ST. LAWRENCE HIGH SCHOOL

A JESUIT CHRISTIAN MINORITY INSTITUTION

STUDY MATERIAL-11

SUBJECT – STATISTICS

1st term

Chapter: DISPERSION **Class: XI**

Topic: Mean Deviation **Date: 15.08.2020**

DISPERSION

PART 2

Definition : Mean deviation about an arbitrary central value 'A' is defined as the arithmetic mean of the absolute deference of the central value 'A' from all the observations.

For ungrouped grouped data

Observations: $x_1 , x_2 , \dots \dots \dots \dots . , x_n$

$$MD_A(x) = \frac{1}{n} \sum_{i=1}^{n} |x_i - A|$$

For grouped data

Observations: $x_1 , x_2 , \dots \dots \dots \dots . , x_n$

Frequency: $f_1 , f_2 , \dots \dots . \dots \dots . , f_n$

$$MD_A(x) = \frac{1}{N} \sum_{i=1}^{n} |x_i - A| f_i$$

Properties:

- Change of base or origin and scale

If $y_i = a + b\, x_i \; \forall\; i = 1(1)n$

Then, $MD_{\bar{y}}(y) = |b|\, MD_{\bar{x}}(x)$

$$MD_{\tilde{y}}(y) = |b|\, MD_{\tilde{x}}(x)$$

$$MD_{\breve{y}}(y) = |b|\, MD_{\breve{y}}(x)$$

In case of grouped data

Proof: By definition

$$MD_{\bar{y}}(y) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \bar{y}|$$

$$= \frac{1}{n}\sum_{i=1}^{n}|a + bx_i - a - b\bar{x}|$$

$$= |b|\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

$$= |b|\ MD_{\bar{x}}(x)$$

In case of grouped data

Proof: By definition

$$MD_{\bar{y}}(y) = \frac{1}{N}\sum_{i=1}^{n}|y_i - \bar{y}|f_i$$

$$= \frac{1}{N}\sum_{i=1}^{n}|a + bx_i - a - b\bar{x}|f_i$$

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

$$= |b|\ MD_{\bar{x}}(x)$$

- Mean deviation is least when taken about median.

Proof: Lemma: For two points P and Q,

$|P - A| + |Q - A|$ is least when $P \leq A \leq Q$.

Take the observations as x_1 , x_2 , … … … … . , x_n

Case 1: Number of observation is odd. n= 2m+1

Rearrange the observations as $x_{(1)}$, $x_{(2)}$, … … … … . , $x_{(2m)}$,$x_{(2m+1)}$

Now using the lemma,

$\left|x_{(1)} - A\right| + \left|x_{(2m+1)} - A\right|$ is least when $x_{(1)} \leq A \leq x_{(2m+1)}$

$\left|x_{(2)} - A\right| + \left|x_{(2m)} - A\right|$ is least when $x_{(2)} \leq A \leq x_{(2m)}$

$\left|x_{(3)} - A\right| + \left|x_{(2m-1)} - A\right|$ is least when $x_{(3)} \leq A \leq x_{(2m-1)}$

.

.

$\left|x_{(m)} - A\right| + \left|x_{(m+2)} - A\right|$ is least when $x_{(m)} \leq A \leq x_{(m+2)}$

$\left|x_{(m+1)} - A\right|$ is least when $A = x_{(m+1)}$

Now $A = x_{(m+1)}$ satisfies all the above constrains.

So $\sum_{i=1}^{2m+1}\left|x_{(i)} - A\right|$ is least when $A = x_{(m+1)}$

ie, $\sum_{i=1}^{2m+1}|x_i - A|$ is least when $A = x_{(m+1)}$

ie, $\frac{1}{2m+1}\sum_{i=1}^{2m+1}|x_i - A|$ is least when $A = x_{(m+1)}$

Again by definition $A = x_{(m+1)}$is the median of the given data.

So mean deviation about median is least for odd number of observations.

Case 2: Number of observation is even. n= 2m

Rearrange the observations as $x_{(1)}$, $x_{(2)}$, … … … … . , $x_{(2m-1)}$,$x_{(2m)}$

Now using the lemma,

$\left|x_{(1)}-A\right|+\left|x_{(2m)}-A\right|$ is least when $x_{(1)} \leq A \leq x_{(2m)}$

$\left|x_{(2)}-A\right|+\left|x_{(2m-1)}-A\right|$ is least when $x_{(2)} \leq A \leq x_{(2m-1)}$

$\left|x_{(3)}-A\right|+\left|x_{(2m-2)}-A\right|$ is least when $x_{(3)} \leq A \leq x_{(2m-2)}$

.

.

$\left|x_{(m)}-A\right|+\left|x_{(m+1)}-A\right|$ is least when $x_{(m)} \leq A \leq x_{(m+2)}$

Now take $A=\frac{x_{(m)}+x_{(m+1)}}{2}$

Now $A=\frac{x_{(m)}+x_{(m+1)}}{2}$ satisfies all the above constrains.

So $\sum_{i=1}^{2m}\left|x_{(i)}-A\right|$ is least when $A=\frac{x_{(m)}+x_{(m+1)}}{2}$

ie, $\sum_{i=1}^{2m}|x_i-A|$ is least when $A=\frac{x_{(m)}+x_{(m+1)}}{2}$

ie, $\frac{1}{2m}\sum_{i=1}^{2m}|x_i-A|$ is least when $A=\frac{x_{(m)}+x_{(m+1)}}{2}$

Again by definition $A=\frac{x_{(m)}+x_{(m+1)}}{2}$ is the median of the given data.

So mean deviation about median is least for even number of observations.

Combining case 1 and case 2 we can say mean deviation is least when taken about median.

- Next result

$$MD_{\bar{x}}(x) = \frac{2}{n}\sum_{i|x_i>\bar{x}}^{n}(x_i-\bar{x})OR\ \ \frac{2}{n}\sum_{i|\bar{x}>x_i}^{n}(\bar{x}-x_i)$$

Proof:

$\sum_{i=1}^{n}(x_i-\bar{x}) = 0 \ \Rightarrow\ \sum_{i|x_i>\bar{x}}^{n}(x_i-\bar{x}) + \sum_{i|\bar{x}>x_i}^{n}(x_i-\bar{x}) = 0$

$\Rightarrow \sum_{i|x_i>\bar{x}}^{n}(x_i-\bar{x}) = \sum_{i|\bar{x}>x_i}^{n}(\bar{x}-x_i)$ ....(1)

Now $MD_{\bar{x}}(x) = \frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|$

$= \frac{1}{n}\ \left(\sum_{i|x_i>\bar{x}}^{n}(x_i-\bar{x}) + \sum_{i|\bar{x}>x_i}^{n}(\bar{x}-x_i)\right)$

$= \ \frac{2}{n}\sum_{i|x_i>\bar{x}}^{n}(x_i-\bar{x})OR\ \ \frac{2}{n}\sum_{i|\bar{x}>x_i}^{n}(\bar{x}-x_i)$ (from (1)).