Assignment 3

79546 - Stephen K. Ng'etich

Contents

1 Pre-requisite

1.1 Load packages

```
# Clear variables
rm(list=ls())

library(readxl)
library(dplyr)
library(tidyverse)
library(lattice)
library(leaps)
library(MASS)
```

1.2 Load dataset

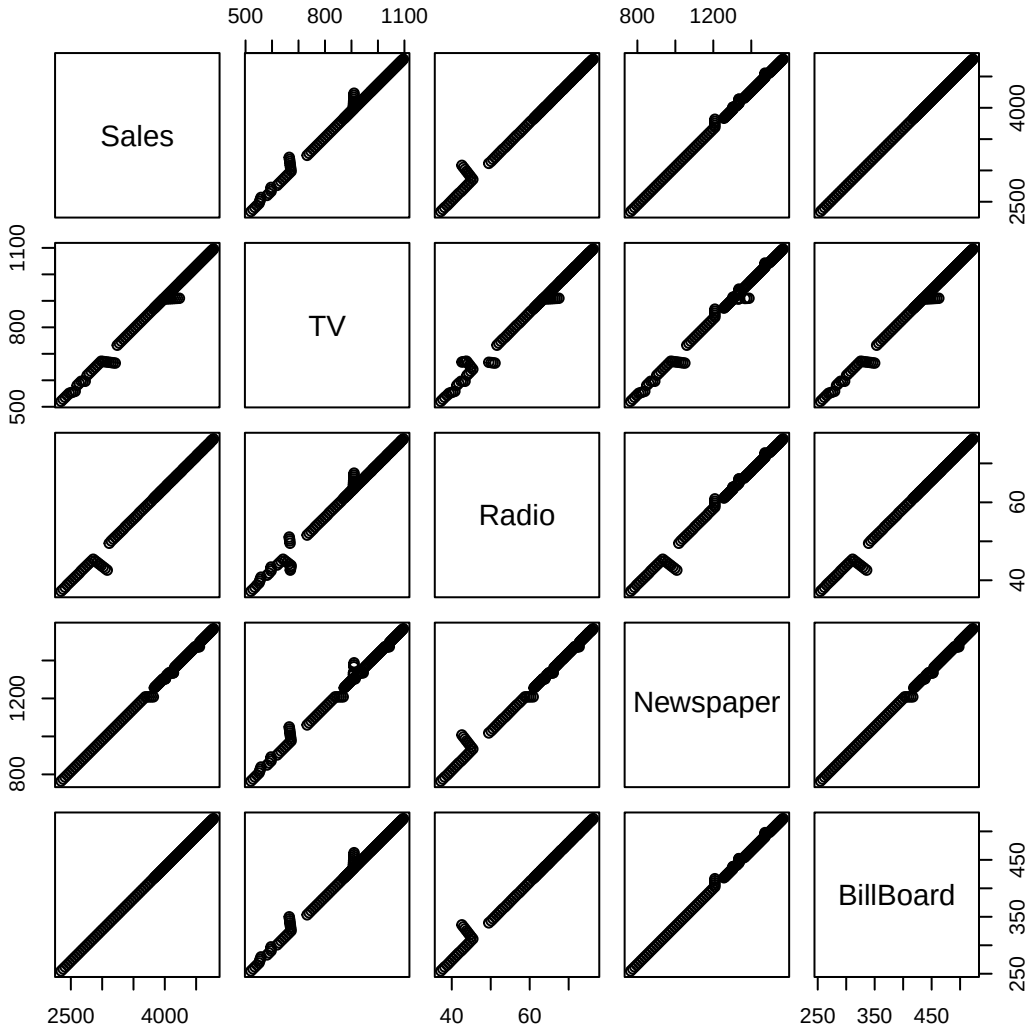
```
# Load Dataset
dataset <- read_excel("dataset/Sales_Dataset.xlsx")
```

2 Exploratory Data Analysis

```
summary(dataset)
```

```
##      Sales            TV             Radio         Newspaper
##  Min.   :2345   Min.   : 520.6   Min.   :37.20   Min.   : 764.9
##  1st Qu.:3425   1st Qu.: 776.0   1st Qu.:54.58   1st Qu.:1121.2
##  Median :4049   Median : 909.2   Median :64.60   Median :1327.2
##  Mean   :3880   Mean   : 879.6   Mean   :61.68   Mean   :1269.5
##  3rd Qu.:4413   3rd Qu.:1009.6   3rd Qu.:70.45   3rd Qu.:1447.3
##  Max.   :4777   Max.   :1095.8   Max.   :76.30   Max.   :1567.4
##    BillBoard
##  Min.   :255.0
##  1st Qu.:373.7
##  Median :442.4
##  Mean   :423.8
##  3rd Qu.:482.4
##  Max.   :522.5
```

```
plot(dataset)
```



From the visual inspection of the plot, the variables are all correlated since the plot shows a linear relationship and heads on one line.

3 Model Selection

3.1 TV and polynomials

```
tv_model = lm(Sales ~poly(TV,degree=2,raw = TRUE), data=dataset)

summary(tv_model)
```

```
##
## Call:
## lm(formula = Sales ~ poly(TV, degree = 2, raw = TRUE), data = dataset)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.800 -25.300 -12.115   2.067 221.541
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       4.997e+01  1.152e+02   0.434    0.665
## poly(TV, degree = 2, raw = TRUE)1  4.598e+00  2.877e-01  15.981   <2e-16 ***
## poly(TV, degree = 2, raw = TRUE)2 -2.677e-04  1.740e-04  -1.538    0.126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48.93 on 133 degrees of freedom
## Multiple R-squared:  0.9948, Adjusted R-squared:  0.9947
## F-statistic: 1.276e+04 on 2 and 133 DF,  p-value: < 2.2e-16
```

Polynomial of TV is not significant

3.2 Newspaper and polynomials

```
np_model = lm(Sales ~poly(Newspaper,degree=2,raw = TRUE), data=dataset)
summary(np_model)
```

```
##
## Call:
## lm(formula = Sales ~ poly(Newspaper, degree = 2, raw = TRUE),
##     data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.650  -8.984  -6.272  -1.243 117.350
##
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           -9.104e+01  4.822e+01  -1.888   0.0612
## poly(Newspaper, degree = 2, raw = TRUE)1  3.241e+00  8.300e-02  39.048   <2e-16
## poly(Newspaper, degree = 2, raw = TRUE)2 -8.595e-05  3.472e-05  -2.475   0.0146
##
## (Intercept)                           .
## poly(Newspaper, degree = 2, raw = TRUE)1 ***
## poly(Newspaper, degree = 2, raw = TRUE)2 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.76 on 133 degrees of freedom
## Multiple R-squared:  0.9992, Adjusted R-squared:  0.9992
## F-statistic: 8.719e+04 on 2 and 133 DF,  p-value: < 2.2e-16
```

Polynomial of newspaper is significant in predicting sales

3.3 Radio and polynomials

```
radio_model = lm(Sales ~poly(Radio,degree=2,raw = TRUE), data=dataset)

summary(radio_model)
```

```
##
## Call:
## lm(formula = Sales ~ poly(Radio, degree = 2, raw = TRUE), data = dataset)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -71.41  -9.63  -0.84   1.99 349.74
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          345.80893  135.38994   2.554   0.0118 *
## poly(Radio, degree = 2, raw = TRUE)1  53.35217    4.83146  11.043   <2e-16 ***
## poly(Radio, degree = 2, raw = TRUE)2   0.06207    0.04177   1.486   0.1397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 53.71 on 133 degrees of freedom
## Multiple R-squared:  0.9938, Adjusted R-squared:  0.9937
## F-statistic: 1.058e+04 on 2 and 133 DF,  p-value: < 2.2e-16
```

polynomial of Radio is not significant since it has a value greater than 0.05

3.4 BillBoard and its polynomials

```
bb_model = lm(Sales ~poly(BillBoard,degree=2,raw = TRUE), data=dataset)

summary(bb_model)
```

```
##
## Call:
## lm(formula = Sales ~ poly(BillBoard, degree = 2, raw = TRUE),
##     data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37399 -0.18976 -0.00166  0.19069  0.39829
##
## Coefficients:
##                                        Estimate Std. Error  t value
## (Intercept)                           2.681e+01  6.708e-01   39.963
## poly(BillBoard, degree = 2, raw = TRUE)1  9.093e+00  3.465e-03 2624.332
## poly(BillBoard, degree = 2, raw = TRUE)2 -2.299e-06  4.348e-06   -0.529
##                                       Pr(>|t|)
## (Intercept)                             <2e-16 ***
```

```
## poly(BillBoard, degree = 2, raw = TRUE)1   <2e-16 ***
## poly(BillBoard, degree = 2, raw = TRUE)2    0.598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2606 on 133 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 4.521e+08 on 2 and 133 DF,  p-value: < 2.2e-16
```

Polynomial of billboard is not statistically significant.

3.5 Selecting the best predictors of Sales.

```
lm_model = lm(Sales ~ . , data=dataset)
summary(lm_model)
```

```
##
## Call:
## lm(formula = Sales ~ ., data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37837 -0.18884 -0.00447  0.18937  0.42574
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.1257607  0.1576845 172.025   <2e-16 ***
## TV          -0.0005214  0.0019283  -0.270    0.787
## Radio       -0.0023525  0.0256175  -0.092    0.927
## Newspaper    0.0029307  0.0035995   0.814    0.417
## BillBoard    9.0836817  0.0121374 748.402   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2622 on 131 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 2.234e+08 on 4 and 131 DF,  p-value: < 2.2e-16
```

Radio, Newspaper and TV has p-value greater than 0.05.This mean that they are not statically significant in predicting Sales.However, BillBoard is the best predictor for predicting sales since its p-values is less than 0.05.The result model equation can be written as

$$Y = 27.125 + 9.0836 \times BillBoard$$

3.6 Interaction of Billboard and Newspaper

```
lm_model1 = lm(Sales ~ BillBoard*Newspaper, data=dataset)
summary(lm_model1)
```

```
##
## Call:
## lm(formula = Sales ~ BillBoard * Newspaper, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38471 -0.19754 -0.00449  0.19625  0.43543
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.668e+01  6.841e-01  39.004   <2e-16 ***
## BillBoard             9.082e+00  1.071e-02 847.939   <2e-16 ***
## Newspaper             3.843e-03  3.807e-03   1.009    0.315
## BillBoard:Newspaper -1.033e-06  1.478e-06  -0.699    0.486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2608 on 132 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 3.011e+08 on 3 and 132 DF,  p-value: < 2.2e-16
```

The interaction between newspaper and Billboard is not significant

3.7 Interaction of Billboard and TV

```
lm_model2 = lm(Sales ~ BillBoard*TV, data=dataset)
summary(lm_model2)
```

```
##
## Call:
## lm(formula = Sales ~ BillBoard * TV, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38153 -0.19010 -0.00126  0.19260  0.40206
##
## Coefficients:
##                Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   2.678e+01  6.444e-01   41.564   <2e-16 ***
## BillBoard     9.093e+00  4.308e-03 2110.654   <2e-16 ***
## TV            9.857e-05  2.206e-03    0.045    0.964
## BillBoard:TV -1.139e-06  2.037e-06   -0.559    0.577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2615 on 132 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 2.994e+08 on 3 and 132 DF,  p-value: < 2.2e-16
```

The interaction between newspaper and Billboard is not significant

3.8 Interaction of Billboard,Newspaper,TV and Radio

```
lm_model2 = lm(Sales ~ BillBoard +Newspaper+BillBoard*TV*Radio*Newspaper, data=dataset)
summary(lm_model2)
```

```
##
## Call:
## lm(formula = Sales ~ BillBoard + Newspaper + BillBoard * TV *
##     Radio * Newspaper, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42573 -0.18945  0.00059  0.20298  0.47646
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.353e+01  2.233e+01   0.606   0.5457
## BillBoard                     1.259e+01  3.849e+00   3.271   0.0014 **
## Newspaper                     3.076e-02  1.079e+00   0.028   0.9773
## TV                            7.697e-02  1.834e-01   0.420   0.6755
## Radio                        -2.472e+01  1.299e+01  -1.903   0.0595 .
## BillBoard:TV                 -4.311e-03  5.191e-03  -0.831   0.4079
## BillBoard:Radio              -2.354e-02  4.371e-02  -0.538   0.5912
## TV:Radio                      4.011e-02  2.016e-02   1.989   0.0489 *
## BillBoard:Newspaper          -1.954e-03  1.466e-03  -1.332   0.1852
## Newspaper:TV                 -6.408e-04  1.412e-03  -0.454   0.6508
## Newspaper:Radio               2.176e-02  1.785e-02   1.219   0.2252
## BillBoard:TV:Radio            1.173e-06  5.238e-05   0.022   0.9822
## BillBoard:Newspaper:TV        3.620e-06  2.055e-06   1.761   0.0807 .
## BillBoard:Newspaper:Radio    -5.626e-07  5.498e-06  -0.102   0.9187
## Newspaper:TV:Radio           -2.390e-05  2.331e-05  -1.025   0.3072
## BillBoard:Newspaper:TV:Radio -6.241e-10  1.118e-09  -0.558   0.5777
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.261 on 120 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 6.011e+07 on 15 and 120 DF,  p-value: < 2.2e-16
```

```
lm_model2 = lm(Sales ~ TV*Radio*Newspaper, data=dataset)
summary(lm_model2)
```

```
##
## Call:
## lm(formula = Sales ~ TV * Radio * Newspaper, data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.283  -2.459  -0.346   2.015  25.914
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)        -5.042e+01  8.256e+01  -0.611   0.5424
## TV                  5.917e-01  3.369e-01   1.757   0.0814 .
## Radio              -1.538e+02  5.411e+00 -28.422   <2e-16 ***
## Newspaper           1.030e+01  2.797e-01  36.842   <2e-16 ***
## TV:Radio            2.237e-01  8.154e-03  27.437   <2e-16 ***
## TV:Newspaper       -1.141e-02  3.751e-04 -30.405   <2e-16 ***
## Radio:Newspaper     4.111e-03  3.389e-03   1.213   0.2273
## TV:Radio:Newspaper  1.186e-06  1.594e-06   0.744   0.4581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.578 on 128 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 2.82e+05 on 7 and 128 DF,  p-value: < 2.2e-16
```

Omitting BillBoard,interaction of TV:Radio and TV:Newspaper are significant in predicting sales

```
lm_model2 = lm(Sales ~ BillBoard + TV + Radio +Newspaper + TV*Radio, data=dataset)
summary(lm_model2)
```

```
##
## Call:
## lm(formula = Sales ~ BillBoard + TV + Radio + Newspaper + TV *
##     Radio, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38378 -0.19913 -0.00658  0.19753  0.43863
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.667e+01  6.589e-01  40.478   <2e-16 ***
## BillBoard    9.082e+00  1.242e-02 731.068   <2e-16 ***
## TV           2.612e-04  2.226e-03   0.117    0.907
## Radio        5.348e-03  2.787e-02   0.192    0.848
## Newspaper    3.418e-03  3.671e-03   0.931    0.354
## TV:Radio    -1.011e-05  1.428e-05  -0.708    0.480
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2627 on 130 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 1.781e+08 on 5 and 130 DF,  p-value: < 2.2e-16
```