Answer on Question #73785 – Math – Statistics and Probability

Consider the following five data points:

X	-1	0	1	2	3
Y	-1	1	2	4	5

a) use regression analysis to calculate by hand the estimated coefficients of the equation $Y = B + AX$

Solution

$$mean\ of\ x = \overline{x} = \frac{\sum x_i}{n}$$

$$mean\ of\ y = \overline{y} = \frac{\sum y_i}{n}$$

$$trend\ line: Y = B + AX, A = \frac{S_{xy}}{S_{xx}}\ , B = \overline{y} - A\overline{x}$$

$$S_{xx} = \frac{\sum(x_i - \overline{x})^2}{n} = \frac{\sum(x_i)^2}{n} - (\overline{x})^2$$

$$S_{yy} = \frac{\sum(y_i - \overline{y})^2}{n} = \frac{\sum(y_i)^2}{n} - (\overline{y})^2$$

$$S_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{\sum x_i \cdot y_i}{n} - \overline{x} \cdot \overline{y}$$

$$mean\ of\ x = \overline{x} = \frac{-1 + 0 + 1 + 2 + 3}{5} = 1$$

$$mean\ of\ y = \overline{y} = \frac{-1 + 1 + 2 + 4 + 5}{5} = 2.2$$

$$S_{xx} = \frac{(-1)^2 + (0)^2 + (1)^2 + (2)^2 + (3)^2}{5} - (1)^2 = 2$$

$$S_{yy} = \frac{(-1)^2 + (1)^2 + (2)^2 + (4)^2 + (5)^2}{5} - (2.2)^2 = 4.56$$

$$S_{xy} = \frac{(-1)(-1) + (0)(1) + (1)(2) + (2)(4) + (3)(5)}{5} - (1)(2.2) = 3$$

$$A = \frac{3}{2} = 1.5$$

$$B = 2.2 - 1.5(1) = 0.7$$

$$Y = 0.7 + 1.5X$$

(b) compute the standard error and the t-statistics for the coefficient of determination of X

Solution

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

the coefficient of determination of X

$$r^2 = \frac{\left(S_{xy}\right)^2}{S_{xx}S_{yy}}$$

$r^2 = \dfrac{(3)^2}{2(4.56)} = 0.986842105$

$r = 0.993399267$

x	y	$\hat{y}$	$\hat{y} - y$	$(\hat{y} - y)^2$
-1	-1	-0.8	0.2	0.04
0	1	0.7	-0.3	0.09
1	2	2.2	0.2	0.04
2	4	3.7	-0.3	0.09
3	5	5.2	0.2	0.04

$SST = \sum (y_i - \overline{y})^2 = 22.8$

$SSR = \sum (\hat{y}_i - \overline{y})^2 = 22.5$

$r^2 = \dfrac{SSR}{SST} = \dfrac{22.5}{22.8} \approx 0.986842105$

$SEE = \sqrt{\dfrac{\sum(\hat{y} - y)^2}{n - 2}}$

$SEE = \sqrt{\dfrac{0.04 + 0.09 + 0.04 + 0.09 + 0.04}{5 - 2}} = \sqrt{0.1} \approx 0.316227766$

Standard error of slope

$SE_A = \dfrac{SEE}{\sqrt{\sum(x - \overline{x})^2}}$

$SE_A = \dfrac{\sqrt{0.1}}{\sqrt{(-1-1)^2 + (0-1)^2 + (1-1)^2 + (2-1)^2 + (3-1)^2}} = 0.1$

The test for the significance of regression for the data in the table

The test is carried out using the $t -$ test on the coefficient A.

$H_0: A = 0$

$H_1: A \neq 0$

The test statistic can be calculated using the following equation:

$$Test\ Statistic: t_0 = \frac{A-0}{SE_A} = \frac{1.5}{0.1} = 15$$

This statistic based on the t distribution with 3 ($n-2=5-2=3$) degrees of freedom can be obtained as follows:

$\alpha = 0.01$

$\alpha/2 = 0.005$

$df = n-2 = 5-2 = 3$

$Critical\ value: = 5.841$

Statistical conclusion:

Reject H_0, there is evidence that A is not equal to zero and that a significant relationship exists between y and x.

The coefficient of non-determination was used in the $t-$test to see if there was significant linear correlation.

The standard error of the estimate is the square root of the coefficient of non-determination divided by its degrees of freedom.

$$s_e = \sqrt{\frac{1-r^2}{n-2}}$$

$$s_e = \sqrt{\frac{1-0.986842105}{5-2}} \approx 0.066226618$$

The test is carried out using the $t-$ test on the coefficient $1-r^2$.

$H_0: 1-r^2 = 1$

$H_1: 1-r^2 < 0$

The test statistic can be calculated using the following equation:

$$Test\ Statistic: t_0 = \frac{(1-r^2)-1}{s_e} = \frac{-0.986842105}{0.066226618} \approx -14.90098898$$

This statistic based on the t distribution with 3 ($n-2=5-2=3$) degrees of freedom can be obtained as follows:

$\alpha = 0.01$

$df = n-2 = 5-2 = 3$

$Critical\ value: = 4.54070$

Statistical conclusion:

Reject H_0, there is evidence that $(1-r^2)$ is not equal to 1 and that a significant relationship exists between y and x.