

Linear-Regression–Carseats-Sales.R

patriciamaya

2020-10-05

```
library(ISLR)

## Warning: package 'ISLR' was built under R version 4.0.2

data("Carseats")
attach(Carseats)

set.seed(123)
indx <- sample(2,nrow(Carseats), replace=T, prob = c(0.8, 0.2))
train <- Carseats[indx ==1, ]
test <- Carseats[indx ==2, ]

#lm - linear model (~)
#lm(num target ~ inputs, data= train)

lmModel <- lm(Sales ~ ., data= train)
summary(lmModel)

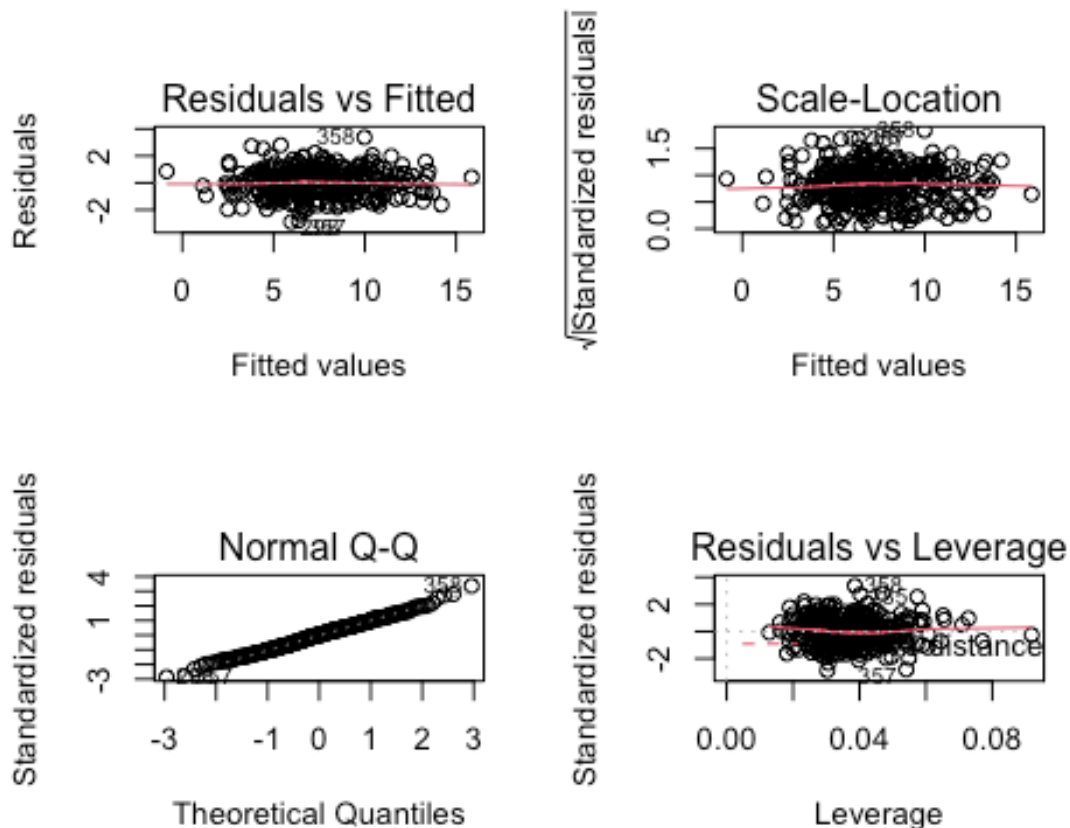
##
## Call:
## lm(formula = Sales ~ ., data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8993 -0.7146  0.0192  0.6676  3.3789
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.843e+00  6.790e-01   8.605 3.75e-16 ***
## CompPrice     9.105e-02  4.495e-03  20.254 < 2e-16 ***
## Income       1.659e-02  2.069e-03   8.017 2.18e-14 ***
## Advertising  1.228e-01  1.222e-02  10.047 < 2e-16 ***
## Population    8.435e-06  4.182e-04   0.020  0.984
## Price       -9.555e-02  2.896e-03 -32.999 < 2e-16 ***
## ShelveLocGood  4.857e+00  1.703e-01  28.511 < 2e-16 ***
## ShelveLocMedium 1.884e+00  1.387e-01  13.583 < 2e-16 ***
## Age          -4.679e-02  3.610e-03 -12.960 < 2e-16 ***
## Education    -1.783e-02  2.249e-02  -0.793  0.428
## UrbanYes     1.947e-01  1.251e-01   1.556  0.121
## USYes       -1.624e-01  1.672e-01  -0.971  0.332
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.024 on 313 degrees of freedom
## Multiple R-squared:  0.876, Adjusted R-squared:  0.8716
## F-statistic: 201 on 11 and 313 DF, p-value: < 2.2e-16

#estimate- coefficient of each variable in the Linear model
#if coefficient is larger-> means variable is more important & sign means
direction of relationship
#if we increase price, sales decrease.
#Pr(>|t|) -p value, whether there is a significant relationship between each
variable & target variable
#we want r-squared to be large-> indicates good fit of model

#Population around the store does NOT affect sales much, also education
level, if urban, and if store
#is in USA.

layout(matrix(c(1,2,3,4), 2, 2))
plot(lmModel)
```



#Residuals vs Fitted - we don't want to see any pattern, points should be scattered at random, should be a horizontal line.
#Scale-Location - variation of residuals, line should be horizontal-> means

variation is the same around all points.
 #Normal Q-Q - check that residuals have normal distribution, should be a 45 degree line.
 #Residual vs Leverage - helps identify influential points.
 #no influential points in our model.

```
predictions <- predict(lmModel, newdata = test)
```

#mean squared error

```
mean(test$Sales-predictions)^2
```

```
## [1] 0.02526576
```

fitted(lmModel) *#predictions for train values*

```
##          1          2          3          6          7          9
10
##  7.2693060 12.4507357  9.2402292  9.6306823  6.2254075  5.9884113
5.7511180
##          12          13          14          15          17          18
19
## 11.8857466  3.6424874 11.9994924  9.7372705  8.4807905 11.8130859
13.4272856
.
.
.
##          389          390          392          393          394          395
396
##  9.5610293  9.0200603  6.2595241  5.6717540  6.3903289  5.8986975
13.1526578
##          397          398          399
##  6.7352176  7.0322852  5.5213980
```

coefficients(lmModel) *#coefficients of regression*

```
##      (Intercept)      CompPrice      Income      Advertising
Population
##  5.842839e+00    9.105120e-02    1.658512e-02    1.227960e-01
8.434767e-06
##          Price  ShelveLocGood ShelveLocMedium      Age
Education
## -9.555257e-02    4.856521e+00    1.884479e+00 -4.679088e-02 -
1.783126e-02
##          UrbanYes      USYes
##  1.946842e-01 -1.624238e-01
```

```
residuals(lmModel) #residuals from actual train values - predicted train values
```

```
##           1           2           3           6           7
9
##  2.230694030 -1.230735704  0.819770826  1.179317660  0.404592547
0.551588688
##           10          12          13          14          15
17
## -1.061118049  0.074253427  0.337512644 -1.039492364  1.432729453 -
0.900790488
.
.
.
##           385          387          388          389          390
392
##  0.686564379 -0.968535721 -0.102111983 -1.421029263 -0.580060304 -
0.159524076
##           393          394          395          396          397
398
## -1.141754048 -0.820328934 -0.548697514 -0.582657813 -0.595217563
0.377714820
##           399
##  0.418602026
```