Class 03 Linear Regression with Supermarket Data

BQOM 2578 | Data Mining

Theresa Wohlever

2025-09-18

Table of contents

Executive Summary **1**
Loading packages 1
Importing data 1
Extra: 10

Executive Summary

Loading packages

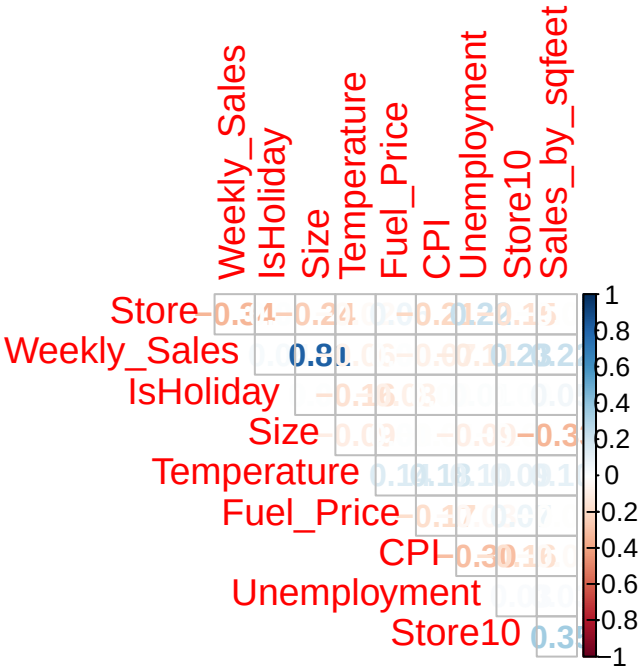
Importing data

Correlation Matrix in R

```
cormat<-df%>%select_if(is.numeric)%>%cor()  # recall that cor() needs numeric data
# install.packages("corrplot")
library(corrplot)
```

```
corrplot 0.95 loaded
```

```
# corrplot(cormat)
# corrplot(cormat, type="upper",diag=FALSE, tl.cex=1.2)
corrplot(cormat, method="number", type="upper",diag=FALSE, tl.cex=1.2)
```



With a cleaned dataset, we can jump straight ahead into running regressions:

Is there a significant difference for Store10 when compared with other stores?

```
m1<-lm(Weekly_Sales~Store10,data=df)
summary(m1)
```

```
Call:
lm(formula = Weekly_Sales ~ Store10, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-817605 -474241  -78613  357205 2791096

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1027591       6928  148.31   <2e-16 ***
Store10       871834      46478   18.76   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 549600 on 6433 degrees of freedom
Multiple R-squared:  0.05186,   Adjusted R-squared:  0.05171
```

```
F-statistic: 351.9 on 1 and 6433 DF,  p-value: < 2.2e-16
```

What is our level of analysis observation? What are our most important variables? Are any of them of the right type? What can we say about them?

```
m2<-lm(Weekly_Sales~Store10+Date,data=df)
summary(lm(Weekly_Sales~Store10+Date,data=df))
```

```
Call:
lm(formula = Weekly_Sales ~ Store10 + Date, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-814944 -475370  -79646  357224 2793471

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 822085.61  359096.19   2.289   0.0221 *
Store10     871833.78   46479.96  18.757   <2e-16 ***
Date            13.57      23.71   0.572   0.5671
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 549600 on 6432 degrees of freedom
Multiple R-squared:  0.05191,   Adjusted R-squared:  0.05161
F-statistic: 176.1 on 2 and 6432 DF,  p-value: < 2.2e-16
```

Let's explore the dataset variable Size:

```
m3<-lm(Weekly_Sales~Store10+Date+Size,data=df)
summary(m3)
```

```
Call:
lm(formula = Weekly_Sales ~ Store10 + Date + Size, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-597590 -226792  -21159  157505 2281051
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.252e+05  1.980e+05  -0.632    0.527
Store10      8.999e+05  2.561e+04  35.133   <2e-16 ***
Date         1.357e+01  1.307e+01   1.039    0.299
Size         7.266e+00  5.982e-02 121.454   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 302900 on 6431 degrees of freedom
Multiple R-squared:  0.7122,    Adjusted R-squared:  0.712
F-statistic:  5304 on 3 and 6431 DF,  p-value: < 2.2e-16
```

```
m3A<-lm(Weekly_Sales~Store10+Size,data=df)
summary(m3A)
```

```
Call:
lm(formula = Weekly_Sales ~ Store10 + Size, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-599871 -225958  -21637  157933 2278676

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.032e+04  8.684e+03   9.249   <2e-16 ***
Store10     8.999e+05  2.561e+04  35.133   <2e-16 ***
Size        7.266e+00  5.982e-02 121.454   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 302900 on 6432 degrees of freedom
Multiple R-squared:  0.7121,    Adjusted R-squared:  0.712
F-statistic:  7955 on 2 and 6432 DF,  p-value: < 2.2e-16
```

We could just look at all the variables, but now the standardized coefficients would be really helpful:

```
mAll <- lm(Weekly_Sales ~ ., data=df)
summary(mAll)
```

```
Call:
lm(formula = Weekly_Sales ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-586242  -76298   -4293   73557  966903

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -7.122e+05  1.318e+05  -5.405 6.71e-08 ***
Date            7.130e+00  9.664e+00   0.738 0.460683
Store          -3.023e+03  1.563e+02 -19.344  < 2e-16 ***
IsHoliday       1.063e+04  6.399e+03   1.661 0.096724 .
TypeB          -2.498e+04  4.962e+03  -5.035 4.92e-07 ***
TypeC          -1.771e+05  8.295e+03 -21.347  < 2e-16 ***
Size            8.148e+00  4.243e-02 192.025  < 2e-16 ***
Temperature    -3.239e+02  9.529e+01  -3.399 0.000681 ***
Fuel_Price     -4.858e+03  6.042e+03  -0.804 0.421435
CPI            -1.026e+03  4.781e+01 -21.451  < 2e-16 ***
Unemployment   -7.535e+03  9.774e+02  -7.709 1.46e-14 ***
Store10         6.145e+04  1.253e+04   4.903 9.65e-07 ***
Sales_by_sqfeet  1.133e+05  7.299e+02 155.292  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 128100 on 6422 degrees of freedom
Multiple R-squared:  0.9486,    Adjusted R-squared:  0.9485
F-statistic:  9867 on 12 and 6422 DF,  p-value: < 2.2e-16
```

Comparing different models and creating an output table:

```r
library(jtools)
library(sjPlot)

# Generate the table using tab_model()

model_table <- tab_model(m1, m2, m3,mAll,dv.labels = c("(1) Store 10", "(2) Store 10 + Date
# Print the table
print(model_table)
```

Standardize Coefficients

Recall that Standardized Coefficients are helpful in identifying key variables in the Linear Regression.

```
# STANDARDIZED COEFFICIENTS can be an important output of a linear regression
#     analysis yet they are not in the lm() function.
# lm.beta() can be called to calculate the standardized coefficients and add
#     that information into the model:  m1b <- lm.beta(m1)
#     where m1 is the output of the lm(), that is m1 <- lm()
#
# After being saved, you can see the results using summary(m1b)
# Alternatively, you can run summary(lm.beta(lm(y~x,df)))
#
# Recall that we already installed and called the library "lm.beta"
#install.packages("lm.beta")
library("lm.beta")
#
# Let's include the standard coeff in our analysis:
summary(lm.beta(m1))
```

```
Call:
lm(formula = Weekly_Sales ~ Store10, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-817605 -474241  -78613  357205 2791096

Coefficients:
             Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept) 1.028e+06           NA  6.928e+03  148.31   <2e-16 ***
Store10     8.718e+05    2.277e-01  4.648e+04   18.76   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 549600 on 6433 degrees of freedom
Multiple R-squared:  0.05186,   Adjusted R-squared:  0.05171
F-statistic: 351.9 on 1 and 6433 DF,  p-value: < 2.2e-16
```

```r
summary(lm.beta(m2))
```

```
Call:
lm(formula = Weekly_Sales ~ Store10 + Date, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-814944 -475370  -79646  357224 2793471

Coefficients:
             Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept) 8.221e+05           NA  3.591e+05   2.289   0.0221 *
Store10     8.718e+05    2.277e-01  4.648e+04  18.757   <2e-16 ***
Date        1.357e+01    6.949e-03  2.371e+01   0.572   0.5671
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 549600 on 6432 degrees of freedom
Multiple R-squared:  0.05191,   Adjusted R-squared:  0.05161
F-statistic: 176.1 on 2 and 6432 DF,  p-value: < 2.2e-16
```

```r
summary(lm.beta(m3))
```

```
Call:
lm(formula = Weekly_Sales ~ Store10 + Date + Size, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-597590 -226792  -21159  157505 2281051

Coefficients:
              Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept) -1.252e+05           NA  1.980e+05  -0.632    0.527
Store10      8.999e+05    2.351e-01  2.561e+04  35.133   <2e-16 ***
Date         1.357e+01    6.949e-03  1.307e+01   1.039    0.299
Size         7.266e+00    8.126e-01  5.982e-02 121.454   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 302900 on 6431 degrees of freedom
Multiple R-squared:  0.7122,    Adjusted R-squared:  0.712
F-statistic:  5304 on 3 and 6431 DF,  p-value: < 2.2e-16
```

```
summary(lm.beta(mAll))
```

```
Call:
lm(formula = Weekly_Sales ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-586242  -76298   -4293   73557  966903

Coefficients:
                 Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept)     -7.122e+05           NA  1.318e+05  -5.405 6.71e-08 ***
Date             7.130e+00    3.651e-03  9.664e+00   0.738 0.460683
Store           -3.023e+03   -6.958e-02  1.563e+02 -19.344  < 2e-16 ***
IsHoliday        1.063e+04    4.804e-03  6.399e+03   1.661 0.096724 .
TypeB           -2.498e+04   -2.146e-02  4.962e+03  -5.035 4.92e-07 ***
TypeC           -1.771e+05   -1.067e-01  8.295e+03 -21.347  < 2e-16 ***
Size             8.148e+00    9.112e-01  4.243e-02 192.025  < 2e-16 ***
Temperature     -3.239e+02   -1.059e-02  9.529e+01  -3.399 0.000681 ***
Fuel_Price      -4.858e+03   -3.951e-03  6.042e+03  -0.804 0.421435
CPI             -1.026e+03   -7.151e-02  4.781e+01 -21.451  < 2e-16 ***
Unemployment    -7.535e+03   -2.505e-02  9.774e+02  -7.709 1.46e-14 ***
Store10          6.145e+04    1.605e-02  1.253e+04   4.903 9.65e-07 ***
Sales_by_sqfeet  1.133e+05    5.653e-01  7.299e+02 155.292  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 128100 on 6422 degrees of freedom
Multiple R-squared:  0.9486,    Adjusted R-squared:  0.9485
F-statistic:  9867 on 12 and 6422 DF,  p-value: < 2.2e-16
```

Also, the tab_model() has an option to display standardized coefficients

```
# Use show.std="std" in tab_model() to have it display the std coeff
```

```
model_table_stdbeta <- tab_model(m1, m2, m3,mAll,dv.labels = c("(1) Store 10", "(2) Store 1


# Print the table
print(model_table_stdbeta)
```

So, let's trim back the variable that includes the answer and the Type (A, B or C):

```
# Remove Sales_by_sqfeet since it includes sales in its formula!
df1 <- select(df,-c(Sales_by_sqfeet, Type))
mNewAll <- lm(Weekly_Sales ~ ., data=df1)
summary(lm.beta(mNewAll))
```

```
Call:
lm(formula = Weekly_Sales ~ ., data = df1)

Residuals:
    Min      1Q  Median      3Q     Max
-656106 -207228  -32318  164263 2285084

Coefficients:
               Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept)  -6.501e+05           NA  3.010e+05  -2.160   0.0308 *
Date          9.003e+01    4.610e-02  2.210e+01   4.074 4.68e-05 ***
Store        -5.189e+03   -1.194e-01  3.081e+02 -16.841  < 2e-16 ***
IsHoliday     7.742e+04    3.499e-02  1.463e+04   5.293 1.24e-07 ***
Size          6.983e+00    7.810e-01  6.034e-02 115.731  < 2e-16 ***
Temperature   3.086e+02    1.009e-02  2.137e+02   1.444   0.1487
Fuel_Price   -6.678e+04   -5.431e-02  1.380e+04  -4.840 1.33e-06 ***
CPI          -1.149e+03   -8.014e-02  1.089e+02 -10.552  < 2e-16 ***
Unemployment -9.615e+03   -3.196e-02  2.217e+03  -4.337 1.47e-05 ***
Store10       7.942e+05    2.074e-01  2.595e+04  30.604  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 293500 on 6425 degrees of freedom
Multiple R-squared:  0.7299,    Adjusted R-squared:  0.7295
F-statistic:  1929 on 9 and 6425 DF,  p-value: < 2.2e-16
```

```r
head(df1)
```

```
        Date Store Weekly_Sales IsHoliday   Size Temperature Fuel_Price
1 2010-02-05     1      1643691         0 151315       42.31      2.572
2 2010-02-05    10      2193049         0 126512       54.34      2.962
3 2010-02-05    11      1528009         0 207499       46.04      2.572
4 2010-02-05    12      1100046         0 112238       49.47      2.962
5 2010-02-05    13      1967221         0 219622       31.53      2.666
6 2010-02-05    14      2623470         0 200898       27.31      2.784
       CPI Unemployment Store10
1 211.0964        8.106       0
2 126.4421        9.765       1
3 214.4249        7.368       0
4 126.4421       13.975       0
5 126.4421        8.316       0
6 181.8712        8.992       0
```

```r
model_table_stdbeta2 <- tab_model(m1, m2, m3,mNewAll,dv.labels = c("(1) Store 10", "(2) Sto


# Print the table
print(model_table_stdbeta2)
```

Extra:

Using Sales by Square Feet instead

```r
n1<-lm(Sales_by_sqfeet~Store10+Date,data=df)
summary(lm.beta(n1))
```

```
Call:
lm(formula = Sales_by_sqfeet ~ Store10 + Date, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7161 -2.0559 -0.3606  1.4580 15.5525

Coefficients:
```

```
             Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept) 6.4507859           NA  1.7233192   3.743 0.000183 ***
Store10     6.6711835    0.3493809  0.2230595  29.908  < 2e-16 ***
Date        0.0001249    0.0128266  0.0001138   1.098 0.272254
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.638 on 6432 degrees of freedom
Multiple R-squared:  0.1222,    Adjusted R-squared:  0.122
F-statistic: 447.8 on 2 and 6432 DF,  p-value: < 2.2e-16
```

```r
#options(scipen=999)
n2<-lm(Sales_by_sqfeet~.,data=df)
summary(lm.beta(n2))
```

```
Call:
lm(formula = Sales_by_sqfeet ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-4.4471 -0.5800  0.0293  0.5241  5.5606

Coefficients:
               Estimate Standardized Std. Error  t value Pr(>|t|)
(Intercept)   5.900e+00           NA  1.033e+00    5.712 1.16e-08 ***
Date          6.385e-05    6.555e-03  7.577e-05    0.843   0.3994
Store         9.842e-03    4.541e-02  1.255e-03    7.844 5.06e-15 ***
Weekly_Sales  6.967e-06    1.397e+00  4.487e-08  155.292  < 2e-16 ***
IsHoliday     3.648e-02    3.306e-03  5.018e-02    0.727   0.4673
TypeB         4.727e-02    8.142e-03  3.898e-02    1.213   0.2253
TypeC         2.037e+00    2.460e-01  6.232e-02   32.678  < 2e-16 ***
Size         -5.842e-05   -1.310e+00  4.634e-07 -126.047  < 2e-16 ***
Temperature   1.764e-03    1.156e-02  7.475e-04    2.359   0.0183 *
Fuel_Price   -4.273e-02   -6.968e-03  4.738e-02   -0.902   0.3671
CPI           6.482e-03    9.064e-02  3.795e-04   17.081  < 2e-16 ***
Unemployment  4.045e-02    2.696e-02  7.682e-03    5.265 1.45e-07 ***
Store10       1.008e+00    5.277e-02  9.764e-02   10.320  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.005 on 6422 degrees of freedom
```

```
Multiple R-squared:  0.8729,    Adjusted R-squared:  0.8726
F-statistic:  3674 on 12 and 6422 DF,  p-value: < 2.2e-16
```

Predictions

```
#We can make predictions using one of our models, for example model 3
tail(df%>%filter(Store10==1)%>%select(Date,Size))
```

```
          Date   Size
138 2012-09-21 126512
139 2012-09-28 126512
140 2012-10-05 126512
141 2012-10-12 126512
142 2012-10-19 126512
143 2012-10-26 126512
```

```
#use predict(model,test dataset) function
#For an individual data-point:
predict(m3,data.frame(Store10=1,Date=ymd("2012-11-3"),Size=126512))
```

```
      1
1906278
```

```
predict(m3,data.frame(Store10=1,Date=ymd("2012-11-10"),Size=126512))
```

```
      1
1906373
```

```
predict(m3,data.frame(Store10=1,Date=ymd("2012-11-17"),Size=126512))
```

```
      1
1906468
```

```
predict(m3,data.frame(Store10=1,Date=ymd("2012-11-24"),Size=126512))
```

```
      1
1906563
```

```
#Later, we'll learn to split datasets in two; one for training and one for testing.
```

Stepwise Regression

```
model <- lm(Weekly_Sales ~ ., data = df)
summary(model)
```

```
Call:
lm(formula = Weekly_Sales ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-586242  -76298   -4293   73557  966903 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -7.122e+05  1.318e+05  -5.405 6.71e-08 ***
Date            7.130e+00  9.664e+00   0.738 0.460683    
Store          -3.023e+03  1.563e+02 -19.344  < 2e-16 ***
IsHoliday       1.063e+04  6.399e+03   1.661 0.096724 .  
TypeB          -2.498e+04  4.962e+03  -5.035 4.92e-07 ***
TypeC          -1.771e+05  8.295e+03 -21.347  < 2e-16 ***
Size            8.148e+00  4.243e-02 192.025  < 2e-16 ***
Temperature    -3.239e+02  9.529e+01  -3.399 0.000681 ***
Fuel_Price     -4.858e+03  6.042e+03  -0.804 0.421435    
CPI            -1.026e+03  4.781e+01 -21.451  < 2e-16 ***
Unemployment   -7.535e+03  9.774e+02  -7.709 1.46e-14 ***
Store10         6.145e+04  1.253e+04   4.903 9.65e-07 ***
Sales_by_sqfeet 1.133e+05  7.299e+02 155.292  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 128100 on 6422 degrees of freedom
Multiple R-squared:  0.9486,    Adjusted R-squared:  0.9485 
F-statistic:  9867 on 12 and 6422 DF,  p-value: < 2.2e-16
```

```
# Perform stepwise regression
#direction can be both, backward or forward
```

```r
#trace can be set to 0 to only display final result or higher to display more information
step_model <- step(model, direction = "backward",trace=999)
```

```
Start:  AIC=151374.7
Weekly_Sales ~ Date + Store + IsHoliday + Type + Size + Temperature +
    Fuel_Price + CPI + Unemployment + Store10 + Sales_by_sqfeet

                  Df  Sum of Sq        RSS    AIC
- Date             1 8.9362e+09 1.0544e+14 151373
- Fuel_Price       1 1.0612e+10 1.0545e+14 151373
<none>                          1.0543e+14 151375
- IsHoliday        1 4.5306e+10 1.0548e+14 151375
- Temperature      1 1.8966e+11 1.0562e+14 151384
- Store10          1 3.9473e+11 1.0583e+14 151397
- Unemployment     1 9.7580e+11 1.0641e+14 151432
- Store            1 6.1436e+12 1.1158e+14 151737
- CPI              1 7.5546e+12 1.1299e+14 151818
- Type             2 8.2257e+12 1.1366e+14 151854
- Sales_by_sqfeet  1 3.9592e+14 5.0136e+14 161406
- Size             1 6.0538e+14 7.1081e+14 163653

Step:  AIC=151373.2
Weekly_Sales ~ Store + IsHoliday + Type + Size + Temperature +
    Fuel_Price + CPI + Unemployment + Store10 + Sales_by_sqfeet

                  Df  Sum of Sq        RSS    AIC
- Fuel_Price       1 2.0940e+09 1.0545e+14 151371
<none>                          1.0544e+14 151373
- IsHoliday        1 4.9536e+10 1.0549e+14 151374
- Temperature      1 1.8621e+11 1.0563e+14 151383
- Store10          1 3.8937e+11 1.0583e+14 151395
- Unemployment     1 1.1176e+12 1.0656e+14 151439
- Store            1 6.1372e+12 1.1158e+14 151735
- CPI              1 7.8218e+12 1.1327e+14 151832
- Type             2 8.2190e+12 1.1366e+14 151852
- Sales_by_sqfeet  1 3.9674e+14 5.0218e+14 161415
- Size             1 6.0538e+14 7.1082e+14 163651

Step:  AIC=151371.3
Weekly_Sales ~ Store + IsHoliday + Type + Size + Temperature +
    CPI + Unemployment + Store10 + Sales_by_sqfeet
```

```
                  Df  Sum of Sq        RSS    AIC
<none>                          1.0545e+14 151371
- IsHoliday        1 5.0566e+10 1.0550e+14 151372
- Temperature      1 2.0189e+11 1.0565e+14 151382
- Store10          1 3.8831e+11 1.0583e+14 151393
- Unemployment     1 1.1241e+12 1.0657e+14 151438
- Store            1 6.1796e+12 1.1163e+14 151736
- CPI              1 8.0969e+12 1.1354e+14 151845
- Type             2 8.2202e+12 1.1367e+14 151850
- Sales_by_sqfeet  1 3.9691e+14 5.0236e+14 161415
- Size             1 6.0695e+14 7.1240e+14 163663
```

```r
#backward starts with everything and drops non-significant values.


# View the summary of the stepwise model
summary(step_model)
```

```
Call:
lm(formula = Weekly_Sales ~ Store + IsHoliday + Type + Size + 
    Temperature + CPI + Unemployment + Store10 + Sales_by_sqfeet, 
    data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-585962  -76074   -4022   73401  965903 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -6.211e+05  1.785e+04 -34.793  < 2e-16 ***
Store           -3.025e+03  1.559e+02 -19.403  < 2e-16 ***
IsHoliday        1.117e+04  6.365e+03   1.755 0.079279 .  
TypeB           -2.512e+04  4.950e+03  -5.074 4.01e-07 ***
TypeC           -1.771e+05  8.291e+03 -21.356  < 2e-16 ***
Size             8.147e+00  4.237e-02 192.294  < 2e-16 ***
Temperature     -3.273e+02  9.331e+01  -3.507 0.000456 ***
CPI             -1.014e+03  4.567e+01 -22.210  < 2e-16 ***
Unemployment    -7.694e+03  9.298e+02  -8.275  < 2e-16 ***
Store10          6.084e+04  1.251e+04   4.864 1.18e-06 ***
Sales_by_sqfeet  1.134e+05  7.290e+02 155.502  < 2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 128100 on 6424 degrees of freedom
Multiple R-squared:  0.9485,    Adjusted R-squared:  0.9485
F-statistic: 1.184e+04 on 10 and 6424 DF,  p-value: < 2.2e-16
```

```
#For more info: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/step
```