \leq P(A) \leq 1$ ii) $P(A) + P(A^c) = 1$ iii) $P(\Phi) = 0$ iv) $P(A) \leq P(B)$ when $A \subset B$
 - v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Numerical problems related to real life situations (Binomial, Geometric and hypergeometric)

UNIT 2: Conditional Probability and Independence

(8L)

- Concepts and definitions of conditional probability, multiplication theorem
 $P(A \cap B) = P(A) \cdot P(B|A)$
- Bayes' theorem (without proof). True positive, false positive and sensitivity of test as application of Bayes' theorem.
- Concept of Posterior probability, problems on posterior probability.
- Concept and definition of independence of two events, example of selection of sample using SRSWOR and SRSWR
- Numerical problems related to real life situations.

UNIT 3: Random Variable (10L)

- Definition of random variable (univariate, bivariate), discrete random variable.
- Definition of probability mass function (p.m.f.) of discrete r.v.
- Cumulative distribution function (c.d.f.) of discrete r.v. and its properties (Characteristic properties only)
- Definition of expectation and variance of discrete r.v., theorem on expectation and variance (statement only), change of origin and scale property.
- Determination of median and mode using p.m.f. only.
- Numerical problems related to real life situations.

UNIT 4 : Standard Discrete Distributions (12L)

- Degenerate distribution
- Discrete Uniform Distribution: definition, mean, variance.
- Binomial Distribution: definition, mean, variance, additive property(statement only), Bernoulli distribution as a particular case with $n = 1$.
- Geometric Distribution (p.m.f $p(x) = pq^x$, $x = 0, 1, 2 \dots$): definition, mean, variance lack of memory property.
- Poisson Distribution: definition, mean, variance, mode, additive property, limiting case of $B(n, p)$
- Numerical problems related to real life situations.

* Only statement of mean and variance, derivation is not expected.

References:

Sr. No.	Title of the book	Name of Author	Publications
1	A First course in Probability	Sheldon Ross.	Pearson Education Inc.
2	Statistical Methods (An Introductory Text)	Medhi J.	New Age International. 1992
3	Modern Elementary Statistics	Freund J.E.	Pearson Publication 2005
4	Probability, Statistics, Design of Experiments and Queuing Theory with Applications of Computer Science	Trivedi K.S.	Prentice Hall of India, New Delhi 2001
5	Fundamentals of Mathematical Statistics (3rd Edition)	Gupta S. C. and Kapoor V. K.	S. Chand and Sons, New Delhi 1987
6	Mathematical Statistics (3 rd Edition)	Mukhopadhyay P.	Books And Allied (P), Ltd. 2015
7	Introduction to Discrete Probability and Probability Distributions	Kulkarni M.B., Ghatpande S.B.	SIPF Academy 2007
8	Programmed Statistics	B.L. Agarwal	New Age International Publishers

Semester I
Paper-III
22-CSST-113: Statistics Practical

No. of Credits: 2

No. of Practicals: 15

Pre-requisites: Knowledge of the topics in theory papers I and II**Objectives:** At the end of the course students are expected to be able

1. The process of collection of data, its condensation and representation for real life data.
2. To use various graphical and diagrammatic techniques and interpret.
3. To compute various measures of central tendency, dispersion, Skewness and kurtosis.
4. To fit the Binomial and Poisson distributions.
5. To compute the measures of attributes.
6. To study free statistical softwares and use them for data analysis in project.

Sr. No.	Title of the Practical	Number of Practical
1	Diagrammatic and graphical representation using EXCEL and data interpretation. (problems on the basis of SET and NET examination in Paper I to be taken)	1
2	Summary statistics for ungrouped data and comparison for consistency using EXCEL.	1
3	Summary statistics for grouped frequency distribution. (Problems based on central tendency, dispersion, measures of skewness: Karl Pearson's and Quartile measure to be covered)	1
4	a) Measure of Skewness. b) Measure of kurtosis	2
5	Fitting of Binomial distribution and Poisson distribution and computation of expected frequencies. (Use the observed and expected frequencies for the next semester χ^2 test) Using M.S.Excel	1
6	Fitting of Poisson distribution and computation of expected frequencies. (Use the observed and expected frequencies for the next semester χ^2 for test.) (Give one data set for fitting both Poisson and Binomial distributions.)	1
7	Measure of attributes. (Two attributes only)	1
8	Study of free statistical softwares and writing a report on it. (individual activity)	3
9	Sample survey Theory and practical (Project(Part-I) -Data collection using survey)	5

Notes:

- 1) For project, a group of maximum 8 students be made.
- 2) All the students in a group are given equal marks for project.
- 3) Different data sets from primary or secondary sources may be collected.

Semester II
Paper-I
22-CSST-121: Methods of Applied Statistics

No. of Credits : 2

No. of lectures: 40

Course Outcomes:

On completion of the course, the students will be able to:

- CO1) understand the concept of correlation of two or more variables and regression of two interrelated variables.
- CO2) analyze concept of Multiple Regression and Multiple & Partial Correlation.
- CO3) solve the problems based on Multiple Regression and Multiple & Partial Correlation.
- CO4) apply above techniques in real life situation.

TOPICS/CONTENTS:

UNIT 1: Correlation (For ungrouped data) (10L)

- Concept of bivariate data, concept of correlation, Positive correlation, Negative correlation, Zero correlation, Scatter diagram, its interpretation.
- Covariance: Definition, Properties
- Karl Pearson's coefficient of correlation, properties of correlation coefficient, Interpretation of correlation coefficient
- Spearman's rank correlation coefficient (formula without ties).
- Numerical problems

UNIT 2: Regression (for ungrouped data) (12L)

- Concept of linear and nonlinear regression, Illustrations, appropriate situations for regression and correlation.
- Linear regression: Fitting of line of regression Y on X using least square method.
- Concept of regression coefficient and interpretation
- Nonlinear regression models: Second degree curve, exponential curves of the type $Y=ab^x$ and $Y=ax^b$
- Coefficient of determination: Concept, its utility.
- Numerical problems related to real life situations

UNIT 3: Multiple Regression, Multiple and Partial Correlation (For Trivariate Data) (10L)

- Concept of multiple regressions, Yule's Notations.
- Fitting of multiple regression planes. [Derivation of equation to the plane of regression of Y on X₁ and X₂ is expected. Remaining two equations to be written analogously.]
- Concept of multiple correlation: Definition of multiple correlation coefficient and its formula.
- Concept of partial correlation: Definition of partial correlation coefficient and its formula.

UNIT4: Queuing Model**(8L)**

- M/M/1: FIFO as an application of exponential distribution, Poisson distribution and Geometric distribution
- Interarrival rate (λ) service rate (μ), traffic intensity ($\rho = \lambda/\mu < 1$)
- Queue discipline, probability distribution of number of customers in queue, average queue length, average waiting time in queue and average waiting time in queuing system

References:

Sr. No.	Title of the book	Name of Author	Publications
1	Introduction to Linear Regression Analysis	Douglas C. Montgomery, , Elizabeth A. Peck, G. Geoffrey Vining,	Wiley
2	Fundamentals of Applied Statistics(3rd Edition)	Gupta and Kapoor	S.Chand and Sons(1987)
3	Fundamentals of Statistics, Vol. 1, Sixth Revised Edition	Goon, A. M., Gupta, M. K. and Dasgupta, B.	The World Press Pvt. Ltd., Calcutta1983
4	Probability, Statistics, Design of Experiments and Queuing Theory with Applications of Computer Science	Trivedi K.S.	Prentice Hall of India, New Delhi 2001

Semester II
Paper-II

CSST122: Continuous Probability Distributions and Testing of Hypotheses

No. of Credits : 2

No. of lectures: 40

Course Outcomes:

The students will be able to:

- CO1) learn the concept of Continuous random variables and the standard continuous probability Distributions like Exponential, Pareto, Normal to different real life situations
- CO2) remember the concept and definitions related to Testing of hypothesis.
- CO3) understand the concept Parametric Tests like Large Sample Test, Small Sample Tests.
- CO4) to obtain the random numbers and pseudo random numbers using Simulation.
- CO5) analyze using some statistical software and verify their theoretical knowledge about different Statistical entities and computations during practical sessions using MS-Excel.

TOPICS/CONTENTS

UNIT 1: Standard Continuous Probability Distributions (10L)

- Continuous r. v., Definition of probability density function (p.d.f.) of continuous r. v. Cumulative distribution function (c.d.f.) of continuous r.v. and their properties.
- Uniform Distribution: statement of p.d.f., mean, variance, nature of probability curve. Theorem (without proof): The distribution function of any continuous r.v. follows $U(0, 1)$ distribution
- Exponential Distribution: statement of p.d.f. of the form $f(x) = (1/\theta) e^{-x/\theta}$, mean, variance, nature of probability curve, lack of memory property.(with proof)
- Parato distribution: Form of pdf $f(x): \alpha/ x^{(\alpha+1)}$; $x \geq 1$, $\alpha > 0$. Mean, variance, symmetry, applications
- Normal Distribution: statement of p.d.f., identification of parameters, nature of probability density curve, standard normal distribution, symmetry, additive property, linear property, computations of probabilities using normal probability table, normal approximation to binomial and Poisson distribution, central limit theorem (statement only), Normal probability plot.
- Numerical problems related to real life situations.

UNIT 2: Concepts and definitions related to testing of hypothesis (4L)

- Concepts of population and sample.
- Definitions: random sample from a probability distribution, parameter, statistic, standard error of estimator.
- Concept of null hypothesis and alternative hypothesis (Research hypothesis), critical region, level of significance, type I and type II error, one sided and two sided tests, Test of hypothesis, p-value.

UNIT 3: Tests of significance**(20L)**

- Large Sample Tests
 - $H_0: \mu = \mu_0$ Vs $H_1: \mu \neq \mu_0$ (two sided tests)
 - $H_0: \mu_1 = \mu_2$ Vs $H_1: \mu_1 \neq \mu_2$ (two sided tests)
 - $H_0: P = P_0$ Vs $H_1: P \neq P_0$ (two sided tests)
 - $H_0: P_1 = P_2$ Vs $H_1: P_1 \neq P_2$ (two sided tests)
 - Numerical problems related to real life situations.
- Test based on F- distribution
 - F-test for testing significance of equality of two population variances.
- Tests based on t – distribution
 - $H_0: \mu_1 = \mu_2$ Vs $H_1: \mu_1 \neq \mu_2$ (two sided tests)
 - Paired t-test.
- Tests based on Chi square distribution
 - Chi-square test for goodness of fit
 - Test for independence of attributes (mxn and 2x2)
- Numerical problems related to real life situations.

UNIT 4: Simulation**(6L)**

- Introduction, concept of simulation , random numbers, pseudo random numbers Box mullar transformation , Advantages , Disadvantages of Simulation. Applications
- Methods of simulation, Linear congruential generator and simulation from Uniform, Exponential and Normal Distribution.

References:

Sr. No.	Title of the book	Name of Author	Publications
1	A First course in Probability	Sheldon Ross.	Pearson Education Inc
2	Statistical Methods (An Introductory Text)	Medhi J.	Age International. 1992
3	Modern Elementary Statistics	Freund J.E.	Pearson Publication. 2005
4	Fundamentals of Mathematical Statistics (3rd Edition)	Gupta S. C. and Kapoor V. K.	Chand and Sons, New Delhi. 1987
5	Mathematical Statistics (3 rd Edition)	Mukhopadhyay P.	Books And Allied (P), Ltd. 2015
6	Simulation Modelling and Analysis	Law A. M. and Kelton W.D.	Tata McGraw Hill 2007
7	Programmed Statistics	B.L. Agarwal	New Age International Publishers
8	Common Statistical Tests Kulkarni M.B	Ghatpande, S.B., Gore S.D.	Satyajeet Prakashan, 1999

Semester II
Paper-III
CSST 123: Statistics Practical

No. of Credits : 2

No. of Practicals: 15

Pre-requisites: Knowledge of the topics in theory papers I and II**Objectives:** At the end of the course students are expected to be able

1. To understand the relationship between two variables using scatter plot.
2. To compute coefficient of correlation, coefficient of regression.
3. To fit various regression models and to find best fit.
4. To fit the Normal distribution.
5. To understand the trend in time series and how to remove it.
6. To apply inferential methods for real data sets.
7. To generate model sample from given distributions.
8. To understand the importance and functions of different statistical organizations in the development of nation.

Sr. No.	Title of the Practical	Number of Practical
1	Linear correlation and Simple linear regression	2
2	Fitting of non-linear regression. (use of scatter plot for explaining the non-linear relationship between two variables) finding the best fit by use of R ² using EXCEL	1
3	Fitting of linear regression model (Simple and Multiple) using EXCEL.	1
4	Fitting of normal distribution and computation of expected frequencies.	1
5	Tests, a) Large sample tests. b) Exact test	2
6	F test, t test, χ^2 test using EXCEL (one problem each with equal and unequal variance)(χ^2 test – for goodness of fit-use fitted problems of Binomial, Poisson and Normal distribution in previous practical problems)	1
7	Simulation : Model sampling from continuous uniform, exponential and normal distributions using Excel	1
8	Write a report on application of some statistical technique in the field of computers.(individual activity)	1
9	Project (Part-II) - Analysis of data collected in semester-I	5
	Total	15

Notes:

- i) For project, a group of maximum 8 students be made.
- ii) All the students in a group are given equal marks for project.
- iii) Students will be asked to use Statistical methods which they have learnt and use of free statistical software for data analysis.
- iv) Project will carry 5 marks as part of internal evaluation.
- v) The formal project report should be prepared by each student and it must be attached in Statistics journal.