P P SAVANI UNIVERSITY

Fourth Semester of B. Tech. Examination

November 2022

SESH2080 Statistics for Machine Learning

21.11.2022, Monday **Time: 01:00 p.m. To 03:30 p.m.** **Maximum Marks: 60**

Instructions:

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

SECTION – I

			CO	BTL
Q - 1	The Wall Street Journal (WSJ) subscriber survey (October 13, 2003) asked 46 questions about subscriber characteristics and interests. State whether each of the following questions provided categorical or quantitative data and indicate the measurement scale appropriate for each. (a) What is your age? (b) Are you male or female? (c) When did you first start reading the WSJ? High school, college, early career, midcareer, late career, or retirement? (d) How long have you been in your present job or position? (e) What type of vehicle are you considering for your next purchase? Nine response categories include sedan, sports car, SUV, minivan, and so on.	**[05]**	**1**	**2**
Q - 2	Consider the following frequency distribution.	**[05]**	**2**	**2/3/6**

Class	Frequency
10-19	10
20-29	14
30-39	17
40-49	7
50-59	2

Construct a cumulative frequency distribution and a cumulative relative frequency distribution.

Construct a histogram and an ogive.

Q - 3	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.	**[10]**	**4**	**1/6**
Q - 3	A data set has a first quartile of 42 and a third quartile of 50. Compute the lower and upper limits for the corresponding box plot. Should a data value of 65 be considered an outlier?	**[05]**	**3**	**2/3**
Q - 5	Calculate the sample correlation of the following data.	**[05]**	**3**	**1/3**

x_i	50	60	70	90	100
y_i	65	51	40	26	8

OR

Q - 5 Calculate the sample correlation of the following data. [05] 3 1/3

x_i	12	10	14	11	12	9
y_i	18	17	23	19	20	15

SECTION – II

Q - 1 How many ways can three items be selected from a group of six items? Use the letters A, B, C, D, E, and F to identify the items, and list each of the different combinations of three items. [05] 5 1/6

Q - 2 Consider the experiment of tossing a coin three times. [05] 5 2/3/6

(a) Develop a tree diagram for the experiment.

(b) List the experimental outcomes.

(c) What is the probability for each experimental outcome?

Q - 3 Suppose that we have two events, A and B, with $P(A) = 0.50$. $P(B) = 0.60$, and $P(A \cap B) = 0.40$. [05] 5 2/3/6

(a) Find P(A|B).

(b) Find P(B|A).

(c) Are A and B independent? Why or why not?

OR

Q - 3 The prior probabilities for events A_1 and A_2 are $P(A_1) = 0.40$ and $P(A_2) = 0.60$. It is also known that $P(A_1 \cap A_2) = 0$. Suppose that $P(B|A_1) = 0.20$ and $P(B|A_2) = 0.05$. [05] 5 2/3/6

(a) Are A_1 and A_2 mutually exclusive? Explain.

(b) Compute $P(A\_1 \cap B)$ and $P(A\_2 \cap B)$.

(c) Compute $P(B)$.

(d) Apply Bayes' theorem to compute $P(A\_1 \mid B)$ and $P(A\_2 \mid B)$.

Q - 4 Military radar and missile detection systems are designed to warn a country of an enemy attack. A reliability question is whether a detection system will be able to identify an attack and issue a warning. Assume that a particular detection system has a 0.90 probability of detecting a missile attack. Use the binomial probability distribution to answer the following questions. [10] 5 1/2/6

(a) What is the probability that a single detection system will detect an attack?

(b) If two detection systems are installed in the same area and operate independently, what is the probability that at least one of the systems will detect the attack?

(c) If three systems are installed, what is the probability that at least one of the systems will detect the attack?

(d) Would you recommend that multiple detection systems be used? Explain.

Q - 5 Suppose we have a multinomial population with four categories: A, B, C, and D. The null hypothesis is that the proportion of items is the same in every category. The null hypothesis is **[05] 6 3/4/5**

$$H_0 =: P_A = P_B = P_C = P_D = 0.25$$

A sample of size 300 yielded the following results.

$$A: 85 \quad B: 95 \quad C: 50 \quad D: 70$$

Use $\chi^2_{0.05} = 7.81473$.

OR

Q - 5 In 1950 in India the mean life expectancy was 50 years. If the life expectancies from a random sample of 11 persons are 58.2, 56.6, 54.2, 50.4, 44.2, 61.9, 57.5, 53.4, 49.7, 55.4, 57.0, does it confirm the expected view. Use $t_{0.01}$ =2.764 for 10 d.o.f. **[05] 6 3/4/5**

****

CO : Course Outcome Number BTL : Blooms Taxonomy Level

Level of Bloom's Revised Taxonomy in Assessment

1: Remember	2: Understand	3: Apply
4: Analyze	5: Evaluate	6: Create