STATISTICS

MEASURES OF CENTRAL TENDENCY AND DISPERSION

SYNOPSIS AND FORMULAE

Measure of Central Tendency

1) Mathematical Averages
   a) Arithmetic mean (A.M.)
   b) Geometric mean (G.M.)
   c) Harmonic mean (H.M.)
2) Averages of Position
   a) Median b) Mode

ARITHMETIC MEAN

i) If a variable x takes values, x_2, … x_2 then the A.M is denoted by $\bar{x}$ and is given by

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_1$$

ii) For an ungrouped frequency distribution

x : x_1 x_2 ……….. x_n

f : f_1 f_2 ………… f_n

mean $\left(\bar{x}\right) = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n} = \frac{1}{N}\sum_{i=1}^{n} f_i x_1$ where $N = \sum_{i=1}^{n} f_i x_1$

iii) Weighted Arithmetic mean : If x takes values x_1 x_2 ….x_n with their respective weights w_1 w_2 … w_n then weighted A.M is given by

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_n x_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

iv) Step Deviation method :

$\bar{x} = a + \sum_{i=1}^{n} f_i\ d_i$ where $d_i = \frac{x_1 - a}{n}$

Where a = assumed mean

N = $\sum$ f = Total frequency

h = length of class interval

v) Combined mean : $\bar{x_1}$ $\bar{x_2}$ … $\bar{x_k}$ be the means of k groups of observations having respective sizes n_1, n_2, … n_k then $\bar{x}$ is the combined mean and is given by

$$\bar{x} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2} + \ldots + n_n \overline{x_n}}{n_1 + n_2 + \ldots + n_n}$$

Note

1) Algebraic sum of deviations of a set of values from their arithmetic means is zero i.e. $\sum (x - \bar{x}) = 0$
2) The sum of the squares of the deviations of a set of values is minimum when taken about mean
3) For frequency distribution : Let the values of the variable be $x_1, x_2 \ldots x_n$ and their corresponding

frequencies be $f_1\ f_2 \ldots.\ F_n$ then $\sum_{i=1}^{n} \sum_{i=1}^{n} f_i\ (x - x_1) = 0$

GEOMETRIC MEAN

If $x_1\ x_2\ x_3 \ldots x_n$ are no observations, none of them being zero them their geometric mean is defined as

$$GM = (x_1\ x_2\ x_3 \ldots x_n\text{-})^{1/n}$$

(or)

$$FM = \frac{anti\log(\log x_1 + \log x_2 + \ldots + \log x_n)}{n}$$

In the case of grouped data, AM of n observations $x_1, x_2, \ldots. x_n$ is given by

$GM = (x_1^{f_1}\ x_2^{f_2} \ldots.\ x_n^{f_n})\frac{1}{N}$ Where $N = \sum_{i=1}^{n} f_i$ (or) GM =- antilog $\left(\frac{\sum_{i=1}^{n} f_i \log(x_1)}{N}\right)$

HARMONIC MEAN

The harmonic mean of an observations $x_1\ x_2 \ldots x_n$ is defined as

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots \frac{1}{x_n}}$$

If $x_1, x_2, x_3 \ldots x_n$ are n observations which occur with frequencies $f_1\ f_2 \ldots f_{n\text{-}}$ respectively then this

HM is given by $HM = \dfrac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)}$

Relation between AM, GM, HM : A.M $\geq$ G.M $\geq$ H.M.

NOTE

Equality sign holds when all the observations are equal

MEDIAN

Median is the middle value of the variable in a set of observations, when the observations are arranged either in ascending order or in descending order of their magnitudes

Individual series : Let n be the descending order

i) Arrange the data in ascending or descending order

ii) If n is odd then median = value of $\left(\frac{n+1}{2}\right)$th observation

iii) If n is even then median = mean of $\frac{n}{2}$th and $\left(\frac{n}{2}+1\right)$th observation

DISCRETE SERIES

i) Arrange the values in ascending or descending order

ii) Prepare cumulative frequency table

iii) If n is odd then median = size of $\left(\frac{n+1}{2}\right)$th term

iv) If n is even then median = size of $\left(\frac{\left(\frac{n}{2}\right)\left(\frac{n+1}{2}\right)}{2}\right)$th term

CONTINUOUS SERIES

i) Prepare cumulative frequency table

ii) Find median class i.e. the class in which $\left(\frac{n}{2}\right)$th observation lies

iii) the median value is given by the formula median = $l+\frac{\left(\frac{n}{2}-c\right)}{f}\times h$

Where l = lower limit of median class

n = total frequency

f = frequency of median class

h = width of median class

c = cumulative frequency of class preceeding the median class.

MODE

1) Mode is the value in series which occurs most frequently

2) In a frequency distribution, mode is the variable which has maximum frequency

Individual series : The variate which is repeated maximum number of times is the mode

Discrete series : Mode is the value of the variate corresponding to the maximum frequency

Continuous series : Fiend the modal class, i.e class which has maximum frequency. The modal class can be determine either by inspection or with help or grouping table

mode = $l+\left(\dfrac{f-f_1}{2f-f_1-f_2}\right)\times h$

Where l = lower limit of modal class

h = width of modal class

f = frequency of modal class

f_1 = frequency of the class preceeding modal class

f_2 = the frequency of the class succeeding modal class

NOTE

i) A distribution in which mean, median and mode coincide is called a symmetrical distribution

ii) If the distribution is moderately skewed then mode = 3(median) – 2(mean)

QUARTILES, DECILES, PERCENTILES

The value of the variable which divides the series when arranged in ascending order or descending order into 4 equal parts is called Quartiles. There are 3 quartiles denoted by Q_1, Q_2, Q_3.

ungrouped data : Arrange the values in ascending or descending order of magnitudes

$Q_1 = \dfrac{(n+1)}{4}$ th term, $Q_3 = \dfrac{3(n+1)}{4}$ th term,

For frequently distribution :

$Q_1 = l+\left(\dfrac{\frac{h}{4}-c}{f}\times h\right)$

$Q_3 = l+\left(\dfrac{\frac{3n}{4}-c}{f}\times h\right)$

l = lower limit of class which particular quartile lies

f = frequency of class interval in which a particular quartile lies

c = cumulative frequency of class preceeding the class in which the particular quartile lies

DECILE

The values of the variable which divides the series when arranged in ascending order or descending order into 10 equal parts is called Decile, there are 9 deciles denoted by $D_1, D_2 \ldots. D_9$

1) For ungrouped data : $D_i = \frac{n \times h}{10}$, I = 1, 2, … 9

2) For grouped data $D_i = l + \left(\frac{\frac{nh}{10} - c}{f} \right) \times h$

PERCENTILE

The values of variable which divides the series when arranged in ascending or descending order into 100 equal parts is called percentile. There are 99 percentiles denoted by p_1 p_2 … p_{99} respectively,

1) For ungrouped data

$$P_i = \frac{n \times h}{100} \text{ where } h = 1, 2, 3 \ldots. 99$$

2) For grouped data

$$P_i = l + \left(\frac{\frac{nh}{100} - c}{f} \right) \times h$$

MEASURE OF DISPERSION

1) Range
2) Quartile deviation (Q.D)
3) Mean deviation (M.D)
4) Standard deviation (S.D)

RANGE

It is the difference between the greatest and the smallest observations of the distribution

Range = Largest observation – Smallest observation

R – L – S

coefficient of range = $\frac{L - S}{L + S}$

QUARTILE DEVIATION (Q.D.)

$$Q.D = \frac{Q_3 - Q_1}{2} \text{ ; coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

MEAN DEVIATION (M.D)

$$M.D = \frac{\sum_{i=1}^{n} f_i \left| x - \bar{x} \right|}{\sum_{i=1}^{n} f_i} \quad \bar{x} \text{ is average (i.e mean or median)}$$

$$\text{Coefficient of M. D} = \frac{Meandeviation}{Correspondinginaverage}$$

STANDARD DEVIATION (S.D.)

It is denoted by σ

i) for ungrouped data $S.D = \frac{\sqrt{\sum_{i=1}^{n} (x_1 - \bar{x})^2}}{N}$

ii) for frequency distribution

$S.D = \frac{\sqrt{\sum f_1 (x_1 - \bar{x})^2}}{N}$ where f_1 is frequency of x_1

NOTE

When values of variable are given in the form of classes then their respective mid points are taken as values of the variable

Coefficient of S.D : For compounding two or more series for variability, the relative measure is called coefficient of variation is used.

$C.V = \frac{\delta}{\bar{x}} \times 100$

variance = square of S.D = s^2

NOTE

If the data is moderately non symmetrical then

$M.D = \frac{4}{5}$ S.D ; $Q.D = \frac{5}{6}$ S.D

relation between A.D, M.D, S.D is

4 S.D = 5 M.D = 6. Q.D