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

Predictions using regression models - Uncertainties in Regression Coefficients.

Exercises for section 7.3: [Text Book Exercise 7.3– Pg. No. [554 – 559]]

1. The following output (from MINITAB) describes the fit of a linear model $y = \beta_o + \beta_1 x + \varepsilon$ that expresses the length of a spring in cm (y) in terms of the load upon it in kg (x). There are $n = 15$ observations.

Predictor	Coef	StDev	T	P
Constant	6.6361	1.1455	5.79	0.000
Load	2.9349	0.086738	33.8	0.000

   a. How many degrees of freedom are there for the Student's t statistics?
   b. Find a 98% confidence interval for β_1
   c. Find a 98% confidence interval for β_o
   d. Someone claims that if the load is increased by 1 kg, that the length will increase by exactly 0.35 cm. Use the given output to perform a hypothesis test to determine whether this claim is plausible.

e. Someone claims that the unloaded (load = 0) length of the spring is more than 1.5 cm. Use the given output to perform a hypothesis test to determine whether this claim is plausible.

[Text Book Exercise – Section 7.3 – Q. No.2 – Pg. No. 555]

Solution:

n = sample size = 15

(a) The degrees of freedom is df = n-2 = 15-2 = 13

(b) c = confidence level = 98% = 0.98

From the MINITAB output,

$$\widehat{\beta_1} = 2.9349$$

$$s_{\beta_1} = 0.086738$$

The critical t-value can be found in the Students T distribution table in the row of df=13 and in the column of α=1-c/2=0.01

t = 2.650

The boundaries of the confidence interval for $\widehat{\beta_1}$ are

$$\widehat{\beta_1} - t \times s_{\beta_1} = 2.9349 - 2.650 \times 0.086738 \approx 2.7050$$

and

$$\widehat{\beta_1} + t \times s_{\beta_1} = 2.9349 + 2.650 \times 0.086738 \approx 3.1648$$

We are 98% confident that the increase in spring length caused by an increase in load of 1 kg is between 2.7050 cm and 3.1648 cm.

(c)

c = confidence level = 98% = 0.98

From MINITAB output,

$$\widehat{\beta_0} = 6.6361$$

$$s_{\beta_0} = 1.1455$$

The critical t-value can be found in the Students T distribution table in the row of df=13 and in the column of α=1-c/2=0.01

t = 2.650

The boundaries of the confidence interval for $\widehat{\beta_0}$ are

$$\widehat{\beta_0} - t \times s_{\beta_0} = 6.6361 - 2.650 \times 1.1455 \approx 3.6005$$

and

$$\widehat{\beta_0} + t \times s_{\beta_0} = 6.6361 + 2.650 \times 1.1455 \approx 9.6717$$

We are 98% confident that the length of the spring is between 3.6005 cm and 9.6717 cm when the load is 0 kg.

(d) There is sufficient evidence to support the claim that the length will not increase by exactly 0.35 cm when the load is increased by 1 kg.

(e)

$$H_0: \beta_0 \leq 1.5$$

$$H_1: \beta_0 > 1.5$$

From MINITAB output,

$$\widehat{\beta_1} = 6.6361$$

$$s_{\widehat{\beta_1}} = 1.1455$$

The value of the test statistic

$$t = \frac{\widehat{\beta_1} - \beta_1}{s_{\widehat{\beta_1}}} = \frac{6.6361 - 1.5}{1.1455} \approx 4.484$$

Find the P-value for df = n-2 = 15-2 = 13

$$P < 0.0005$$

If the P-value is less than or equal to the significance level, then the null hypothesis is rejected:

$P < 0.05$ => Reject H_0

There is sufficient evidence to support the claim that the unloaded length of the spring is more than 1.5 cm.

2. In the manufacture of synthetic fiber, the fiber is often "set" by subjecting it to high temperatures. The object is to improve the shrinkage properties of the fiber. In a test of 25 yarn specimens, the relationship between temperature in ◦C (x) and shrinkage in % (y) was summarized by the least-squares line $y = -12.789 + 0.133x$. The total sum of squares was $\sum(y_i - \bar{y})^2 = 57.313$, and the estimated error variance was s^2 = 0.0670. Compute the coefficient of determination r^2.

   [Text Book Exercise – Section 7.3 – Q. No.14 – Pg. No. 558]

   Solution:

   Given

$$s^2 = 0.0670,$$

   The total sum of squares = $\sum(y_i - \bar{y})^2 = 57.313,$

$$n = 25$$

$$s^2 = \frac{Error\ sum\ of\ squares}{n-2}$$

$$\rightarrow Error\ sum\ of\ squares = (n-2)s^2 = (25-2)(0.0670) \approx 1.541$$

We know that

Residual sum of squares = total sum of squares - error sum of squares
= 57.313 - 1.541 = 55.772

$$r^2 = \frac{Residual\ sum\ of\ squares}{Total\ sum\ of\ squares} = \frac{55.722}{57.313} \approx 0.9731$$