PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit-5 - Power of Test and Simple Linear Regression

QUESTION BANK - SOLVED

The Least squares line:

Exercises for section 7.2: [Text Book Exercise 7.2 – Pg. No. [536 – 539]]

1. A least-squares line is fit to a set of points. If the total sum of squares is $\sum(y_i - \bar{y})^2 = 181.2$, and the error sum of squares is $\sum(y_i - \hat{y})^2 = 33.9$, compute the coefficient of determination r^2.

   [Text Book Exercise – Section 7.2 – Q. No.4 – Pg. No. 536]

   Solution:
   Given
   Total sum of squares =$\sum(y_i - \bar{y})^2 = 181.2$,
   Error sum of squares = $\sum(y_i - \hat{y})^2 = 33.9$,

   We know that
   Total sum of squares = Regression sum of squares + Error sum of squares

   Hence Regression sum of squares= 181.2-33.9 = 147.3

   Also we know that

$$r^2 = \frac{Regression\ sum\ of\ squares}{Total\ sum\ of\ squares} = \frac{147.3}{181.2} \approx 0.8129$$

2. Curing times in days (x) and compressive strengths in MPa (y) were recorded for several concrete specimens. The means and standard

deviations of the x and y values were $\bar{x} = 5,\ s_x = 2, \bar{y} = 1350, s_y = 100$. The correlation between curing time and compressive strength was computed to be $r = 0.7$. Find the equation of the least-squares line to predict compressive strength from curing time.

[Text Book Exercise – Section 7.1 – Q. No.12 – Pg. No. 538]

Solution:

Given $\bar{x} = 5,\ s_x = 2, \bar{y} = 1350, s_y = 100, r = 0.7$

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x$$

The slope is the product of the correlation coefficient and the ratio of the standard deviations.

$$\widehat{\beta_1} = r\frac{s_y}{s_x} = 0.7\frac{100}{2} \approx 35$$

Next, we determine the estimate of y-intercept

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1} x = 1350 - 35(5) = 1175$$

The equation of the least square line is

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x$$

Hence,

$$\hat{y} = 1175 + 35x$$