P P SAVANI UNIVERSITY

Second Semester of M.Sc. CS Examination

August 2022

SESH7030 Statistical Methods for Data Science

08.08.2022, Monday **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 separate answer sheets.
3. Make suitable assumptions and draw neat figures wherever required.
4. Use of scientific calculator is allowed.

SECTION – I

Q - 1 Answer any two of the following. **[05]**

(i) Define sample, population and mutually exclusive event.

(ii) Define variable and regression line.

(iii) What is Chebyshev's theorem?

Q - 2 (a) Consider the following data. **[05]**

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.

Q - 2 (b) 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. **[05]**

(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.

OR

Q - 2 (a) Compute the mean, median, mode, Q_1, Q_2 and Q_3 of 87, 96, 85, 96, 85, 57, 49, 35, 14, 3, 9, 71, 3, 5, 1, 6, 2, 5 *and* 99. **[05]**

Q - 2 (b) Calculate the sample correlation of the following data. **[05]**

x_i	30	50	40	55	30	25	60	25	50	55
y_i	28	25	25	23	30	32	21	35	26	25

Q - 3 (a) Consider a sample with data values of 10, 20, 12, 17, and 16. Compute the z-score for each of the five observations. **[05]**

Q - 3 (b) Consider a sample with data values of 27, 25, 20, 15, 30, 34, 28, and 25. Provide the five-number summary for the data. Also show the box plot for the data. **[05]**

OR

Q - 3 (a) Consider a sample with data values of 53, 55, 70, 58, 64, 57, 53, 69, 57, 68 and 53. Compute the range, Interquartile range, variance and standard deviation. **[05]**

Q - 3 (b) 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]**

Q - 4 Florida Power & Light (FP&L) Company has enjoyed a reputation for quickly fixing its electric system after storms. However, during the hurricane seasons of 2004 and 2005, a **[05]**

new reality was that the company's historical approach to emergency electric system repairs was no longer good enough (The Wall Street Journal, January 16, 2006). Data showing the days required to restore electric service after seven hurricanes during 2004 and 2005 follow.

Hurricane	Days to Restore Service
Charley	13
Frances	12
Jeanne	8
Dennis	3
Katrina	8
Rita	2
Wilma	18

Based on this sample of seven, compute the following descriptive statistics:
(a) Mean, median, and mode
(b) Range and standard deviation

SECTION – II

Q - 1 Answer any two of the following. **[05]**

(i) Write addition law for mutually exclusive event.

(ii) Define Normal distribution and give proper example.

(iii) What is difference between t-test and F-test?

Q - 2 (a) A decision maker subjectively assigned the following probabilities to the four outcomes of an experiment: **[05]**

$$P(E_1) = 0.10, P(E_2) = 0.15, P(E_3) = 0.40 \text{ and } P(E_4) = 0.20.$$

Are these probability assignments valid? Explain.

Q - 2 (b) 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]**

(a) Find $P(A|B)$.
(b) Find $P(B|A)$.
(c) Are A and B independent?

OR

Q - 2 (a) Consider a binomial experiment with $n = 10$ and $p = 0.10$. **[05]**

(a) Compute $f(0)$.
(b) Compute $f(2)$.

Q - 2 (b) Consider a Poisson distribution with $\mu = 3$. **[05]**

(a) Write the appropriate Poisson probability function.
(b) Compute $f(2)$.
(c) Compute $f(1)$.
(d) Compute $P(x \geq 2)$.

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 $P(B|A_2) = 0.05$. **[10]**

(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 $P(A_1|B)$ and $P(A_2|B)$.

OR

Q - 3 Suppose that we have a sample space $S = \{E_1, E_2, \ldots, E_7\}$. **[10]**
where $E_1, E_2, \ldots, E_7$ denote the sample points. The following probability assignments apply: $P(E_1) = 0.05$, $P(E_2) = 0.20$, $P(E_3) = 0.20$, $P(E_4) = 0.25$, $P(E_5) = 0.15$, $P(E_6) = 0.10$ and $P(E_7) = 0.05$.
Let $A = \{E_1, E_4, E_6\}, B = \{E_2, E_4, E_7\}, C = \{E_2, E_3, E_5, E_7\}$.

(a) Find $P(A), P(B)$and $P(C)$.

(b) Find $A \cup B$ and $P(A \cup B)$.

(c) Find $A \cap B$ and $P(A \cap B)$.

(d) Are events A and C mutually exclusive?

(e) Find B^c and $P(B^c)$.

Q - 4 Can we conclude that the two population variances are equal for the following data of Officers passed out from a 'state' and 'private' university. **[05]**

STATE	1234	5678	1235	8521	1452	2653
PRIVATE	2563	6589	7631	889	145	7548

****