Lab Assignment 4

Jeevan ~ 14/08/2019

Problem: A soft drink bottler is analyzing the vending machine service routes in his distribution system. He is interested in predicting the amount of time required by the route driver to service the vending machines in an outlet. This service activity includes stocking the machine beverage products and minor maintenance or housekeeping. The industrial engineer responsible for the study has suggested that the two most important variables affecting the delivery time (Y) are the number of cases of product stocked (X1), and the distance walked by the route driver (X2). The engineer has collected 25 observations on delivery time, which are shown below in the table:

ObservationNo	DeliveryTimeY	NumberOfCasesX1	DistanceX2
1	16.68	7	560
2	11.5	3	220
3	12.3	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.5	5	605
11	40.33	16	688
12	21	10	215
13	13	4	255
14	19.75	6	462
15	24	9	448
16	29	10	776
17	15.35	6	200
18	19	7	132
19	9.5	3	36
20	35.1	17	770
21	17.9	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

Fit a multiple linear regression to the above data that involving the following steps by using R and interpret the outputs.

1. *Estimate the values of the regression coefficients of a multiple linear regression model and establish the relationship between the dependent and independent variables.*

```
mlrm = lm(DeliveryTimeY~NumberOfCasesX1+DistanceX2)
mlrm
##
## Call:
## lm(formula = DeliveryTimeY ~ NumberOfCasesX1 + DistanceX2)
##
## Coefficients:
##     (Intercept)  NumberOfCasesX1       DistanceX2
##         2.32347          1.61537          0.01442
```

Here we the regression coefficients have values –

1.61537 & 0.01442

----

2. *Obtain the predicted values of Y.*

```
predicted_dt=fitted.values(mlrm)
predicted_dt
```

```
##         1         2         3         4         5         6         7
## 21.704696 10.341368 12.071427  9.938340 14.178288 18.388750  7.140107
##         8         9        10        11        12        13        14
## 16.658692 71.833732 19.122720 38.088456 21.576899 12.461342 18.676441
##        15        16        17        18        19        20        21
## 23.320722 29.664922 14.899146 15.534153  7.688612 40.886036 20.495612
##        22        23        24        25
## 56.001086 23.349556 24.401356 10.947541
```

----

3. *Find the values of residuals.*

```
error_terms=residuals(mlrm)
error_terms
```

```
##          1          2          3          4          5          6
## -5.0246957  1.1586320  0.2285735  4.9416599 -0.4282885 -0.2787501
##          7          8          9         10         11         12
##  0.8598928  1.1713084  7.4062681  2.3772799  2.2415445 -0.5768987
##         13         14         15         16         17         18
##  0.5386578  1.0735592  0.6792781 -0.6649225  0.4508538  3.4658465
##         19         20         21         22         23         24
##  1.8113885 -5.7860360 -2.5956121 -3.6810862 -4.5995562 -4.5713561
##         25
## -0.1975409
```

----

4. *Test the significance of the regression coefficients at 5% of level of significance.*

```r
summary(mlrm)
```

```
##
## Call:
## lm(formula = DeliveryTimeY ~ NumberOfCasesX1 + DistanceX2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7860 -0.6649  0.4509  1.1713  7.4063
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.323473   1.095005   2.122 0.045348 *
## NumberOfCasesX1 1.615374   0.170466   9.476 3.18e-09 ***
## DistanceX2      0.014417   0.003607   3.997 0.000608 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.254 on 22 degrees of freedom
## Multiple R-squared:  0.9597, Adjusted R-squared:  0.9561
## F-statistic: 262.3 on 2 and 22 DF,  p-value: 4.497e-16
```

Since p-value is less than 0.05, we reject the null hypothesis and accept the alternative hypothesis.

---

5. *Obtain the 95 percent and 99 percent confidence interval for partial regression coefficients and intercept term.*

```r
confint(mlrm, level = .95)

confint(mlrm, level = .99)
```

```
##                       2.5 %     97.5 %
## (Intercept)     0.052570316 4.59437486
## NumberOfCasesX1 1.261848071 1.96889950
## DistanceX2      0.006935857 0.02189845
```

```
0.5 %     99.5 %
(Intercept)     -0.763080606 5.41002578
NumberOfCasesX1  1.134870535 2.09587703
DistanceX2       0.004248763 0.02458555
```

---

6. *Is the overall regression model significant at 5% level of significance.*

```
anova(mlrm)
```

```
## Analysis of Variance Table
##
## Response: DeliveryTimeY
##                  Df Sum Sq Mean Sq F value    Pr(>F)
## NumberOfCasesX1   1 5386.0  5386.0 508.556 < 2.2e-16 ***
## DistanceX2        1  169.2   169.2  15.972 0.0006082 ***
## Residuals        22  233.0    10.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here,

<u>Number of cases:</u>

HO = There is no significant difference between the number of cases and delivery time

H1 = There is a significant difference between the number of cases and delivery time.

<u>Distance:</u>

H0 = There is no significant difference between the distance travelled and distance time

H1 = There is a significant difference between distance and delivery time.

Therefore,

We reject the null hypothesis and accept the alternate hypothesis and since the p – value i.e. ***0.0006082*** < 0.05 and therefore we can say that there is a significant difference between the distance travelled and delivery time.

---

7. *Predict the Delivery time for the number of cases =8 and distance =275 and obtain the 99% CI for the predicted delivery time*

```
newdata = data.frame(NumberOfCasesX1=8,DistanceX2=275)

predict(mlrm,newdata)

##        1
## 19.21118

ci=predict(mlrm, newdata, interval="confidence", level=.99)
ci

##        fit      lwr     upr
## 1 19.21118 17.08006 21.3423
```