Chapter (13)

Linear Regression and Correlation

(Examples)

The Coefficient of coefficient

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{(n-1) S_x S_y}$$

$$S_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}}$$

Example (1)

If we want compared between students scales in the statistics(Y) and math (X), we selected random sample by size 8 students, they are marks follows:

Y	**1**	**2**	**4**	**4**	**5**	**7**	**8**	**9**
X	1	3	4	6	8	9	11	14

Compute the S_x , S_y, r

Solution:

Y	**1**	**2**	**4**	**4**	**5**	**7**	**8**	**9**	$\sum Y =$ **40**
X	**1**	**3**	**4**	**6**	**8**	**9**	**11**	**14**	$\sum X =$ **56**
$Y - \bar{Y}$	**−4**	**−3**	**−1**	**−1**	**0**	**2**	**3**	**4**	$\sum (Y - \bar{Y}) =$ **0**
$X - \bar{X}$	**−6**	**−4**	**−3**	**−1**	**1**	**2**	**4**	**7**	$\sum X - \bar{X} =$ **0**
$(Y - \bar{Y})^2$	**16**	**9**	**1**	**1**	**0**	**4**	**9**	**16**	$\sum (Y - \bar{Y})^2 =$ **56**
$(X - \bar{X})^2$	**36**	**16**	**9**	**1**	**1**	**4**	**16**	**49**	$\sum (X - \bar{X})^2 =$ **132**
$(X - \bar{X})(Y - \bar{Y})$	**24**	**12**	**3**	**1**	**0**	**4**	**12**	**28**	$\sum (X - \bar{X})(Y - \bar{Y}) =$ **84**

$$\bar{Y} = \frac{40}{8} = 5 \quad \bar{X} = \frac{56}{8} = 7$$

$$S_{xy} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{n-1} = \frac{84}{7} = 12$$

$$S_X = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{\frac{132}{7}} = \sqrt{18.86} = 4.34$$

$$S_y = \sqrt{\frac{\sum(Y-\bar{Y})^2}{n-1}} = \sqrt{\frac{56}{7}} = \sqrt{8} = 2.83$$

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{(n-1)S_x S_y} = \frac{84}{7\times 4.34\times 2.83} = \frac{84}{85.9754} = 0.977$$

The relationship (Positive & Strong)

Spearman’s coefficient

This coefficient is applied when the two variables are quantitative & quantitative (ordinal) . It’s called **“Rank correlation coefficient”.** The formula used to calculate “rs” is given by:

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

Example (2):

The following data grades of (11) students in the two subjects

No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Sub (1)	C	F	D	C	A	B	C	A	F	D	B
Sub (2)	B	D	D	B	A	A	C	B	D	F	C

Compute the Spearman’s coefficient.

Solution:

The ranks

Sub (1)	**A**	**A**	**B**	**B**	**C**	**C**	**C**	**D**	**D**	**F**	**F**
Rank(1)	1.5	1.5	3.5	3.5	6	6	6	8.5	8.5	10.5	10.5
Sub (2)	A	A	B	B	B	C	C	D	D	D	F
Rank(2)	1.5	1.5	4	4	4	6.5	6.5	9	9	9	11

No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Total**
Sub (1)	C	F	D	C	A	B	C	A	F	D	B	
Sub (2)	B	D	D	B	A	A	C	B	D	F	C	
R(1)	6	10.5	8.5	6	1.5	3.5	6	1.5	10.5	8.5	3.5	
R(2)	4	9	9	4	1.5	1.5	6.5	4	9	11	6.5	
D	2	1.5	−0.5	2	0	2	−0.5	−2.5	1.5	−2.5	−3	
D^2	4	2.25	0.25	4	0	4	0.25	6.25	2.25	6.25	9	38.5

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)} \qquad = 1-[6(38.5)/11(120)] = 0.83$$

Example (3):

If x independent variable (height) and y dependent variable (weight) Compute the Spearman's coefficient.

X	**160**	**165**	**165**	**170**	**162**
Y	56	64	60	66	56

Solution:

X	Y)	R(X)	R(Y)	D	D^2
160	56	1	1.5	-0.5	0.25
165	64	3.5	4	-0.5	0.25
165	60	3.5	3	0.5	0.25
170	66	5	5	0	0
162	56	2	1.5	0.5	0.25
					$\sum d^2 = 1$

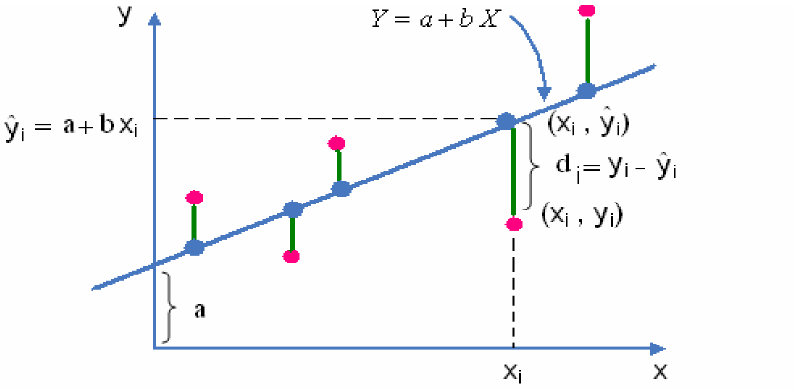
$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$= \quad 1 - \frac{6(1)}{5(25-1)} = 1 - \frac{6}{120} = 1 - 0.05 = 0.95$$

Regression

The **straight line equation for simple regression** is:

$$Y = \alpha + \beta X$$



A plot of the data point (in red), the least squares line of best fit (in blue), and the deviations (in green)

$$\hat{Y} = \alpha + \beta X$$

$$\beta = r\frac{S_y}{S_x}$$

$$\alpha = \overline{Y} - \beta\overline{X}$$

Example (**4**):

Recall example (1)

-Estimated regression Y on X.

- Find $\hat{Y}$ if X = 12.

-Find the (R^2)

Solution:

$$\hat{Y} = \alpha + \beta X$$

$$\beta = r\frac{S_y}{S_x} = 0.977\left(\frac{2.83}{4.34}\right) = 0.64$$

$$\alpha = \overline{Y} - \beta\overline{X}$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{40}{8} = 5$$

$$\bar{X} = \frac{\sum X}{n} = \frac{56}{8} = 7$$

$$\alpha = \bar{Y} - \beta\bar{X} = 5 - 0.64(7) = 5 - 4.48 = 0.52$$

$$\hat{Y} = \alpha + \beta X = 0.52 + 0.64X$$

$$\hat{Y} = 0.52 + 0.64X$$

If X=12

$$\hat{Y} = 0.52 + 0.64(12) = 8.2$$

The Coefficient of Determination (R^2)

The Coefficient of Determination (R^2):

$$R^2 = \frac{\hat{\beta}\sum(X - \bar{X})(Y - \bar{Y})}{\sum(Y - \bar{Y})^2}$$

$$R^2 = r_{xy}^2$$

Example (**5**):

Y	X	$(X - \bar{X})$	$(Y - \bar{Y})$	$(X - \bar{X})(Y - \bar{Y})$	$(Y - \bar{Y})^2$
1	1	-6	-4	24	16
2	3	-4	-3	12	9
4	4	-3	-1	3	1
4	6	-1	-1	1	1
5	8	1	0	0	0
7	9	2	2	4	4
8	11	4	3	12	9
9	14	7	4	28	16
40	56	0	0	84	56

$$R^2 = \frac{\hat{\beta}\sum(X-\bar{X})(Y-\bar{Y})}{\sum(Y-\bar{Y})^2} = 0.64(84)\ /\ 56\ =\ 0.96\ = 96\%$$

$$R^2 = r_{xy}^2 = 0.98^2\ =\ 0.96\ =\ 96\%$$

The independent variable X explained 96% from the variation in the dependent variable Y, other variables (called random variables) explained only 4%.

Example (6):

If : $r = 0.88\ , S_y = 4.84\ \ , S_x = 3.99$

Find $\hat{\beta}$

Solution:

$$\hat{\beta} = R\frac{S_y}{S_x} = (0.88)\frac{4.84}{3.99} = 1.07$$

As X increases Y increases

Example (7):

If $\hat{Y} = 6 - 1.2X$

Find $\bar{y}$ If $\bar{x} = 4$

Solution:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$6 = \bar{Y} - (-1.2)\,4$$

$$6 = \bar{Y} + 4.8$$

$$\bar{Y} = 6 - 4.8 = 1.2$$