23STMIL121: Descriptive Statistics for Data Science

Offered to: All UG Honours Programs   **Course Type:** Minor 1 (Core -TH)

Year of Introduction: 2023-24   **Year of offering: 2023 - 2024**

Semester: II   60 Hrs   **Credits: 3**

Course Prerequisites : Basic knowledge in Mathematics.

Course Description:
This course helps the students to familiarize with the ways in which we talk about descriptive statistics, uncertainty and estimate their situations in which probability arises. Also this course aims at providing basic knowledge about theoretical and applications of attributes.

Course Objectives:
1) To compute various measures of central tendency, dispersion, skewness and kurtosis.
2) To find the probabilities of events.
3) To get the knowledge regarding qualitative factors

Course Outcomes:		
Course Outcome	Upon successful completion of this course, students should have the knowledge and skills to:	**Program Outcomes Mapping**
CO 1	To learn different types measures of central tendency and dispersion.	PO1
CO 2	To understand the relationship between raw moments and central moments and also concepts of skewness and kurtosis.	PO2
CO3	Understand the basic concepts of probability and to find probabilities of various events.	PO6
CO 4	Get the knowledge in respect of usage in day-to-day life in decision making in the face of uncertainty	PO6
CO 5	Analyze the qualitative data	PO6

CO-PO MATRIX

COURSE CODE	CO-PO	PO1	PO2	PO3	PO4	PO5	PO6
23STMIL121	**CO1**	3					
	CO2		3				
	CO3						3
	CO4						3
	CO5						3

Syllabus

Unit	Learning Units	Lecture Hours
I	**Data** – Types of data, truncated data, censored data. **Mean** – Combined mean and corrected mean. **Dispersion** – Mean deviation, Standard deviation, Combined Standard deviation and Coefficient of variation – Problems.	12
II	**Moments:** Central and non-central moments and their inter-relationships, Sheppard's corrections for moments for grouped data and problems. **Skewness:** Definition, measures of skewness by Karl Pearson's, Bowley's formulae and based on moments and problems. **Kurtosis:** Definition, measures of kurtosis based on moments and problems.	12
III	**Attributes** Attributes Notations, Class, Order of class frequencies, Ultimate class frequencies, Consistency of data, Conditions for consistency of data for 2 and 3 attributes only , Independence of attributes, Association of attributes and its measures, Relationship between association and colligation of attributes, Contingency table: Square contingency, Mean square contingency, Coefficient of mean square contingency, Tschuprow's coefficient of contingency – Problems	12
IV	**Probability – I** Terminology - Random experiments, trial, sample space, mutually exclusive , exhaustive , equally likely, favorable, conditional and independent events. Definitions-Mathematical, Statistical and Axiomatic definitions of probabilities. Addition law of probabilities for two, three and n events. Boole's inequalities and problems.	12
V	**Probability – II** Conditional Probability- multiplication law of probability for two and n events. Pairwise independent events and conditions for mutual independence of n events and Baye's theorem and its applications. Introductory concept of Geometric probability.	12

Text Book:

Fundamentals of Mathematical Statistics, 11th Edition, 2010, S. C. Gupta and V. K. Kapoor, Sultan Chand & Sons, New Delhi

Reference Books:

1. B.A/B.Sc. Second Year Statistics(2010) , Telugu Akademi, Hyderabad.
2. Mathematical Statistics with Applications, 2009, K.M.Ramachandran and Chris P.Tsokos Academic Press(Elsevier), Haryana .
3. Probability and Statistics, Volume I & II, D. Biswas, New central book Agency (P) Ltd, NewDelhi.
4. An outline of Statistical theory, Volume II,3rd Edition,2010(with corrections) A.M.Goon,M.K. Gupta, B.Dasgupta ,The World Press Pvt.Ltd., Kolakota.

    Sanjay Arora and BansiLal:. New Mathematical Statistics, SatyaPrakashan , New Delhi.

23STMIL121: **Descriptive Statistics for Data Science**   **Course Type:** Minor -1 (Th)
Semester II
Time: 3 hours   Maximum Marks: 70

Section – A

Answer the following Questions   **5 x 4M = 20Marks**

1. (a) Explain the concept of censored data.   (L-2, CO-1)

                              (OR)

    (b) Two groups of students revealed the following results in the semester end examinations
        as follows.   (L-2, CO-1)

Groups	Number of students	Mean
A	25	73.2
B	28	71.8

    Calculate the combined mean for two groups.

2. (a) Show that for discrete distributions $\beta_1 > 1$.   (L-2, CO-2)

                              (OR)

    (b) The first four moments of a distribution about the value 5 are -4, 22,-117 and 560. Find
        the corresponding moments about the mean and also comment on the nature of the data.

                                                                      (L-2, CO-2)

3. (a) State and prove addition theorem of probability for two events.   (L-1, CO-4)

                              (OR)

    (b) Define axiomatic definition of probability.   (L-1, CO-4)

4. (a) If A and B are independent events, then prove that   (L-1, CO-5)

        (i) $\bar{A}$ and B   (ii) $\bar{A}$ and $\bar{B}$ are also independent.

                              (OR)

    (b) State and prove multiplication theorem of probability.   (L-1, CO-5)

5. (a) Write the conditions for consistency of data for 3 attributes.   (L-1, CO-3)

                              (OR)

    (b) Define yule's coefficient of association and write its properties.   (L-1, CO-3)

Section – B

Answer the following Questions   **5 x 10M = 50Marks**

6. (a) Define data. Explain the types of collection of data.   (L-2, CO-1)

(OR)

(b) For a group of 200 candidates, the mean and standard deviation of scores were found to be 40 and 15 respectively. Later on it was discovered that the scores 43 and 35 were misread as 34 and 53 respectively. Find the corrected mean and standard deviation corresponding to the correct figures.   (L-2, CO-1)

7. (a) Define moments. Establish the relationship between the moments about mean (Central moments) in terms of moments about any arbitrary point and vice versa.   (L-2, CO-2)

(OR)

(b) The scores in statistics of 250 candidates appearing at an examination have mean = 39.72, variance = 97.80, 3rd central moment and 4th central moments are -114.18 and 28,396.14. It was later found on scrutiny that the score 61 of a candidate has been wrongly recorded as 51. Make necessary corrections in the given values of the mean and central moments.   (L-2, CO-2)

8. (a) State and prove Boole's inequality.   (L-1, CO-4)

(OR)

(b) State and prove the addition theorem of probability for n events.   (L-1, CO-4)

9. (a) For two events A and B, prove that   (L-1, CO-5)

     (i) $P(\bar{A} \cap B) = P(B) - P(A \cap B)$ (ii) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

     (iii) If $B \subset A$ then $P(A \cap \bar{B}) = P(A) - P(B)$ (iv) If $A \subset B$ then $P(\bar{A} \cap B) = P(B) - P(A)$

(OR)

(b) State and prove Baye's theorem of probability.   (L-1, CO-5)

10. (a) Derive the relation between yule's coefficient of association and colligation.   (L-3, CO-3)

(OR)

(b) Out of 200 students that appeared for M.B.A examination , 80 were married . Among 60 students who failed , 24 were married . Calculate the coefficient of association between marriage and failure in the examination.   (L-3, CO-3)

***