Correlation Coefficient

If large values of x (length) occur with large value of strength (y),

$x \uparrow \quad y \uparrow \qquad\qquad x \downarrow \quad y \downarrow$

$x > \bar{x} \quad y > \bar{y} \qquad\qquad x < \bar{x} \quad y < \bar{y}$

$(x_i - \bar{x}) > 0 \; (y_i - \bar{y}) > 0 \qquad (x_i - \bar{x}) < 0 \; (y_i - \bar{y}) < 0$

$(x_i - \bar{x})(y_i - \bar{y}) > 0 \qquad (x_i - \bar{x})(y_i - \bar{y}) > 0$

$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$ will have a large value.

otherwise, it will have cancellation by + & − values and hence have lower value.

$$\sum_{i=1}^{n} x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x}\bar{y}$$

$$= \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_i x_i - \bar{x} \sum_i y_i + \bar{x}\bar{y} \sum_i$$

$$= \sum_{i=1}^{n} x_i y_i - \bar{y} . n . \bar{x} - \cancel{\bar{x} . n . \bar{y}} + \cancel{\bar{x}\bar{y} . n}$$

$$= \sum_{i=1}^{n} x_i y_i - \cancel{n} . \frac{\sum x_i}{\cancel{n}} . \frac{\sum y_i}{n} = \sum_{i=1}^{n} x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

Scaling factor

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

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

if ~~x~~ $y_i = m x_i$ $\qquad \bar{y} = \frac{\sum y_i}{n} = \frac{\sum m x_i}{n} = m\bar{x}$

$$r = \frac{\sum(x_i-\bar{x})(mx_i-m\bar{x})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(mx_i-m\bar{x})^2}} = \frac{m\sum(x_i-\bar{x})^2}{m\sqrt{\left(\sum(x_i-\bar{x})^2\right)\left(\sum(x_i-\bar{x})^2\right)}}$$

$$= 1$$

r is called the sample correlation coefficient.

if $y_i = -m x_i$

$$r = -1$$