STATISTICS

Measures of Central Tendency or Averages

A measure of central tendency or an average is defined as a representative value and it stands for the entire set of observations. The important averages are

i) Arithmetic mean (AM or $\bar{x}$)

ii) Geometric mean (G.M)

iii) Harmonic mean (H.M)

iv) Median

v) Mode

Arithmetic mean (A.M or $\bar{x}$)

i) For a raw data A.M is given by

$$\bar{x} = \frac{\sum x_i}{n}$$

ii) In a frequency distribution the A.M. is given by

$$\bar{x} = \frac{\sum f_i x_i}{N}$$ where

$x_i =$ Value of variable in an ungraped f.d. and it is mid point (class mark) in a continuous f.d.

N = Total frequency ie $N = \sum f_i$

Properties of A.M.

1. The algebraic sum of deviation of a set of observation from their mean is zero.i.e. $\sum\left(x_i - \bar{x}\right) = 0$

2. The sum of squares of deviations from the A.M is the minimum i.e. $\sum\left(x_i - \bar{x}\right)^2$ is the minimum

3. The A.M is dependent of change of scale and origin

$$AM\left(x_i\right) = \bar{x}$$

$$A.M\left(x_i \pm k\right) = \bar{x} \pm k \left(\text{change of origin}\right)$$

$$A.M\left(k\ x_i\right) = \bar{x}k\left(\text{change of scale}\right)$$

$$A.M.\left(\frac{ax+b}{c}\right) = \frac{a\bar{x}+b}{c}\left(\text{change of scale and origin}\right)$$

4. A.M of an A.P. = $\dfrac{\text{first term} + \text{last term}}{2}$

5. Combined A.M.

If $\bar{x}_1, \bar{x}_2 \ldots\ldots \bar{x}_k$ are means of k samples of size n_1, n_2.....n_k then combined A.M is given by $\bar{x} = \dfrac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i}$

6. **Weighted A.M**

In statistics, weight means the importance given to be observations

Let $x_1, x_2 \ldots. x_n$ be the observations with corresponding weights w_1, w_2,....w_n, then weighted A.M is given by

$$\bar{X}_w = \frac{\sum w_i x_i}{\sum w_i}$$

Geometric mean (G.M.)

G.M. of $x_1, x_2, x_3 \ldots. x_n = n^{th}$ root of product

GM $= \left(x_1.x_2.x_3 \ldots. x_n\right)^{1/n}$

$$= \left(\prod x_i\right)^{1/n}$$

log $G.M. = \frac{1}{n}\log\left(\prod x_i\right) = \frac{1}{n}\log\left(x_1.x_2 \ldots x_n\right) = \frac{1}{n}\sum \log\ x_i$

$$\log G.M. = \frac{\sum \log x_i}{n}$$

Result : log of G.M. = A.M of log of observations

Result : G.M of a GP = $\sqrt{\text{first tem} \times \text{last term}}$

Hermonic mean - H.M.

$x_1, x_2, x_3 .... x_n -$ observations

$$\frac{1}{x_1}, \frac{1}{x_2} - - - - - \frac{1}{x_n} \text{reciprocals}$$

$$\frac{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + - - - \frac{1}{x_n}}{n} - \text{A.M.}$$

Reciprocal of A.M. of the reciprocals of observations

$$H.M = \frac{n}{\sum\left(\frac{1}{x_i}\right)}; \text{H.M.of } a, b = \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2ab}{a+b} \text{H.M.of } a, b, c = \frac{3abc}{ab + bc + ac}$$

Median: Median is the middle most item when the set is in ascending or decreasing order. For a set in ascending order

Median $= \left(\frac{n+1}{2}\right)^{th}$ item when the data are in ascending order

Mode: Mode is the most frequently appearing item in a set. A set can be unimodal, Bimodal, trimodal etc.

Measures of Dispersion

Many sets can have the same average. Hence in order to check the effectiveness of an average as a measure of central tendency. Some secondary measure are required such measures are called measures of dispersion. Thus a measure of dispersion is the total deviation of observations from an average.

A measure of dispersion expressed in the units of observations is called an **absolute measure of dispersion. A Relative measure of dispersion** is a pure number free from units.

Absolute measures of Dispersion

1) Range
2) Mean Deviation (M.D)
3) Root mean square deviation (RMSD)
4) Standard deviation (SD or σ)

5) Variance $\left(\sigma^2\right)$

Relative measures of dispersion

1) Coefficient of range

2) Coefficient of mean deviation

3) Coefficient of standard deviation

4) Coefficient of variance (C.V)

1) Range and coefficient of Range

Let X_{maxi} be the highest item and X_{mini} be the lowest item in a set.

Range = $X_{maxi} – X_{mini}$

Coe. of Range = $\frac{X_{max\,i} - X_{min\,i}}{X_{maxi} + X_{min\,i}}$

Ex: The highest of some plants are given below 5 cm 12 cm 8 cm 3 cm 9cm

$X_{max\,i} = 12\text{cm} \;\; X_{min\,i} = 3\text{cm}$

Range = 12 cm – 3 cm = 9 cm (has units)

Coef. of Range = $\frac{12\text{cm} - 3\text{cm}}{12\text{cm} + 3\text{cm}} = \frac{9\text{cm}}{15\text{cm}} = \frac{3}{5}\left(\text{Free from units}\right)$

2) Mean Deviation: Mean devidation is the A.M. of absolute deviations of observations from an average A (where A can be any of the average like mean, median, mode etc. For a raw data mean deviation is given by $\text{M.D} = \frac{\sum\left|x_i - A\right|}{n}$ and in a f.d. $\text{M.D} = \frac{\sum f_i\left|x_i - A\right|}{N}$ where $N = \sum f_i$

Note

Here A is any average. It can be mean, median, mode etc.

Result : Sum of absolute deviations is minimum from median and hence mean deviation is minimum from median

Ex: Consider the observations 5,7,10,15,20, We find mean deviation from median = 10 $\therefore A = 10$

Mean deviation = $\frac{\sum\left|x_i - A\right|}{n}$

$$= \frac{|5-10| + |7-10| + |10-10| + |15-10| + |20-10|}{5}$$

$$= \frac{5+3+0+5+10}{5} = \frac{23}{5}$$

Coefficient of mean deviation

The coefficient of M.D is defined as coe. of M.D= $\frac{\text{mean deviation}}{\text{average A}}$

In the above example the coefficient of M.D= $\frac{\left(\frac{23}{5}\right)}{10} = \frac{23}{50}$

Root mean square deviation (RMSD)

Root mean square deviation is square root of Arithmetic mean of the squares of deviations of observations from an average. A where A can be any average. For a raw data it is given by

$$\text{RMSD} = \sqrt{\frac{\sum (x_i - A)^2}{n}}$$

In a f.d. RMSD is given by

$$\text{RMSD} = \sqrt{\frac{\sum f_i (x_i - A)^2}{N}}$$

Where $N = \sum f_i$ and A is any average

Result

Sum of square deviations is minimum from A.M. and hence RMSD is minimum when A = Median

Standard Deviation (SD or σ)

Standard deviation is the most important absolute measure of dispersion. It is the minimum value of RMSD or in otherwords the RMSD will become S.D when $A = \overline{x}$. Thus S.D can be defined as follows.

Definition of S.D

The S.D is the square root of the A.M. of the squares of deviations of observations from A.M $\overline{x}$. For a raw data the S.D is given by

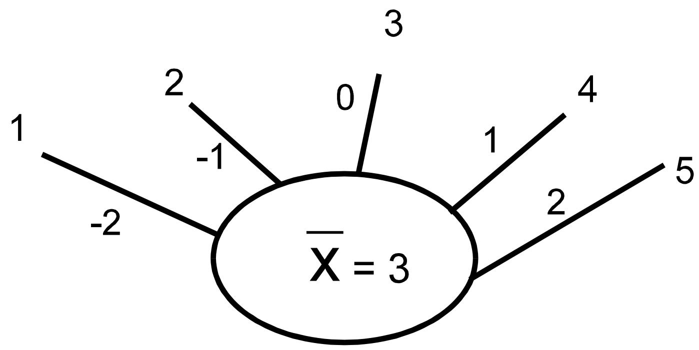
$$\sigma = \sqrt{\frac{\sum \left(x_i - \overline{x}\right)^2}{n}}$$

In a f.d. the standard deviation is given by

$$\sigma = \sqrt{\frac{\sum f_i \left(x_i - \overline{x}\right)^2}{N}} \text{ where } N = \sum f_i$$

Example : Find s.D of 1,2,3,4,5

$\overline{x} = 3$



$$\Sigma\left(x_i - \overline{x}\right)^2 = (-2)^2 + (-1)^2 + 0^2 + (1)^2 + (2)^2 = 10$$

$$\sigma = \sqrt{\frac{\Sigma\left(x_i - \overline{x}\right)^2}{n}} = \sqrt{\frac{10}{5}} = \sqrt{2}$$

Variance $\left(\sigma^2\right)$

The square of S.D is defined as variance. It is denoted by σ^2. If S.D is 2 variance is $\left(\text{S.D}\right)^2 = \sigma^2 = 4$. However S.D. should be the non-negative square root of variance

Coefficient of S.D

Coe. of S.D= $\frac{\text{standard deviation}}{\text{Arithmetic mean}} = \frac{\sigma}{x}$

Computation of S.D

1) For a raw data the S.D can be calculated by $\sigma = \sqrt{\frac{\Sigma x_i^{\,2}}{n} - \left(\overline{x}\right)^2}$. The variane is given by

$$\sigma^2 = \frac{\Sigma x_i^2}{n} - \left(\overline{x}\right)^2$$

2) In a f.d. the S.D is given by $\sigma = \sqrt{\frac{\Sigma f_i x_i^{\,2}}{N} - \left(\overline{x}\right)^2}$

where $N = \Sigma f_i$ and $\overline{x} = \frac{\Sigma f_i x_i}{N}$. The variance is given by

$$\sigma^2 = \frac{\sum f_i x_i^{\,2}}{N} - \left(\overline{x}\right)^2$$

Results

1) The S.D is independent of change of origin. Let S.D (x) = σ. Then $S.D\left(X \pm K\right) = \sigma$

2) The S.D is dependent of change of scale. Let S.D $\left(X\right) = \sigma$. Then $S.D\left(KX\right) = \left|K\right|\sigma$

$$S.D\left(\frac{ax+b}{c}\right) = \left|\frac{a}{c}\right|\sigma$$

3) Range $\simeq 6\sigma$

4) M.D $\simeq \frac{4}{5}\sigma$

5) The sum of squares of observations is given by

$$\sum x_i^{\,2} = n\left(\sigma^2 + \left(\overline{x}\right)^2\right)$$

Combined S.D

	Set I	Set II
no of items	n_1	n_2
A.M	$\overline{x}_1$	$\overline{x}_2$
S.D	σ_1	σ_2

The S.D of combined set of $n_1 + n_2$ items is given by

$$\sigma = \sqrt{\frac{n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)}{n_1 + n_2}}$$ where

$d_1 = \overline{x}_1 - \overline{x}$ and $d_2 = \overline{x}_2 - \overline{x}$ and $\overline{x} = \frac{n_1 \overline{x}_1 + n_2 \overline{x}_2}{n_1 + n_2}$. The method can be extended for more sets

Result

1) S.D of I[st] 'n' natural numbers is $\sigma = \sqrt{\frac{n^2 - 1}{12}}$

2) Variance of I[st] 'n' natural numbers is $\sigma^2 = \frac{n^2 - 1}{12}$

3) S.D of Ist n odd natural numbers is $\sigma = \sqrt{\frac{n^2 - 1}{3}}$

4) S.D of Ist n even natural numbers is $\sigma = \frac{n^2 - 1}{3}$

Coefficient of variance (C.V)

The C.V of a set is given by

$$C.V = \frac{S.D}{AM} \times 100 = \frac{\sigma}{x} \times 100$$

The C.V is used to compare the variability among different sets of observations irrespective of the unit of measurement. The set having the least C.V said to be more consistent. homogenous, uniform or stable set.