Q) Show that ~~th~~ r is independent of effect of change of origin ~~but~~ ~~no~~ & scale.

let $x_1, x_2 \ldots x_n$ be a set of observation
let us change the origin of 'x' to 'a' & set scale by 'h'
where a & h are arbitary constant
& $(h>0)$

x	a
2	-1
3	0
4	1

The we have

$$u = \frac{x-a}{h}$$

$$uh = x-a$$

$\frac{x-a}{h}$ $\frac{1}{h}$

$$x = a + uh \text{ —— (i)}$$

Now
multiplng & ÷ by Σf_i

$$\frac{\Sigma f x}{\Sigma f} = \frac{a \cancel{\Sigma f}}{\cancel{\Sigma f}} + h \frac{\Sigma f u}{\Sigma f}$$

$\frac{\Sigma f x}{\Sigma f} =$

$$\bar{x} = a + h\bar{u} \text{ —— (ii)}$$

$\frac{\Sigma f u}{\Sigma f} =$

now subtracting 1 & 4

$$x - \bar{x} = a + uh - (a + \bar{u}h)$$

$$= \cancel{a} + uh - \cancel{a} - \bar{u}h$$

$$\Rightarrow x - \bar{x} = uh - \bar{u}h$$

$\Rightarrow x - \bar{x} = h(u - \bar{u})$ — (iii)

Similarly for Y

Let $y_1, y_2 \ldots y_n$

Let us change the origin of y to b's scale by K.

where b & K are arbitrary constants $(K > 0)$

Then we have

$$V = \frac{Y - b}{K}$$

$$\Rightarrow VK = Y - b$$

$$\Rightarrow Y = b + VK \text{ — (iv)}$$

* & ÷ both sides by Σf we will get

$$\Rightarrow \bar{Y} = b + \bar{V}K \text{ — (v)}$$

Subtracting 1 & 4

$$\Rightarrow Y - \bar{Y} = \cancel{b} + VK - \cancel{b} - \bar{V}K$$

$$\Rightarrow Y - \bar{Y} = K(V - \bar{V}) \text{ — (VI)}$$

$$\frac{\sum \{(u-\bar{u})h \quad (v-\bar{v})k\}}{\frac{1}{n}\left[\sqrt{\sum(u-\bar{u})^2h^2} \cdot \sqrt{\sum\{v-\bar{v}\}^2k^2}\right]}$$

$$r = \frac{\cancel{hk}\sum[(u-\bar{u})(v-\bar{v})]}{\cancel{hk}\left[\sqrt{\sum(u-\bar{u})^2} \cdot \sqrt{\sum(v-\bar{v})^2}\right]}$$

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

$$r = \frac{\mathrm{cov}(u,v)}{\sigma_u \sigma_v}$$

$$r_{xy} = r_{uv}$$

Thus the correlation of coefficient is independent of change of origin & scale.

$\sqrt{\frac{1}{n}\sum(x-\bar{x})^2} = \sigma_x$

$\frac{1}{n}\sum(x-\bar{x})^2 = \sigma_x^2$

$\sqrt{\frac{1}{n}}$

$(A+B)^2$

$A^2+2AB+B^2$

$(A-B)^2$

$A^2-2AB+B^2$

Now putting the value of (iii) & (vi) in

$$r = \frac{\frac{1}{n}\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\frac{1}{n}\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2\frac{1}{n}}}$$

$$\sqrt{\frac{1}{n}\sum(x-\bar{x})^2} = \sigma_x$$

$$r = \frac{\frac{1}{n}\sum\left[\{(u-\bar{u})h\ (v-\bar{v})k\}\right]}{\frac{1}{n}\left[\sqrt{\sum\{(u-\bar{u})^2h^2\}}\cdot\sqrt{\sum\{(v-\bar{v})^2k^2\}}\right]}$$

$$\frac{1}{n}\sum(x-\bar{x})^2 \quad \sigma_x^2$$

$$r = \frac{\cancel{hk}\sum[(u-\bar{u})(v-\bar{v})]}{\cancel{hk}\left[\sqrt{\sum(u-\bar{u})^2}\cdot\sqrt{\sum(v-\bar{v})^2}\right]}$$

$$\sqrt{\frac{1}{n}} \quad \sigma_x^2 \quad \sigma_x$$

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

$$r = \frac{\mathrm{cov}(u,v)}{\sigma_u\sigma_v}$$

$$(A+B)^2$$

$$A^2+2AB+B^2$$

$$r_{xy} = r_{uv}$$

$$(A-B)^2$$

$$A^2-2AB+B^2$$

Thus the ~~r~~ correlation of coefficient is independent of change of origin & scale.

2012

Show that $(-1 \leq r \leq +1)$

Proof) Let X & Y be 2 variates, where the value of X tends to $x_1, x_2, \dots x_n$ & value of Y is $y_1, y_2 \dots y_n$.
Also let σ_x & σ_y be the S.D of x & y respectively.

Now

Let us consider the sum of the sq of

$$\sum\left(\frac{x-\bar{x}}{\sigma_x} \pm \frac{y-\bar{y}}{\sigma_y}\right)^2 \geq 0$$

$$\Rightarrow \sum\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2 \pm 2\left(\frac{x-\bar{x}}{\sigma_x}\right)\left(\frac{y-\bar{y}}{\sigma_y}\right)\right\} \geq 0$$

$$\Rightarrow \frac{\sum(x-\bar{x})^2}{\sigma_x^2} + \frac{\sum(y-\bar{y})^2}{\sigma_y^2} \pm 2\frac{\sum(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} \geq 0$$

$\div$ both side by n

$$\Rightarrow \frac{1}{\sigma_x^2}\frac{\sum(x-\bar{x})^2}{n} + \frac{1}{\sigma_y^2}\cdot\frac{\sum(y-\bar{y})^2}{n} \pm \frac{2}{\sigma_x\sigma_y}\frac{\sum(x-\bar{x})(y-\bar{y})}{n} \geq 0$$

$$\Rightarrow \frac{1}{\sigma_x^2}\sigma_x^2 + \frac{1}{\sigma_y^2}\cdot\sigma_y^2 \pm \frac{2}{\sigma_x\sigma_y}\,\mathrm{cov}(x,y) \geq 0$$

$$\Rightarrow 1 + 1 \pm 2r \geq 0 \qquad \left[\,2\frac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y}\right.$$

$$= 2 \pm 2r \geq 0$$

now

either

$2 + 2r \geq 0$

$r \geq -1$

or

$2 - 2r \geq 0$

$2 \geq 2r$

$1 \geq r$

$\therefore -1 \leq r \leq 1$

RANK Correlation

Spearman's Rank correlation of coefficient is defined as

$$R_{12} = 1 - \frac{6 \Sigma D_{12}^2}{N(N^2-1)} \qquad D = R_1 - R_2$$

or

$$R = 1 - \frac{6 \Sigma D^2}{N^3 - N}$$

where

R = Rank coefficient of correlation
D = Difference of Rank betw paired items in 2 series
N = No. Pair of items

Features

① The sum of the differences of ranks betw 2 variables shall be zero '0'. Symbolicaly $\Sigma D = 0$

② Spearman's correlation coefficient is distribution free or non parametric.

(3) The spearman's correlation coefficient is nothing but Karl Pearson's correlation coefficient betw ranks

Property 3

If 2 variates are independent then they are un-correlated.

Let us consider 2 independent variates x & y. Since the variates are independent so from the theorem of expectation, we have:-

$$E(xy) = E(x) \cdot E(y) \quad \text{—— (1)}$$

Now,

$$r_{x,y} = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{E(xy) - E(x)E(y)}{\sigma_x \cdot \sigma_y} \quad \{ Cov(xy) = E(xy) - E(x)E(y)$$

$$= \frac{E(x) \cdot E(y) - E(x)E(y)}{\sigma_x \cdot \sigma_y} \quad \text{—— from (1)}$$

$$= 0$$

//

Thus is 2 variates are independent then they are uncorrelated.

Limitation of correlation coefficient:

Regression (unit-IV)

Regression analysis reveals average relationship betw 2 variables, & thus makes possible estimation or prediction. Regression analysis is used for estimating or predicting the unknow value of a variable for a given ~~valt~~ known value of another variable. More precisely if x & y are 2 related variables then regression analysis helps us to estimates the value of Y for given value of x & vice versa.

According to prof. Blair, "Regression is a mathematical measure of the average relationship between 2 or more variables in terms of original units of the data.

Dependent & Independent variable

The regressional analysis is used for estimating the unknown value of 1 variable from the known value of the other variable. The variable whose value is to be predicted is called as the dependent variable or explained variable. On the other hand, the variable with the help of which prediction is done is called the independent variable or explanatory variable

Types of Regression

If Regression analysis is confined to the study of 2 variables only then it is known as simple regression.

On the other hand the regression analysis that studies more than 2 variables at the same time is called multiple regression.

In case of simple regression if the relation betw the dependent & independent variable follows a straight line pattern then the regression is called as linear regression. However if the relation is expressed in the form of a curve, then that type of regression is called ~~curvilin~~ curvilinear regression.

Simple linear regression / Regression lines

In case of 2 variables, X & Y, we shall have 2 regression lines, i.e regression of X on Y & regression of Y on X. The regressio the independent variable or explanatory variable

Types of Regression

If Regression analysis is confined to the study of 2 variables only then it is known as simple regression.

On the other hand the regression analysis that studies more than 2 variables at the same time is called multiple regression

In case of simple regression if the relation betw the dependent & independent variable follows a straight line pattern then the regression is called as linear regression. How ever if the relation is expressed in the form of a curve, then that type of regression is called ~~curvilin~~ curvilinear regression.

Simple linear regression / Regression lines

In case of 2 variables, x & y, we shall have 2 regression lines, i.e regression on x on y & regression of y on x. The regression line of x on y gives the most probable value of x for given value of y. Similarly the regression line of y on x gives the most probable value of y for given value of x. However, when there is a

Limitation of correlation coefficient

Regression (unit-4)

Regression analysis reveals average relationship b/w 2 variables, & thus makes possible estimation or prediction. Regression analysis is used for estimating or predicting the unknown value of a variable from the given yet known value of another variable. More precisely if x & y are 2 related variables then regression analysis helps us to estimate the value of Y for given value of x & vice versa.

According to Prof. Blair, "Regression is a mathematical measure of the average relationship between 2 or more variables in terms of original units of the data."

Dependent & Independent variable

The regressional analysis is used for estimating the unknown value of 1 variable from the known value of the other variable. The variable whose value is to be predicted is called as the dependent variable or explained variable. On the other hand, the variable with the help of which prediction is done is called perfect correlation betw x & Y ($r = \pm 1$) then the regression lines will co-incides, i.e we have only one line. otherwise, the regression lines intersect each other at point $(\bar{x}, \bar{Y})$.

The regression eqn of x on Y is given by ~~to~~

~~r_{xy}~~

$$X - \bar{X} = b_{xY}(Y - \bar{Y})$$

Here

$\bar{X}$ & $\bar{Y}$ are the A.M. of x & Y respectively

b_{xy} = is the regression coefficient of x on Y.

& is given by

$$b_{xy} = r\frac{\sigma_x}{\sigma_Y}$$

Similarly

The regression eqn of Y on x is given by

$$Y - \bar{Y} = b_{Yx}(x - \bar{x})$$

where,

$$b_{yx} = r\frac{\sigma_Y}{\sigma_x}$$

Difference betⁿ correlation & Regression

1) The correlation coefficient measures the degree of associations between the variables, where as the objective of regression analysis is to estb the mathematical relationship betⁿ the variables.

2) Correlation cannot be used for the purpose of prediction, where as regression analysis is basically used for prediction purpose.

3) The coefficient of correlation varies betⁿ ± 1 where as the regressn eqn varies from $-\alpha$ to $+\alpha$

4) In correlation analyses there is no concept of independent & dependent variable but in regression analysis 1 has to decide which variable is to be taken as the dependent & which one as the independent variable.

5) The correlation is independent of change of scale & origin. But regressn coefficient are independent of change of origin but not of scale.

6) There may be non sense correlation betⁿ 2 variables which is purely due to chance & has no practicle relevance. However there is nothing like non sense regressn.

Properties of Regression Coefficient

1) The range of the regression coefficient is from $-\alpha$ to $+\alpha$

2) The correlation coefficient betw 2 variable is equal to the geomatric mean of the Regressn coefficient. Thus. $r = \sqrt{\frac{b_{xy}}{b_{yx}}}$

$$r = \sqrt{b_{xy} \times b_{yx}}$$

3) The sign of regression coefficients & correlatn coefficients are always the same.

4) The A.M of the 2 regression coefficients is greater then or equal to the values of the correlation coefficient.

5) If 1 of the regression coefficient in a simple linear regression is greater then (1) unity, the other will be less then (1) unity.

i.e $b_{xy} > 1$, $b_{yx} < 1$

&

if $b_{yx} > 1$, $b_{xy} < 1$

6) Regression coefficient is independent of the change of origin but not of scale.

Property Proof

Regression coefficient is independent of change of origin but not of scale.

Let $x_1, x_2, x_3 \ldots x_n$ be a set of observations & $y_1, y_2 \ldots y_n$ be a ~~set~~ another set of "

Then regression coefficient of x on y =

$$b_{xy} = r \frac{\sigma_x}{\sigma_y}$$

$$= \frac{cov(x,y)\,\cancel{\sigma_x}}{\cancel{\sigma_x} \cdot \sigma_y \; \sigma_y}$$

$$= \frac{cov(x,y)}{\sigma_y^2}$$

$$= \frac{\cancel{\frac{1}{n}} \Sigma (x-\bar{x})(y-\bar{y})}{\cancel{\frac{1}{n}} \Sigma (y-\bar{y})^2}$$

$$= \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (y-\bar{y})^2} \quad \text{—— (i)}$$

Now let us change the origin of x to a & scale by h where a & h are arbitrary constants ($h > 0$). Then we have

$$u_1 = \frac{x_i - a}{h}$$

$$\Rightarrow u_1 h = x_i - a$$

$$\Rightarrow x_i = a + u_1 h \quad \text{—— (ii)}$$

* A $\div$ $\frac{\Sigma f_i}{}$ to eqn (2) we get

$$\Rightarrow \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f_i a}{\Sigma f_i} + h \frac{\Sigma f_i u_i}{\Sigma f_i}$$

$$\Rightarrow \overline{X_i} = a + h\,\overline{u_i} \quad \text{—— (III)}$$

from II – III

$$x_i - \overline{X_i} = \cancel{a} + u_1 \cdot h - \cancel{a} - h\overline{u_1}$$

$$x_1 - \overline{X_1} = h(u_1 - \overline{u_1}) \quad \text{—— (IV)}$$

Similarly

Let ~~the~~ us change the origin of y to 'b' & scale by 'k' where ~~b &~~ v & k are arbitrary constants. $(k>0)$. Then, we have

$$v_1 = \frac{y_1 - b_1}{k}$$

$$\Rightarrow v_1 k = y_i - b_i$$

$$\Rightarrow y_i = b_i + v_i k \quad \text{—— (V)}$$

* A $\div$ Σf_i to eq (V)

$$\Rightarrow \frac{\Sigma f_i y_i}{\Sigma f_i} = \frac{\Sigma f_i}{\Sigma f_i} b + k \frac{\Sigma f_i v_i}{\Sigma f_i}$$

$$\Rightarrow \overline{Y} = b + k\,\overline{v} \quad \text{—— (VI)}$$

Probability

...periment: It is an ex...
...ducted repeatedly...
...ditions doesn't...
...y know ...
...at the ...
...exac...

correlatn coefficient
regression coefficient
relation bet a...
regression coeff...

Now from (V) – (VI)

& $Y' - \bar{Y}' = k(V_i - \bar{V})$ —— (VII)

Now substituting eqn (IV) & (VII) in eqn (I) we get.

$$= \frac{\Sigma\{[h(U_i-\bar{U}) \cdot k(V_i-\bar{V})]\}}{\Sigma\{k(V_i-\bar{V})\}^2}$$

$$= \frac{\cancel{\Sigma hk\{\Sigma(U-\bar{U})(V-\bar{V})\}}}{\cancel{k^2 \Sigma\, \Sigma(V-\bar{V})^2}}$$

$$= \frac{\cancel{h\, \Sigma(U_1-\bar{U})}}{\cancel{k\, \Sigma(}}$$

$$= \frac{hk\{\Sigma(U_i-\bar{U})(V_i-\bar{V})\}}{k^2\, \Sigma(V_i-\bar{V})^2}$$

$$= \frac{h}{k}\, \frac{\Sigma(U_i-\bar{U})(V_i-\bar{V})}{\Sigma(V_i-\bar{V})^2}$$

$$= \frac{h}{k}\, b_{uv}$$

Thus regressn eqn is not independent of change of scale.

Proof

correlatn coefficient is G.M betw the 2 regression coefficient or Deduce the relation betn correlatn coefficient & regressn coefficient.

Solu

we know the regressn coefficient or for the line x on y is given by

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} \quad \text{--- (i)}$$

Similarly the regressn eqn of y on x is

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} \quad \text{--- (ii)}$$

from i & ii

$$\Rightarrow b_{xy} \times b_{yx} = r\frac{\sigma_x}{\sigma_y} \times r\frac{\sigma_y}{\sigma_x}$$

$$\Rightarrow b_{xy} \times b_{yx} = r^2$$

$$= r^2 = \sqrt{b_{xy} \times b_{yx}}.$$

Let $b_{xy} > 1$

$$\frac{1}{b_{xy}} \quad \text{--- (1)}$$

we know

$$b_{xy} \times b_{yx} \leq 1$$

$$b_{xy} \times b_{yx} \leq 1$$

$$b_{yx} \leq \frac{1}{b_{xy}} < 1$$