

Statistical Modelling for Machine Learning (313307)

MSBTE Diploma · K-Scheme · Model Question Paper — Set A

Time: 3 Hours

Max. Marks: 70

Instructions:

- All questions are compulsory.
- Answer each next main question on a new page.
- Illustrate your answers with neat sketches wherever necessary.
- Figures to the right indicate full marks.
- Assume suitable data, if necessary.
- Use of non-programmable electronic pocket calculator is permissible.
- Mobile phone, pager and any other electronic communication devices are not permissible in the examination hall.

Q.1 Attempt any FIVE of the following:

10

- Find the lower quartile (Q_1) and the upper quartile (Q_3) of the following data:
23, 15, 31, 27, 19, 42, 35, 28, 17, 38, 25.
- Find the harmonic mean (H.M.) of 4, 6, 8, 12, 24.
- For a distribution, mean = 45, mode = 41.5 and standard deviation = 7. Find Karl Pearson's coefficient of skewness.
- Find the coefficient of correlation between x and y if $n = 10$, $\Sigma x = 50$, $\Sigma y = 80$, $\Sigma x^2 = 290$, $\Sigma y^2 = 700$ and $\Sigma xy = 430$.
- A die is thrown twice. Find the probability that the sum of the numbers is 9 or both numbers are the same (a doublet).
- Construct the forward difference table for $f(x) = 2x^3 - x^2 + 3$ taking $x = 0, 1, 2, 3, 4$ and show that the third differences are constant.
- Define null hypothesis and alternative hypothesis.

Q.2 Attempt any THREE of the following:

12

- Calculate the median of the following distribution of marks:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	5	9	14	20	16	10	6

- Find the mode of the following data graphically (histogram) and verify it by formula:

Class	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	6	11	19	27	21	9

- c) Calculate Bowley's coefficient of skewness for the following data:

C.I.	0–10	10–20	20–30	30–40	40–50	50–60
Frequency	5	12	20	18	10	5

- d) Karl Pearson's coefficient of skewness of a distribution is -0.4 , its mean is 60 and standard deviation is 15. Find the mode and the median of the distribution.

Q.3 Attempt any THREE of the following:

12

- a) Calculate Karl Pearson's coefficient of skewness for the following ages of employees:

Age (years)	20–25	25–30	30–35	35–40	40–45	45–50	50–55
Frequency	6	14	22	30	16	8	4

- b) The first four moments of a distribution about the value 4 are 1, 6, 10 and 120. Find the moments about the mean, β_1 and β_2 , and comment on the skewness and kurtosis.
- c) Calculate Karl Pearson's coefficient of correlation for the following data and state the type of correlation:

x	4	6	8	10	12	14
y	12	15	14	20	24	23

- d) Fit a straight line $y = a + bx$ to the following data by the method of least squares and estimate y when $x = 7$:

x	1	2	3	4	5	6
y	3	7	8	12	14	17

Q.4 Attempt any THREE of the following:

12

- a) Marks obtained by 8 students in Statistics (x) and Python (y) are given below. Calculate Spearman's rank correlation coefficient.

x	45	62	38	70	55	49	81	66
y	52	60	41	68	49	58	75	72

- b) Find the two regression lines for the following data and estimate y when $x = 6$:

x	1	3	5	7	9
y	15	12	10	8	5

- c) A card is drawn at random from a well-shuffled pack of 52 cards. Find the probability that it is (i) a king or a heart, (ii) a red card or a face card.
- d) A bag contains 5 white and 8 red balls. Two balls are drawn one after the other without replacement. Find the probability that (i) both are red, (ii) one is white and the other is red.

- e) Three machines M_1 , M_2 and M_3 produce 30%, 45% and 25% of the total items of a factory. Of their output 3%, 2% and 4% respectively are defective. An item is chosen at random and found defective. Find the probability that it was produced by machine M_3 .

Q.5 Attempt any TWO of the following:

12

- a) (i) Using Lagrange's interpolation formula, find $f(5)$ from the table:

x	1	3	4	6
f(x)	5	13	22	50

- (ii) Using Lagrange's formula, find the polynomial $f(x)$ passing through (0, 2), (1, 3) and (3, 11). Hence find $f(2)$.

- b) The population of a town (in thousands) is given below:

Year	1971	1981	1991	2001	2011
Population	40	48	59	72	88

- (i) Estimate the population in 1976 using Newton's forward interpolation formula.
(ii) Estimate the population in 2006 using Newton's backward interpolation formula.

- c) (i) From the following table, find $f(7)$ by extrapolation:

x	1	2	3	4	5
f(x)	2	9	28	65	126

- (ii) Prove that (a) $E = 1 + \Delta$, (b) $\nabla = 1 - E^{-1}$.

Q.6 Attempt any TWO of the following:

12

- a) The number of books issued by a college library on six days of a week is given below. Test whether the issue of books is uniformly distributed over the days at 5% level of significance. (χ^2 for 5 d.f. at 5% = 11.07)

Day	Mon	Tue	Wed	Thu	Fri	Sat
Books issued	72	84	66	90	78	90

- b) (i) A sample of 10 students gave the following test scores: 48, 52, 47, 55, 51, 46, 49, 53, 44, 50. Can it be regarded as taken from a population with mean 47? (t for 9 d.f. at 5% = 2.262)
(ii) A coin is tossed 600 times and heads turn up 330 times. Test the hypothesis that the coin is unbiased at 5% level of significance.

- c) In an experiment on vaccination of 400 persons the following results were obtained. Use χ^2 -test to find whether vaccination and attack of the disease are independent. (χ^2 for 1 d.f. at 5% = 3.84)

	Attacked	Not attacked	Total
Vaccinated	25	175	200
Not vaccinated	45	155	200
Total	70	330	400

Note: This is a **model question paper prepared by Ananda Technologies** only for practice and exam preparation, based on the pattern of previous MSBTE question papers and the K-scheme syllabus. It is **not an official MSBTE paper**, and it does **not** mean that these questions will come in the exam. Answers follow standard textbook points; marks shown are indicative. Free for students — anandatechnologies.in