

No. of Printed Pages : 7

MST-002

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

Term-End Examination

June, 2015

00898

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

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

1. State whether the following statements are *True* or *False*. Give reason in support of your answer. $5 \times 2 = 10$
- (a) Mode is the best measure of central tendency.
 - (b) The coefficient of correlation will have positive sign when X does not change with changes in Y.

- (c) If two variables are not correlated, the lines of regression become perpendicular to each other.
- (d) The regression coefficient of Y on X is 3.2 and that of X on Y is 0.8.
- (e) Attributes A and B are said to be independent, if $(AB) < \frac{(A)(B)}{N}$.

2. (a) The average salary of male employees in a firm was ₹ 52,000 and that of female employees was ₹ 42,000. The mean salary of all the employees was ₹ 50,000. Find the percentage of male and female employees in the firm. 5

- (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 median is 38, find the values of upper and lower quartiles. 5

OR

- (a) An incomplete frequency distribution is given as follows :

| <i>Class Interval</i> | <i>Frequency</i> |
|-----------------------|------------------|
| 10 – 20 | 12 |
| 20 – 30 | 30 |
| 30 – 40 | ? |
| 40 – 50 | 65 |
| 50 – 60 | ? |
| 60 – 70 | 25 |
| 70 – 80 | 18 |
| Total | 229 |

Given that the median value is 46, determine the missing frequencies of the distribution.

6

- (b) The first of the two samples has 100 items with mean 15 and standard deviation 3. If the whole group has 250 items with mean 15.6 and standard deviation $\sqrt{13.44}$, find the standard deviation of the second group.

4

3. (a) Fit a second degree curve to the following data related to the profit of a certain company :

| <i>Year</i> | <i>Profit (in ₹ lakhs)</i> |
|-------------|--------------------------------|
| 1960 | 125 |
| 1962 | 140 |
| 1964 | 165 |
| 1966 | 195 |
| 1968 | 230 |

Estimate the profit in the year 1975.

4

- (b) The coefficient of rank correlation of the marks obtained by 10 students in Maths and Statistics was found to be 0.5. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 3 instead of 7. Find the corrected rank correlation coefficient.

6

OR

- (a) A computer while calculating correlation coefficient between two variables X and Y from 25 pairs of observations obtained the following results :

$$\Sigma X = 125, \Sigma X^2 = 650, \Sigma Y = 100, \Sigma Y^2 = 460, \\ \Sigma XY = 508$$

It was later found that the operator had copied down two pairs (6, 14) and (8, 6) as (X, Y) while the correct values were (8, 12) and (6, 8) respectively. Obtain the correct value of $r(X, Y)$.

7

- (b) Explain the method of least squares.

3

4. (a) . The equations of two regression lines obtained in a data analysis are as follows :

$$3X + 12Y = 19$$

$$3Y + 9X = 46$$

- Obtain (i) mean values of X and Y; and
(ii) the value of correlation coefficient.

7

- (b) Explain what are regression lines. Why are there two such lines ?

3

OR

- (a) For 50 students of a class, the regression equation of marks in Statistics (X) on marks in Mathematics (Y) is $3Y - 5X + 180 = 0$. The mean marks in Mathematics are 44 and the variance of marks in Statistics is $(9/16)^{\text{th}}$ of the variance of marks in Mathematics. Find the mean marks in Statistics and the coefficient of correlation between marks in the two subjects. 7

- (b) In a trivariate distribution, $r_{23.1} = 0.69$, $r_{12} = 0.8$ and $r_{13} = 0.6$. Compute r_{23} . 3

5. (a) What are the possible types of association between attributes? Explain the distinction between them. 5

- (b) Given that $(AB) = 150$, $(A\beta) = 230$, $(\alpha B) = 260$ and $(\alpha\beta) = 2340$. Find the frequencies of A and B. Also obtain the value of N. 5

OR

The following table shows the IQ values of 1000 students at a college according to the economic conditions of their parents :

| <i>Economic Condition</i> | <i>IQ level</i> | | <i>Total</i> |
|---------------------------|-----------------|------------|--------------|
| | <i>High</i> | <i>Low</i> | |
| Rich | 460 | 140 | 600 |
| Poor | 240 | 160 | 400 |
| Total | 700 | 300 | 1000 |

Use the coefficient of contingency to determine the amount of association between economic condition and IQ level. It is given that the value of C_{\max} is 0.707.

10

No. of Printed Pages : 7

MST-002

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

Term-End Examination

00444

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

No. of Printed Pages : 7

MST-002

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

Term-End Examination

00107

June, 2016

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

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

1. State whether the following statements are *true* or *false*. Give reasons in support of your answers. $5 \times 2 = 10$

- (a) The weighted mean of n natural numbers is $\frac{(2n+1)}{3}$, if weights are the corresponding numbers.
- (b) The values of regression coefficients b_{yx} and b_{xy} are $\frac{1}{5}$ and 10 respectively.

- (c) If Y is proportional to X, then $r(X, Y) = 1$.
- (d) In case of consistent data, no class frequency can be negative.
- (e) If 25 is subtracted from each value of X and Y and then divided by 10, the new b'_{yx} is $2\frac{1}{2}$ times of b_{yx} .

2. (a) The annual rates of growth of output of a factory in 5 years are 5.0, 7.5, 2.5, 5.0 and 10.0, respectively. What is the compound rate of growth of output per annum for the period?
- (b) The number of employees, average daily wages and the variance of the wages for two factories are given as :

| | Factory A | Factory B |
|-----------------------------|-----------|-----------|
| Number of employees | 50 | 100 |
| Average daily wage (₹) | 120 | 85 |
| Variance of daily wages (₹) | 09 | 16 |

- (i) In which factory is there a greater variation in the distribution of daily wages?

- (ii) Suppose in factory B, the daily wages of an employee were wrongly noted as ₹ 120 instead of ₹ 100. What would the correct variance for factory B be? 6

OR

The data of monthly profit (in lacs) of various companies is given as follows :

| Monthly Profit (in lacs) | Number of Companies |
|-----------------------------|------------------------|
| Less than 7.5 | 04 |
| 7.5 – 12.5 | 10 |
| 12.5 – 17.5 | 20 |
| 17.5 – 22.5 | 36 |
| 22.5 – 27.5 | 16 |
| 27.5 – 32.5 | 12 |
| 32.5 – 37.5 | 02 |

Determine whether the distribution is symmetric and mesokurtic by calculating the moments, β_1 and β_2 . 10

3. (a) A computer while calculating the correlation coefficient between the variables X and Y obtained the following results :

$$N = 30, \Sigma X = 120, \Sigma Y = 90, \Sigma X^2 = 600, \\ \Sigma Y^2 = 250, \Sigma XY = 335$$

It was, however, later discovered at the time of checking that it had copied wrongly, two pairs of observations (X, Y) as (8, 10) and (12, 7) while the correct values were (8, 12) and (10, 8) respectively. Obtain the correct value of the correlation coefficient between X and Y.

7

- (b) The coefficient of rank correlation of the marks obtained by 10 students in Statistics and Accountancy was found to be 0.8. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was taken as 7 instead of 9. Find the correct value of the rank correlation coefficient.

3

OR

- (a) In order to find the correlation coefficient between X and Y from 12 pairs of observations, the following calculations were made :

$$\Sigma X = 30, \Sigma Y = 5, \Sigma X^2 = 670, \Sigma Y^2 = 285, \Sigma XY = 334$$

On subsequent verification it was found that the pair (X = 11, Y = 4) was copied wrongly while the correct value was (X = 10, Y = 14). Find the correct value of the correlation coefficient.

7

- (b) An examination of eight applicants for a clerical post was taken by a firm. From the marks obtained by the applicants in the Accountancy and Statistics papers, compute the rank correlation coefficient to check how far the competence of the applicants in both subjects go together ?

3

| Applicant : | A | B | C | D | E | F | G | H |
|------------------------|----|----|----|----|----|----|----|----|
| Marks in Accountancy : | 15 | 20 | 28 | 12 | 40 | 60 | 20 | 80 |
| Marks in Statistics : | 40 | 30 | 50 | 30 | 20 | 10 | 30 | 60 |

4. (a) The regression equation of Y on X and X on Y are $Y = 2X$ and $6X - Y = 4$, respectively and the second moment of X about the origin is 3. Find (i) the mean values of X and Y; (ii) the value of the correlation coefficient, and (iii) the standard deviation of Y. 7
- (b) Suppose a data analyst has found $r_{12} = 0.96$, $r_{13} = 0.36$ and $r_{23} = 0.78$ for a given set of values of X_1 , X_2 and X_3 . Examine whether their computations may be said to be free from error. -6 3

OR

Given the following bivariate data. :

| | | | | | | | | |
|-----|----|---|---|---|---|---|---|---|
| X : | -1 | 5 | 3 | 2 | 1 | 1 | 7 | 3 |
| Y : | -6 | 1 | 0 | 0 | 1 | 2 | 1 | 5 |

- (i) Fit a regression line of Y on X and find Y, if $X = 10$.
- (ii) Fit a regression line of X on Y and find X, if $Y = 2.5$. 10

5. (a) The following ultimate class frequencies are given as :

$$(ABC) = 60, (AB\gamma) = 250, (A\beta C) = 80,$$

$$(A\beta\gamma) = 350, (\alpha BC) = 75, (\alpha B\gamma) = 650,$$

$$(\alpha\beta C) = 55 \text{ and } (\alpha\beta\gamma) = 8200.$$

Find the frequencies of all positive classes. 6

- (b) What is meant by a class frequency of (i) first order, and (ii) third order ? How would you express a class frequency of first order in terms of class frequencies of third order ?

4

OR

A company is interested in determining the strength of association between the communicating time of their employees and the level of stress-related problems observed on job. A study of 116 assembly workers reveals the following :

| <i>Communicating Time</i> | <i>Stress</i> | | | <i>Total</i> |
|---------------------------|---------------|-----------------|------------|--------------|
| | <i>High</i> | <i>Moderate</i> | <i>Low</i> | |
| Under 20 minutes | 9 | 5 | 18 | 32 |
| 20 – 50 minutes | 17 | 8 | 28 | 53 |
| Over 50 minutes | 18 | 6 | 7 | 31 |
| Total | 44 | 19 | 53 | 116 |

Use the coefficient of contingency to determine the amount of association between the communicating time and stress, given $C_{\max} = 0.816$.

10

No. of Printed Pages : 7

MST-002

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

Term-End Examination

December, 2016

01167

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

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

1. State whether the following statements are *True* or *False*. Give reasons in support of your answers. $5 \times 2 = 10$

- (a) The standard deviation of 10 values of a data is 50. If each value is divided by 5, then the new standard deviation would become 10.
- (b) If the correlation coefficient is 0, then the regression coefficient would also become 0.

(c) The geometric mean of regression coefficients is equal to twice the correlation coefficient.

(d) If $\frac{(A)}{N} = x$, $\frac{(B)}{N} = 2x$ and $\frac{(C)}{N} = 3x$ and $\frac{(AB)}{N} = \frac{(BC)}{N} = \frac{(AC)}{N} = y$, then neither x nor y can exceed $\frac{1}{4}$, if the data is consistent.

(e) If $N = 100$, $(A) = 60$, $(B) = 40$ and $(AB) = 24$, then the attributes A and B are positively associated.

2. (a) A toy factory has assigned a group of 4 workers to complete an order of 1400 toys of a certain type. The productive rates of the four workers are given below :

| <i>Worker</i> | <i>Productive Rate</i> |
|---------------|------------------------|
| A | 4 minutes per toy |
| B | 6 minutes per toy |
| C | 10 minutes per toy |
| D | 15 minutes per toy |

Find the average minutes per toy devoted by the group of workers.

4

- (b) The first four moments of a distribution about the value 5 of the variable are 2, 20, 40 and 50, respectively. Show that the mean is 7. Also check whether the distribution is symmetric and mesokurtic.

6

OR

For two firms, A and B, belonging to the same industry, the following details are available :

| | <i>Firm A</i> | <i>Firm B</i> |
|-----------------------------|---------------|---------------|
| Number of Employees | 100 | 200 |
| Average Wage | ₹ 240 | ₹ 170 |
| Standard Deviation of Wages | ₹ 6 | ₹ 8 |

- (a) Which firm pays out a larger amount as weekly wages ?
- (b) Which firm shows greater variability in the distribution of the wages ?
- (c) Find the average weekly wages and standard deviation of the wages of all employees of both firms.

10

3. (a) In order to find the correlation coefficient between two variables X and Y from 12 pairs of observations, the following calculations were made :

$$\Sigma X = 30, \Sigma Y = 5, \Sigma X^2 = 670, \Sigma Y^2 = 285 \text{ and } \Sigma XY = 334.$$

On subsequent verification it was found that the pair (X = 11, Y = 4) was copied wrongly, while the correct value was (X = 10, Y = 14). Find the correct value of the correlation coefficient.

7

- (b) Two housewives, Mrs. Neena and Mrs. Meena, asked to express their preferences for different kinds of detergents, gave the following replies :

| <i>Detergent</i> | <i>Meena</i> | <i>Neena</i> |
|------------------|--------------|--------------|
| A | 1 | 1 |
| B | 4 | 2 |
| C | 2 | 4 |
| D | 3 | 3 |
| E | 5 | 7 |
| F | 7 | 8 |
| G | 6 | 6 |
| H | 8 | 5 |
| I | 9 | 9 |
| J | 10 | 10 |

To what extent do the preferences of these two ladies go together ?

3

OR

- (a) The profits of ₹ y of a certain company in the x^{th} year of its existence are given by

X: 1 2 3 4 5

Y: 1250 1400 1650 1950 2300

Show that the best fit equation of the parabola on the given data is

$$Y = 1140 + 72X + 32 \cdot 15X^2.$$

7

- (b) The coefficient of rank correlation of the marks obtained by 10 students in Statistics and Accountancy was found to be 0.8. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was taken as 7 instead of 9. Find the correct value of the rank correlation coefficient.

3

4. (a) Two lines of regression are given by

$$x + 2y - 5 = 0 \text{ and } 2x + 3y - 8 = 0 \text{ and } \sigma_x^2 = 12.$$

Find (i) the mean values of x and y , (ii) the correlation coefficient between x and y , and (iii) the value of standard deviation of y .

7

- (b) In a certain investigation, the following results were obtained for a given set of values of X_1 , X_2 and X_3 :

$$r_{12} = 0.8, r_{13} = 0.2 \text{ and } r_{23} = -0.5.$$

Do you think that the computations are free from error?

3

OR

A company wants to assess the impact of R&D expenditure on its annual profit. The following table presents the information for the last eight years :

| <i>Years</i> | <i>R&D Expenditure (in lakhs)</i> | <i>Annual Profit (in thousands)</i> |
|--------------|---|---|
| 2001 | 2 | 20 |
| 2002 | 3 | 25 |
| 2003 | 5 | 34 |
| 2004 | 4 | 30 |
| 2005 | 10 | 60 |
| 2006 | 5 | 41 |
| 2007 | 7 | 42 |
| 2008 | 9 | 45 |

Estimate the regression equation and predict the annual profit for 2012 for an allocated sum of ₹ 1 crore as R&D expenditure.

10

5. (a) Among the adult population of a certain town, 50% of the population is male, 60% are wage-earners and 50% are 45 years of age or above. 10% of the males are not wage-earners and 45% of the males are under 45. Can we infer anything about what percentage of population of 45 years of age or above are wage-earners ?

4

- (b) In a sample of 1000 children, 400 came from higher income group and the rest from lower income group. The number of delinquent children in these groups was 50 and 200, respectively. Investigate if there is any association between delinquency and income groups.

6

OR

For the given data, study the association between the temperament of Brothers and Sisters by computing the coefficient of contingency and interpret the results :

10

| | | Temperament of Sisters | | | Total |
|---------------------------------|--------------|------------------------|--------------|--------|-------|
| | | Quick | Good Natured | Sullen | |
| Temperament of Brothers | Quick | 850 | 571 | 580 | 2001 |
| | Good Natured | 618 | 593 | 455 | 1666 |
| | Sullen | 540 | 456 | 457 | 1453 |
| | Total | 2008 | 1620 | 1492 | 5120 |
| Given that $C_{\max} = 0.816$. | | | | | |

No. of Printed Pages : 5

MST-002

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

Term-End Examination

June, 2017

02262

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

- (i) *Attempt all questions. Questions no. 2 to 5 have internal choices.*
- (ii) *Use of scientific calculator is allowed.*
- (iii) *Use of Formulae and Statistical Tables Booklet is allowed.*
- (iv) *Symbols have their usual meaning.*

1. State whether the following statements are *True* or *False*. Give reason in support of your answer. $5 \times 2 = 10$
- (a) For a given set of data; if $AM = 4$ and $HM = 8$, then GM will be 16.
 - (b) Standard deviation of the data : 10, 10, 10, 10, 10 is 0.1.
 - (c) In a population of size N , we define three attributes A, B, C such that $(ABC) = 60$, $(\alpha B C) = 75$, $(AB \gamma) = 250$, $(AB) = 10$. The data is consistent.
 - (d) Suppose x is measured in cm and y is measured in kg. If we want to fit $y = a + bx$, then the unit of b will be cm/kg.
 - (e) If $r(x, y) = 0$, then x and y have no relationship.

2. (a) For the data 5, 8, 12, 15, 20, 30,

what should be the value of A so that

$$\sum_{i=1}^6 (x_i - A)^2 \text{ is minimum, as compared to}$$

any other value ?

2

- (b) A cyclist travels the first 5 km at a speed of 20 km/h and the next 3 km at a speed of 10 km/h. Find the average speed of the cyclist.

2

- (c) The scores of two batsmen in 5 innings are as follows :

6

| | | | | | |
|-----------|----|----|----|----|----|
| Batsman A | 10 | 30 | 50 | 70 | 90 |
| Batsman B | 10 | 20 | 30 | 40 | 50 |

Which one is more consistent ?

OR

Comment on the symmetry and peakness of the following data :

10

| | |
|---------|----|
| CI | f |
| 0 – 10 | 10 |
| 10 – 20 | 40 |
| 20 – 30 | 20 |
| 30 – 40 | 0 |
| 40 – 50 | 10 |
| 50 – 60 | 40 |
| 60 – 70 | 16 |
| 70 – 80 | 14 |

3. It is suggested that the relation between y and x is of the type $y = ab^x$. Obtain the best fit equation of the curve, for the following data : 10

| | | | | | |
|---|---|---|---|----|----|
| x | 2 | 4 | 6 | 8 | 10 |
| y | 1 | 3 | 6 | 12 | 24 |

OR

The ages of husbands and wives at the time of marriage are noted and the following data is obtained :

| Age of Husband | Age of Wife | | | | | Total |
|----------------|-------------|---------|---------|---------|---------|-------|
| | 10 - 20 | 20 - 30 | 30 - 40 | 40 - 50 | 50 - 60 | |
| 15 - 25 | 6 | 3 | - | - | - | 9 |
| 25 - 35 | 3 | 16 | 10 | - | - | 29 |
| 35 - 45 | - | 10 | 15 | 7 | - | 32 |
| 45 - 55 | - | - | 7 | 10 | 4 | 21 |
| 55 - 60 | - | - | - | 4 | 5 | 9 |
| Total | 9 | 29 | 32 | 21 | 9 | 100 |

Find the degree of linear relationship between the age of husband and wife. 10

4. A research scholar collected some data for his purpose in hand and obtained many results. Due to heavy rain, all the collected data and the results obtained were destroyed except a piece of paper where two lines of regressions were written as :

$$6x + 15y = 27, \quad 6x + 3y = 15$$

On the basis of this information, 4+2+2+2

- (i) State which line is x on y and which is y on x.
- (ii) Obtain the mean of x and y variables.
- (iii) Find $r(x, y)$.
- (iv) Find the estimated value of y for $x = 2$.

OR

Explain the difference between multiple correlation and partial correlation. Interpret the meaning of $r_{1.23}$ and $r_{12.3}$. Obtain

- (i) the correlation coefficient between x_2 and x_3 after removing the linear effect of x_1 and
- (ii) the correlation coefficient between x_1 and joint effect of x_2 and x_3 on x_1 , for the following

data : 2+2+3+3

| | | | | | | | |
|-------|----|----|----|----|----|----|----|
| x_1 | 20 | 15 | 25 | 26 | 28 | 40 | 38 |
| x_2 | 12 | 13 | 16 | 15 | 23 | 15 | 28 |
| x_3 | 13 | 15 | 12 | 16 | 14 | 18 | 14 |

5. In an observation of 100 cases it was found that the number of unmarried students was 40, the number of failed students in the examination was 55 and the number of married students who failed in the examination was 30. From the information given above find $2+2+2+2+2$
- the number of married students,
 - the number of students who passed the examination,
 - the number of married students who passed,
 - the number of unmarried students who passed,
 - the number of unmarried students who failed.

OR

A company is interested in determining the strength of association between the communicating time of their employees and the level of stress-related problems observed on job.

A study of 116 assembly line workers reveals the following data : 10

| Communicating Time | Stress | | | |
|--------------------|--------|----------|-----|-------|
| | High | Moderate | Low | Total |
| Under 20 minutes | 9 | 5 | 18 | 32 |
| 20 – 50 minutes | 17 | 8 | 28 | 53 |
| Over 50 minutes | 18 | 6 | 7 | 31 |
| Total | 44 | 19 | 53 | 116 |

Assume that $C_{\max} = 0.816$.

No. of Printed Pages : 5

MST-002

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

Term-End Examination

December, 2017

00232

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

- (i) Attempt **all** questions.
- (ii) Questions no. 2 to 5 have internal choices.
- (iii) Use of scientific calculator is allowed.
- (iv) Formulae and Tables Booklet for PGDAST is allowed.
- (v) Symbols have their usual meanings.

1. State whether the following statements are *True* or *False*. Give reasons in support of your answers.

$5 \times 2 = 10$

- (a) If D_i denotes the i^{th} decile, $i = 1, 2, \dots, 9$, then 70% observations are greater than or equal to D_7 and 30% observations are less than or equal to D_7 .

- (b) The range of the data given below is 60.

| CI | f |
|---------|----|
| 10 – 20 | 0 |
| 20 – 30 | 5 |
| 30 – 40 | 10 |
| 40 – 50 | 10 |
| 50 – 60 | 5 |
| 60 – 70 | 0 |

- (c) For the data 4, 7, 9, 5, 10, the value of A is

5.5, such that $\sum_{i=1}^5 |x_i - A|$ is minimum.

- (d) If X is measured in metres and Y is measured in kg, then the unit of $r(X, Y)$ is kg/m.
- (e) If $(AB) = 150$, $(\alpha B) = 260$, $(A\beta) = 230$, then $(A) = 410$.

2. (a) For the data 4, 7, 9, 5, 8, 15, 11, 10, 12, 18,

if we are interested in a value which is greater than or equal to 40% observations and less than or equal to 60% observations, which measure should be applied and what is the required value ?

- (b) For the data 10, 27, 40, 60, 33, 30, 10; find the square root of the mean of squares of the deviations from the mean of this data. Also find out the unit of this measure, if the unit of the given data is km.

- (c) Assume that the data given below is in ascending order :

$$4, 7, 9, x + 3, 2x, 17, 19, 20$$

Find the value of x if the median of the data is 13.5.

4+4+2

OR

Weekly wages of workers (in "000" rupees) are given in the the following table :

| Weekly wages | Frequency |
|--------------|-----------|
| 10 – 12 | 1 |
| 12 – 14 | 3 |
| 14 – 16 | 7 |
| 16 – 18 | 12 |
| 18 – 20 | 12 |
| 20 – 22 | 4 |
| 22 – 24 | 3 |

By using appropriate measures, comment on the skewness and kurtosis of the distribution of the data.

10

3. For the data

| | | | | | |
|---|---|---|---|---|----|
| x | 6 | 7 | 8 | 9 | 11 |
| y | 5 | 4 | 3 | 2 | 1 |

fit a linear curve between y (response) and x (predictor), using the principle of least squares.

Also estimate the response if predictor is 5.

10

OR

X and Y are two characteristics which cannot be measured directly. Using a criterion, scores are given to these characteristics as

| | | | | | | | |
|---|----|----|----|----|----|----|-----|
| X | 70 | 70 | 80 | 80 | 80 | 90 | 100 |
| Y | 90 | 90 | 90 | 80 | 70 | 60 | 50 |

Find the association between qualitative characteristics X and Y.

10

4. Height of fathers and their sons in inches are given below :

| | | | | | | | | |
|-------------------|----|----|----|----|----|----|----|----|
| Height of Fathers | 65 | 66 | 67 | 67 | 68 | 69 | 70 | 71 |
| Height of Sons | 66 | 68 | 65 | 69 | 74 | 73 | 72 | 70 |

What is the estimated average height of the son corresponding to the height 68.5 inches of the father ?

10

OR

From the given data in the following table, estimate the value of X_1 for $X_2 = 45$ and $X_3 = 8$:

10

| | | | | | |
|-------|---|---|---|---|---|
| X_1 | 1 | 2 | 3 | 4 | 5 |
| X_2 | 3 | 4 | 5 | 6 | 7 |
| X_3 | 4 | 5 | 6 | 7 | 8 |

5. (a) Given the following class frequencies, do you find any inconsistency in the data ?
 $(A) = 300, (B) = 150, (\alpha\beta) = 110, N = 500$
- (b) A number of school children were examined for the presence or absence of certain defects of which three chief descriptions were noted. Let A denote development defects; B, nerve sign and C, low nutrition. Given the following ultimate frequencies, find the frequencies of the class defined by the presence of the defects :

$$(ABC) = 60, (\alpha BC) = 75, (AB\gamma) = 250,$$

$$(\alpha B\gamma) = 650, (A\beta C) = 80, (\alpha\beta C) = 55,$$

$$(A\beta\gamma) = 350, (\alpha\beta\gamma) = 8200$$

5+5

OR

- (a) In a survey of 1000 children, 811 liked pizza, 752 liked chowmein and 418 liked burger; 570 liked both pizza and chowmein; 356 liked both pizza and burger; 348 liked both chowmein and burger; 297 liked all three. Test the consistency of the data.
- (b) 1660 candidates appeared for a competitive examination. Of these, 425 were successful; 252 had attended a coaching class and of them, 150 were successful. Is there any association between success and utility of the coaching class ?

5+5

No. of Printed Pages : 8

MST-002

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

Term-End Examination

June, 2018

01915

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

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

1. State whether the following statements are *True* or *False*. Give reasons in support of your answers.

$5 \times 2 = 10$

- (a) If $x_1, x_2, x_3, \dots, x_n$ and $y_1, y_2, y_3, \dots, y_n$ are the variate values of two variables X and Y , and their geometric means are G_1 and G_2 , respectively, then geometric mean of $(x_i/y_i); i = 1, 2, \dots, n$ will be (G_1/G_2) .

(b) If each value of X is divided by 2 and of Y is multiplied by 2, then b'_{yx} will be same as b_{yx} .

(c) If X and Y are two independent variables and the variables $U = X + Y$ and $V = X - Y$ then the $r(U, V) = \frac{\sigma_x^2 - \sigma_y^2}{\sigma_x^2 + \sigma_y^2}$.

(d) If $(A) = 90$, $(AB) = 40$, $N = 150$ and $(\beta) = 80$ then $(\alpha\beta) = 30$.

(e) The mean and standard deviation of a set of values are 25 and 5, respectively. If a constant value 5 is added to each value, the coefficient of variation of the new set of values is equal to 10%.

2. (a) The numbers 3·2, 5·8, 7·9 and 4·5 have frequencies Y, (Y + 2), (Y - 3) and (Y + 6), respectively. If the arithmetic mean is 4·876, find the value of Y and write the whole series.

4

- (b) The following is the distribution of age (in years) of 800 workers :

| <i>Age Group</i> | <i>No. of Workers</i> |
|------------------|-----------------------|
| 20 – 25 | 50 |
| 25 – 30 | 70 |
| 30 – 35 | 100 |
| 35 – 40 | 180 |
| 40 – 45 | 150 |
| 45 – 50 | 120 |
| 50 – 55 | 70 |
| 55 – 60 | 60 |

Find (i) Median, (ii) Quartile Deviation, and (iii) Coefficient of Quartile Deviation.

6

OR

- (a) In the given data, two frequencies are missing and its mean is found to be 1.46.

| <i>No. of Accidents (x)</i> | <i>Frequencies (f)</i> |
|-----------------------------|------------------------|
| 0 | 46 |
| 1 | ? |
| 2 | ? |
| 3 | 25 |
| 4 | 10 |
| 5 | 5 |
| Total | 200 |

Find the missing frequencies.

6

- (b) Suppose 2, 6, 9, 5, 4 are 5 observations on a variable X with standard deviation 2.32. A new variable Y is obtained by multiplying each observation on X by 3. Further, another variable Z is obtained by dividing each observation on X by 2. Then find the effect of these changes in obtained variables in terms of standard deviation and explain the results.

4

3. (a) Suppose X and Y are the two variables having the correlation coefficient 0.85. The following are the values they have :

| X | Y |
|----|----|
| 10 | 40 |
| 30 | 30 |
| 50 | 70 |
| 60 | 80 |

If two new variables X' and Y' are obtained by adding 50 to each value of X and 100 to each value of Y, respectively, calculate the correlation coefficient between X' and Y' using the above data. Also compare the results.

5

- (b) Calculate the intensity of relation between the ranks of the marks, obtained by 10 candidates in an interview allotted by two experts, given as follows :

5

| <i>Expert-1</i> | <i>Expert-2</i> |
|-----------------|-----------------|
| 48 | 13 |
| 33 | 13 |
| 40 | 24 |
| 09 | 06 |
| 16 | 15 |
| 16 | 04 |
| 65 | 20 |
| 24 | 09 |
| 16 | 06 |
| 57 | 19 |

OR

- (a) The value of Spearman's rank correlation coefficient of a set of non-repeating values was found to be $\frac{2}{3}$. The sum of the squares of difference between the corresponding ranks was 55. Find the number of pairs.

4

- (b) Calculate Karl Pearson's coefficient of correlation between X and Y for the following data :

$$N = 12, \Sigma X = 120, \Sigma Y = 130, \Sigma (X - 8)^2 = 50, \\ \Sigma (Y - 10)^2 = 200 \text{ and } \Sigma (X - 8)(Y - 10) = 50.$$

6

4. (a) The following table shows the information as :

| Statistical Measures | Advertisement Expenditure (X) (₹ Lakhs) | Sales (Y) (₹ Lakhs) |
|----------------------|--|------------------------|
| Mean | 20 | 100 |
| Standard Deviation | 03 | 12 |

$r(X, Y) = 0.8$. Then find

- (i) the expected advertising expenditure of the company if sale is ₹ 125 lakhs, and
- (ii) the expected sales of the company if the advertising expenditure is ₹ 32 lakhs. 6
- (b) Given the following data : 4

$$r_{12} = 0.8, r_{13} = 0.6 \text{ and } r_{23} = 0.4$$

then find

(i) $r_{12.3}$

(ii) $r_{13.2}$

(iii) $r_{23.1}$

(iv) $R_{1.23}$

OR

- (a) In a statistical study relating to the prices (in ₹) of two shares, X and Y, the following two regression lines were found as

$$8X - 10Y + 70 = 0$$

$$20X - 9Y - 65 = 0$$

The standard deviation of X = 3,

then find

(i) the values of \bar{X} and \bar{Y} ,

(ii) $r(X, Y)$, and

(iii) standard deviation of Y.

7

- (b) Suppose a computer has found for a given set of values of X_1, X_2 and X_3 : $r_{12} = 0.90$, $r_{13} = 0.30$ and $r_{23} = 0.70$. Examine whether these computations are error free.

3

5. (a) An investigation of 23713 households was made in an urban and rural mixed locality. Of these 1618 were farmers, 2015 well to do and 770 families were having at least one graduate. Of these graduate families 335 were those of farmers and 428 were well to do; also 587 well to do families were those of farmers and out of them only 156 were having at least one graduate. Obtain all the ultimate class frequencies.

6

(b) Can vaccination be regarded as a preventive measure for smallpox from the given data :

- (i) Of 1482 persons in a locality exposed to smallpox, 368 in all were attacked, and
- (ii) Of 1482 persons, 343 had been vaccinated and of these only 35 were attacked.

4

OR

(a) Test whether there is any association between the heights of parents and that of their offsprings for the following data :

7

| Offsprings Heights | Parents Heights | | | | Total |
|--------------------|-----------------|-------------|-------------|------------|-------|
| | Above 5'8" | 5'6" – 5'8" | 5'3" – 5'5" | Below 5'3" | |
| Above 5'8" | 20 | 30 | 20 | 02 | 72 |
| 5'6" – 5'8" | 14 | 125 | 85 | 12 | 236 |
| 5'3" – 5'5" | 03 | 140 | 165 | 125 | 433 |
| Below 5'3" | 03 | 37 | 68 | 151 | 259 |
| Total | 40 | 332 | 338 | 290 | 1000 |

(b) 50% of items have characteristics A and B both, 35% have A but not B, 25% have B but not A. Show that there must be some misprints in this report.

3

No. of Printed Pages : 7

MST-002

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

Term-End Examination

00042

December, 2018

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

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

1. State whether the following statements are *True* or *False*. Give reasons in support of your answers.

$5 \times 2 = 10$

- (a) There are $(n + 1)$ observations in a sample. If \bar{x}_1 is the mean of the first n observations and \bar{x}_2 is the mean of the last n observations, then $\bar{x}_2 = \bar{x}_1$.

- (b) If the two lines of regression are coincident, the relation between the regression coefficients is $b_{YX} \cdot b_{XY} = 1$.
- (c) If X and Y are two independent variables with $\text{Var}(X) = 25$ and $\text{Var}(Y) = 15$, then correlation coefficient between $V = X + Y$ and $V = X - Y$ is equal to 0.25.
- (d) If $(A) = 90$, $(AB) = 40$, $N = 150$ and $(\beta) = 80$, then $(\alpha B) = 50$.
- (e) The mean and standard deviation of X are 25 and 5, respectively. If a constant 10 is multiplied to each value of X, the CV remains unchanged.
2. (a) The price of a commodity increased by 5% from 1999 to 2000, 8% from 2000 to 2001 and 77% from 2001 and 2002. The average price increase from 1999 to 2002 is quoted as 26% and not 30%. Why? 4

- (b) The mean and standard deviation of 20 items are found to be 10 and 2, respectively. At the time of checking, it was observed that one item 8 was incorrect. Find the mean and standard deviation if (i) the value is omitted, (ii) it is replaced by 12. 6

OR

- (a) The runs scored by two batsmen in eight matches are as follows :

| <i>Batsman A</i> | <i>Batsman B</i> |
|------------------|------------------|
| 27 | 0 |
| 16 | 100 |
| 39 | 80 |
| 45 | 5 |
| 101 | 60 |
| 80 | 40 |
| 40 | 10 |
| 52 | 121 |

Who is a better run scorer ? Also find which of the two batsmen is more consistent in scoring.

7

- (b) The geometric mean of 10 observations on a certain variable was calculated as 16.2. It was later discovered that one of the observations was wrongly read as 12.9, in fact it was 21.9. Apply the appropriate correction and calculate correct geometric mean.

3

3. (a) Fit an equation of the form $y = ab^x$ to the following data :

5

| | | | | | |
|---|-----|-------|-------|-------|-------|
| x | 2 | 3 | 4 | 5 | 6 |
| y | 144 | 172.8 | 207.4 | 248.5 | 298.5 |

- (b) Suppose X and Y are the two variables having the correlation 0.85 on the following values :

| | | | | |
|---|----|----|----|----|
| X | 10 | 30 | 50 | 60 |
| Y | 40 | 30 | 70 | 80 |

Then two new variables X' and Y' are obtained by multiplying each observation of X and Y by 3 and 4, respectively. Find the correlation between X' and Y' using the given data. Also, interpret the result.

5

OR

- (a) The following data relating to the heights of the fathers and their sons (in inches) are as follows :

| <i>Height of Fathers</i> | <i>Height of Sons</i> |
|--------------------------|-----------------------|
| 68 | 65 |
| 68 | 64 |
| 69 | 67 |
| 72 | 69 |
| 65 | 64 |
| 59 | 60 |
| 62 | 59 |
| 67 | 68 |
| 61 | 60 |
| 71 | 64 |

What conclusion do you draw from this data in regard to the relation between the heights of fathers and sons ?

5

- (b) The coefficient of rank correlation of the marks obtained by 10 students in Statistics and Accountancy was found to be 0.2. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 9 instead of 7. Find the correct value of coefficient of rank correlation. 3
- (c) Explain the method of least squares. 2
4. (a) The equation of two regression lines are given as :
- $$Y = X + 5 \text{ and } 16X = 9Y + 95$$
- If standard deviation of Y is 4, then find the values of (i) Means of X and Y, (ii) $r(X, Y)$, and (iii) Standard deviation of X. 7
- (b) Examine if it is possible to get the following values of $r_{12} = 0.91$, $r_{13} = 0.33$ and $r_{23} = 0.81$ from some set of the experimental data. 3

OR

For the following data given in the table below :

| | | | | | | |
|---------|----|----|----|----|----|----|
| X_1 : | 4 | 6 | 7 | 19 | 13 | 15 |
| X_2 : | 15 | 12 | 8 | 6 | 4 | 3 |
| X_3 : | 30 | 24 | 20 | 14 | 10 | 4 |

find the regression equation of X_1 on X_2 and X_3 .
Also estimate the value of X_1 for $X_2 = 6$ and $X_3 = 8$.

10

5. (a) If in an urban district 817 per thousand of the women between 20 and 25 years of age were noted as occupied at a census and 263 per thousand were noted as married or widowed, what is the lowest proportion per thousand of the married or widowed that must have been occupied ?

3

(b) For 1000 school boys, test whether there is any association between the General ability and Mathematical ability from the following data :

7

| Mathematical Ability | General Ability | | | Total |
|----------------------|-----------------|------|------|-------|
| | Good | Fair | Poor | |
| Good | 44 | 22 | 04 | 70 |
| Fair | 265 | 257 | 178 | 700 |
| Poor | 41 | 91 | 98 | 230 |
| Total | 350 | 370 | 280 | 1000 |

OR

(a) A fundamental set of class frequencies are given below :

$$N = 275, (A) = 132, (B) = 112, (C) = 129$$

$$(AB\gamma) = 27, (A\beta C) = 39, (\alpha BC) = 37, (ABC) = 20.$$

Calculate all (i) positive, and (ii) ultimate class frequencies.

5

- (b) In a college of 1000 students, equal number of students came from each of higher and lower income groups. The number of boys in these groups was 100 and 400, respectively. Calculate the coefficient of association between gender and income group.

5



www.ignouassignmentguru.com

631081

No. of Printed Pages : 7

MST-002

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

Term-End Examination

June, 2019

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 Hours

Maximum Marks : 50

Note : Question No. 1 is compulsory. Attempt any four questions out of the remaining Question Nos. 2 to 7. Use of scientific calculator (non-programmable) is allowed. Use of formulae and statistical tables booklet for PGDAST programme is allowed. Symbols have their usual meanings.

-
-
1. State whether the following statements are True or False. Give reasons in support of your answers : 2×5=10
- (a) In a frequency curve of scores, the mode was found to be higher than the mean.

[2]

MST-002

This shows that the distribution is negatively skewed.

- (b) If variance of X is 3 and $Y = 5X - 3$, then variance of Y is 12.
- (c) If $b_{xy} = -0.9$, $b_{yx} = -0.4$, then $r(x, y) = +0.6$.
- (d) If we have 15 attributes, then total number of classes of order 3 is 3660.
- (e) If $10X - Y + 5 = 0$ is line of regression of X on Y, then for $X = 3$, the estimated value 35 of Y is the best estimated value.

2. (a) Comment on shape (symmetry) and flatness of the data given below by calculating appropriate descriptive statistics measures : 8

3, 13, 2, 10, 7, 11, 18, 1, 11, 10, 5, 7, 9, 8, 20, 9, 12, 2, 17, 5.

- (b) Dimensions of a cuboid are 40 cm, 80 cm, 20 cm. Using a descriptive statistical tool, find the edge of a cube which enclose the same volume as that of the cuboid. 2

[3]

3. (a) X and Y are associated by the relation $Y = aX^b$. Estimate values of a and b to obtain best fit curve from the following information: 5

| X | Y |
|----|----|
| 6 | 9 |
| 2 | 11 |
| 10 | 12 |
| 5 | 8 |
| 8 | 7 |

- (b) Using suitable measure of descriptive statistics, find strength and direction of linear relationship between X and Y from the following available information: 5

| X | Y |
|----|----|
| 10 | 90 |
| 20 | 85 |
| 30 | 80 |
| 40 | 60 |
| 50 | 45 |

[4]

MST-002

4. (a) On the basis of ranks given in the following table, to what extent the knowledge of the students in statistics and mathematics is related : 2

| Rank in Statistics | Rank in Mathematics |
|--------------------|---------------------|
| 1 | 2 |
| 2 | 4 |
| 3 | 1 |
| 4 | 5 |
| 5 | 3 |
| 6 | 8 |
| 7 | 7 |
| 8 | 6 |

- (b) From the data given below, study the association between temperament of brothers and sisters : 8

| Temperament of Brothers | Temperament of Sisters | | | |
|-------------------------|------------------------|--------------|--------|-------|
| | Quick | Good Natured | Sullen | Total |
| Quick | 850 | 571 | 580 | 2001 |
| Good Natured | 618 | 593 | 455 | 1666 |
| Sullen | 540 | 456 | 457 | 1453 |
| Total | 2008 | 1620 | 1492 | 5120 |

(A-18)

5. From the table given below find out : 10

(i) Least square regression equation of X_1 on X_2 and X_3 .

(ii) Estimated value of X_1 for $X_2 = 45$ and $X_3 = 8$.

| X_1 | X_2 | X_3 |
|-------|-------|-------|
| 1 | 3 | 4 |
| 2 | 4 | 5 |
| 3 | 5 | 6 |
| 4 | 6 | 7 |
| 5 | 7 | 8 |

6. (a) Among adult population of a certain town, 50% of the population is male, 60% are wage earners and 50% are 45 years of age or over. 10% of the males are not wage earners and 40% of the males are under 45. Can we infer anything about what percentage of the population of 45 years of age or over are wage earners? 5

(b) In an examination, at which 600 candidates appeared, boys outnumbered girls by 16% of all candidates. Number of passed candidates exceeded the number of failed candidates by 310. Boys failing in the examination numbered 88. Find the Yule's coefficient of association between sex and result in the examination. 5

7. (a) For the data given below : 4
- 4, 7, 20, 8, 35, 15, 21, 19, 13, 22, 25, 29,
25, 23, 8, 22, 18, 11, 33, 10
calculate :

(i) Average of squared deviations from its mean.

(ii) A value which divides the data into two equal halves.

(b) The line of regression of marks in Statistics (X) on marks of Economics (Y) for a class of 50 students is :

$$3Y - 5X + 180 = 0.$$

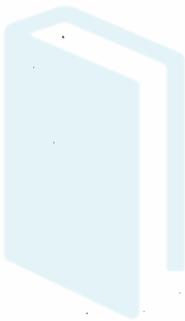
[7]

MST-002

Mean marks in Economics is 44 and the variance of marks in Statistics is $\frac{9}{16}$ of the variance of marks in Economics. 6

Find :

- (i) mean marks in Statistics
- (ii) coefficient of correlation between the marks of two subjects.



ignou
ASSIGNMENT GURU

www.ignouassignmentguru.com

MST-002

3,000

(A-18)

No. of Printed Pages : 3

MST-002

POST GRADUATE DIPLOMA IN APPLIED STATISTICS (PGDAST)

Term-End Examination

December, 2019

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note : (i) Question No. 1 is compulsory.

(ii) Attempt any four questions from the remaining questions No. 2 to 7.

(iii) Use of Scientific Calculator (Non-programmable) is allowed.

(iv) Use of Formulae and Statistical Tables Booklet for PGDAST programme is allowed.

(v) Symbols have their usual meanings.

1. State whether the following statements are True or False. Give reasons in support of your answers. 5x2=10
- (a) Two distributions, with the same mean, standard deviation and coefficient of skewness, must have the same peakedness.
- (b) In simple regression analysis, the quantity that gives the amount by which Y (response variable) changes for a unit change in X (regressor variable) is called the Y intercept of the regression line.
- (c) The amount of variability in dependent variable that is explained by the independent variable is called correlation.
- (d) There can be two attributes A and B such that $(AB) > (A)$, where A and B have been observed from the same population.
- (e) We are given a regression equation between two variables. From this equation we can estimate whether the association is linear or non-linear.
2. Comment on the shape and peakedness of the data given below : 10
48, 55, 48, 66, 39, 50, 54, 49, 51, 39, 51, 67, 47, 35, 66
3. (a) By using the method of Least Squares, find estimated values of a and b, where X and Y have a relationship of the type $Y = ab^X$ from the following data : 8
- | | | | | | |
|---|---|---|---|----|----|
| X | 2 | 4 | 6 | 8 | 10 |
| Y | 1 | 3 | 6 | 12 | 24 |
- (b) Explain what information is obtained by correlation and regression analysis. 2

4. For the given data, find out : 10
- (a) Regression equation of x_1 on x_2 and x_3 .
- (b) Estimate the value of x_1 for $x_2=6$ and $x_3=8$.

| | | | | |
|-------|---|---|----|----|
| x_1 | 2 | 6 | 8 | 10 |
| x_2 | 4 | 5 | 9 | 12 |
| x_3 | 4 | 6 | 10 | 12 |

5. (a) In a Girl's High School, there were 200 students. Their results in the quarterly, half yearly and annual examinations were as follows : 5
- 85 passed the quarterly examination.
 80 passed the half yearly examination.
 94 passed the annual examination.
 28 passed all the three and 40 failed in all the three.
 25 passed the first two and failed in the annual examination.
 43 failed the first two but passed the annual examination.

Find how many students passed atleast two examinations.

- (b) A company is interested in determining the strength of association between the communicating time of their employees and the level of stress-related problem observed on job. A study of 116 assembly line workers reveals the following information : 5

| Communicating time (in minutes) ↓ | Stress | | | |
|---|--------|----------|-----|-------|
| | High | Moderate | Low | Total |
| Under 20 | 9 | 5 | 18 | 32 |
| 20 - 50 | 17 | 8 | 28 | 53 |
| Over 50 | 18 | 6 | 7 | 31 |
| Total | 44 | 19 | 53 | 116 |

Find strength of association between above considered attributes.

6. (a) For the following data : 5
- 11, 9, 14, 10, 9, 14, 6, 13, 12, 6, 7, 13, 7, 17, 4, 10, 8, 10, 15, 12, 4
- Find A, B and C such that

(i) $\frac{1}{21} \sum_{i=1}^{21} (x_i - A) = 0$

(ii) $\frac{1}{21} \sum_{i=1}^{21} |x_i - B|$ is minimum

(iii) $\frac{1}{21} \sum_{i=1}^{21} (x_i - C)^2$ is least

- (b) In a competitive examination, 200 graduates appeared. Following facts were noted : 5

No. of boys = 139

No. of Science graduate girls who failed to qualify for interview = 25

No. of Arts graduate girls who qualified for interview = 30

No. of Arts graduate girls failed to qualify for interview = 18

Test the consistency of the data.

7. (a) Compute the correlation ratio for the following bivariate data : 5

| $x \backslash y$ | 5 | 10 | 15 | 20 | 25 |
|------------------|---|----|----|----|----|
| 5 | 8 | 8 | 4 | 0 | 0 |
| 10 | 7 | 15 | 15 | 1 | 0 |
| 20 | 0 | 6 | 1 | 15 | 11 |
| 25 | 0 | 0 | 5 | 10 | 8 |

- (b) From the following data, obtain the correlation coefficient between (i) x_1 and x_3 after removing linear effect of x_2 on them and (ii) x_1 and joint effect of x_2 and x_3 . 5
- $r_{12} = 0.30$, $r_{13} = -0.26$ and $r_{23} = -0.41$.

www.ignouassignmentguru.com

No. of Printed Pages : 7

MST-002

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

Term-End Examination

June, 2020

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 Hours

Maximum Marks : 50

Note : (i) Question No. 1 is compulsory.

(ii) Attempt any four questions from the remaining (Question Nos. 2 to 7).

(iii) Use of scientific calculator (non-programmable) is allowed.

(iv) Use of formulae and statistical tables booklet for PGDAST programme is allowed.

(v) Symbols have their usual meanings.

[2]

MST-002

1. State whether the following statements are True or False. Give reasons in support of your answers : 2 each

(a) Average rainfall of a city from Monday to Saturday is 0.3 inches. Due to heavy rainfall of 1.7 inches on Sunday the average rainfall increased to 0.5 inches.

(b) If 25% of the items are less than 20 and 25% are more than 40, then quartile deviation is 20.

(c) If $X' = 2X$ and $Y' = Y + 3$ and $r(X, Y) = 0.80$, then $r(X', Y') = 0.40$.

(d) If $b_{XY} = -0.9$ and $b_{YX} = -0.4$ then $r(X, Y) = -0.6$.

(e) If $(AB) = 150$, $(\alpha B) = 260$, $(A\beta) = 230$, then $(B) = 490$.

2. (a) The frequency distribution of marks obtained by 55 students in Statistics is as follows :

7

| Class | No. of Students |
|-------|-----------------|
| 10—14 | 03 |
| 15—19 | 07 |
| 20—24 | 16 |
| 25—29 | 12 |
| 30—34 | 09 |
| 35—39 | 05 |
| 40—44 | 03 |

Calculate :

- (i) 1st quartile
- (ii) IIIrd quartile
- (iii) 8th decile
- (iv) 75th percentile
- (v) Coefficient of Quartile Deviation.
- (b) The mean and standard deviation of a variable of 100 items were found to be 60

P. T. O.

and 10, respectively. At the time of calculations two items were wrongly taken as 5 and 45 instead of 30 and 20. Calculate the corrected mean and standard deviation.

3

3. Calculate Karl Pearson's co-efficient of correlation between X and Y from the following bivariate frequency distribution of 140 pairs of X and Y :

10

| Y \ X | 10—20 | 20—30 | 30—40 | 40—50 |
|-------|-------|-------|-------|-------|
| 10—20 | 20 | 26 | — | — |
| 20—30 | 8 | 14 | 37 | — |
| 30—40 | — | 4 | 18 | 3 |
| 40—50 | — | — | 4 | 6 |

4. We are given the following data on three variables X_1 , X_2 and X_3 : 10

| X_1 | X_2 | X_3 |
|-------|-------|-------|
| 64 | 57 | 08 |
| 71 | 59 | 10 |
| 53 | 49 | 06 |
| 67 | 62 | 11 |
| 55 | 51 | 08 |
| 58 | 50 | 07 |
| 77 | 55 | 10 |
| 57 | 48 | 09 |
| 56 | 52 | 10 |
| 51 | 42 | 06 |
| 76 | 61 | 12 |
| 68 | 57 | 09 |

Find :

- (i) the least square regression equation of X_1 on X_2 and X_3 .
- (ii) estimate value of X_1 for given values of $X_2 = 54$ and $X_3 = 9$.

5. The following table gives the distribution of students and also of regular players among them, according to age in completed : 10

| Years | Age of Minor | | | Age of Major | | |
|-----------------|--------------|-----|-----|--------------|-----|----|
| | 15 | 16 | 17 | 18 | 19 | 20 |
| Age in years | | | | | | |
| No. of Students | 250 | 200 | 150 | 120 | 100 | 80 |
| Regular Players | 200 | 150 | 90 | 48 | 30 | 12 |

Calculate the coefficient of association between majority and playing habit, on the assumption that majority is attained in 18th year.

6. (a) For a distribution, the mean is 10, variance is 16, γ_1 (gamma 1) is +1 and β_2 (beta 2) is 4. Obtain the first four moments about the origin.

[7]

MST-002

(b) The coefficient of rank correlation of the marks obtained by 10 students in Statistics and Accountancy was found to be 0.4. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 5 instead of 6. Find the corrected value of coefficient of rank correlation. 4

7. (a) For a bivariate data, the equation of regression lines $4X - 7 = 35$ are $Y = 4X - 35$ and $9X - Y = 135$, find : 6

(i) the mean value of X and Y.

(ii) the value of $r(X, Y)$.

(iii) the value of σ_X if $\sigma_Y = 12$.

(b) Check whether A and B are independent, positively associated or negatively associated in the following case : 4

$(AB) = 256$, $(\alpha B) = 768$, $(A\beta) = 48$ and $(\alpha\beta) = 144$.

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

Term-End Examination

February, 2021

MST-002 : DESCRIPTIVE STATISTICS

Time : 3 hours

Maximum Marks : 50

Note :

- (i) Question no. 1 is **compulsory**.
- (ii) Attempt any **four** questions from the remaining questions no. 2 to 7.
- (ii) Use of scientific (non-programmable) calculator is allowed.
- (iii) Use of Formulae and Statistical Tables Booklet for PGDAST is allowed.
- (iv) Symbols have their usual meanings.

1. State whether the following statements are *True* or *False*. Give reasons in support of your answers.

$5 \times 2 = 10$

- (a) If the arithmetic mean of the numbers 3·2, 5·8, 7·9 and 4·5 with their corresponding frequencies Y , $(Y + 2)$, $(Y - 3)$ and $Y + 6$ is 4·876, then the value of Y is 5.

- (b) The sum of squares of deviations for 10 observations taken from their mean 50 is 250. The coefficient of variation is 10%.
- (c) If the sum of the product of deviations of X and Y values from their respective means is zero, then the $r(x, y)$ will be -1 .
- (d) If $b_{xy} = -\frac{4}{3}$ and $b_{yx} = -\frac{1}{12}$, the value of r will be $+\frac{1}{3}$.
- (e) The data is consistent if $N = 1000$, $(A) = 600$, $(B) = 500$ and $(AB) = 50$.

2. (a) The following is the distribution of age of 80 workers :

| Age Group | No. of Workers |
|-----------|----------------|
| 20 – 25 | 5 |
| 25 – 30 | 7 |
| 30 – 35 | 10 |
| 35 – 40 | 18 |
| 40 – 45 | 15 |
| 45 – 50 | 12 |
| 50 – 55 | 7 |
| 55 – 60 | 6 |

Find Quartile deviation.

- (b) The number of runs scored by two batsmen in consecutive eight matches are given below :

| | | | | | | | | |
|-------------|----|-----|----|----|-----|----|----|-----|
| Batsman A : | 27 | 16 | 39 | 45 | 101 | 80 | 40 | 52 |
| Batsman B : | 0 | 100 | 80 | 5 | 60 | 40 | 10 | 121 |

Find who is a better run scorer. Also find which of the two batsmen is more consistent in scoring.

7

3. With 10 observations, each on two variables X and Y, the following data were observed :

$$\bar{X} = 12, \sigma_x = 3, \bar{Y} = 15, \sigma_y = 4 \text{ and } r = 0.5$$

However, on subsequent verification, it was found that one value of X (=15) and one value of Y (=13) were wrongly taken as 16 and 18 respectively. Find the correct value of correlation coefficient.

10

4. Find the multiple linear regression equation of X_1 on X_2 and X_3 from the data relating to three variables given below :

| | | | | | | |
|---------|----|----|----|----|----|----|
| X_1 : | 4 | 6 | 7 | 9 | 13 | 15 |
| X_2 : | 15 | 12 | 8 | 6 | 4 | 3 |
| X_3 : | 30 | 24 | 20 | 14 | 10 | 4 |

Also estimate the best value of X_1 for $X_2 = 4$ and $X_3 = 10$.

10

5. 800 candidates comprising both boys and girls appeared in an examination. The boys outnumbered the girls by 15% of the total. The number of candidates who passed exceeded the number failed by 480. Equal number of boys and girls failed in the examination. Prepare a 2×2 table and find the coefficient of association. 10

6. (a) The mean annual salary of all employees in a company is ₹ 25,000. The mean salary of male and female employees is ₹ 27,000 and ₹ 17,000 respectively. Find the percentage of males and females employed by the company. 3

(b) For a bivariate data, the two regression equations are $8Y = 6X$ and $Y = 3X$. Find (i) means of X and Y, (ii) $r(X,Y)$, and (iii) value of σ_y if value of $\sigma_x = 4$. 7

7. (a) Calculate the coefficient of rank correlation for the following data : 5

| | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| X : | 80 | 78 | 75 | 75 | 68 | 57 | 60 | 59 |
| Y : | 110 | 111 | 114 | 114 | 114 | 116 | 115 | 117 |

(b) Check whether A and B are independent, positively associated or negatively associated in case of the following data : 5

(A) = 490, (AB) = 294, (α) = 570 and ($\alpha\beta$) = 380.