uwa6xv_M04_HW

Alanna Hazlett

2024-02-20

Problem 1

Fit an appropriate linear regression and answer the following questions

```
copier<-read.table(file="copier.txt", header = TRUE)
result<-lm(Minutes~Serviced,data=copier)
```

(a)
Obtain the 95% confidence interval for the slope, beta 1.

```
confint(result,level=0.95)[2,]
```

```
##    2.5 %   97.5 %
## 14.06101 16.00949
```

The 95% confidence interval for the slope is (14.06, 16.01).
(b)
Suppose a service person is sent to service 5 copiers. Obtain an appropriate 95% interval that predicts the total service time spent by the service person.

```
newdata<-data.frame(Serviced=5)
predict(result,newdata,level=0.95,interval="prediction")
```

```
##        fit      lwr      upr
## 1 74.59608 56.42133 92.77084
```

The predicted amount of time for a service person to service 5 copiers, with 95% confidence, is (56.42, 92.77) minutes.

(c)
What is the value of the residual for the first observation? Interpret this value contextually.

```
summary(result)
```

```
##
## Call:
## lm(formula = Minutes ~ Serviced, data = copier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7723  -3.7371   0.3334   6.3334  15.4039
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5802     2.8039  -0.207    0.837
## Serviced     15.0352     0.4831  31.123   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.914 on 43 degrees of freedom
## Multiple R-squared:  0.9575, Adjusted R-squared:  0.9565
## F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16
```

$$\hat{y} = -0.5802 + 15.0352x$$

$$\hat{y} = -0.5802 + 15.0352(2)\hat{y} = 29.4902$$

$$e_i = y_i - \hat{y}_i$$

$$e_i = 20 - 29.4902 = -9.4902\ minutes$$

The value of our residual states that the actual time it took to service 2 copiers took about 9 minutes less than our regression predicted.

Problem 2

A substance used in biological and medical research is shipped by airfreight to users in cartons of 1000 ampules. The data consist of 10 shipments. The variables are number of times the carton was transferred from one aircraft to another during the shipment route (transfer), and the number of ampules found to be broken upon arrival (broken).

(a)

Carry out a hypothesis test to assess if there is a linear relationship between the variables of interest.

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

The test statistic is:

$$t = \frac{\hat{\beta}_1 - 50}{se(\hat{\beta}_1)} = \frac{4 - 0}{0.4690} = 8.528785$$

The critical value is:

```
qt(0.975,df=10-2)
```

```
## [1] 2.306004
```

Our test statistic is greater than our critical value, so we reject the null hypothesis. Our data supports the alternative hypothesis that the slope does not equal zero. This shows that there is a linear association between our variables.

(b)

Calculate the 95% confidence interval that estimates the unknown value of the population slope.

$$CI = \hat{\beta}_1 \pm t_{0.975,8} * se(\hat{\beta}_1)$$

$$CI = 4 - (2.306004 * 0.4690) = 2.9185$$

$$CI = 4 + (2.306004 * 0.4690) = 5.0815$$

The 95% Confidence Interval estimating the unknown slope is (2.9185, 5.0815).

(c)

The consultant believes the mean number of broken ampules when no transfers are made is different from 9.

Perform hypothesis test.

$$H_0 : \beta_0 = 9$$
$$H_a : \beta_0 \neq 9$$

The test statistic is:

$$t = \frac{\hat{\beta}_0 - 9}{se(\hat{\beta}_0)} = \frac{10.2 - 9}{0.6633} = 1.8091$$

The critical value is:

```
qt(0.975,df=10-2)
```

```
## [1] 2.306004
```

Our test statistic is less than our critical value, so we fail to reject our null hypothesis that when there are no transfers made that there are 9 broken ampules.

(d)

Calculate a 95% confidence interval for mean number of broken ampules and a 95% prediciton interval for number of broken ampules when transfers is 2.

Confidence Interval

$$CI = \hat{\mu}_{y/x_0} \pm t_{0.975,8} * s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$$
$$\hat{\mu}_{y/x_0} = 10.2 + (4 * 2) = 18.2$$
$$CI = 18.2 \pm (t_{0.975,8})2.306004) * (1.483)\sqrt{\frac{1}{10} + \frac{(2-1)^2}{10}} = 18.2 \pm 1.529383$$

CI = (16.6706,19.7294)

Prediction Interval

$$PI = \hat{y}_0 \pm t_{0.975,8} * s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$$
$$= 18.2 \pm (2.306004)(1.483)\sqrt{1 + \frac{1}{10} + \frac{(2-1)^2}{10}}$$

PI = (14.4538,21.9462)

(e)

What happens to CI and PI when number of transfers is 1?

* The center of our interval is shifted, it will be shifted left, to a smaller value, as the mu and y hat are affected by x.
* The span will become smaller. The entire last element of the square root will equate to zero, which leads a smaller value from the center of our interval to the end of of our interval on each side mu and y hat.

(f)

What is the value of the F Statistic?

$$F = \frac{MS_R}{MS_{res}} = \frac{160}{2.2} = 72.7272$$

(g)

Calculate R^2 and give an explanation in context.

$$R^2 = \frac{SS_R}{SS_T} = \frac{160}{(160 + 17.6)} = 0.900900$$

0.9 is the proportion of variance in broken ampules that can be explained by the number of transfers. This indicates a good fit of our data to our model.