variance is independent of origin but dependent on the scale of measurement.

Proof:- Let $x_1, x_2, \ldots, x_n$ be a set of n values of a variable x. The arithmatic mean of x is

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

and variance of x is

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

If these values are transformed to a new set of values $y_1, y_2, \ldots, y_n$ such that

$$y_i = \frac{x_i - a}{h}.$$

$$\Rightarrow h y_i = x_i - a.$$

$$\Rightarrow x_i = a + h y_i \quad \text{——} (\mathrm{I})$$

$$\Rightarrow \bar{x} = a + h\bar{y} \quad \text{——} (\mathrm{II})$$

Hence from (I) & (II)

$$x_i - \bar{x} = h(y_i - \bar{y})$$

$$\sum (x_i - \bar{x})^2 = \sum h^2 (y_i - \bar{y})^2$$

$$\Rightarrow \frac{1}{n}\sum (x_i - \bar{x})^2 = h^2 \sum \frac{(y_i - \bar{y})^2}{n}$$

Squaring both sides and then taking sum

$\Rightarrow S_x^2 = h^2 S_y^2$

So variance is independent on origin but dependent on scale of measurement.

: Correlation :-

Correlation means the relationship (linear) among two or more variables.

Correlation coefficient :- correlation coefficient is a quantitative measure of the direction and strength of linear relationship between numerically measured variables.

It is usually denoted by r.

Let x and y are two numerical variables

$$r = \frac{\text{Sum of Products }(X, Y)}{\sqrt{\text{Sum of square }(X) \cdot \text{Sum of square }(Y)}}.$$

$$= \frac{SP(X,Y)}{\sqrt{SS(X)\cdot SS(Y)}}$$

$$= \frac{\Sigma (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma (x_i-\bar{x})^2 \cdot \Sigma (y_i-\bar{y})^2}}$$

$$= \frac{\Sigma x_i y_i - \bar{x}\Sigma y_i - \bar{y}\Sigma x_i + n\bar{x}\bar{y}}{\sqrt{\left\{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right\}\left\{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}\right\}}}$$

$$= \frac{\Sigma x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}}{\sqrt{\left\{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right\}\left\{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}\right\}}}$$

$$= \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\sqrt{\cdots}}$$

Interpretation of r :-

$r = 0 \rightarrow$ variables are uncorrelated

$r = +1 \rightarrow$ There are perfect Positive correlation

$r = -1 \rightarrow$ There are perfect negative "

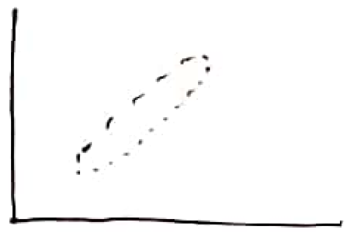
r close to $+1 \rightarrow$ there are strong positive correlation

r close to $-1 \rightarrow$ There are strong negative correlation

r close to $0 \rightarrow$ there are weak correlati

$r = +1$

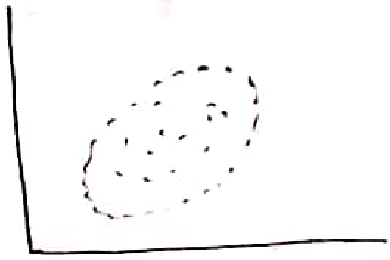
$r = -1$



r close to $+1$

r close to -1

$r = 0$



r close to ze
(+)

r close to zero (negative

The value of r lies between -1 to 1.
i.e. $-1 \leq r \leq 1$

Proof :- we can write

$$\left(\frac{x_i - \bar{x}}{s_x} \pm \frac{y_i - \bar{y}}{s_y} \right)^2 \geq 0$$

$$\Rightarrow \frac{(x_i - \bar{x})^2}{s_x^2} + \frac{(y_i - \bar{y})^2}{s_y^2} \pm \frac{2(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \geq 0$$

$$\Rightarrow \frac{\Sigma (x_i - \bar{x})^2}{s_x^2} + \frac{\Sigma (y_i - \bar{y})^2}{s_y^2}$$

$$\pm 2 \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \geq 0$$

$$\frac{\Sigma (x_i - \bar{x})^2 / n}{s_x^2 / n} + \frac{\Sigma (y_i - \bar{y})^2 / n}{s_y^2 / n}$$

$$\pm \frac{2 \Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{s_x^2 s_y^2}} \geq 0$$

$$\Rightarrow \frac{s_x^2}{s_x^2/n} + \frac{s_y^2}{s_y^2/n} \pm \frac{2\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{\sum(x_i-\bar{x})^2}{n}\sum\frac{(y_i-\bar{y})^2}{n}}}$$

$$\Rightarrow n+n \pm 2n \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2(y_i-\bar{y})^2}}$$

$$\Rightarrow n+n \pm 2nr \geq 0$$

$$\Rightarrow 2n \pm 2nr \geq 0 \Rightarrow 2n(1 \pm r) \geq$$

$$\Rightarrow 1 \pm r \geq 0. \Rightarrow r \geq -1,\ r \leq 1.$$

$$\therefore -1 < r < 1$$

(Proved)

X	5	10	15	20	25
Y	2	6	10	15	20

Find r and Interprete your result.

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
$\sum x_i$	$\sum y_i$	$\sum x_i y_i$	$\sum x_i^2$	$\sum y_i^2$

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left\{\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right\}\left\{\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right\}}}$$

$$r = 0.99$$

i.e. there is strong positive correlation between X and Y.

correlation coefficient is independent on both origin and scale of measurement

Proof:- Let x and y are two variables takes on values $x_1, x_2 \dots x_n$ and $y_1, y_2 \dots y_n$ respectively. Then

$$r_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(x_i - x)^2 (y_i - \bar{y})^2}}.$$

Then us transform x and y as follows

$$u_i = \frac{x_i - a}{h}, \quad v_i = \frac{y_i - b}{k}.$$

$$\Rightarrow x_i = a + hu_i, \quad y_i = b + kv_i$$

$$\Rightarrow \bar{x} = a + h\bar{u}, \quad \bar{y} = b + k\bar{v}$$

$$x_i - \bar{x} = h(u_i - \bar{u}), \quad y_i - \bar{y} = k(v_i - \bar{v})$$

$$r_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 \Sigma (y_i - \bar{y})^2}}$$

$$= \frac{\Sigma h(u_i - \bar{u}) \Sigma k(v_i - \bar{v})}{\sqrt{\Sigma h^2 (u_i - \bar{u})^2 \Sigma k^2 (v_i - \bar{v})^2}}$$

$$= \frac{hk \sum (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{h^2 k^2 \sum (u_i - \bar{u})^2 \sum (v_i - \bar{v})^2}}$$

$$= \frac{hk \sum (u_i - \bar{u})(v_i - \bar{v})}{hk \sqrt{\sum (u_i - \bar{u})^2 \sum (v_i - \bar{v})^2}}$$

$$= \frac{\sum (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum (u_i - \bar{u})^2 \sum (v_i - \bar{v})^2}}$$

$= r_{uv}$.

$\therefore r_{xy} = r_{uv}$

(Proved).