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
I	CSST-1101	Statistical Methods -I	2
	CSST-1102	Probability and some discrete probability distributions	2
	CSST-1103	Practical-I	2
II	CSST-1201	Statistical Methods- II	2
	CSST-1202	Statistical Testing of Hypothesis and Use of R Software	2
	CSST-1203	Practical-II	2

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

Academic Year 2019-2020

Class : F.Y. B. Sc. (Computer Science) (Semester- I)
Paper Code: CSST-1101
Paper : I Title of Paper: Statistical Methods I
Credit : 2 credits No. of lectures: 36

A) Learning Objectives:

1. To compute various measures of central tendency, dispersion, skewness and kurtosis.
2. Compute the correlation coefficient for bivariate data and interpret it.

B) Learning Outcome:

The main outcome of this course is to acquaint students with initial description of the data as part of a more extensive statistical analysis by using some elementary statistical methods.

TOPICS/CONTENTS:

UNIT1: Data Representation (6L)

1.1 Definition, importance, scope and limitations of statistics w.r.to computer science.

1.2 Data Condensation: Types of data (Primary and secondary), Attributes and variables, discrete and Continuous variables, classification and construction of frequency distribution.

1.3 Graphical Representation: Histogram, Frequency polygon, Frequency curve, Ogive Curves, Steam and leaf chart.

1.4 Numerical problems related to real life situations.

UNIT2: Measures of central tendency (8L)

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 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)

2.4 Partition Values: Quartiles, Percentiles, Deciles, Box Plot.

2.5 Numerical problems related to real life situations.

UNIT3: Measures of Dispersion (6L)

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, combined variance & standard deviation, merits and demerits.

3.4 Numerical problems related to real life situations.

UNIT4: Moments, Skewness and Kurtosis (8L)

4.1 Raw and central moments: definition for ungrouped and grouped data (only first four moments), relation between central and raw moments upto fourth order. (without proof)

4.2 Measures of Skewness: Types of skewness, Pearson's and Bowley's coefficient of skewness, Measures of skewness based on moments.,

4.3 Measures of Kurtosis: Types of kurtosis, Measures of kurtosis based on moments

4.4 Numerical problems related to real life situations

UNIT 5: Correlation (For bivariate raw data) (8L)

5.1 Concept of bivariate data, scatter diagram, concept of correlation, positive correlation, negative correlation, zero correlation.

5.2 Karl Pearson's coefficient of correlation, properties of correlation coefficient, interpretation of correlation coefficient.

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

5.4 Numerical problems

References:

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

2 An Introductory Statistics, Kennedy and Gentle.

3 Statistical Methods, G.W. Snedecor, W.G. Cochran, John Wiley & sons, 1989.

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

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

Academic Year 2019-2020

Class : F.Y. B. Sc. (Computer Science) (Semester- I)
Paper Code: CSST-1102
Paper : II Title of Paper: Probability and some discrete probability distributions

Credit : 2 credits No. of lectures: 36

A) Learning Objectives:
The main objective of this course is to acquaint students with some basic concepts of probability, axiomatic theory of probability, concept of random variable, probability distribution (univariate) and some standard continuous distributions.

B) Learning Outcome:
Students are expected to be able,
1) To distinguish between random and non-random experiments.
2) To find the probabilities of various events.
3) To obtain probability distribution of univariate continuous random variables.
4) To use distributions in real life situations.

TOPICS/CONTENTS:

Unit-1. Sample space and Events: (6L)

1.1 Concepts of experiments, deterministic and nondeterministic experiments.

1.2 Definitions: Sample space, Types of sample space, Event, Types of Events: Elementary event, Complementary event, Sure event, Impossible event.

1.3 Concept of occurrence of an event, Equally-likely events

1.4 Algebra of events (Union, Intersection, Complementation).

1.5 Definitions of Mutually exclusive events, Exhaustive events.

1.6 Algebra of events and its representation of events in set theory notation:

Occurrence of the following events:

a) at least one of the given events

b) none of the given events

c) all of the given events

d) mutually exclusive events

e) mutually exhaustive events

f) exactly one event out of the given events.

1.7 Numerical problems related to real life situations.

Unit-2. Probability: (13L)

2.1 Concept of Multiplication principle and Permutation and Combination

2.2 Classical Probability: Classical definition of probability, examples ,Probability model, probability of an event, examples. Axiomatic definition of probability.

Proof of the results:

i) $0 \leq P(A) \leq 1$,

ii) $P(A)+P(A^c)=1$

iii) $P(\Phi) = 0$

iv) If $A \subset B$, $P(A) \leq P(B)$

v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Addition theorem of probability) and its generalization (Statement only).

Numerical problems related to real life situations.

2.3 Conditional Probability

- Concepts and definitions of Conditional Probability
- Definition of conditional probability of an event.
- Multiplication theorem for two events. Examples.
- Partition of sample space.
- Idea of Posterior probability, Statement and proof of Bayes' theorem, examples on Baye's theorem.

2.4 Independence of Events

- Concept of Independence of two events.
- Proof of the result that if A and B are independent then, i) A and B^c, ii) A^c and B iii) A^c and B^c are independent.
- Pair wise and Mutual Independence for three events.
- Numerical problems related to real life situations.

UNIT 3: Discrete random Variable (4L)

3.1 Definition of random variable and discrete random variable

3.2 Definition of probability mass function, distribution function and its properties

3.3 Definition of expectation and variance, theorems on expectation

3.4 Numerical problems related to real life situations

UNIT 4: Standard Discrete distribution **(13L)**

4.1 Uniform Distribution : definition, mean, variance

4.2 Bernoulli Distribution : definition, mean, variance, additive property

4.3 Binomial Distribution : definition, mean, variance, additive property

4.4 Poisson Distribution : definition, mean, variance, mode, additive property, limiting case of B(n, p)

4.5 Numerical problems related to real life situations

References:

1. Bhat B. R., Srivenkatramana T and Madhava Rao K. S. (1997): Statistics: a Beginner's Text, Vol. II, New Age International (P) Ltd.
2. Goon A. M., Gupta M. K., Das Gupta B. (1999): Fundamentals of Statistics, Vol.II, World Press, Calcutta.
3. Hogg R. V. and Crag R. G.: Introduction to Mathematical Statistics Ed.4.
4. Gupta and Kapoor : Fundamentals of Mathematical Statistics, Sultan Chand and Sons, New Delhi.
5. Meyer P.L.(1970): Introductory Probability and Statistical Applications, Edition Wesley.
6. Rohatgi V. K. and Saleh A. K. Md. E. (2002): An Introduction to probability and statistics. John wiley & Sons (Asia)
7. Gupta and Kapoor : Fundamentals of Applied Statistics, Sultan Chand and Sons, New Delhi.

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

Academic Year 2019-2020

Class : F.Y. B. Sc. (Computer Science) (Semester- I)
Paper Code: CSST-1103
Paper : III Title of Paper: Practical-I
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 computation ability with interpreting summary Statistics

B) Learning Outcome:

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

i) Represent statistical data diagrammatically and graphically.

ii) Compute various measures of central tendency, dispersion, moments, skewness and kurtosis.

iii) Compute correlation coefficient, regression coefficients and to interpret the results.

iv) Interpret summary Statistics of computer output.

INDEX

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

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) Learning Outcome:

Students are expected to be able,

1) To apply discrete probability distributions studied in this course in different situations.

2) Know some standard discrete probability distributions with real life situations.

3) How to fit the regression model to the given bivariate data

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.

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

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) Learning Objectives:
Students are expected to be able
1. To testing of statistical hypothesis in real life.
2. To handle with statistical software.
3. To differentiate between parametric and non-parametric test.

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 Ho: $\mu = \mu_o$ Vs H1: $\mu \neq \mu_o$, $\mu < \mu_o$, $\mu > \mu_o$ (One sided and two sided tests)
3.2 Ho: $\mu_1 = \mu_2$ Vs H1: $\mu_1 \neq \mu_2$, $\mu_1 < \mu_2$, $\mu_1 > \mu_2$ (One sided and two sided tests)
3.3 Ho: $P = P_o$ Vs H1: $P \neq P_o$, $P < P_o$, $P > P_o$ (One sided and two sided tests)
3.4 Ho: $P_1 = P_2$ Vs H1: $P_1 \neq P_2$, $P_1 < P_2$, $P_1 > P_2$ (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 Ho: $\mu = \mu o$ Vs H1: $\mu \neq \mu o$, $\mu < \mu o$, $\mu > \mu o$ (One sided and two sided tests)

4.2 Ho: $\mu 1 = \mu 2$ Vs H1: $\mu 1 \neq \mu 2$, $\mu 1 < \mu 2$, $\mu 1 > \mu 2$ (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

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

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) Learning Outcome:

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

i) Compute regression coefficients and to interpret the results.

ii) Analyze the data with respect to Bivariate discrete and continuous distributions .

INDEX