unit 2

Dispersion

DATE____
PAGE____

Map

Dispersion = amount of deviations | spread of the observations

Range = largest − smallest value

Coefficient = $\frac{L-S}{L+S} \times 100$

Mean Deviation = $\varepsilon \left| \frac{x_i - A}{n} \right|$ about mean

= $\varepsilon \frac{|x_i - x \text{ median}|}{n}$ about median

coefficient = $\frac{\text{Mean deviation about Mean | Median}}{\text{Mean | Median}} \times 100$

Quartile Deviation = $\frac{Q_3 - Q_1}{2}$

Coefficient = $\frac{QD}{\frac{Q_3 - Q_1}{2}} \times 100$ / or $\frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

Standard Deviation = $\sqrt{\varepsilon \left(\frac{x_i - \bar{x}}{n} \right)^2}$

Variance = square of SD

coefficient of variation = $\frac{SD}{AM} \times 100$

DATE
PAGE

Combined SD :

$$\sqrt{\frac{n_1\left(SD_1^2 + D_1^2\right) + n_2\left(SD_2^2 + D_2^2\right)}{n_1 + n_2}}$$

Correcting incorrect SD :

$$SD_r = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\bar{X}_r\right)^2}$$

Relationship : $4SD = 5MD = 6QD$

unit - 2
Dispersion

* Dispersion for a given set of observations the may be defined as amount of deviation of the observations.

* way of describing how spread out a set of data is

Absolute (easy)
- dependent on the unit of variable under considera-tion.
- calculated with unit
- different ways -
  (1) Range (2) Quartile deviation (3) Mean deviation (4) Standard deviation

Relative measure
- unit free. imp for comparison
- for comparing two or more distributions, relative measures are considered.
- different ways - coefficients

I Range = largest value - smallest value

* discrete series (ungrouped) = consider value of x. largest - smallest.

grouped = consider class limits not × boundaries

DATE
PAGE

If not given in ques = assumption = continuous

continious series = convert in class boundaries

Range = largest class boundary − smallest class boundary

coefficient of range

$\frac{L-S}{L+S} \times 100$ (in percentages)

Properties

(sign of b does not matter)

$y = \underbrace{a}_{\text{origin shift}} + \underbrace{b}_{\text{scaling}}x$

- not affected by [(+) (−)] origin & shift
- only affected by scaling ((÷),(×))

$R_y = |b| \times R_x$

- only magnitude considered

Ideal measure for characteristics

→ properly defined
→ easy to comprehend
→ simple to compute
→ includes all observations
→ unaffected by sampling fluctuations
→ amendable to some further mathematical treatment.

PAGE

Merits of Range

→ minimum time → easiest

Demerits of Range

→ not based on all observations

→ poor measure. considers only extreme values. tells nothing about the distribution of no. in between.

→ values varies from sampling fluctuations

→ not for open-end classes/ distribution

→ not suitable for further mathematical treatment.

Mean Deviation (subtract)

Average (AM) of the deviations of given values from a measure of central tendency

(mean of absolute deviations (| |) from any central tendency - mean/ median)

Mean deviation = $\frac{\Sigma |x_i - A|}{n}$

→ include the deviation of each and every observations.

(*) Best method. mean deviation from median

$$\frac{\Sigma|X_i - X_{med}|}{n} = \text{minimum}$$

i.e. the sum of the deviations of items from median is least.

(*) discrete series $\frac{\Sigma f|(x_i - \bar{x})|}{\Sigma f}$

$$\frac{\Sigma f_i|x_i - x_{med}|}{\Sigma f_i}$$

(*) continuous series (just take value of x_i = class mark/mid value)

(about mean) →

(about median) → $L + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$

$$\frac{\Sigma f_i|x_i - x_{med}|}{\Sigma f_i}$$

Coefficient of Mean Deviation

$$\frac{\text{Mean deviation about Mean}}{\text{Mean}} \times 100$$

or median. (instead of Mean)

Properties

- Mean deviation → minimum value when the deviations are taken from the median.
- MD of y = $|b|$ MD of x

Merits

→ easy and simple → rigidly defined

→ based on all observations

→ less affected by extreme observations

Demerits

→ Algebraic signs are ignored while taking deviations

→ not suitable for further mathematical treatment.

→ cannot be calculated for open end classes

Quartile Deviation

Semi Inter - Quartile Range

Inter quartile range is an absolute measure of dispersion defined by formula =

$Q_3 - Q_1$

Quartile deviation (semi inter quartile range)

$$= \frac{Q_3 - Q_1}{2}$$

Coefficient of Q_D $= \dfrac{Q_D}{\frac{Q_3 + Q_1}{2}} \times 100$

Simply $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

It is a pure number independent of units of measurement and can be used to compare two distributions expressed in different.

For symmetrical distribution

$Q_3 - Q_2 = QD$ —①

$Q_2 - Q_1 = QD$ —②

(Median) $Q_2 = \frac{Q_3 + Q_1}{2}$ solving ① & ②

also $\Rightarrow \dfrac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} \times 100$] coefficient of Q_D (if we divide by 2)

$= \frac{Q_D}{\text{Median}} \times 100$

(only in symmetrical distribution)

Properties

- $y = a + bx \quad Q_{D_y} = |b| \times Q_{D_x}$.

Merits

- only measure that can be used in open end classes.
- less affected due to sampling fluctuations (consider only 50%)

used to study variability in the central half part of the data (specially used for this)

Demerits

not based on all observations (only 50%)

not suitable for further mathematical treatment

It is affected considerably by sampling fluctuations.

Standard Deviation

Positive square root of AM of the squares of deviations of the observations from their AM.

denoted by Greek letter (sigma)

Formula = $\sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

(root mean square deviation from AM)

also Formula : $\sqrt{\frac{\Sigma x_i^2}{n} - (\overline{x})^2}$

Frequency distribution

$$\sigma = \sqrt{\frac{\Sigma f_i (x_i - \overline{x})^2}{\Sigma f_i}}$$

or

$$= \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2}$$

Variance

square of standard deviation

$$\frac{\Sigma (x_i - \overline{x})^2}{n}, \quad \frac{\Sigma x_i^2}{n} - (\overline{x})^2$$

$$\frac{\Sigma f_i (x_i - \overline{x})}{\Sigma f_i}, \quad \frac{\Sigma f_i x_i^2}{\Sigma f_i} - (\overline{x})^2$$

For any two numbers SD is always

$SD = \frac{|a-b|}{2}$ [Half of the range]

DATE
PAGE

(*) Coefficient of Variance $= \frac{SD}{AM} \times 100$

Correcting Incorrect Values of Means and Standard Deviation

Given $= \bar{X}_w, X_r, X_w, SD_w, N$

(1) $\bar{X}_r = \bar{X}_w + \frac{x_r - x_w}{N}$

(2) $SD_w = \sqrt{\frac{\Sigma x_w^2}{n} - (\bar{X}_w)^2}$

(3) Find $\Sigma x_w^2 = ?$

(*) (Use above to find $\Sigma X r^2$)

(4) $\Sigma x_r^2 = \Sigma x_w^2 - x_w^2 + x_r^2$

(*) $SD_r = \sqrt{\frac{\Sigma x_r^2}{n} - (\bar{X}_r)^2}$

Combined Standard Deviation

$$= \sqrt{\frac{n_1 (SD_1^2 + D_1^2) + n_2 (SD_2^2 + D_2^2)}{n_1 + n_2}}$$

where $D_1 = \overline{X}_1 - \overline{X}_{CM}$

$D_2 = \overline{X}_2 - \overline{X}_{CM}$ and so on.

Coefficient of variation

$$= \frac{SD}{\overline{X}} \times 100$$

- Used for comparison of two data (pure no.)

- A distribution for which the coefficient of variation is $\boxed{\text{smaller}}$ is said to be less variable / more consistent / more uniform / more stable / more homogenous and vice versa

Properties of Standard Deviation (ideal measure)

- suitable for further mathematical treatment
- Related as : $y = a + bx$

(independent of origin but not scale)

$$SD_y = |b|\, SD_x$$

$$\text{Variance}_y = b^2\, \text{variance}_x$$

- when all values are equal (say k) then $\boxed{SD = 0}$

(this property also present in range, mean deviation)

- SD of first n natural numbers = (consecutive terms)

$$\sqrt{\frac{n^2-1}{12}}$$

Relationship between SD, MD, QD

$$4\,SD = 5\,MD = 6\,QD$$