

**GUJARAT TECHNOLOGICAL UNIVERSITY****BE - SEMESTER-V (NEW) EXAMINATION – WINTER 2022****Subject Code:3154301****Date:09-01-2023****Subject Name:Statistics and Exploratory Data Analysis****Time:10:30 AM TO 01:00 PM****Total Marks:70****Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Simple and non-programmable scientific calculators are allowed.

**MARKS**

- Q.1** (a) If  $p(A) = 0.3$ ,  $p(B) = 0.2$  and  $p(A \cap B) = 0.1$  then determine (i)  $p(A \cup B)$  (ii)  $p[(A \cup B)']$  **03**
- (b) Explain different type of data measurement scales. **04**
- (c) State Baye's Theorem. An insurance company insured 2000 scooter drivers, 4000 car drivers, and 6000 truck drivers. The probability of an accident is 0.01, 0.03 and 0.15 in the respective category. One of the insured drivers meets with an accident. What is the probability that he is a scooter driver. **07**

- Q.2** (a) What is the need of Skewness and Kurtosis. **03**
- (b) The number of message sent per hour over a computer network has the following distribution **04**

x = No. of message	10	11	12	13	14	15
f(x)	0.08	0.30	0.15	0.20	0.07	0.20

Determine the mean and S.D of the number of message sent per hour.

- (c) The female student in an undergraduate engineering core course reported their heights to the nearest inch. The data are: 62,64,66,67,65,68,61,65,67,65,64,63,67,68,64,66, 68,69,65,67,62,66,68,67,66,65,69,65,70,65,67,68,65, 63,64,67,67. Then Calculate : **07**
- (i) Sample mean and variance of height.
  - (ii) Construct a stem and leaf diagram for the height data
  - (iii) what is the median height of this group of female engineering student

**OR**

- (c) Determine  $Q_1$ ,  $Q_3$ ,  $Q_5$  and inter quartile range for the following data **07**

Marks	10-15	15-20	20-25	25-30	30-35	35-40
Number of student	20	15	31	22	10	2

- Q.3** (a) Write properties of normal distribution. **03**
- (b) From the following data, obtain the two regression line **04**

Sales(x)	100	98	78	85	110	93	80
Purchase(y)	85	90	70	72	95	81	74

- (c) A manufacturer, who produce medicine bottles, find that 0.1% of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug manufacturer buys 100 boxes from the producer of bottles. Using poisson distribution, find how many boxes will contain (i) no defective bottles and (ii) at least 2 defective bottles **07**

**OR**

- Q.3** (a) Explain multivariate categorical variable. **03**

- (b) Using method of least squares, find the best fitting straight line to the given data **04**

x	1	2	3	4	5
y	1	3	5	6	5

- (c) A factory produces 1000 bulbs in a day. The bulb have an average life of 2000 burning hours with standard deviation of 200 hours. What number of bulbs. assuming the hours are distribution normally, are expected to burn (i) more than 2400 hours (ii) between 1900 and 2300 hours. **07**  
(Consider  $p(x < 2) = 0.9972$ ,  $p(0 < z < 1.5) = 0.4332$  and  $p(0 < z < 0.5) = 0.1915$ )

- Q.4** (a) Compare bar graph, box- plot and histogram with respect to their applicability in data visualization. **03**

- (b) What do you mean by Exploratory data Analysis? How t-test is useful for EDA **04**

- (c) Explain Probability Density Function (PDF) and Cumulative Distribution Function(CDF) of a continuous random variable with suitable example. **07**

**OR**

- Q.4** (a) Define Categorical data in EDA. **03**

- (b) Define covariance and what is the importance of covariance in data analysis? Explain it with example **04**

- (c) Which are the basic activities we performed as a part of data science pipelines? Summarize and explain in brief. **07**

- Q.5** (a) How do you calculate maximum likelihood estimation. **03**

- (b) Compare linear regression and multiple regression. **04**

- (c) How is chi square test is different from student t- distribution explain with suitable examples. **07**

**OR**

- Q.5** (a) Explain standard error in Hypothesis testing. **03**

- (b) A tyre company claim the lives of tyres have mean 43500km with standard deviation of 3500km. A change in the production process is believed to result in better product. A test sample of 100 new tyres has a mean life of 44500km. Test that the new product significantly better than the old one. **04**

(c) The following mistakes per pages were observed in a book:

07

Number of mistakes per page	0	1	2	3	4
Number. of Pages	211	90	19	5	0

Fit a Poisson distribution and test the goodness of fit

\*\*\*\*\*