

1) Show that σ is independent of effect of change of origin but in a scale.

Let $x_1, x_2, \dots, x_n$ be a set of observations
 let us change the origin of x to a
 and scale by h
 where a & h are arbitrary constant

$$h (h > 0)$$

Then we have

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

$$uh = x - a$$

$$x = a + uh \quad \text{--- (1)}$$

Now

multiplying 1 by $\sum f_i$

$$\sum f_i x = \sum f_i a + h \sum f_i u$$

$$\bar{x} = a + h \bar{u} \quad \text{--- (2)}$$

now subtracting 1 & 2

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

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

$$2) x - \bar{x} = uh - \bar{u}h$$

$x - a$	3
2	-1
3	0
4	1

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

$$\frac{\sum f_i x}{\sum f_i} = \frac{\sum f_i a}{\sum f_i} + h \frac{\sum f_i u}{\sum f_i}$$

$$\Rightarrow x - \bar{x} = k(u - \bar{u}) \text{ --- (iii)}$$

Similarly for y

Let $y_1, y_2, \dots, y_n$

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

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

Then we have

$$v = \frac{y - b}{k}$$

$$\Rightarrow vk = y - b$$

$$\Rightarrow y = b + vk \text{ --- (iv)}$$

* $\div$ both sides by k we will get

$$\Rightarrow \bar{y} = b + \bar{v} k \text{ --- (v)}$$

subtracting (v) from (iv)

$$\Rightarrow y - \bar{y} = b + vk - b - \bar{v} k$$

$$\Rightarrow y - \bar{y} = k(v - \bar{v}) \text{ --- (vi)}$$

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

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

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

$$\rho = \frac{\text{cov}(u,v)}{\sigma_u \sigma_v}$$

May 2 1894

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

Now putting the value of σ_u & σ_v in

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

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

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

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

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

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

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

$$r = \frac{\text{Cov}(u, v)}{\sigma_u \sigma_v}$$

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

$$(A+B)^2 = A^2 + 2AB + B^2$$

$$(A-B)^2 = A^2 - 2AB + B^2$$

Thus the correlation 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

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

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

$$\Rightarrow 1 + 1 \pm 2r \geq 0$$

$$\Rightarrow 2 \pm 2r \geq 0$$

now

either

$$2 + 2r \geq 0$$

$$r \geq -1$$

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

or

$$2 - 2r \geq 0$$

$$2 \geq 2r$$

$$1 \geq r$$

$$\left[\frac{2 \text{cov}(x, y)}{\sigma_x \sigma_y} \right]$$

RANK Correlation

Spearman's Rank correlation coefficient is defined as

$$R_{12} = 1 - \frac{6 \sum D_{12}^2}{N(N^2 - 1)}$$

$$D = R_1 - R_2$$

or

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

where

R = Rank coefficient of correlation

D = Difference of rank bet^w paired items
in 2 series

N = No. of items

Features

- ① The sum of the differences of ranks bet^w 2 variables shall be zero (0). Symbolically
 $\sum D = 0$
- ② Spearman's correlation coefficient is distributed free or non parametric.
- ③ The Spearman's correlation coefficient is nothing but Karl Pearson's correlation coefficient bet^w ranks

Property 3

If 2 variates are independent then they are uncorrelated.

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

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

Now,

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

$$= \frac{E(xy) - E(x)E(y)}{\sigma_x \cdot \sigma_y} \quad \left\{ \text{COV}(xy) = E(xy) - E(x)E(y) \right\}$$

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

$$= 0$$

//
Thus if 2 variates are independent then they are uncorrelated.

Limitation of correlation coefficient:

Regression (unit-IV)

Regression analysis reveals average relationship betⁿ 2 variables, & thus makes possible estimation or prediction. Regression analysis is used for estimating or predicting the unknown value of a variable for a given or 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. A

According to prof. Blain, "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 regression 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. How ever if the relation is expressed in the form of a curve, then that type of regression is called 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 and Y & regression of Y on X . The regression

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 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 (various)

Regression analysis removes average relation ship between 2 variables, & thus makes precise estimation or prediction. Regression analysis is used for estimating or predicting the unknown value of a variable from a given or known value of another variable. Thus frequently 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. A

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

Dependent & Independent Variables

The regression analysis is used for estimating the unknown value of y from the known value of the other variable x .

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

Types of Regression

1. Simple

2. Multiple

3. Stepwise

4. Logistic

5. Linear

6. Non-linear

7. Polynomial

8. Quadratic

9. Cubic

10. Quartic

11. Quintic

12. Sextic

13. Septic

14. Octic

15. Nonic

16. Decic

17. Undecic

18. Duodecic

19. Tredecic

20. Quattuordecic

21. Quindecic

22. Sexdecic

23. Septendecic

24. Octodecic

25. Nonadecic

26. Vigintic

perfect correlation between x & y ($r = \pm 1$) then the regression lines will co-incide, 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

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

Here $\bar{x}$ & $\bar{y}$ are the A.M. of x & y respectively

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

It is given by

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

Similarly

The regression eqⁿ of y on x is given by

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

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, whereas the objective of regression analysis is to establish the mathematical relationship betⁿ the variables.
- 2) Correlation cannot be used for the purpose of prediction, whereas regression analysis is basically used for prediction purpose.
- 3) The coefficient of correlation varies betⁿ ± 1 whereas the regression eqⁿ varies from $-\infty$ to $+\infty$.
- 4) In correlation analysis there is no concept of independent & dependent variable but in regression analysis I 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 regression 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 practical significance, however there is nothing like non sense regression.

Properties of Regression coefficient

- 1) The range of the regression coefficient is from -1 to $+1$.
- 2) The correlation coefficient betⁿ 2 variables is equal to the geometric mean of the Regression coefficient. Thus $r = \sqrt{\frac{b_{xy}}{b_{yx}}}$
$$r = \sqrt{b_{xy} \times b_{yx}}$$
- 3) The sign of regression coefficients & correlation coefficients are always the same.
- 4) The A.M of the 2 regression coefficients is greater than or equal to the value of the correlation coefficient.
- 5) If 1 of the regression coefficient in a simple linear regression is greater than (1) unity, the other will be less than (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 $x_1, x_2, x_3, \dots, x_n$ be a set of observations

$y_1, y_2, \dots, y_n$ be a set of another set of y .

Then regression coefficient of x on $y =$

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

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

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

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

$$= \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (y - \bar{y})^2} \quad \text{--- (1)}$$

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

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

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

$$\Rightarrow x_i = a + u_1 h \quad \text{--- (2)}$$

* 4 $\div \frac{\sum f_i}{N}$ to eqⁿ we get

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

$$\Rightarrow \bar{X}_i = a + h \bar{u}_i \quad \text{--- (iii)}$$

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

$$x_i - \bar{X}_i = a + u_i h - a - h \bar{u}_i$$

$$x_i - \bar{X}_i = h(u_i - \bar{u}_i) \quad \text{--- (iv)}$$

Similarly

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

$$v_i = \frac{y_i - b}{k}$$

$$\Rightarrow v_i k = y_i - b$$

$$\Rightarrow y_i = b + v_i k \quad \text{--- (v)}$$

* 4 $\div \sum f_i$ to eq (v)

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

$$\Rightarrow \bar{Y} = b + k \bar{v} \quad \text{--- (vi)}$$

Probability

Experiment: It is an empirical method. Repeated trials don't know the outcome.

prob
correlation coefficient
regression coefficient
collation coefficient
regression coefficient

Now from (i)-(vi)

$$y_i - \bar{y} = k(x_i - \bar{x}) \quad \text{--- (vii)}$$

Now substituting eqn (vii) & (viii) in eqn (i) we get:

$$= \frac{\sum \{ [k(x_i - \bar{x}) - k(x_i - \bar{x})] \}}{\sum \{ k(x_i - \bar{x}) \}^2}$$

$$= \frac{\sum \{ k \} \sum (x_i - \bar{x})(x_i - \bar{x})}{k^2 \sum (x_i - \bar{x})^2}$$

$$= \frac{h \sum (x_i - \bar{x})}{k \sum (x_i - \bar{x})^2}$$

$$k \sum (x_i - \bar{x})$$

$$= \frac{h \sum (x_i - \bar{x})(x_i - \bar{x})}{k \sum (x_i - \bar{x})^2}$$

$$= \frac{h}{k} \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$$= \frac{h}{k} \text{ b.u.v}$$

Thus regression eqn is not independent of change of scale.

$$1 - \frac{1}{n} \sum x_i^2 = \frac{1}{n} \sum x_i^2$$

$$1 \geq \frac{1}{n} \sum x_i^2$$

Proof

correlation coefficient is G.M bet^w the 2 regression coefficient or reduce the relation betⁿ correlation coefficient & regression coefficient.

Soln

we know the regression coefficient for the line n on y is given by

$$b_{ny} = \frac{\sum (x_n - \bar{x})(y - \bar{y})}{\sum (y - \bar{y})^2} \quad \text{--- (1)}$$

Similarly the regression eqⁿ of y on n is

$$b_{yn} = r \frac{\sigma_y}{\sigma_n} \quad \text{--- (2)}$$

from 1 & 2

$$\Rightarrow b_{ny} \times b_{yn} = \left(r \frac{\sigma_y}{\sigma_n} \right) \times \left(r \frac{\sigma_n}{\sigma_y} \right)$$

$$= r^2 \frac{(\sigma_y)^2 (\sigma_n)^2}{(\sigma_y)^2 (\sigma_n)^2} = r^2$$

$$r^2 = \sqrt{b_{ny} \times b_{yn}}$$

Let $b_{ny} > 1$

$$\frac{1}{b_{ny}} \quad \text{--- (3)}$$

are known

$$b_{ny} \times b_{yn} \leq 1$$

$$b_{ny} \times b_{yn} \leq 1$$

$$b_{yn} \leq \frac{1}{b_{ny}} < 1$$