

## SET II SOLUTION

- Q 1A-1  
 1 False - Ratio scale  
 2 False - Five point summary  
 3 False - Range  
 4 True  
 5 False Mean > Median > Mode.

1b-1 Qualitative data which cannot be expressed numerically. Example. Number of students having green eyes.  
 2 Classification of data into two distinct classes is known as dichotomous classification

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - cf\right) \times h_1}{f} \quad Q_3 = l_3 + \frac{\left(\frac{3N}{4} - cf\right) \times h_3}{f}$$

$$SKB = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$MD = \frac{\sum (x - A)^2}{n} \quad MD \text{ from Mode} = \frac{\sum |x - A|}{n} \quad \text{Coefficient of MD} = \frac{MD \text{ from } A}{A} \quad \text{where } A = \text{mode}$$

Q 2a Questionnaire - Method of collecting primary data. - Sent by  
 Schedule - method of collecting primary data through  
 enumerators. Data is personally collected.  
 Questionnaire is a slow method as compared to  
 Schedule which is faster & timely method

b Data already collected by some other agency is known as Secondary data. Published & unpublished sources of collecting secondary data.

c Data representing quality or characteristic.  
 Positive Association -  $(AB) > \frac{(A)(B)}{N}$

Negative Association -  $AB < \frac{(A)(B)}{N}$

Independent -  $AB = \frac{(A)(B)}{N}$

d  $r = -0.594$ ; Quinine is useful in checking malaria.

2

(2)

Quantiles are values which divide the data into a number of parts. They are also known as partition values. Three quartiles are  $Q_1, Q_2, Q_3$  & Deciles & Percentiles.

$$Q_i = l_1 + \frac{\frac{CN}{4} - CF}{f} \times (l_2 - l_1)$$

Similarly  $D_i$  &  $P_i$   $i$ th Decile & Percentile with all terms in usual notation.

b Geometric Mean =  $G = (x_1 x_2 \dots x_n)^{1/n}$

For grouped data =  $G = A \log \left[ \frac{\sum f \log x}{N} \right]$

Harmonic Mean =  $H = \frac{n}{\sum (1/x)} = \frac{\sum f}{\sum (f/x)}$  for grouped data.

$$A = \frac{a+b}{2} \quad G = \sqrt{ab} \quad H = \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2ab}{a+b}$$

$$A \times H = ab = G^2 \quad G = \sqrt{ab} = \sqrt{A \times H}$$

Geometric Mean used in data where weights are assigned as per importance of items & Harmonic mean is used for averaging speed and prices.

c  $AM = 51.18$  Median = 51 Mode = 51

d Mode for grouped frequency distribution =

$$l_1 + \frac{(l_2 - l_1)(f_1 - f_0)}{(f_1 - f_0) + (f_1 - f_2)}$$

- All terms in normal notations

Merits - Easy to understand, simple to calculate

Not affected by extreme values

Can be calculated for open end classes

Can be determined graphically

Demerit - Not rigidly defined

Not based on all observations

Not capable of further mathematical treatment

Indeterminate if Modal class is at extreme.

2

3

Q4 a Mean deviation is A.M of the absolute deviation of the observations from any suitable constant A.

$$MD = \frac{1}{n} \sum |x-A| = \frac{1}{N} \sum f|x-A| \text{ for grouped data.}$$

$$\text{Relative Measure} = \text{Coefficient of M.D from A} = \frac{\text{M.D from A}}{A}$$

Significance - Range of G.D not based on all observations & don't A can be mean median or mode. Indicate scatter of values.

b.  $SDx = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$   $u_i = x_i - a$

Find  $SDu$  &  $SDx = c \cdot SDu$

c Kurtosis - Degree of flattening of the curve. Figure showing platykurtic, mesokurtic and leptokurtic curve. Explaining characteristics of each.

d Positive skewness : Mean > Median > Mode

$$Q_3 - Q_2 > Q_2 - Q_1 \text{ \& } \mu_3 > 0$$

Negative skewness - Mean < median < mode

$$Q_3 - Q_2 < Q_2 - Q_1 \text{ \& } \mu_3 < 0$$

Box & whisker plot is a graphical technique used in explanatory data analysis. It shows a five-point summary for a univariate data set giving a quick impression of the distribution. The five point summary involves least and largest value and three quartiles. It reveals at a glance whether data is symmetric. If  $Q_2$  falls at the centre then the data is symmetrical. Useful for comparing several groups of members.

Q5 a Time series - data pertaining to time or period Ex. Weekly data of sales; Geographic data that which related to the location of a place. Ex. Temperature of 4 major cities.

4

b. Title, Table number, Stubs & Captions, Body, Source -  
note & Footnote are main parts of a Statistical table.

c. Raw moments -  $M_{r'} = E(x^r)$   $r = 1, 2, 3, 4$

Central moment -  $M_r = E(x - \bar{x})^r$   $r = 1, 2, 3, 4$

d. Variance =  $\frac{\sum (x_i - \bar{x})^2}{n}$   $S.D. = \sqrt{\text{Variance}}$

Coefficient of variation =  $\frac{S.D.}{\text{Mean}} \times 100$

e.  $P_6 = l_1 + \left( \frac{6N - cf}{f} \right) \times l_2 - l_1$  Explaining all terms involved

$P_{71} = l_1 + \left( \frac{71N - cf}{100} \right) \times l_2 - l_1$

f. Mode is calculated graphically by first plotting a histogram for the given data set. The opposite vertices of the highest frequency are then joined to their neighbouring vertices. The intersection of the two lines is thereby the mode of the distribution.

g. Range =  $L - S = 16$  Coefficient of Range =  $\frac{L - S}{L + S} = 0.23$

h. Association of attributes -

Attributes are qualitative characteristics & their association shows their relationship with each other. i.e. change in one affects the other & vice versa. If change in one does not affect the other then they are independent.

e.g. Colour of Father & Sons eyes may be related.

i. Median of grouped frequency distribution:

$P_{50} = l_1 + \left( \frac{\frac{N}{2} - cf}{f} \right) \times l_2 - l_1$ ; Explaining the terms

5

PAGE NO. (5)  
DATE

Merits of Median - Easy to calculate & understand  
It is rigidly defined, can be computed for a distribution  
with open end classes, can be determined graphically. not  
affected by sampling fluctuations.

Demerits - Difficult to arrange data in ascending or  
descending order, Not based on all the data. not capable  
of further mathematical treatment.