

**POST GRADUATE DIPLOMA IN
APPLIED STATISTICS (PGDAST)**

Term-End Examination

December, 2015

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

- (i) *Question no. 1 is compulsory.*
- (ii) *Questions no. 2 to 5 have the internal choices.*
- (iii) *Use of scientific calculator is allowed.*
- (iv) *Formulae and Table Booklet for PGDAST is allowed. Symbols have their usual meaning.*

1. State whether the following statements are *true* or *false*. Give reasons for your answer. $5 \times 2 = 10$

- (a) If $N = 50$, $(A) = 35$, $(B) = 25$ and $(AB) = 15$, then the attributes A and B are said to be independent.
- (b) Both regression lines of Y on X and X on Y do not intersect at all.

- (c) If $r(X, Y) > 0$, then as X increases, Y decreases.
- (d) The sum of deviations of all values taken by a distribution from their mean is 4.6.
- (e) The regression coefficient of Y on X is 4.0 and that of X on Y is 0.6.

2. (a) Define the weighted mean of a set of numbers.

2

(b) Find the missing information from the following data :

8

	Number	Standard Deviation	Mean
Group I	50	6	113
Group II	?	7	?
Group III	90	?	115
Combined	200	7.745	116

OR

- (a) The following table shows the distribution of 100 families according to their expenditure per week. A number of families corresponding to two of the expenditure groups are missing in the table. The median and mode are given to be ₹ 25 and ₹ 24.

Find the missing frequencies of the data :

7

<i>Expenditure</i>	<i>No. of Families</i>
0 – 10	14
10 – 20	?
20 – 30	27
30 – 40	?
40 – 50	15

- (b) In a frequency distribution, the coefficient of skewness based on the quartiles is 0.6. If the sum of the upper and lower quartiles is 100 and the median is 38, find the values of the upper and lower quartiles.

3

3. How do you define the term "line of best fit" ?
Show that the line of best fit to the following
data is given by $Y = -0.5X + 8$:

10

X	Y
6	5
7	5
7	4
8	5
8	4
8	3
9	4
9	3
10	3

OR

Describe different types of correlation between two variables. In two sets of variables X and Y with 50 observations each, the following data were observed :

$$\Sigma X = 500, \Sigma Y = 300, \Sigma X^2 = 5450, \Sigma Y^2 = 2000$$

and $r(X, Y) = 0.3$.

But on subsequent verification it was found that one value of X (= 10) and that of Y (= 6) were inaccurate and hence removed. With the remaining 49 pairs of values, how is the original value of $r(X, Y)$ affected ?

10

4. (a) Find the most likely price of an item in Delhi corresponding to its price of ₹ 70 at Chennai from the following data :

	Chennai	Delhi
Average Price (₹)	65	67
Standard Deviation	2.5	3.5

The correlation coefficient between the prices of the item in the two cities is given to be 0.8.

6

- (b) Explain the concept of multiple correlation in terms of total and partial correlation coefficients. 4

OR

- (a) The equations of two regression lines are as follows :

$$3X + 12Y = 19 \text{ and } 9X + 3Y = 46.$$

Obtain

- (i) the value of correlation coefficient, and
(ii) mean values of X and Y. 7
- (b) Explain the concept of partial correlation analysis. 3

5. Find whether A and B are independent, positively associated or negatively associated in each of the following cases : 10

- (i) $N = 1000$, $(A) = 470$, $(B) = 620$ and $(AB) = 320$
- (ii) $(A) = 490$, $(AB) = 294$, $(\alpha) = 570$, and $(\alpha B) = 380$
- (iii) $(AB) = 256$, $(\alpha B) = 768$, $(A\beta) = 48$, and $(\alpha\beta) = 144$

OR

The following contingency table presents the analysis of 300 persons according to skin colour and eye colour. Determine the amount of association between them.

<i>Eye Colour</i>	<i>Skin Colour</i>			<i>Total</i>
	<i>Fair</i>	<i>Wheatish</i>	<i>Dark</i>	
Blue	30	10	40	80
Grey	40	20	40	100
Brown	50	30	40	120
Total	120	60	120	300

It is given that the value for $C_{\max}$ of the 3×3 Contingency table is 0.816.

10