

Statistical Modelling for Machine Learning (313307)

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

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 median and the mode of: 12, 15, 11, 15, 18, 20, 15, 13, 11, 16.
- Find the geometric mean of 3, 6, 24 and 48.
- The two regression coefficients are $b_{yx} = 0.8$ and $b_{xy} = 0.45$. Find the coefficient of correlation.
- For a distribution, mean = 50, median = 48 and standard deviation = 12. Find Karl Pearson's coefficient of skewness.
- A bag contains 4 red, 6 blue and 5 green balls. One ball is drawn at random. Find the probability that it is red or green.
- Using Newton's forward interpolation formula, find $f(14)$ from:

x	10	20	30
f(x)	15	25	41

- Define standard error. Write the standard error of (i) sample mean, (ii) sample proportion.

Q.2 Attempt any THREE of the following:

12

- Find the median of the following distribution of daily wages:

Wages (₹ hundred)	20–29	30–39	40–49	50–59	60–69	70–79
Frequency	5	12	20	28	15	7

- Find the mode of the following distribution by drawing a histogram, and check it by formula:

Class	0–5	5–10	10–15	15–20	20–25	25–30	30–35
Frequency	3	8	14	22	17	9	4

- c) Calculate the quartile deviation and the coefficient of quartile deviation from the following data:

Profit less than (₹ lakh)	10	20	30	40	50	60
No. of companies	6	18	40	70	88	100

- d) For a distribution, Bowley's coefficient of skewness is 0.6, the sum of the upper and lower quartiles is 130 and the median is 59. Find the quartiles and the coefficient of quartile deviation.

Q.3 Attempt any THREE of the following:

12

- a) Calculate Karl Pearson's coefficient of skewness for the following distribution of weights (kg):

Weight	40–45	45–50	50–55	55–60	60–65	65–70	70–75
Frequency	5	12	18	25	20	13	7

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

Class	0–20	20–40	40–60	60–80	80–100	100–120
Frequency	8	15	27	25	15	10

- c) Calculate the first four central moments, β_1 and β_2 for the following distribution and state the type of kurtosis:

x	0	1	2	3	4
f	5	10	20	10	5

- d) Fit a second-degree parabola $y = a + bx + cx^2$ to the following data by the method of least squares and estimate y at $x = 5$:

x	0	1	2	3	4
y	2	3	6	11	19

Q.4 Attempt any THREE of the following:

12

- a) Find both the regression lines for the following data. Estimate y when $x = 10$ and x when $y = 7$.

x	3	5	6	8	9	11
y	2	3	4	6	5	8

- b) Two judges ranked 10 contestants in a coding competition as follows. Find Spearman's rank correlation coefficient and comment.

Contestant	A	B	C	D	E	F	G	H	I	J
Judge I	3	5	8	4	7	10	2	1	6	9
Judge II	6	4	9	8	1	2	3	10	5	7

- c) Draw a scatter diagram and calculate Karl Pearson's coefficient of correlation for:

x	2	4	5	6	8	11
y	18	12	10	8	7	5

- d) (i) A card is drawn from a pack of 52 cards. Find the probability that it is a black card or a king.
(ii) Two cards are drawn one after the other without replacement. Find the probability that both are hearts.
- e) A disease affects 2% of a population. A test detects the disease correctly in 95% of persons who have it, but also gives a positive result for 4% of healthy persons. A person chosen at random tests positive. Find the probability that the person really has the disease.

Q.5 Attempt any TWO of the following:

12

- a) (i) Using Lagrange's formula, find $f(4)$ from:

x	2	3	5	6
f(x)	3	7	21	31

- (ii) Using Lagrange's formula, find the polynomial through $(-1, 2)$, $(0, 1)$ and $(2, 5)$. Hence find $f(1)$.
- b) (i) The sales (₹ lakh) of a company are given below. Estimate the sales for 2022 using Newton's backward interpolation formula.

Year	2015	2017	2019	2021	2023
Sales	12	15	20	27	36

- (ii) Prove that $\Delta \nabla = \nabla \Delta = \Delta - \nabla$.
- c) From the following table, find $f(0)$ and $f(7)$ by extrapolation:

x	1	2	3	4	5
f(x)	3	9	19	33	51

Q.6 Attempt any TWO of the following:

12

- a) A die is thrown 120 times with the following results. Test whether the die is unbiased at 5% level of significance. (χ^2 for 5 d.f. at 5% = 11.07)

Face	1	2	3	4	5	6
Frequency	15	22	18	26	17	22

- b) (i) The IQs of 10 randomly selected students are: 104, 98, 110, 95, 102, 108, 99, 112, 101, 106. Do these data support the assumption of a population mean IQ of 100? (t for 9 d.f. at 5% = 2.262)
- (ii) A sample of 150 students of college A has mean marks 62 with S.D. 8, and a sample of 120 students of college B has mean marks 59 with S.D. 10. Is the difference between the means significant at 5% level?
- c) The following table shows the programming language preferred by 200 AI & ML students. Test whether the preference is independent of gender at 5% level of significance. (χ^2 for 2 d.f. at 5% = 5.991)

Gender	Python	R	Java	Total
Male	60	25	15	100
Female	50	30	20	100
Total	110	55	35	200

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