

**Anekant Education Society's
Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati**

Autonomous

Course Structure for F. Y. B. Sc. (Computer Science) STATISTICS

Semester	Paper Code	Title of Paper	No. of Credits
II	CSST-1201	Statistical Methods- II	2
	CSST-1202	Statistical Testing of Hypothesis and Use of R Software	2
	CSST-1203	Practical-II	2

Programme Outcomes:

PO1. Computer Knowledge : Apply the knowledge of mathematics, statistics and computer fundamentals to IT application

PO2. Design / Development of solution: Design solution for IT applications using latest technologies and develop and implement the solutions using various latest language.

PO3. Modern tool usage : Create, select and apply appropriate techniques, resources and modern engineering and IT tools including prediction and modelling to complex IT applications with an understanding of the limitations.

PO4. Environment and sustainability : Understand the impact of the IT analyst solutions in societal and environmental contexts, and demonstrate the knowledge and need for sustainable development.

PO5. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO6. Individual and Team work : Function effectively as an individual, and as a member or leader in diverse team, and in multidisciplinary settings.

PO7. Innovation, employability and Entrepreneurial skills : Identify opportunity, pursue that opportunity to create value and wealth for the betterment of the individual and society at large. Develop the capacity to study and research independently that will help to develop skills for transition to employment in hardware/software companies.

SYLLABUS (CBCS) FOR F. Y. B. Sc. (Computer Science) STATISTICS
(w.e. from June, 2019)
2019 pattern
Academic Year 2019-2020

Class : F.Y. B. Sc. (Computer Science) (Semester- II)

Paper Code: CSST-1201

Paper : I

Title of Paper: Statistical Methods II

Credit : 2 credits

No. of lectures: 36

A) Learning Objectives:

1. The main objective of this course is to acquaint students with concept of discrete random variable and its probability distribution.
2. Fit linear regression to two variables and multiple regression (for trivariate data).
3. To study the discrete random variables and their distributions and also some standard discrete probability distributions with real life situations.

B) Course Outcome:

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

- CO1.** To apply discrete probability distributions studied in this course in different situations.
- CO2.** Know some standard discrete probability distributions with real life situations.
- CO3.** How to fit the regression model to the given bivariate data.
- CO4.** Understand the basic concept of linear regression and fit linear regression model to the bivariate data, interpretation of coefficients, and prediction of outcomes.
- CO5.** Fit quadratic and exponential curves to the bivariate data to investigate relation between two variables.
- CO6.** Describe bivariate random variable, joint distribution function, joint probability mass function, marginal and conditional distributions, mathematical expectation, and independence of random variables, and apply their properties in problem-solving.
- CO7.** Apply distribution's in different real life situations

TOPICS/CONTENTS:

UNIT 1: Regression (for ungrouped data)

(5L)

1.1 Regression, illustrations, appropriate situations for regression and correlation

1.2 Linear regression

1.3 Fitting of straight line using least squares method

1.4 Properties of regression coefficients : $b_{xy} \cdot b_{yx} = r^2$, $b_{xy} * b_{yx} \leq 1$, $b_{xy} = r (\sigma_x / \sigma_y)$
and $b_{yx} = r (\sigma_y / \sigma_x)$, coefficient of determination.

1.5 Numerical problems related to real life situations

UNIT 2: Multiple Regression and Multiple, partial Correlation (For Trivariate Data) (8L)

2.1 Concept of multiple regressions, Yule's Notations.

2.2 Fitting of multiple regression planes.

2.3 Partial regression coefficients, interpretations.

2.4 Concept of multiple correlation: Definition of multiple correlation coefficient and its formula..

2.5 Concept of partial correlation. Definition of partial correlation coefficient and its formula.

UNIT 3: Time series (6L)

3.1 Meaning and utility

3.2 Components of time series

3.3 Additive and multiplicative models

3.4 Methods of estimating trend : moving average method, least squares method and exponential smoothing method

3.5 Numerical problems related to real life situations

Unit-4. Continuous Random Variable: (6L)

4.1 Definition of continuous random variable(r.v.)

4.2 Probability density function (p.d.f.)

4.3 cumulative distribution function (c.d.f.)

4.4 Calculation of Mean, Mode ,Median,Variance,Standard deviation for Continuous random variable.

4.5 Numerical problems related to real life situations.

Unit-5. Standard Continuous Probability Distributions: (11L)

5.1 Uniform Distribution: statement of p.d.f., mean, variance, nature of probability curve.

5.2 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.

5.3 Normal Distribution: statement of p.d.f., identification of parameters, nature of probability density curve, standard normal distribution, symmetry, distribution of $aX+b$, $aX+bY+c$ where X and Y are independent normal variables, computations of

probabilities using normal probability table, normal approximation to binomial and Poisson distribution , central limit theorem (statement only), normal probability plot.
5.4 Numerical problems related to real life situations.

References:

- 1 Statistical Methods, G.W. Snedecor, W.G. Cochran, John Wiley & sons, 1989.
- 2 Statistical Methods, J. Medhi, New Age International, 1992.
- 3 Introduction to Linear Regression Analysis, Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, Wiley
- 4 Time Series Methods, Brockwell and Davis, Springer, 2006.
- 5 Time Series Analysis, 4th Edition, Box and Jenkin, Wiley, 2008.
- 6 Introduction to Discrete Probability and Probability Distributions, Kulkarni M.B., Ghatpande S.B, 2007.

Programme Outcomes and Course Outcomes Mapping :

Course Outcomes	Programme Outcomes(POs)						
	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	-	-	1	-	-	-	-
CO2	-	-	1	-	-	-	-
CO3	2	-	1	-	-	-	-
CO4	-	3	-	-	-	-	-
CO5	-	-	-	-	-	-	3
CO6	-	-	-	-	-	-	-
CO7	-	-	-	-	-	2	-

PO1. Computer Knowledge

CO3. How to fit the regression model to the given bivariate data (Weightage: 2)

Justification: The process of fitting a regression model often involves using computer-based tools for analysis. Understanding how to fit a regression model implies computer knowledge.

PO2. Design / Development of solution:

CO4. Understand the basic concept of linear regression and fit linear regression model to the bivariate data, interpretation of coefficients, and prediction of outcomes (Weightage: 3)

Justification: Designing and developing solutions often involve analyzing data and making predictions. Understanding and applying linear regression contribute directly to solution development.

PO3. Modern tool usage

CO6. Describe bivariate random variable, joint distribution function, joint probability mass function, marginal and conditional distributions, mathematical expectation, and independence of random variables, and apply their properties in problem-solving (Weightage: 3)

Justification: Describing and applying various probability distributions and statistical concepts require the use of modern statistical tools and methods.

PO4. Environment and sustainability

No direct mapping found in the provided information.

PO5. Ethics

No direct mapping found in the provided information.

PO6. Individual and Team work

CO7. Apply distribution's in different real-life situations (Weightage: 2)

Justification: Applying probability distributions in real-life situations may involve both individual and team work, especially when solving problems that require a comprehensive understanding of distributions.

PO7. Innovation, employability, and Entrepreneurial skills

CO5. Fit quadratic and exponential curves to the bivariate data to investigate the relation between two variables (Weightage: 3)

Justification: Investigating relationships between variables using different curve types requires innovative thinking and contributes to employability and entrepreneurial skills.

**SYLLABUS (CBCS) FOR F. Y. B. Sc. (Computer Science) STATISTICS
(w.e. from June, 2019)
2019 pattern**

Academic Year 2019-2020

Class : F.Y. B. Sc. (Computer Science) (Semester- II)
Paper Code: CSST-1202
Paper : II
Title of Paper: Statistical Testing of Hypothesis and Use of R Software
Credit : 2 credits No. of lectures: 36

A) Learning Outcome:

The main outcome of this course is to use statistical software and testing of hypothesis.

B) course Outcomes:

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

- CO1.** Use testing of statistical hypothesis in real life.
- CO2.** Handle with statistical software.
- CO3.** Differentiate between parametric and non-parametric test.
- CO4.** Assess the plausibility of a hypothesis by using sample data.
- CO5.** Understand that decision is not to reject H_0 when H_0 is true (correct decision).
- CO6.** Assess whether a difference between two samples represents a real difference between the populations from which the samples were taken.
- CO7.** Handling everything from data manipulation and visualization to statistical analysis.

TOPICS/CONTENTS:

Unit-1 Introduction to R-software

(5L)

1.1 Introduction to R, features of R, getting help in R.

1.2 Vectors and vector arithmetic:

- (a) Creating of vector using functions `c`, `seq`, `rep`.
- (b) Arithmetic operations on vectors using operations `+`, `-`, `*`, `/`, `^`.
- (c) Numerical functions: `log`, `sort`, `max`, `min`, `unique`, `range`, `length`, `var`, `prod`, `sum`, `summary`, `fivenum`, etc.
- (d) Accessing vectors.

1.3 Data frames: Creation using `data.frame`, `subset` and `transform` commands.

1.4 `p`, `q`, `d`, `r` functions.

Unit-2 Concepts and definitions related to Testing of Hypothesis (6L)

- 2.1 Definitions: population, statistic, SRSWR, SRSWOR, random sample from a probability distribution, parameter, statistic, standard error of estimator, sampling distributions.
- 2.2 Concept of null hypothesis and alternative hypothesis, critical region, level of significance, type I and type II error, power, one sided and two sided tests, p-value.

Unit-3 Large Sample Tests (7L)

- 3.1 $H_0: \mu = \mu_0$ Vs $H_1: \mu \neq \mu_0, \mu < \mu_0, \mu > \mu_0$ (One sided and two sided tests)
- 3.2 $H_0: \mu_1 = \mu_2$ Vs $H_1: \mu_1 \neq \mu_2, \mu_1 < \mu_2, \mu_1 > \mu_2$ (One sided and two sided tests)
- 3.3 $H_0: P = P_0$ Vs $H_1: P \neq P_0, P < P_0, P > P_0$ (One sided and two sided tests)
- 3.4 $H_0: P_1 = P_2$ Vs $H_1: P_1 \neq P_2, P_1 < P_2, P_1 > P_2$ (One sided and two sided tests)
- 3.5 Numerical problems related to real life situations.

Unit-4 Test based on t distribution (7L)

- 4.1 $H_0: \mu = \mu_0$ Vs $H_1: \mu \neq \mu_0, \mu < \mu_0, \mu > \mu_0$ (One sided and two sided tests)
- 4.2 $H_0: \mu_1 = \mu_2$ Vs $H_1: \mu_1 \neq \mu_2, \mu_1 < \mu_2, \mu_1 > \mu_2$ (One sided and two sided tests)
- 4.3 Paired t-test.
- 4.4 Test of significance of correlation coefficient for bivariate raw data.
- 4.5 Test of significance of regression coefficients for bivariate raw data.
- 4.6 Numerical problems related to real life situations.

Unit-5 Test based on Chi-Square distribution (4L)

- 5.1 Chi square test for goodness of fit
- 5.2 Test for independence of attributes (m X n contingency table)
- 5.3 Test for significance of variation for a population.
- 5.4 Numerical problems related to real life situations.

Unit-6 Simulation (7L)

- 6.1 Introduction to Simulation, merits and demerits and pitfall.
- 6.2 Pseudo-random number generator ,requisites of a good random number generator, Testing these requirements by using various test of hypothesis using Run test, goodness of fit test, Sign test etc.
- 6.3 Model Sampling from uniform and exponential distribution.
- 6.4 Model sampling from Normal distribution using Box-Muller transformation.
- 6.5 Numerical problems related to real life situations.

References:

- 1) Statistical Methods (An Introductory Text), Medhi J., New Age International,(1992).
- 2) Statistical Methods, G.W. Snedecor, W.G. Cochran, John Wiley & sons, 1989.
- 3) Fundamentals of Applied Statistics (3rd Edition), Gupta and Kapoor, S.Chand and Sons, New Delhi, 1987.
- 4) Purohit, S.G.; Gore, S.D. and Deshmukh, S.R. (2015). Statistics using R, second edition. Narosa Publishing House, New Delhi.

- 5) Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye:
Probability & Statistics for Engineers & Scientists

Programme Outcomes and Course Outcomes Mapping :

Course Outcomes	Programme Outcomes (POs)						
	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	-	-	-	-	-	-	-
CO2	3	-	3	-	-	-	-
CO3	-	-	-	-	-	-	-
CO4	-	-	-	-	2	-	-
CO5	-	-	-	-	-	-	2
CO6	-	-	-	-	-	-	-
CO7	-	3	-	-	-	2	-

PO1. Computer Knowledge

CO2. Handle with statistical software (Weightage: 3)

Justification: Statistical software involves computer-based tools for data analysis. Handling statistical software requires computer knowledge to effectively utilize these tools for hypothesis testing and data manipulation.

PO2. Design / Development of solution:

CO7. Handling everything from data manipulation and visualization to statistical analysis (Weightage: 3)

Justification: Designing and developing solutions often involve data manipulation, visualization, and statistical analysis. This CO encompasses a comprehensive approach to handling various aspects of data-related problems.

PO3. Modern tool usage

CO2. Handle with statistical software (Weightage: 3)

Justification: Statistical software is a modern tool extensively used for data analysis, and proficiency in using such tools aligns with the modern tool usage outcome.

PO4. Environment and sustainability

No direct mapping found in the provided information.

PO5. Ethics

CO4. Assess the plausibility of a hypothesis by using sample data (Weightage: 2)

Justification: Ethical considerations are crucial when assessing hypotheses based on sample data. Ensuring the validity and reliability of data aligns with ethical principles.

PO6. Individual and Team work

CO7. Handling everything from data manipulation and visualization to statistical analysis (Weightage: 2)

Justification: Handling diverse aspects of data analysis may involve individual tasks as well as collaborative efforts. This CO addresses both individual and team-oriented skills.

PO7. Innovation, employability, and Entrepreneurial skills

CO6. Assess whether a difference between two samples represents a real difference between the populations from which the samples were taken (Weightage: 2)

Justification: The ability to assess differences between samples and populations contributes to innovation and employability skills, as it involves critical thinking and problem-solving.

**SYLLABUS (CBCS) FOR F. Y. B. Sc. (Computer Science) STATISTICS
(w.e. from June, 2019)
2019 pattern**

Academic Year 2019-2020

Class : F.Y. B. Sc. (Computer Science) (Semester- II)
Paper Code: CSST-1203
Paper : III Title of Paper: Practical-II
Credit : 2 credits No. of lectures: 40

Pre requisites: Knowledge of the topics in the theory papers.

A) Learning Objectives:

The main objective of this course is to acquaint students with concept of discrete bivariate random variable and its probability distribution.

B) Course Outcome:

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

- CO1.** Compute regression coefficients and to interpret the results.
- CO2.** Analyse the data with respect to bivariate discrete and continuous distributions.
- CO3.** Get helps organizations understand the underlying causes of trends or systemic patterns over time.
- CO4.** Describes the variability of the distances between sample means and the population mean for t distribution.
- CO5.** Used to compare observed results with expected results.
- CO6.** Use statistical tools like Fitting of linear regression model (non-linear regression, Model sampling with the help of R-Software.
- CO7.** Use statistical tools like correlation, F test, t test, χ^2 test, and Time Series with the help of MS-Excel.

INDEX

Sr.No.	Title of Experiments	Page No.	Remark
1.	Linear Regression		
2.	a) Multiple and Partial Correlation and Regression b) Multiple and Partial Correlation and Regression Using R		
3.	Time Series Analysis		
4.	Large Sample Tests		
5.	Test Based on t distribution		
6.	Test Based on Chi-Square distribution		
7.	(a) Fitting of Normal distribution (b) Fitting of Normal Distribution using R		
8.	Model Sampling from Uniform and Exponential distribution		
9.	Model Sampling from Exponential Distribution Normal Distribution (Box Muller Transformation etc.) using R		
10.	Project(1): Analysis of data collected in sem-I		

Programme Outcomes and Course Outcomes Mapping :

Course Outcomes	Programme Outcomes (POs)						
	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	-	3	-	-	-	-	-
CO2	-	-	-	-	-	2	-
CO3	-	-	-	-	-	-	3
CO4	-	-	-	-	-	-	-
CO5	-	-	-	-	-	-	-
CO6	3	-	3	-	-	-	-
CO7	-	-	-	-	-	2	-

PO1. Computer Knowledge

CO6. Use statistical tools like Fitting of linear regression model (non-linear regression, Model sampling with the help of R-Software) (Weightage: 3)

Justification: This CO involves using statistical tools, including R-Software, which requires computer knowledge. Proficiency in these tools aligns with the computer knowledge outcome.

PO2. Design / Development of solution:

CO1. Compute regression coefficients and to interpret the results (Weightage: 3)

Justification: Designing and developing solutions often involve analyzing data and interpreting results. Computing regression coefficients is a key step in data analysis and solution development.

PO3. Modern tool usage

CO6. Use statistical tools like Fitting of linear regression model (non-linear regression, Model sampling with the help of R-Software) (Weightage: 3)

Justification: The use of modern statistical tools, including R-Software, is directly aligned with the modern tool usage outcome.

PO4. Environment and sustainability

No direct mapping found in the provided information.

PO5. Ethics

No direct mapping found in the provided information.

PO6. Individual and Team work

CO2. Analyse the data with respect to bivariate discrete and continuous distributions (Weightage: 2)

Justification: Analyzing data involves both individual and team work, especially when dealing with bivariate distributions that may require collaborative efforts.

PO7. Innovation, employability, and Entrepreneurial skills

CO3. Get helps organizations understand the underlying causes of trends or systemic patterns over time (Weightage: 3)

Justification: Understanding underlying causes and patterns over time requires innovative thinking and contributes to employability and entrepreneurial skills.