MD2201: Data Science

Name of the student: Bhavin Patil **Roll No. 78**

Div: D **Batch: B-3**

Date of performance:

Experiment No.5

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.
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$

Code:

```
#Simple Linear Regression
cat("Simple Linear Regression\n")
dataset <- read.csv("Toy_sales_csv.csv")
l1 <- lm(Unitsales~Price, dataset)
print(summary(l1))
```

```
#Scatter Plot
library(ggplot2)
p <- ggplot(dataset, aes(Price, Unitsales))+geom_point()+geom_smooth(method = "lm", formula = y~x,
    col="red", se=F)
print(p)

#Predicted Values

cat("\n\nPredicted Values\n")
pred1 <- predict(l1)
print((pred1))

#Error Values

cat("\n\nError Values\n")
er <- dataset$Unitsales - pred1
print(er)


#Multiple Linear Regression
cat("\n------------------------------------------------------------\n")
cat("\n\n\nMultiple Linear Regression\n")
l2 <- lm(Unitsales ~ Price + Adexp + Promexp, dataset)
print(summary(l2))

#Predicted Values
cat("\n\nPredicted Values\n")
pred2 <- predict(l2)
print(pred2)

s <- data.frame(Price = c(9.1,8.1), Adexp = c(52,50), Promexp = c(61,60))
pred2 <- predict(l2, s)
print(pred2)
```

Results:

Simple Linear Regression

Call:

lm(formula = Unitsales ~ Price, data = dataset)

Residuals:

Min	1Q	Median	3Q	Max
-3967.4	-1488.3	673.4	1529.6	2739.0

.Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 114215.1 6695.9 17.057 3.60e-14 ***

Price -4913.7 821.9 -5.978 5.13e-06 ***

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

Residual standard error: 1997 on 22 degrees of freedom

Multiple R-squared: 0.619, Adjusted R-squared: 0.6017

F-statistic: 35.74 on 1 and 22 DF, p-value: 5.125e-06

Predicted Values

1 2 3 4 5 6 7 8 9 10 11 12 13

71219.97 70040.68 77362.13 78590.56 77853.50 72448.40 72939.77 75396.64 78590.56 71465.66 72939.77 74413.89 72939.77

14 15 16 17 18 19 20 21 22 23 24

77853.50 74905.27 73431.15 74413.89 73922.52 70040.68 74954.40 72448.40 75396.64 74954.40 73676.83

Error Values

1 2 3 4 5 6 7 8 9 10

2739.0295 1503.3239 1224.8708 1773.4391 917.4981 -462.4022 1945.2251 -2051.6384 -1931.5609 414.3431

11 12 13 14 15 16 17 18 19 20

658.2251 479.1070 -3936.7749 688.4981 -2362.2657 815.8524 1839.1070 -1340.5203 -1018.6761 1245.5970

21 22 23 24

-2747.4022 1608.3616 -3967.4030 1966.1660

Multiple Linear Regression

Call:

lm(formula = Unitsales ~ Price + Adexp + Promexp, data = dataset)

Residuals:

Min	1Q	Median	3Q	Max
-2808.2	-263.8	256.3	667.6	2434.5

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-25096.8	24859.6	-1.010	0.324773	
Price	-5055.3	526.4	-9.603	6.22e-09	***
Adexp	648.6	209.0	3.103	0.005602	**
Promexp	1802.6	392.8	4.589	0.000178	***

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

Residual standard error: 1275 on 20 degrees of freedom

Multiple R-squared: 0.8588, Adjusted R-squared: 0.8377

F-statistic: 40.56 on 3 and 20 DF, p-value: 1.085e-08

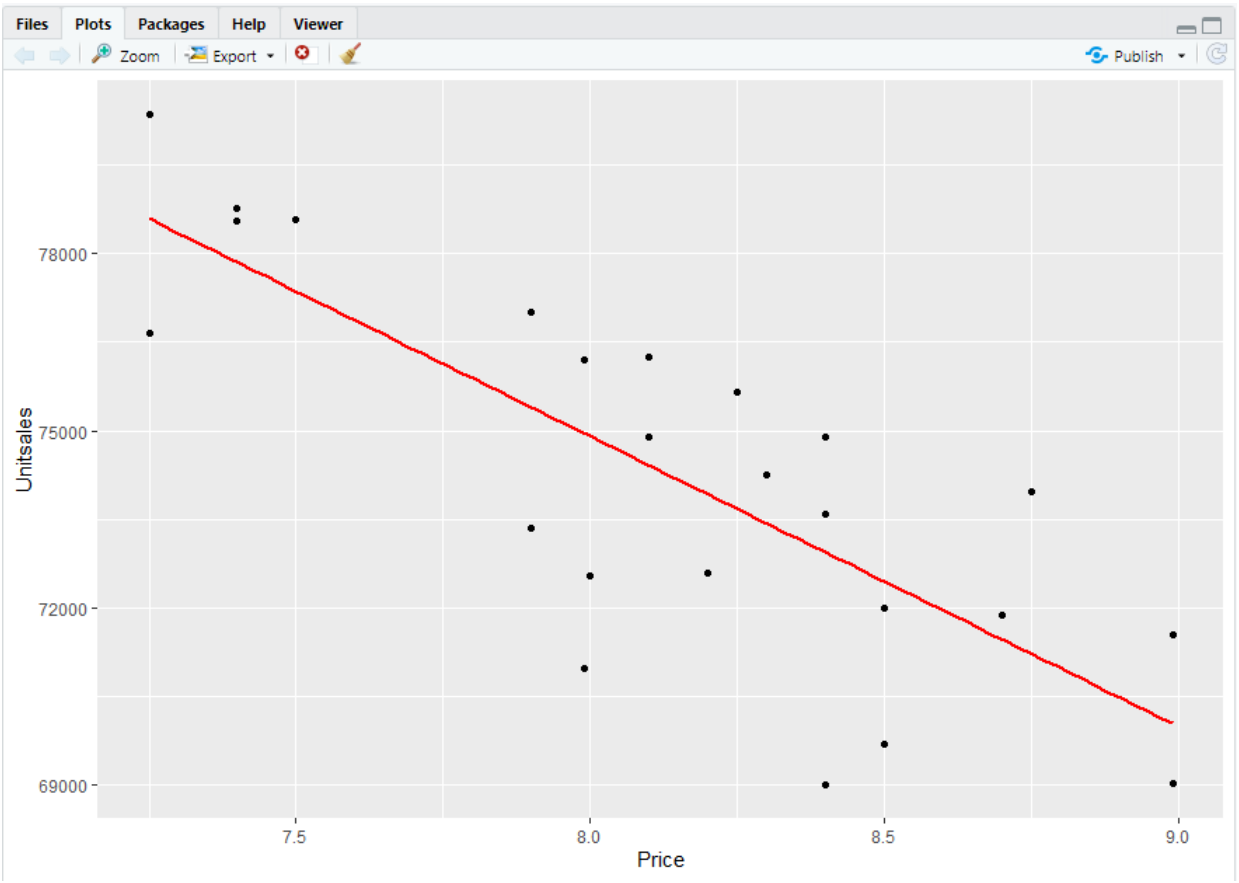
Predicted Values

1	2	3	4	5	6	7	8	9	10	11	12	13
73319.71	70866.02	76152.52	79699.84	78369.44	72725.92	73842.82	75795.30	77517.91	71055.98	73439.68	75094.58	71811.17

14	15	16	17	18	19	20	21	22	23	24
78309.53	74875.93	73289.60	76512.83	71965.73	69297.73	76252.82	69420.80	77246.94	70650.41	74685.76

1	2
72587.31	74542.75

scatter plot



Conclusion: In this assignment we have successfully created a simple linear regression model and multiple linear regression model with predicted values and error values and a scatter plot showing the regression line for Unitsales for Toys.