Module - 1

Arithmetic mean

→ Individual series

- Direct method, $\bar{x} = \frac{\Sigma x}{n}$
- Shortcut method, $\bar{x} = A + \frac{\Sigma d}{n}$, $d = x - A$

→ Discrete series

- Direct method, $\bar{x} = \frac{\Sigma fx}{N}$, $N = \Sigma f$
- Shortcut method, $\bar{x} = A + \frac{\Sigma fd}{N}$; $d = x - A$, $N = \Sigma f$

→ Continuous series

- Direct method, $\bar{x} = \frac{\Sigma fm}{N}$, x = mid-value.
- Shortcut, $\bar{x} = A + \frac{\Sigma fd}{N}$, $d = x - A$, A = assume

Mean / Quartiles / Percentile.

→ Individual series

Q_1 = value of $\left(\frac{n+1}{4}\right)^{th}$ item

Q_3 = value of $\left(\frac{3(n+1)}{4}\right)^{th}$ item.

M_d = value of $\left(\frac{n+1}{2}\right)^{th}$ item.

D_i = i^{th} decile = value of $\left(\frac{i(n+1)}{10}\right)^{th}$ item.

P_i = i^{th} percentile = value of $\left[\frac{i(n+1)}{100}\right]^{th}$ item.

→ Discrete Series

Q_1 = value of $\left(\frac{N+1}{4}\right)^{th}$ item

Q_3 = value of $\left(\frac{3(N+1)}{4}\right)^{th}$ item

M_d = value of $\left(\frac{N+1}{2}\right)^{th}$ item

D_i = i^{th} decile = value of $\left(\frac{i(N+1)}{10}\right)^{th}$ item

P_i = i^{th} percentile = value of $\left(\frac{i(N+1)}{100}\right)^{th}$ item

N = total frequency.

→ Continuous Series

$$Q = l + \frac{\frac{N}{4} - cf}{f} \times h. \qquad Q_3 = l + \frac{\frac{3N}{4} - cf}{f} \times h$$

$$M_d = l + \frac{\frac{N}{2} - cf}{f} \times h.$$

$$D_i = l + \frac{\frac{iN}{10} - cf}{f} \times h$$

$$P_i = l + \frac{\frac{iN}{100} - cf}{f} \times h$$

l = lower limit of class in which particular partition value lies

N = total frequency

cf = cf preceeding class in which particular partition value lies

f = corresponding frequency of class in which particular partion valuelie.

h = class size

Mode:

-) Discrete series

Mode = the item having max frequency.

-) Continuous series

$$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - 2} \times h$$

f_1 = max freq

f_0 = frequency preceeding modal class

f_2 = frequency following modal class

h = class length

L = lower limit of modal

Range

→ Range $= \frac{L-S}{2}$

→ Coefficient of range $= \frac{L-S}{L+S}$.

Quartile deviation

→ Quartile deviation $= \frac{Q_3-Q_1}{2}$

→ Coefficient of Q.D $= \frac{Q_3-Q_1}{Q_3+Q_1}$

Mean deviation

→ Individual series

Mean deviation from mean $= \frac{\Sigma |X-\bar{x}|}{n}$

Mean deviation from median $= \frac{\Sigma |X-M_d|}{n}$

Mean deviation from mode $= \frac{\Sigma |X-M_o|}{n}$

→ Discrete / Continuous series.

Mean deviation from mean $= \frac{\Sigma f|X-\bar{x}|}{n}$

Mean deviation from median $= \frac{\Sigma f|X-M_d|}{n}$

Mean deviation from mode $= \frac{\Sigma f|X-M_o|}{n}$

X denotes given item but in continuous series it stands for mid-value.

Coefficient of mean deviation from mean,
$$= \frac{\text{Mean deviation from mean}}{\text{Mean}}.$$

Coefficient of mean deviation from median,
$$= \frac{\text{Mean deviation from median}}{\text{Median}}.$$

Standard Deviation

→ Individual series

$$S.D\ (\sigma) = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

Shortcut,

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} \Rightarrow d = x - A$$

→ Discrete/continuous.

$$S.D\ (\sigma) = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}} = \sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N}\right)^2}$$

Shortcut,

$$S.D(\sigma) = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}, \quad d = x - A$$

Step-deviation,

$$SD\ (\sigma) = \sqrt{\frac{\Sigma f d'^2}{N} - \left(\frac{\Sigma f d'}{N}\right)^2} \times h, \quad d' = \frac{x-A}{h}$$

Variation

Variance $= (\text{Standard deviation})^2 = \sigma^2$

Coefficient of SD $= \frac{\sigma}{\bar{x}}$

Coefficient of variation $= \frac{\sigma}{\bar{x}} \times 100$

Skewness

(i) Karl Person's coeff of Skewness

$$Sk(P) = \frac{\bar{x} - m_o}{\sigma}$$

$$= \frac{3(\bar{x} - m_d)}{\sigma}$$

(ii) Bowley's coeff of skewness

$$Sk(B) = \frac{Q_3 + Q_1 - 2m_d}{Q_3 - Q_1}$$

Correlation

i) Correlation coeff $r = \frac{Cov(x,y)}{\sqrt{Var\,x}\ \sqrt{Var\,y}}$.

$$Cov(x,y) = \frac{1}{n} \Sigma (x-\bar{x})(x-\bar{y})$$

ii) $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\ \sqrt{\Sigma y^2}}$, $x = X - \bar{X}$, $y = Y - \bar{Y}$

iii) $r = \frac{\Sigma xy}{n\,\sigma x\,\sigma y}$

iv) $r = \frac{n\,\Sigma uv - \Sigma u \cdot \Sigma v}{\sqrt{n\,\Sigma u^2 - (\Sigma u)^2}\ \sqrt{n\,\Sigma v^2 - (\Sigma v)^2}}$

$u = X - A$, $v = Y - B$

v) $r = \frac{n\,\Sigma xy - n\,\bar{x}\cdot\bar{y}}{\sqrt{\Sigma x^2 - n(\bar{x})^2}\ \sqrt{\Sigma y^2 - n(\bar{y})^2}}$

Regression.

(i) eqn of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

where, $b_{yx} = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}} = \frac{r\sigma_y}{\sigma_x}$.

REMARKS.

A) • $-3 \le S_K(P) \le 3$

$-1 \le S_K(B) \le 1$

→ if $S_K = 0$, symmetrical

→ if $S_K > 0$, positively skewed

→ if $S_K < 0$, negatively skewed.

B) $-1 \le r \le 1$

$r_{xy} = r_{yx}$

→ correlation coefficient bet^n two variable is independent of change of origin and scale.

C) → Regression coeff are independent of change of origint but not scale

→ $b_{yx} \cdot b_{xy} \le 1$

→ $r = \sqrt{b_{yx} \cdot b_{xy}}$, sign of r must be same as sign of both b_{yx} and b_{xy}.