MD2201: Data Science

Name of the student: Rajeev Tapadia **Roll No.: 67**

Div:C **Batch:3**

Date of performance:

Experiment No.4

Title: Regression .

Aim: i. To construct a simple linear regression model
ii. To construct a multiple linear regression model.

Software used: Programming language R.

Data Set: Toy Sales Dataset

Code Statement:

1. **Simple Linear Regression**
   - i. Consider the Toy sales data set.
   - ii. Apply simple linear model considering response as Unit sales and explanatory variable as Price.
   - iii. Plot the scatter plot and draw the regression.
   - *iv.* What are values of R-square and residual standard error? *(Write in conclusion)*
   - v. Display all predicted values from the designed model and the corresponding values of error.

```
vi.   model <- lm(formula = sales ~ price, data = dataset)
vii.  abline(model, col = "red")
viii. summary(model)$r.squared
ix.   summary(model)$sigma
x.    predicted <- predict(model, newdata = dataset)
xi.   sales - predicted
```

```
> summary(model)$r.squared
[1] 0.6189902
> summary(model)$sigma
[1] 1997.153
> predicted <- predict(model, newdata = dataset)
> sales - predicted
        1         2         3         4         5         6         7
 2739.0295  1503.3239  1224.8708  1773.4391   917.4981  -462.4022  1945.2251
        8         9        10        11        12        13        14
-2051.6384 -1931.5609   414.3431   658.2251   479.1070 -3936.7749   688.4981
       15        16        17        18        19        20        21
-2362.2657   815.8524  1839.1070 -1340.5203 -1018.6761  1245.5970 -2747.4022
       22        23        24
 1608.3616 -3967.4030  1966.1660
```

2. Multiple Linear regression:

i. Consider Toy sales data set.
ii. Consider all variables to fit the regression model.
iii. Compare the R-square of SLR with MLR. *(Write in conclusion)*
iv. Which of the variable is more significant? Why? *(Write in conclusion)*
v. Can you reject Null hypothesis for promotion expenditure variable? *(Write in conclusion)*
vi. Which scenario from the following you will select to be applied to get maximum number of Unit sales? *(Write in conclusion)*
   a. Price=9.1$, Adexp=52,000$, Promexp=61,000$
   b. Price=8.1$, Adexp=50,000$,Promexp=60,000$

```
c. full_model <- lm(Unitsales ~ Price + Adexp + Promexp, data =
   dataset)
d. summary(full_model)
e.
f. coef_summary = summary(full_model)$coefficients
g. names(coef_summary[which(coef_summary[,4] < 0.05), 1])
h.
i. coef_summary[grep("Promexp", names(coef_summary)), 4]
j.
k. predict(full_model, newdata = data.frame(Price = 9.1, Adexp =
   52, Promexp = 61))
l. predict(full_model, newdata = data.frame(Price = 8.1, Adexp =
   50, Promexp = 60))
```

```
> summary(full_model)$r.squared
[1] 0.8588447
> coef_summary = summary(full_model)$coefficients
> names(coef_summary[which(coef_summary[,4] < 0.05), 1])
[1] "Price"   "Adexp"   "Promexp"
> coef_summary[grep("Promexp", names(coef_summary)), 4]
numeric(0)
> predict(full_model, newdata = data.frame(Price = 9.1, Ade
       1
72587.31
> predict(full_model, newdata = data.frame(Price = 8.1, Ade
       1
74542.75
```

Conclusion:

A linear regression analysis was performed on the Toy Sales data. Simple linear regression with Price as the explanatory variable showed a relationship with Unit Sales. However, including all variables in a multiple linear regression model explained a larger portion of the variance in Unit Sales (compared to R-squared of simple regression). It was found the R squared value of MLR > SLR. Finally, by analyzing the coefficients of the full model, we can predict which scenario (given options) is likely to lead to maximum unit sales.