P P SAVANI UNIVERSITY

Fourth Semester of B. Tech. Examination

May 2022

SESH2080 Statistics for Machine Learning

18.05.2022, Wednesday **Time: 09:00 a.m. To 11:30 a.m.** **Maximum Marks: 60**

Instructions:

1. The question paper comprises of two sections.
2. Section I and II must be attempted in same answer sheet.
3. Make suitable assumptions and draw neat figures wherever required.
4. Use of scientific calculator is allowed.

SECTION – I

Q - 1 Answer the Following Questions. **(Attempt any two)** **[10]**

(i) State whether each of the following variables is categorical or quantitative and indicate its measurement scale.

(a) Annual sales.
(b) Soft drink size (small, medium, large).
(c) Employee classification (GS1 through GS18).
(d) Earnings per share.
(e) Method of payment (cash, check, credit card).

(ii) Consider the following data.

8.9	10.2	11.5	7.8	10.0	12.2	13.5	14.1	10.0	12.2
6..8	9.5	11.5	11.2	14.9	7.5	10.0	6.0	15.8	11.5

(a) Construct a dot plot.
(b) Construct a frequency distribution.
(c) Construct a percent frequency distribution.

(iii) Consider a sample with the given data values: 89,78,96,97,63,52,41,47,58,69,97, 97. Compute following for given data.

(a) Sample variance.
(b) Sample standard deviation.
(c) Coefficient of variation.

Q - 2 Answer the Following Questions. **(Attempt any two)** **[10]**

(i) Draw Box plot with "Five number summary" for the following data.
8408, 1374, 1872, 8879, 2459, 11413, 608, 14138, 6452, 1850, 2818, 1356, 10498, 7478, 4019, 4341, 739, 2127, 3653, 5794, 8305.

(ii) Consider a sample with a mean of 500 and a standard deviation of 100. What is the z-scores for the following data values: 520, 650, 500, 450, and 280?

(iii) Consider a sample with a mean of 30 and a standard deviation of 5. Use Chebyshev's theorem to determine the percentage of the data within each of the following ranges:

(a) 20 to 40
(b) 15 to 45
(c) 22 to 38
(d) 18 to 42
(e) 12 to 48

Q - 3 Answer the Following Questions. **(Attempt any two)** **[10]**

(i) Calculate the correlation coefficient between the demand and the corresponding price of a commodity for the following data.

Demand in Quintals	65	66	67	67	68	69	70	72
Price in rupees per kg	67	68	65	68	72	72	69	71

(ii) Explain the types of correlation with examples.

(iii) Compute the mean, median and mode and Q_1, Q_2, Q_3 for the following data:
53,55,70,58,64,57,53,69,57,68,53.

SECTION-II

Q - 1 A husband and wife appeared in an interview for two vacancies in an office. The probability of the husband's selection is $\frac{1}{7}$ and that of the wife's selection is $\frac{1}{5}$. Find the probability that **[05]**

(a) Both of them are selected.
(b) Only one of them is selected.
(c) None of them is selected.
(d) At least one of them is selected.

OR

Q - 1 A bag A contains 2 white and 3 red balls, and a bag B contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that the red ball is drawn from the bag B. **[05]**

Q - 2 The mean and variance of a binomial distribution are 4 and $\frac{4}{3}$ respectively. Find $P(X \geq 1)$. **[05]**

Q - 3 If 2% of lightbulbs are defective, find the probability that; **[05]**

(a) At least one is defective.
(b) Exactly 7 are defective. Also, find $P(1 < X < 8)$ in a sample of 100.

Q - 4 Write the Ten characteristics of the normal distribution. **[05]**

OR

Q - 4 The random variable x is known to be uniformly distributed between 10 and 20. **[05]**

(a) Show the graph of the probability density function.
(b) Compute $P(x < 15)$.
(c) Compute $P(12 \leq x \leq 18)$.
(d) Compute $E(x)$.
(e) Compute $Var(x)$.

Q - 5 One of the questions on the Business Week Subscriber Study was, "In the past 12 months, when traveling for business, what type of airline ticket did you purchase most often?" The data obtained are shown in the following contingency table. [Use $\chi^2_{0.05} = 5.99$] **[05]**

Type of Ticket	Type of Flight	
	Domestic Flights	International Flights
First class	29	22
Business/executive class	95	121
Full fare economy/coach class	518	135

Q - 6 An ambulance service company claims that on an average it takes 20 minutes between a call for an ambulance and the patient's arrival at the hospital. If in 6 calls the time taken (between a call and arrival at hospital) are 27,18,26,15,20,32. Can the company's claim be accepted? [Use $t_{0.1} = 1.476$]. **[05]**

****