Maths & Stats Pre-Sessional Tutorial

Topic 6: Regression Analysis - Solutions

Exercise 1

A large consumer goods company has been studying the effect of advertising on total profits. As part of this study, data on advertising expenditures (X) and total sales (Y) were collected for a 5-month period and are as follows:

(x, y) : (10,100) (15,200) (7,80) (12,120) (14,150).

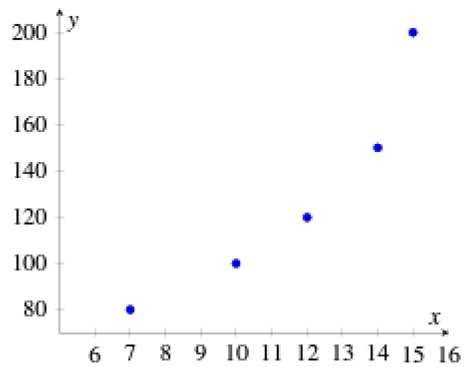
a) Plot the data.

b) Does the plot provide evidence that advertising has a positive effect on sales?

c) Knowing that Cov(x, y) = 140 and s $Cov\ (X, Y) = 140$ and $s_X^2 = 10.3$, compute the regression coefficients $\widehat{\beta_0}$ and $\widehat{\beta_1}$.

Solution:

a)



b) It appears that advertising has a positive effect on sales. Note that correlation does not imply causation. Other factors could have been changing at the same time that advertising changed. For example, prices of competitive goods, or tastes and preferences of consumers, or the number of buyers in the market could have been changing.

c)

(c) $Cov(x, y) = 140$ and $s_x^2 = 10.3$ imply that:

$$b_1 = \frac{140}{10.3} = 13.5922.$$

Then, compute the mean of x and y:

$$\bar{x} = 11.6 \quad \bar{y} = 130$$

and find $b_0 = \bar{y} - b_1\bar{x} = 130 - 13.5922 \cdot 11.6 = -27.669$.

Exercise 2

An aircraft company wanted to predict the number of worker-hours necessary to finish the design of a new plane. Relevant explanatory variables were thought to be the plane's top speed, its weight, and the number of parts it had in common with other models built by the company. A sample of 27 of the company's planes was taken, and the following model was estimated:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

where y_i is the design effort, in millions of worker-hours; x_{1i} is the plane's top speed, in miles per hour; x_{2i} is plane's weight, in tons; x_{3i} is the percentage number of parts in common with other models.

The estimated regression coefficients were as follows:

$$\widehat{\beta_1} = 0.661$$

$$\widehat{\beta_2} = 0.065$$

$$\widehat{\beta_3} = -0.018$$

Provide a definition of estimate and interpret the estimates above.

Solution:

The estimated regression slope coefficients are interpreted as follows:

$\widehat{\beta_1}$ =0.661. All else equal, an increase in the plane's top speed by one mph will increase the expected number of hours in the design effort by an estimated 0.661 million, or 661 thousand working hours.

$\widehat{\beta_2}$ =0.065. All else equal, an increase in the plane's weight by one ton will increase the expected number of hours in the design effort by an estimated 0.065 million, or 65 thousand working hours.

$\widehat{\beta_3}$ = −0.018. All else equal, an increase in the percentage of parts in common with other models will result in a decrease in the expected number of hours in the design effort by an estimated 0.018 million, or 18 thousand working hours.

Exercise 3

Consider the following linear model estimated for an industrial sector:

$$\hat{y} = 10 + 5x_1 + 4x_2 - 2x_3$$

where: *y* is profit (in thousand $) for the firm, x_1 is the number of workers, x_2 is the average number of years of education of the firm's workforce, and x_3 is the number of competitors.

a) Compute $\hat{y}$ when $x_1 = 20,\ x_2 = 11\ and\ x_3 = 10$
b) Compute $\hat{y}$ when $x_1 = 15,\ x_2 = 14\ and\ x_3 = 20$

Solution:

a) $\hat{y} = 10 + 5 * 20 + 4 * 11 - 2 * 10 = 134$

b) $\hat{y} = 10 + 5 * 15 + 4 * 14 - 2 * 20$

Exercise 4

Consider the following regression of the percentage change in the Dow Jones index in a year on the percentage change in the index over the first 5 trading days of the year:

$$\hat{y} = 12.942 - 2.034x$$

n = 13, and the standard error of the slope is 1.378.

(a) Find and interpret a 95% confidence interval for the slope of the population regression line.

(b) Test at the 10% significance level, against a two-sided alternative, the null hypothesis that the slope of the population regression line is 0

Solution:

a)

First, using the table provided, find the relevant value for $t_{n-2,\alpha/2}$. In this case, $t_{11,0.025} = 2.201$. Hence, the 95% confidence interval is the following:

$$[-2.034 - 2.201(1.378);\ -2.034 + 2.201(1.378)]$$

$$[-5.0673;\ +0.9991]$$

b)

The hypothesis to test is the following: $\mathbf{H}_0 : \beta = 0$; $\mathbf{H}_1 : \beta \neq 0$. Compute the t-statistic as follows:

$$t = \frac{-2.034}{1.378} = -1.48.$$

Since, $t_{11,0.05} = 1.796$, we do not reject $\mathbf{H}_0$ at the 10% level. Indeed, $-1.48 > -1.796$.