Statistics for Data Science - 1

Sample question paper for Diploma Qualifier

1. What can be said about the correlation coefficient r of x and y where $y = x^2 + 3x + 5$, x takes the values of the first twenty positive integers? [3 marks]

   (a) $r = 1$

   (b) $0 < r < 1$

   (c) $-1 < r < 0$

   (d) $r = -1$

2. Let X be a discrete random variable having the following probability mass function:[3 marks]

$$P(X = x) = k \times \frac{{}^{3}C_{x}}{{}^{3}C_{(3-x)}}$$

   where x is taking the value 0, 1, and 2. Find the value of k.
   0.5

3. Twelve cards are drawn simultaneously at random from an ordinary pack of 52 cards. Find the probability that exactly 2 are ace cards. [3 marks]

   (a) $\frac{{}^{4}C_{2} \times {}^{48}C_{10}}{{}^{52}C_{12}}$

   (b) $\frac{{}^{4}C_{2}}{{}^{52}C_{12}}$

   (c) $\frac{1}{8}$

   (d) $\frac{1}{2}$

4. Students from 2 colleges X and Y have participated in a competition, where 30% of the participants are from college X and 70% of the participants are from college Y. If 50% and 40% of the participants from colleges X and Y respectively are boys, choose the correct statement(s) from the following: [3 marks]

   (a) There is a 53% chance that a randomly selected participant is a boy.

   (b) There is a 43% chance that a randomly selected participant is a boy.

(c) If a boy is randomly selected, the probability that he is from college Y is 0.283.

(d) If a boy is randomly selected, the probability that he is from college Y is 0.651.

5. Which of the following variables is/are categorical? [2 marks]

   (a) Height (in metres)

   (b) Aadhar card number

   (c) Pan card number

   (d) Passenger Name Reservation (PNR) number

   (e) Stop watch time (in seconds)

   (f) Pin code

   Answer: b, c, d, f

6. Let the variance and probability of success, p, for a binomial random variable X be 1.2 and 0.4 respectively. Find the cumulative distribution function of the random variable $Y = X(X-2)$. [6 marks]

   (a)
$$F(y) = \begin{cases} 0 & \text{for } y < \text{-}1 \\ 0.2592 & \text{for } -1 \leq y < 0 \\ 0.68256 & \text{for } 0 \leq y < 3 \\ 0.91296 & \text{for } 3 \leq y < 8 \\ 0.98976 & \text{for } 8 \leq y < 15 \\ 1 & \text{for } y \geq 15 \end{cases}$$

   (b)
$$F(y) = \begin{cases} 0 & \text{for } y < \text{-}1 \\ 0.42336 & \text{for } -1 \leq y < 0 \\ 0.68256 & \text{for } 0 \leq y < 3 \\ 0.91296 & \text{for } 3 \leq y < 8 \\ 0.98976 & \text{for } 8 \leq y < 15 \\ 1 & \text{for } y \geq 15 \end{cases}$$

   (c)
$$F(y) = \begin{cases} 0 & \text{for } y < 0 \\ 0.2592 & \text{for } 0 \leq y < -1 \\ 0.68256 & \text{for } -1 \leq y < 3 \\ 0.91296 & \text{for } 3 \leq y < 8 \\ 0.98976 & \text{for } 8 \leq y < 15 \\ 1 & \text{for } y \leq 15 \end{cases}$$

(d)

$$F(y) = \begin{cases} 0 & \text{for } y < 0 \\ 0.2592 & \text{for } 0 \leq y < -1 \\ 0.68256 & \text{for } -1 \leq y < 3 \\ 0.91296 & \text{for } 3 \leq y < 8 \\ 0.98976 & \text{for } 8 \leq y < 15 \\ 1 & \text{for } y \geq 15 \end{cases}$$

Answer: a

7. There are 5 vacant positions in the film certification board. A total of 200 people from the film industry applied for the positions. If a person can take more than one position, then in how many ways can the 5 vacant positions be filled? [2 marks]

   a. ${}^{200}C_5$

   b. ${}^{200}P_5$

   c. 5^{200}

   d. 200^5

8. The variance of a binomial random variable X is plotted against varying values of p with n kept constant. It is observed that the maximum value of variance is 10, then what is the value of n? [2 marks]

   (a) 15

   (b) 10

   (c) 20

   (d) 40

   Answer: d

9. From the options, choose the outliers, if any, for the following dataset:
   9.5, 10.5, 11, 12, 12.1, 12.4, 12.5, 13, 13.6, 13.8, 13.9, 14, 15.2, 16.3, 17.3. [3 marks]

   a. 16.3

   b. 17.3

   c. 10.5

   d. 9.5

10. If the median of the dataset x_i, where $i = 1, 2, 3, \cdots, n$, is 13, what is the median of the dataset $2x_i - 1$, where $i = 1, 2, 3, \cdots, n$? [2 marks]
    Answer: 25

11. In a subject of total duration 5 weeks, there are weekly graded tests. The probability that a student gets pass mark in a weekly test is related to the marks obtained in previous weekly test. If a student gets pass marks in previous week's test then there is 80% probability that the student will get pass mark in current week's test, else if a student fails to get pass mark in previous week's test then there is a 40% chance that the student will get pass marks in current week's test. (Assume that every student gets pass mark in the first week graded test.) If the student gets pass marks in exactly 4 weeks out of the 5 graded tests, what is the probability that the student failed in the 4^{th} week? [5 marks]

(a) $\frac{1}{15}$
(b) $\frac{1}{10}$
(c) $\frac{1}{5}$
(d) $\frac{2}{5}$

Answer: c

12. Contingency Table 1.E.1 summarizes two categorical variables: status of completion of a course and the three colleges from which the course has been taken by students. [2 marks]

	College A	College B	College C
Completed	420	420	480
Not Completed	180	280	120

Table 1.E.1: Course completion data

A student is selected at random. What is the probability that the student has completed the course given that he/she is from college B? (Correct up to 1 decimal points)
Answer: 0.6 Accepted range: 0.55 - 0.65

13. A company has predicted that the next year's profit will follow the probability distribution shown in Table 1.E.2. The random variable X denote the profit in million rupees. Loss is denoted by a negative profit.

X (Profit in million rupees)	-0.5	0	1	2	3
$P(X = x)$	0.2	0.15	0.25	0.2	0.2

Table 1.E.2: Distribution of company's next year profit.

But the company cannot retain all the profits to itself since it has to share 5% of the profit to its investors. The amount that the company retains is given by $Y = 0.95X$. Find the standard deviation (in million rupees) of Y. [5 marks]

a. 1.57

b. 1.42

c. 1.19

d. 1.25

14. Priyanka selects three numbers randomly from the set of natural numbers from 1 to n. Find the probability that she will select a triplet (i.e., three consecutive numbers), $n \geq 3$. [3 marks]

(a) $\frac{1}{n-2}$

(b) $\frac{(n-2)}{{}^{n}C_3}$

(c) $\frac{1}{{}^{n}C_3}$

(d) $\frac{2}{{}^{n}C_3}$

Answer: b

15. If the population standard deviation of first $2n$ natural numbers (excluding zero) is s_1, the population standard deviation of next $2n$ natural numbers is s_2, and the population standard deviation of first $4n$ natural numbers (excluding zero) is s_3. Which of the following statements is/are true? [6 marks]

(a) $s_1 > s_2$

(b) $s_2 = s_1$

(c) $s_2 < s_3$

(d) $s_1 < s_3$

(e) $s_1 = s_2 > s_3$

Answer: b, c, d