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	UCSST121	Statistical Methods- II	2
	UCSST122	Continuous Probability Distributions and Testing of Hypothesis	2
	UCSST123	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 usinglatest technologies and develop and implement the solutions using various latest language.

PO3. Modern tool usage : Create, select and apply appropriate techniques, resourcesand 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 knowledgeand 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.(CS) STATISTICS
(2022 Pattern)
(w. e. from June, 2022)

Name of the Programme : B.Sc. Computer Science
Program Code : USCO
Class : F.Y.B.Sc. (Computer Science)
Semester : II
Course Name : Statistical Methods II
Course Code : UCSST121
No. of lectures : 36
Credit : 2 credits

Course Outcomes:

The students will acquire knowledge about the;

CO1. Relationship between two variable.(correlation)

CO2. Concept of regression (Simple and multiple)

CO3. Forecasting (predicting future values of the time series variable).

CO4. basic concept of R-Software.

CO5. Understand the basic concept of linear regression and fit linear regression model to the bivariate data, interpretation of coefficients, and prediction of outcomes.

CO6. Fit quadratic and exponential curves to the bivariate data to investigate relation between two variables.

CO7. Identify and distinguish between the various components of time series data, such as trend, seasonality, and residual components.

TOPICS/CONTENTS:

UNIT 1: Correlation (For ungrouped data) (07L)

1.1Concept of bivariate data, scatter diagram, its interpretation, concept of correlation, Positive correlation, negative correlation, zero correlation.

1.2 Karl Pearson's coefficient of correlation, properties of correlation coefficient, Interpretation of correlation coefficient, coefficient of determination with interpretation.

1.3 Spearman's rank correlation coefficient (formula with and without ties).

1.4Numerical problems

UNIT 2: Regression (for ungrouped data) (10L)

2.1Concept of linear and nonlinear regression.

2.2 Illustrations, appropriate situations for regression and correlation

2.3 Linear regression: Fitting of both lines of regression using least square method.

2.4 Concept of regression coefficients.

2.5 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 $b_{yx} = r\,(\sigma_y / \sigma_x)$.

2.6 Nonlinear regression models: Second degree curve, exponential curves of the type $Y=ab^x$ and $Y=ax^b$.

2.7 Numerical problems related to real life situations

UNIT3: Multiple Regression and Multiple, partial Correlation (For Trivariate Data) (8L)

3.1 Concept of multiple regressions, Yule's Notations.

3.2 Fitting of multiple regression planes.[Derivation of equation to the plane of regression of X_1on X_2 and X_3 is expected. Remaining two equations to be written analogously.]

3.3 Concept of partial regression coefficients, interpretations.

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

3.5 Concept of partial correlation. Definition of partial correlation coefficient and Its formula.

UNIT4: Time series (6L)

4.1 Meaning and utility

4.2 Components of time series

4.3 Additive and multiplicative models

4.4 Methods of estimating trend, moving average method, least squares method and exponential smoothing method(with graph and interpretation).

4.5 Numerical problems related to real life situations

Unit 5 Fundamentals of R-Software (5L)

5.1 Introduction to R, features of R, starting and ending R session, getting help in R, R commands and case sensitivity.

5.2 Vectors and vector arithmetic

a) Creation of vectors using functions c, seq, rep

b) Arithmetic operations on vectors using operators +, - , * , / , ^.

c) Numerical functions: log10, log, sort, max, min, unique, range, length, var, prod, sum, summary, fivenum etc.

d) Accessing vectors

5.3 Data frames: creation using data. Frame, subset and transform commands, Resident data sets: Accession and summary,p, q, d, r functions.

References:

1 Introduction to Linear Regression Analysis, Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, Wiley

2 Time Series Methods, Brockwell and Davis, Springer, 2006.

3 Time Series Analysis, 4th Edition, Box and Jenkin, Wiley, 2008.

4 Fundamentals of Applied Statistics(3rd Edition), Gupta and Kapoor, S. Chand and Sons,
New Delhi, 1987.

5 Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). The World Press Pvt. Ltd., Calcutta

6 Statistical Computing Using R- Software, Vishwas R. Pawgi, Third Edition

Programme Outcomes and Course Outcomes Mapping :

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

PO1. Computer Knowledge:

CO.1 Represent statistical data diagrammatically and graphically. (Weightage: 3)

Justification: The ability to represent statistical data graphically requires a strong understanding of both statistical concepts and computer fundamentals. Utilizing tools like MS-Excel involves knowledge of both mathematics and computer applications.

CO.5 To use statistical tools like graphical representation, summary statistics of data with the help of MS-Excel. (Weightage: 3)

Justification: This directly aligns with applying statistical knowledge in the context of IT applications, specifically using Microsoft Excel as a tool to analyze and represent data.

PO2. Design/Development of solution:.

CO.2 Compute various measures of central tendency and dispersion. (Weightage: 2)

Justification: Calculating central tendency and dispersion is crucial in designing solutions for IT applications, as it provides insights into the distribution and characteristics of data, aiding in informed decision-making.

PO3. Modern tool usage:.

CO.4 Interpret summary statistics of computer output. (Weightage: 3)

Justification: Interpretation of summary statistics from computer output is directly related to the modern tool usage aspect. Understanding the output is essential in making informed decisions about the IT application.

CO.6 Examine spreadsheet concepts and explore the Microsoft Office Excel environment. (Weightage: 3)

Justification: Knowledge of spreadsheet concepts and tools like Microsoft Excel is fundamental in modern tool usage for analysis, modeling, and decision-making in complex IT applications.

PO4. Environment and sustainability:

CO.7 Gather information in a measured and systematic manner to ensure accuracy and facilitate data analysis. (Weightage: 2)

SYLLABUS (CBCS) FOR F. Y. B. Sc.(CS) STATISTICS
(2022 Pattern)
(w. e. from June, 2022)

Name of the Programme : B.Sc. Computer Science
Program Code : USCO
Class : F.Y.B. Sc. (Computer Science)
Semester : II
Course Name : Continuous Probability Distributions and Testing of Hypothesis
Course Code : UCSST122
No. of lectures : 36
Credit : 2 credits

Course Outcomes:

Students are expected to be able to,

CO1. Understand standard continuous probability distributions and their application.

CO2. Apply the idea of estimation and Tests of Hypothesis

CO3. Apply testing of hypothesis in real life situations

CO4. Use R-Software for computation of Probability and testing of Hypothesis.

CO5. Assess the plausibility of a hypothesis by using sample data.

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: Standard Continuous Probability Distributions (8L)

1.1 Uniform Distribution: statement of p.d.f., mean, variance, nature of probability curve. Theorem (without proof): The distribution function of any continuous r.v. if it is Invertible follows U(0, 1) distribution

1.2 Exponential Distribution: statement of p.d.f. of the form $(x) = (1/\theta)\, e(-x/\theta)$, mean, Variance, nature of probability curve, lack of memory property.(with proof)

1.3 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.

1.4 Numerical problems related to real life situations.

UNIT 2: Statistical Inference: The idea of estimation and Tests of Hypothesis (4L)

2.1 Concepts of population and sample.

2.2 Definitions: random sample from a probability distribution, parameter, statistic, Standard error of estimator.

2.3 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: Parametric Tests (18L)

3.1 Large Sample Tests

3.1.1 Ho: $\mu = \mu o$ Vs H_1: $\mu \neq \mu o$, $\mu < \mu o$, $\mu > \mu o$ (One sided and two sided tests)

3.1.2 Ho: $\mu_1 = \mu_2$ Vs H_1: $\mu_1 \neq \mu_2$, $\mu_1 < \mu_2$, $\mu_1 > \mu_2$ (One sided and two sided tests)

3.1.3 Ho: $P = Po$ Vs H_1: $P \neq Po$, $P < Po$, $P > Po$ (One sided and two sided tests)

3.1.4 Ho: $P_1 = P_2$ Vs H_1: $P_1 \neq P_2$, $P_1 < P_2$, $P_1 > P_2$ (One sided and two sided tests)

3.1.5 Numerical problems related to real life situations.

3.2 Test based on F- distribution

3.2.1 F-test for testing significance of equality of two population variances.

3.3 Tests based on t – distribution

3.3.1 Ho: $\mu = \mu o$ Vs H_1: $\mu \neq \mu o$, $\mu < \mu o$, $\mu > \mu o$ (One sided and two sided tests)

3.3.2 Ho: $\mu_1 = \mu_2$ Vs H_1: $\mu_1 \neq \mu_2$, $\mu_1 < \mu_2$, $\mu_1 > \mu_2$ (One sided and two sided tests)

3.3.3 Paired t-test.

3.4 Tests based on Chi square distribution

3.4.1 Chi-square test for goodness of fit

3.4.2 Test for independence of attributes (mxn and 2x2)

3.5 Numerical problems related to real life situations.

UNIT 4: Use of R-Software in statistical inference (6L)

4.1 Probability computation using R.

4 .2 Model sampling from standard Continuous distributions

4.3 Tests of hypothesis using R.

References

1. A First course in Probability, Sheldon Ross. Pearson Education Inc.
2. Statistical Methods (An Introductory Text), Medhi J. 1992, New Age International.
3. Modern Elementary Statistics, Freund J.E. 2005, Pearson Publication.
4.Probability, Statistics, Design of Experiments and Queuing Theory with Applications of Computer Science, Trivedi K.S. 2001, Prentice Hall of India, New Delhi.

5.Gupta S. C. and Kapoor V. K.1987 Fundamentals of Mathematical Statistics(3rd Edition) S.Chand and Sons, New Delhi.
6. Mukhopadhyay P. 2015, Mathematical Statistics (3rd Edition), Books And Allied (P), Ltd.
7. Programmed Statistics, B.L. Agarwal, New Age International Publishers.
8. Common Statistical Tests Kulkarni M.B., Ghatpande, S.B.,Gore S.D.1999 Satyajeet Prakashan,
9. Statistical Computing Using R- Software., Vishwas R. Pawgi, Third Edition

Programme Outcomes and Course Outcomes Mapping :

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

PO1. Computer Knowledge

CO4. Use R-Software for computation of Probability and testing of Hypothesis (Weightage: 3)

Justification: Utilizing R-Software for probability computation and hypothesis testing requires computer knowledge, making it strongly related to this PO.

PO2. Design / Development of solution:

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

Justification: Handling various aspects of data manipulation, visualization, and statistical analysis contributes to the design and development of solutions.

PO3. Modern tool usage

CO4. Use R-Software for computation of Probability and testing of Hypothesis (Weightage: 3)

Justification: Using R-Software aligns with modern tool usage for statistical analysis and hypothesis testing.

PO4. Environment and sustainability

No direct mapping found in the provided information.

PO5. Ethics

CO5. Assess the plausibility of a hypothesis by using sample data (Weightage: 2)

Justification: Assessing the plausibility of a hypothesis involves ethical considerations related to the validity and reliability of sample data.

PO6. Individual and Team work

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

Justification: Handling various tasks from data manipulation to statistical analysis may involve both individual and team work.

PO7. Innovation, employability, and Entrepreneurial skills

CO2. Apply the idea of estimation and Tests of Hypothesis (Weightage: 3)

Justification: Applying estimation and hypothesis testing involves innovative thinking and contributes to employability and entrepreneurial skills.

.

SYLLABUS (CBCS) FOR F. Y. B. Sc.(CS) STATISTICS
(2022 Pattern)

(w. e. from June, 2022)

Name of the Programme : B.Sc. Computer Science
Program Code : USCO
Class : F.Y.B. Sc. (Computer Science)
Semester : II
Course Name : Practical – II
Course Code : UCSST123
No. of lectures : 36
Credit : 2 credits

Course Outcomes:

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

CO1. Use statistical tools like Fitting of linear regression model (non-linear regression, Model sampling with the help of R-Software.

CO2. Use statistical tools like correlation, F test, t test, χ^2 test, and Time Series with the help of MS-Excel.

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. Compare observed results with expected results.

CO6. Compute regression coefficients and to interpret the results.

CO7. Analyse the data with respect to bivariate discrete and continuous distributions.

Sr. No.	Title of the Practical
1	Linear correlation and regression (use of scatter plot for explaining the linear Relationship between two variables) using Excel.
2	Fitting of non-linear regression. (use of scatter plot for explaining the nonlinear relationship between two variables) using R-Software.
3	Fitting of normal distribution and computation of expected frequencies.
4	Fitting of the linear regression model (Simple and Multiple) and non-linear Regression models and finding the best fit by using R-Software..
5	Model sampling from continuous uniform, exponential and normal Distributions using R-Software.
6	Large sample tests.
7	F test, t test, χ^2 test using EXCEL (one problem each with equal and unequal variance)(χ^2 test – for the goodness of fit-use fitted problems of Binomial, Poisson and Normal distribution in previous practical problems)
8	Time Series- Estimation of trend by using the method of moving averages using Excel.
9	Write a report on the application of some statistical techniques in the field of computers. (Individual activity)
10	Project (Part-II) - Analysis of data collected in semester - I

Programme Outcomes and Course Outcomes Mapping :

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

PO1. Computer Knowledge

CO1. Use statistical tools like Fitting of linear regression model (non-linear regression, Model sampling with the help of R-Software) (Weightage: 3)

Justification: Using statistical tools like R-Software for regression models requires computer knowledge, making it strongly related to this PO.

PO2. Design / Development of solution:

CO6. Compute regression coefficients and to interpret the results (Weightage: 3)

Justification: Computing regression coefficients is a key step in the design and development of solutions, contributing to the understanding of relationships in data.

PO3. Modern tool usage

CO1. Use statistical tools like Fitting of linear regression model (non-linear regression, Model sampling with the help of R-Software) (Weightage: 3)

Justification: Utilizing modern statistical tools like R-Software aligns with modern tool usage for data analysis and modeling.

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

Justification: Analyzing data with respect to bivariate distributions may involve both individual and team efforts, contributing to individual and team work.

PO7. Innovation, employability, and Entrepreneurial skills

CO2. Use statistical tools like correlation, F test, t test, $\chi 2$ test, and Time Series with the help of MS-Excel (Weightage: 3)

Justification: Using statistical tools like MS-Excel for various tests involves innovative thinking and contributes to employability and entrepreneurial skills.