CHAPTER 13
STATISTICS

NOTES

- ***Measures of Central Tendency***

  1) Mean 2) Median 3) Mode

- ***Types of Mean***

  1) Arithmetic Mean (A.M.)
  2) Geometric Mean (G.M.)
  3) Harmonic Mean (H.M.)

- ***Devices to determine the Arithmetic Mean of a grouped data***

  1) Direct Method

  $$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i$$

  2) Assumed Mean Method (Change of Origin Method)

  $$\bar{x} = a + \frac{1}{N}\Sigma f_i d_i \text{ , where } d_i = x_i - a$$

  3) Step Deviation Method (Change of Origin & Scale Method)

  $$\bar{x} = a + h\bar{u}$$

Derivation:

Let us consider a new variate vale u_i such that $u_i = \frac{x_i - a}{h}$.

Now, $u_i = \frac{x_i - a}{h}$

$$\Rightarrow u_i . h = x_i - a$$

$$\Rightarrow hu_i = x_i - a$$

$$\Rightarrow f_i hu_i = f_i x_i - af_i$$

$$\Rightarrow \Sigma f_i hu_i = \Sigma f_i x_i - \Sigma af_i$$

$$\Rightarrow h\Sigma f_i u_i = \Sigma f_i x_i - a\Sigma f_i$$

$$\Rightarrow \frac{1}{N} h\Sigma f_i u_i = \frac{1}{N}\Sigma f_i x_i - \frac{1}{N} a\,\Sigma f_i$$

$$\Rightarrow h\frac{1}{N}\Sigma f_i u_i = \bar{x} - \frac{a}{N} \times \mathrm{N}$$

$$\Rightarrow h\bar{u} = \bar{x} - a$$

$$\therefore \bar{x} = a + h\bar{u}$$

- ***Preparation of table to find A.M. by Direct Method***

Class	Frequency (f_i)	Mid-Value(x_i)	$f_i.x_i$

- ***Preparation of table to find A.M. by Assumed Mean Method***

Class	Mid-value ($\boldsymbol{x_i}$)	$\boldsymbol{d_i = x_i - a}$	Frequency ($\boldsymbol{f_i}$)	$\boldsymbol{f_i.d_i}$

- ***Preparation of table to find A.M. by Step Deviation Method***

Class	Mid-value (x_i)	$u_i = \frac{x_i - a}{h}$	Frequency (f_i)	$f_i.u_i$

❖ ***Geometric Mean:*** $\boldsymbol{G = (\prod x_i^{f_i})^{\frac{1}{N}}}$

❖ ***Harmonic Mean:*** $\boldsymbol{\frac{1}{H} = \frac{1}{N}\sum \frac{f_i}{x_i}}$

❖ ***Defect of Geometric Mean and Harmonic Mean***

The defect of geometric mean and harmonic mean are as follows:

(i) Inconveniences in calculating the numerical values.

(ii) Lack of versatility.

Note:

Arithmetic Mean is taken in most cases to represent the mean of the data because GM and HM have the following disadvantages:

(i) Inconveniences in calculating the numerical values.

(ii) Lack of versatility.

❖ ***Median***

The median is the value of the variate such that half of the total number of variates have their values less than or equal to it and the other half have their values greater than or equal to it.

- ***Formula:*** $\boldsymbol{M = l + \frac{\frac{N}{2} - c}{f} \times h}$ (for a grouped frequency distribution)

where, l = lower limit of the median class

N = size of the sample or population

c = cumulative frequency of the class just before the median class

f = frequency of the median class

h = width of the median class

- ***Preparation of table to find Median of a grouped data***

Class	Frequency (f_i)	cumulative frequency

❖ *Mode*

The mode of a data is the value of the variate for which there is maximum frequency.

- ***Formula:*** $$\boldsymbol{Mode = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h}$$ (for a grouped frequency distribution)

  where, l = lower limit of the modal class

  f_m = frequency of the modal class (frequency having the maximum value)

  f_1 = frequency of the class just before the modal class

  f_2 = frequency of the class just after the modal class

  h = width of the modal class

- ***Preparation of table to find Mode of a grouped data***

Class	***Frequency*** (f_i)

❖ *Relation among the three measures of Central Tendency*

Karl Pearson's Emperical Formula:-

Mean – Mode =3(Mean – Median)

i.e. Mode = 3Median – 2Mean

Note:

In an ideal case (Normal Distribution), the mean, median and mode of a distribution coincide at the same point.

❖ *Measures of Location (Or Partition Values Or Quantiles)*

Some of the important Partition Values are

1) Quartiles 2) Deciles 3) Percentiles

- ***Quartiles:*** The values of the variate which divide the total frequency distribution into four equal parts are called Quartiles.

  There are three values of Quartiles. They are:

  The first or Lower Quartile (Q_1)***:*** The value of the variate such that one fourth of all the values of the variate in the data have their values less than it while three fourth have their values greater than it is called the first or Lower Quartile. It is denoted by Q_1.

  The second Quartile (Q_2): It is same to the Median.

  The third Quartile (Q_3)***:*** The value of the variate such that three fourth of all the values of the variate in the data have their values less than it while one fourth have their values greater than it is called the third or Upper Quartile. It is denoted by Q_3.

Combined Formula for Quantiles:

$$Q_i = l + \frac{\frac{iN}{4} - c}{f} \times h \qquad (i = 1, 2, 3)$$

where, l = lower limit of the quartile class

N = sum of all the frequencies

c = cumulative frequency of the class just before the quartile class

f = frequency of the quartile class

h = width of the quartile class

- ***Deciles:*** The values of the variate which divide the total frequency distribution into ten equal parts are called Deciles.

  There are nine values of Quartiles.

  Combined Formula for Deciles:

$$D_i = l + \frac{\frac{iN}{10} - c}{f} \times h \qquad (i = 1, 2, 3, \ldots\ldots, 9)$$

where, l= lower limit of the decile class

N = sum of all the frequencies

c = cumulative frequency of the class just before the decile class

f = frequency of the decile class

h = width of the decile class

- ***Percentiles:*** The values of the variate which divide the total frequency distribution into one hundred equal parts are called Percentiles.

  There are ninety nine values of Percentiles.

  Combined Formula for Percentiles:

$$P_i = l + \frac{\frac{iN}{100} - c}{f} \times h \qquad (i = 1, 2, 3, \ldots\ldots, 99)$$

where, l=lower limit of the percentile class

N = sum of all the frequencies

c = cumulative frequency of the class just before the percentile class

f = frequency of the percentile class

h = width of the percentile class

Note: Median $= Q_2 = D_5 = P_{50}$

Preparation of tables for finding Quartiles, Deciles and Percentiles are same to that of Medians.

****