22 December 2023 08:36 PM

Dispersion

Absolute Measures of Dispersion

1. Range
2. Mean deviation
   a. M.D about $\bar{x}$
   b. M.D about median (Me)
3. Variance
4. Standard Deviation
5. Quartile Deviation (also known as semi-interquartile range)

Relative Measures of Dispersion

1. Coefficient of variation (C.V.).
2. coeff of MD
3. coeff of Q.D.

① Range = Max value − Min value.

Prop: If x variable is changed by origin 'a' and scale 'b'

such that $y = a + bx$

Then, $\boxed{\text{Range } y = |b| \text{ Range } x}$

Two variables x and y are related as follows,

Ex: Two variables x and y are related as follows,
$\boxed{3y + 4x = 9}$ and range of $x = 3$
Then compute the value of range of y.

Soln $3y + 4x = 9$

or $3y = 9 - 4x \Rightarrow y = \frac{9-4x}{3}$

$\Rightarrow y = 3 - \frac{4}{3}x$

range of $x = 3$

If we compare it with $y = a + bx$

then $b = -4/3$

We know range $y = |b|$ range x

Range $y = \frac{4}{\cancel{3}} \times \cancel{3} = 4$ (ans)

2. Mean Deviation about mean ($\bar{x}$)

a) without frequency, $MD_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$

(b) with frequency, $MD_{\bar{x}} = \frac{1}{\Sigma f}\sum_{i=1}^{n}|x_i - \bar{x}|f_i$

where $\bar{x} = \frac{1}{\Sigma f}\sum_{i=1}^{n} x_i f_i$

Prop: if $y = a + bx$

then $MD_{\bar{y}} = |b| \, M.D_{\bar{x}}$ ✓

MD about median, $MD_{Me} = \frac{1}{n}\sum_{i=1}^{n}|x_i - Me|$

and $MD_{Me} = \frac{1}{\Sigma f}\sum_{i=1}^{n}|x_i - Me|f_i$

Ef

(*) Prove that M.D about median is least.

Suppose we are given n number of observations $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ arranged in ascending order.

Here we have to prove that

$$MD = \frac{1}{n} \sum |x_i - A| \text{ is minimum,}$$

when A = median.

it means

$$x_1 \le x_2 \le x_3 \le \cdots \le x_n .$$

<u>Case 1</u> : Let n be odd number of obs and $n = 2m+1$

if we have

$$x_{(1)} \le x_{(2)} \le x_{(3)} \le \cdots \le \boxed{x_{(m)} \le x_{(m+1)} \le x_{(m+2)}} \cdots \le x_{(2m)} \le x_{(2m+1)}$$

Now, $|x_1 - A| + |x_{2m+1} - A|$ is least when

$$x_1 \le A \le x_{2m+1}$$

Similarly $|x_2 - A| + |x_{2m} - A|$ is least when

$$x_2 \le A \le x_{2m}$$

Likewise, $|x_m - A| + |x_{m+1} - A|$ is least when

$$x_m \le A \le x_{m+1}$$

finally. $|x_{m+1} - A|$ is least when $A = x_{m+1}$

... order series

Finally. $|x_{m+1} - A|$ is least when $A = x_{m+1}$

We can see from the ascending order series that x_{m+1} is the middlemost obs.
and we know that median is the middle most obs in case of odd no. of observation

$\therefore$ Median $= x_{m+1} = A$

$\therefore$ $MD_A = \frac{1}{n} \Sigma |x_i - A| = \frac{1}{n} \Sigma |x_i - Me|$

is minimum when $A = Me = x_{m+1}$

Case 2 : Let n be even number of observations
ie $n = 2m$.

ie we have $x_1 \leq x_2 \leq x_3 \leq \dots \leq \boxed{x_m \leq x_{m+1}} \leq x_{m+2} \leq \dots \leq x_{2m}$

Now $|x_1 - A| + |x_{2m} - A|$ is least when $x_1 \leq A \leq x_{2m}$
$|x_2 - A| + |{2m-1} - A|$ is least when $x_2 \leq A \leq x_{2m-1}$

$\vdots$
Likewise we have $|x_m - A| + |x_{m+1} - A|$ is least when $x_m \leq A \leq x_{m+1}$

Now from the data x_m and x_{m+1} are the middle most ~~obs~~ observations
In case of even number of obs median $= \frac{x_m + x_{m+1}}{2} = A$

middle ... of obs
In case of even member of obs
median $= \frac{x_m + x_{m+1}}{2} = A$

ie median lies between x_m and x_{m+1}

$\therefore$ if $A =$ median, then $MD_{Me} = \frac{1}{n}\sum|x_i - Me|$
will be least.

(Proved)

Note
$\boxed{MD_{\bar{x}} > MD_{Me}}$

③ **Variance and Standard Deviation**

$$\boxed{\text{variance, } v(x) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \frac{1}{n}\sum(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \frac{1}{n}\sum x_i^2 - 2\bar{x}\cdot\frac{1}{n}\sum x_i + \frac{1}{n}\sum\bar{x}^2$$

$$\boxed{\sigma^2 = \frac{1}{n}\sum x_i^2 - (\bar{x})^2}$$

Standard Deviation, $\sigma = \sqrt{var(x)}$

or $\sigma = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$

or $\sigma = \sqrt{\frac{1}{n}\sum x_i^2 - \bar{x}^2}$

With frequency

S.d, $\sigma = \sqrt{\frac{1}{\sum f_i} \sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}$

$= \sqrt{\frac{1}{\sum f_i} \sum x_i^2 f_i - \bar{x}^2}$

where $\boxed{\bar{x} = \frac{1}{\sum f_i} \sum x_i f_i}$

$\therefore$ variance, $\sigma^2 = \left[\frac{1}{\sum f_i} \sum x_i^2 f_i - \bar{x}^2\right]$

Properties ① If all observations are same then variance and s.d is zero.

Let $x_i = c$ for all values of x and $i = 1(1)n$.

Then, $\bar{x} = \frac{1}{n} \sum x_i = \frac{1}{n} \sum c = \frac{n}{n} c = c$

$\therefore$ variance, $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i^2) - \bar{x}^2$

$= \frac{1}{n} \sum c^2 - (c)^2$

$= \frac{1}{\cancel{n}} \times \cancel{n} c^2 - c^2$

$= 0$

$\therefore$ s.d $= \sqrt{\text{variance}} = 0$ ✓ (Proved)

(2) Variance and s.d depends only on change in scale. It is independent of change in origin.

Let us change all 'n' the values of x by origin 'a' and scale 'b'

such that $y_i = a + bx_i$

then variance of y, $\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$

or, $\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(\cancel{a} + bx_i - \cancel{a} - b\bar{x})^2$

$\sigma_y^2 = b^2 \frac{1}{n}\sum(x_i - \bar{x})^2$

$\sigma_y^2 = b^2 \sigma_x^2$

$\boxed{\sigma_y = |b| \sigma_x}$

Proved.

Combined variance or combined s.d.

from two groups or data set of observations

... of obs and

from two groups ...

say n_1 & $\bar{x}_1$ are number of obs and mean of group 1.

again n_2 & $\bar{x}_2$ are no. of obs & mean of group 2

such that $\bar{X} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$ is the combined mean.

$\therefore$ The combined variance is given by

$$s^2 = \frac{1}{(n_1+n_2)}(n_1 s_1^2 + n_2 s_2^2) + \frac{n_1(\bar{x}_1-\bar{x})^2 + n_2(\bar{x}_2-\bar{x})^2}{(n_1+n_2)}$$

or,

$$\boxed{s^2 = \frac{1}{(n_1+n_2)}(n_1 s_1^2 + n_2 s_2^2) + \frac{n_1 n_2}{(n_1+n_2)^2}(\bar{x}_1-\bar{x}_2)^2}$$

(only in case of 2 groups)

$\rightarrow$ Generalise the formula for r groups we get,

$$s^2 = \frac{\sum^{r} n_i s_i^2}{} \quad \frac{\sum^{r} n_i(\bar{x}_i - \bar{x})^2}{}$$

$$s^2 = \frac{\sum_{i=1}^{r} n_i s_i^2}{\sum_{i=1}^{r} n_i} + \frac{\sum_{i=1}^{r} n_i (\bar{x}_i - \bar{x})^2}{\sum_{i=1}^{r} n_i}$$

, where $\bar{x} = \frac{\sum_{i=1}^{r} n_i \bar{x}_i}{\sum_{i=1}^{r} n_i}$ (grand mean).

⑤ Quartile Deviation (semi-interquartile Range)

$$Q.D = \frac{Q_3 - Q_1}{2}$$

where Q_3 is third Quartile

Q_1 is the ~~...~~ first Quartile.

Q For two values a and b of a variable x the mean and s.d are 25 and 4 resp. Find a and b.

mean, $\bar{x} = 25$

$s.d, \sigma = 4$

$\sigma^2 = 16$

mean, $\bar{x} = 25$

$$\frac{a+b}{2} = 25$$

$$\boxed{a+b = 50} \text{ — } 1$$

$$(a+b)^2 = 2500$$

$$a^2+b^2+2ab = 2500$$

$$1282 + 2ab = 2500$$

$$2ab = 2500 - 1282$$

$$2ab = 1218 \text{ — } 3$$

Now $(a-b)^2 = a^2+b^2 - 2ab$

$$= 1282 - 1218$$

$$(a-b)^2 = 64$$

$$\therefore a-b = \sqrt{64} = 8 \text{ — } 4$$

Adding 1 and 4 we get,

$$a+\cancel{b} = 50$$

$$a-\cancel{b} = 8$$

$$(+)$$

$$2a = 58$$

$$\boxed{a = \frac{58}{2} = 29}$$

$$\therefore \boxed{b = 50 - a = 50 - 29 = 21}$$

$S.D = 4$

$$\sigma^2 = 16$$

$$\frac{a^2+b^2}{2} - 25^2 = 16$$

$$\frac{a^2+b^2}{2} = 16 + 25^2$$

$$a^2+b^2 = 1282 \text{ — } 2$$

$\therefore \; b = 30 - a - 8 - 6$

Calculate s.d, md (about mean & median)

Q

Marks	(f) Number of Students	cf	x_i	$x_i f_i$	$\lvert x_i - Me \rvert f_i$	$x_i^2 f$
10-20	4	4	15			
$\boxed{20-30}$	6	10	25			
30-40	5	15	35			
40-50	5	20	45			
	$\Sigma f = 20 = N$			$\Sigma x_i f_i =$	$\Sigma \lvert x_i - Me \rvert f_i =$	$\Sigma x_i^2 f =$

$\sigma = \sqrt{\frac{1}{n} \Sigma x_i^2 f_i - \bar{x}^2}$

$\bar{x} = \frac{1}{\Sigma f} \Sigma x_i f_i$