Assignment 3

Getrude Gichuhi

2022-05 - 14

Use the dataset attached to do various extensions of Linear regression.

Use R Markdown for your submissions.

Ensure you use sales as the response variable for model fitting.

Try variety of extensions i.e. Interaction, polynomial, combinations of both interaction and polynomial.

You can choose any of the covariates. Ensure you interpret your results.

```
#install.packages("readxl")
```

Loading Libraries

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.1.3
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.2     v stringr 1.4.0
## v tidyr   1.1.3     v forcats 0.5.1
## v readr   2.1.2
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3

## Warning: package 'readr' was built under R version 4.1.3

## Warning: package 'forcats' was built under R version 4.1.3

## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lattice)
```

Load the Dataset

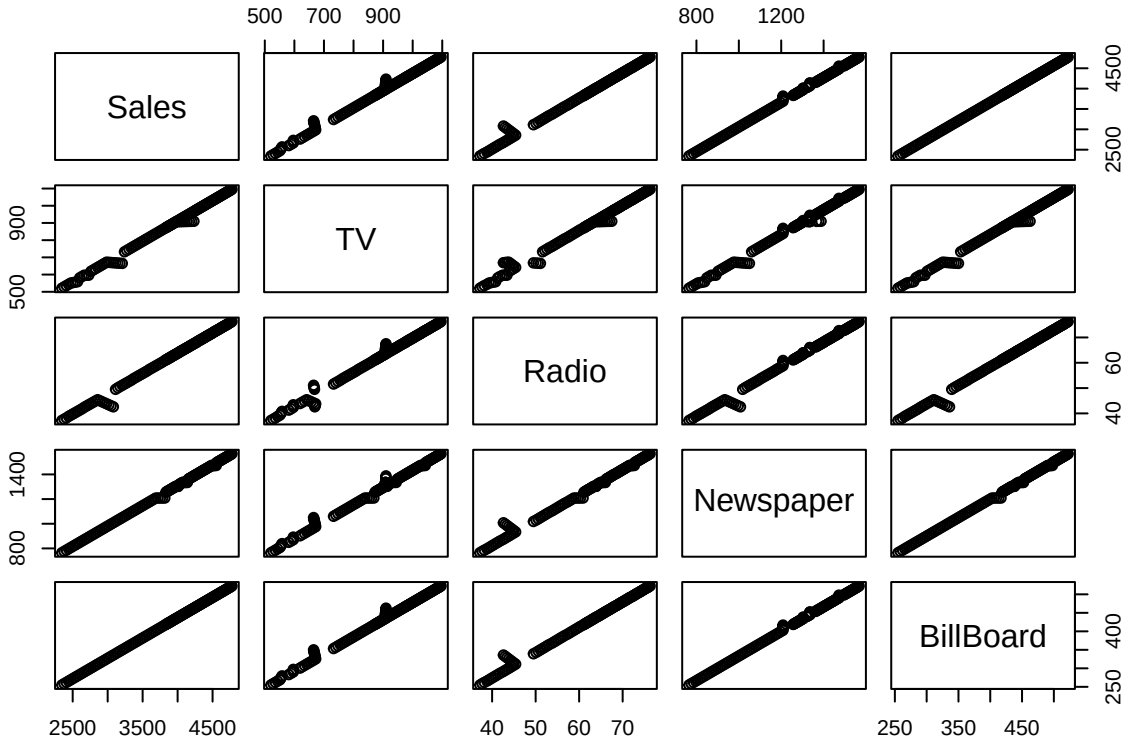
```
df <- read_excel("Sales_Dataset.xlsx")
print(head(df))
```

```
## # A tibble: 6 x 5
##   Sales    TV Radio Newspaper BillBoard
##   <dbl> <dbl> <dbl>     <dbl>     <dbl>
## 1  2345  521.  37.2      765.      255
## 2  2377  528.  37.7      775.      258.
## 3  2409  536.  38.2      786       262
## 4  2441  543.  38.8      796.      266.
## 5  2473  551.  39.3      807.      269
## 6  2505  553.  39.8      818.      273.
```

```
summary(df)
```

```
##      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(df)
```



Interaction Models

```
lm_Sales = lm(Sales~ .,data=df)
summary(lm_Sales)
```

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

Interaction of TV and Radio

```
lm_Sales1 = lm(Sales ~ TV* Radio, data=df)
summary(lm_Sales1)
```

```
##
## Call:
## lm(formula = Sales ~ TV * Radio, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.148 -18.837  -7.762   2.620 203.225
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 91.188626  93.910865   0.971    0.333
## TV           2.433081   0.243423   9.995  < 2e-16 ***
## Radio       28.517608   3.169465   8.998 2.17e-15 ***
## TV:Radio    -0.001965   0.002032  -0.967    0.335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38.19 on 132 degrees of freedom
## Multiple R-squared:  0.9969, Adjusted R-squared:  0.9968
## F-statistic: 1.4e+04 on 3 and 132 DF,  p-value: < 2.2e-16
```

interaction of TV and Newspaper

```
lm_Sales2 = lm(Sales ~ TV* Newspaper, data=df)
summary(lm_Sales2)
```

```
##
## Call:
## lm(formula = Sales ~ TV * Newspaper, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.842 -10.277  -5.689   0.975  98.608
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -7.077e+01  4.323e+01  -1.637   0.1040
## TV            7.042e-01  1.351e-01   5.213 6.99e-07 ***
## Newspaper     2.741e+00  9.309e-02  29.441  < 2e-16 ***
## TV:Newspaper -1.281e-04  4.507e-05  -2.842   0.0052 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.52 on 132 degrees of freedom
## Multiple R-squared:  0.9993, Adjusted R-squared:  0.9993
## F-statistic: 6.663e+04 on 3 and 132 DF,  p-value: < 2.2e-16
```

Interaction of TV and Billboard

```r
lm_Sales3 = lm(Sales ~ TV* BillBoard, data=df)
summary(lm_Sales3)
```

```
##
## Call:
## lm(formula = Sales ~ TV * BillBoard, data = df)
##
## 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 ***
## TV            9.857e-05  2.206e-03    0.045    0.964
## BillBoard     9.093e+00  4.308e-03 2110.654   <2e-16 ***
## TV:BillBoard -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
```

Interaction of Tv, Billboard, Radio and Newspaper

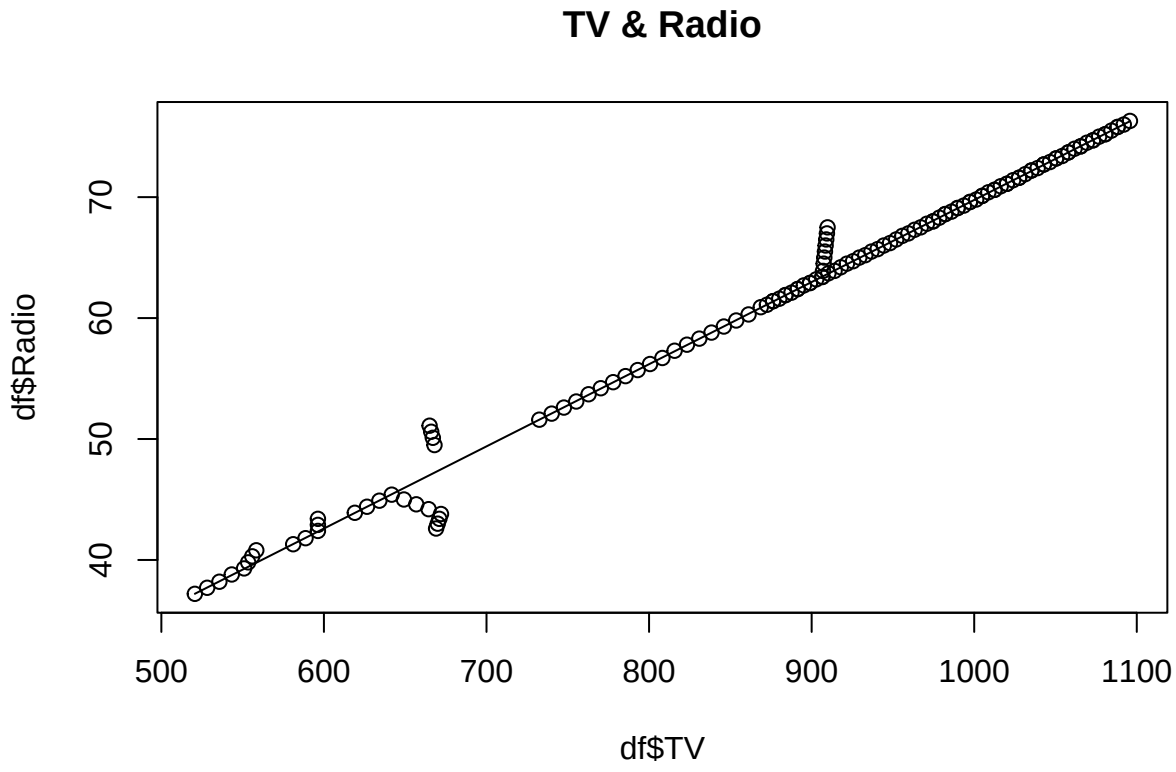
```r
lm_sales4 = lm(Sales ~ TV*BillBoard*Radio*Newspaper, data=df)
summary(lm_sales4)
```

```
##
## Call:
## lm(formula = Sales ~ TV * BillBoard * Radio * Newspaper, data = df)
##
## 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
## TV                            7.697e-02  1.834e-01   0.420   0.6755
## BillBoard                     1.259e+01  3.849e+00   3.271   0.0014 **
## Radio                        -2.472e+01  1.299e+01  -1.903   0.0595 .
## Newspaper                     3.076e-02  1.079e+00   0.028   0.9773
## TV:BillBoard                 -4.311e-03  5.191e-03  -0.831   0.4079
## TV:Radio                      4.011e-02  2.016e-02   1.989   0.0489 *
## BillBoard:Radio              -2.354e-02  4.371e-02  -0.538   0.5912
## TV:Newspaper                 -6.408e-04  1.412e-03  -0.454   0.6508
## BillBoard:Newspaper          -1.954e-03  1.466e-03  -1.332   0.1852
## Radio:Newspaper               2.176e-02  1.785e-02   1.219   0.2252
## TV:BillBoard:Radio            1.173e-06  5.238e-05   0.022   0.9822
## TV:BillBoard:Newspaper        3.620e-06  2.055e-06   1.761   0.0807 .
```

```
## TV:Radio:Newspaper           -2.390e-05  2.331e-05  -1.025   0.3072
## BillBoard:Radio:Newspaper    -5.626e-07  5.498e-06  -0.102   0.9187
## TV:BillBoard:Radio:Newspaper -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
```

Polynomial

```
scatter.smooth(x=df$TV, y=df$Radio, main="TV & Radio")
```




```
tv_lm = lm(Sales ~poly(TV,degree=2,raw = TRUE), data=df)

summary(tv_lm)
```

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

The significance level has a 99.47% chance, which falls close to the empirical rule, but not at the 99.7%.

```
News_lm = lm(Sales ~poly(Newspaper,degree=2,raw = TRUE), data=df)

summary (News_lm)
```

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

The significance level is at 99.92%.

```
radio_lm = lm(Sales ~poly(Radio,degree=2,raw = TRUE), data=df)

summary(radio_lm)
```

```
##
## Call:
```

```
## lm(formula = Sales ~ poly(Radio, degree = 2, raw = TRUE), data = df)
##
## 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
```

```
Bill_lm = lm(Sales ~poly(BillBoard,degree=2,raw = TRUE), data=df)

summary(Bill_lm)
```

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

```
mse_tv<-lm(Sales ~ TV, data=df)

mean(mse_tv$residuals^2)
```

```
## [1] 2383.077
```

```
#summary(mse_tv)
```

```
mse_radio<-lm(Sales ~ Radio, data=df)

mean(mse_radio$residuals^2)
```

```
## [1] 2868.057
```

```
mse_news<-lm(Sales ~ Newspaper, data=df)

mean(mse_news$residuals^2)
```

```
## [1] 360.0615
```

```
mse_Bill<-lm(Sales ~ BillBoard, data=df)

mean(mse_Bill$residuals^2)
```

```
## [1] 0.06657789
```

The best polynomial regression to use is Billboard since it has a small error and the adjusted R-Squared is equal to 1.