# U.G. 1st Semester Examination - 2023

## STATISTICS

### [MINOR]

**Course Code : BSTAMEA11T**

**Course Title : Descriptive Statistics**

**[NEP-20]**

Full Marks : 40 Time : 2 Hours

*The figures in the right-hand margin indicate marks.*

*Candidates are required to give their answers in their own words as far as practicable.*

1. Answer any **ten** questions: 1×10=10

a) What are the types of 'Ogive'?

b) What is meant by population and sample?

c) Define class frequency and total frequency.

d) Point out two advantages of cumulative frequency distribution.

e) What is the relation between mean, median and mode?

f) Why 'quartile deviation' is known as 'semi-interquartile range'?

g) If the variables are decreased or increased by the same proportion, what will be the effect on standard deviation?

h) The first two moments of a distribution about the value 4 are –1.5 and 2.7. Find the moment about zero.

i) Explain the term 'Kurtosis' used in connection with the frequency distribution of a continuous variable.

j) What is a Scatter Diagram?

k) If x and y satisfy the relationship $y = -5+6x$, what is the product-moment correlation?

l) Define classmark of a distribution.

m) What are regression lines?

n) What are the 'quantitative' and 'qualitative' data?

o) What is meant by 'association of attributes'?

2. Answer any five questions : $2 \times 5 = 10$

a) Why is the Lorenz curve convex?

b) The arithmetic mean and variance of n values $x_1, x_2, \ldots\ldots, x_n$ are 0 and $\sigma^2$ respectively. Also y is a variable defined by $y = x^2$. Prove that $\bar{y} = \sigma^2$.

c) The numbers 3.2, 5.8, 7.9 and 4.5 have frequencies $x, (x+2), (x-3)$ and $(x+6)$ respectively. If the arithmetic mean is 4.876, find the value of x.

d) Prove that the co-efficient of correlation is the geometric mean of the co-efficients of regression.

e) Write down two advantages and disadvantages of mode.

f) What is 'histogram' for a frequency distribution?

g) Illustrate : Box plot.

h) Show that the central moments are invariant under change of origin, but not under change of scale.

3. Answer any **two** questions : 5×2=10

a) Calculate the coefficient of skewness based on quartiles from the following :

| | |
|---|---|
| More than 0 | 5474 |
| More than 10 | 5426 |
| More than 20 | 5259 |
| More than 30 | 5023 |
| More than 40 | 4475 |
| More than 50 | 3712 |
| More than 60 | 2718 |
| More than 70 | 1406 |
| More than 80 | 764 |
| More than 90 | 370 |
| More than 100 | 160 |
| More than 110 | 39 |

b) Find the equation of the line of regression of x on y for the following data :

| x | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 |
|---|---|---|---|---|---|---|---|
| y | 5.3 | 5.7 | 6.3 | 7.2 | 8.2 | 8.7 | 8.4 |

c) Draw a cumulative frequency graph and estimate the number of persons between the ages 30–32 from the following table :

| Age | 20-25 | 25-30 | 30-35 | 35-40 |
|---|---|---|---|---|
| No. of persons | 50 | 70 | 100 | 180 |

| Age | 40-45 | 45-50 | 50-55 | 55-60 |
|---|---|---|---|---|
| No. of persons | 150 | 120 | 70 | 59 |

CF- 799

4+1

4. Answer any **one** question : 10×1=10

a) i) Calculate Pearson's coefficient of correlation from the following data :

| X | 43 | 44 | 46 | 40 | 44 | 42 | 45 | 42 | 38 | 40 | 42 | 57 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 29 | 31 | 19 | 18 | 19 | 27 | 27 | 29 | 41 | 30 | 26 | 10 |

ii) If $y_i = \frac{x_i - c}{d}$ (i=1,2,.....,n) where c and d are constants, prove that $\bar{x} = c + d\bar{y}$.

5+5

b) i) In a certain examination 10 students obtained the following marks in Mathematics and Statistics.

Find Spearman's rank correlation coefficient.

| Roll No. of Students | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks in Mathematics | 85 | 42 | 75 | 68 | 45 | 63 | 60 | 90 | 58 | 62 |
| Marks in Statistics | 90 | 30 | 82 | 45 | 32 | 65 | 40 | 88 | 66 | 73 |

ii) Fit a straight line by the method of least squares to the following data :

| Age | 21 | 42 | 38 | 64 | 53 | 61 | 47 |
|---|---|---|---|---|---|---|---|
| Absence (no of days) | 4 | 14 | 10 | 38 | 19 | 34 | 17 |

Estimate the probable number of days absent when age is 40 years. 5+(4+1)

c) i) If $n_1, \sigma_1$ and $\bar{x}_1$ be the number of observations, standard deviation and mean of a set of observations, and $n_2, \sigma_2$ and $\bar{x}_2$ be those for a second set of observations, prove that the standard deviation $\sigma$ of the combined set of $(n_1+n_2)$ observation is given by,

Find Spearman's rank correlation coefficient.

| Roll No. of Students | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks in Mathematics | 85 | 42 | 75 | 68 | 45 | 63 | 60 | 90 | 58 | 62 |
| Marks in Statistics | 90 | 30 | 82 | 45 | 32 | 65 | 40 | 88 | 66 | 73 |

ii) Fit a straight line by the method of least squares to the following data :

| Age | 21 | 42 | 38 | 64 | 53 | 61 | 47 |
|---|---|---|---|---|---|---|---|
| Absence (no of days) | 4 | 14 | 10 | 38 | 19 | 34 | 17 |

Estimate the probable number of days absent when age is 40 years. 5+(4+1)

c) i) If $n_1, \sigma_1$ and $\bar{x}_1$ be the number of observations, standard deviation and mean of a set of observations, and $n_2, \sigma_2$ and $\bar{x}_2$ be those for a second set of observations, prove that the standard deviation $\sigma$ of the combined set of $(n_1+n_2)$ observation is given by,

$$(n_1+n_2)\sigma^2 = n_1\sigma_1^2 + n_2\sigma_2^2 + n_1d_1^2 + n_2d_2^2$$

where, $d_1 = \bar{x}_1 - m$, $d_2 = \bar{x}_2 - m$ and $m = (n_1\bar{x}_1 + n_2\bar{x}_2)/(n_1+n_2)$.

ii) The expenditure of 1000 families is given below :

| Expenditure (Rs.) | 40-59 | 60-79 | 80-99 | 100-119 | 120-139 |
|---|---|---|---|---|---|
| No. of Families | 50 | ? | 500 | ? | 50 |

The median and mean for the distribution are both Rs.87.50. Calculate the missing frequencies. 6+4

---