ISLR Chapter3

Linear Regression

ACO

2025-01-04

Contents

Simple Linear Regression

```r
library(ISLR2)
```

```r
advertising = read.csv(
  file = 'Advertising.csv', row.names = 1)
head(advertising)
```

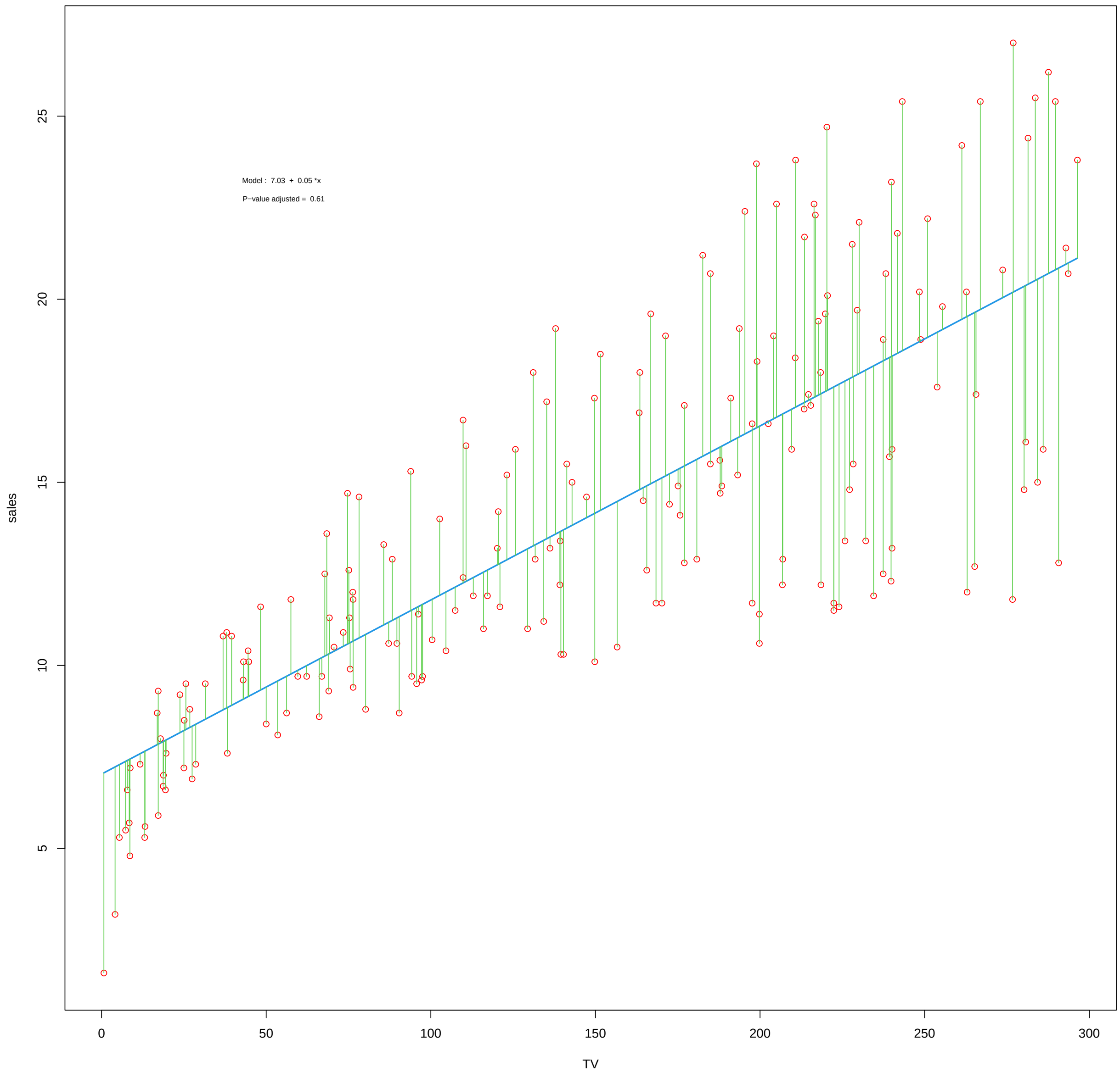
```
##      TV radio newspaper sales
## 1 230.1  37.8      69.2  22.1
## 2  44.5  39.3      45.1  10.4
## 3  17.2  45.9      69.3   9.3
## 4 151.5  41.3      58.5  18.5
## 5 180.8  10.8      58.4  12.9
## 6   8.7  48.9      75.0   7.2
```

Estimating Coefficients

```r
# Figure 3.1
plot(sales ~ TV, data = advertising, col = 'red')
slm_model = lm(sales ~ TV, data = advertising)
# https://r-graph-gallery.com/44-polynomial-curve-fitting.html
poly_predict <- predict( slm_model )
ix <- sort(advertising$TV,index.return=T)$ix
lines(advertising$TV[ix], poly_predict[ix], col=4, lwd=2 )
#
apply(
  cbind(#https://stackoverflow.com/questions/23494232/regression-line-to-data-points-how-to-create-vert
    advertising$TV, advertising$TV, advertising$sales,
    predict(slm_model)),1,
  function(coords){
    lines(coords[1:2],coords[3:4], col = 3)
    })
```

```
## NULL
```

```r
# I add the features of the model to the plot
coe_ff <- round(slm_model$coefficients , 2)
text(55, 23 ,
     paste(
       "Model : ",coe_ff[1] , " + " ,
       coe_ff[2] , "*x"  , "\n\n" , "P-value adjusted = ",
       round(summary(slm_model)$adj.r.squared,2)), cex = 0.6)
```




```r
advert_2 = advertising[c('TV', 'sales')]
x_bar = mean(advert_2$TV)
y_bar = mean(advert_2$sales)
advert_2['x-x_bar'] = advert_2$TV - x_bar
advert_2['y-y_bar'] = advert_2$sales - y_bar
head(advert_2)
```

```
##      TV sales   x-x_bar y-y_bar
## 1 230.1  22.1   83.0575  8.0775
## 2  44.5  10.4 -102.5425 -3.6225
## 3  17.2   9.3 -129.8425 -4.7225
## 4 151.5  18.5    4.4575  4.4775
## 5 180.8  12.9   33.7575 -1.1225
```

```
## 6   8.7   7.2 -138.3425 -6.8225
```

```r
beta_1 = sum(advert_2$`x-x_bar` * advert_2$`y-y_bar`
             )/ sum(advert_2$`x-x_bar`^2)
beta_0 = y_bar - beta_1 * x_bar
#
beta_1; beta_0
```

```
## [1] 0.04753664
```

```
## [1] 7.032594
```

Assessing the Accuracy of the Coefficient Estimates

```r
RSS = sum((advert_2$sales - beta_0 - beta_1 * advert_2$TV)^2)
RSS
```

```
## [1] 2102.531
```

```r
#
n = nrow(advert_2)
n
```

```
## [1] 200
```

```r
RSE = sqrt(RSS/(n - 2))
RSE
```

```
## [1] 3.258656
```

```r
#
SE_beta_0 = sqrt(
  RSE * RSE * (
  (1/n)+((x_bar^2)/sum(advert_2$`x-x_bar`^2)))
  )
SE_beta_1 = sqrt(RSE * RSE / sum(advert_2$`x-x_bar`^2))
#
SE_beta_0; SE_beta_1
```

```
## [1] 0.4578429
```

```
## [1] 0.002690607
```

```r
#
c(beta_0 - 2*SE_beta_0, beta_0 + 2*SE_beta_0)
```

```
## [1] 6.116908 7.948279
```

```r
c(beta_1 - 2*SE_beta_1, beta_1 + 2*SE_beta_1)
```

```
## [1] 0.04215543 0.05291785
```

```r
#
t_beta_0 = (beta_0 - 0)/SE_beta_0
t_beta_1 = (beta_1 - 0)/SE_beta_1
#
t_beta_0; t_beta_1
```

```
## [1] 15.36028
```

```
## [1] 17.66763
```

```r
TSS = sum(advert_2$`y-y_bar`^2)
R_SQUARED = 1 - (RSS/TSS)
R_SQUARED
```

```
## [1] 0.6118751
```