Final Project

MULTIPLE REGRESSION PROJECT

Course:

Analytical Methods for Business

QMB6304.003F22

* **Source of data**: Self-gathered
* **Description of the Data set used**:

We used self-gathered data in our Data set. Our data set has the data of a company’s order shipment details. This dataset has information on 120 samples of days on which the shipments were ordered and arrived late. We are emphasizing the number of orders which arrived late. By comparing the late orders and factors like the number of employees absent on the day the orders were shipped, the total number of orders shipped on that day, and the carrier of the shipments on that day, we can deduce the possible reason for the orders to arrive late. By using multiple regression models, we can find the best fit for our data which helps in analyzing the possible reasons and the impact of each of these factors.

The variables in the data set are:

1. **Late\_ship** : Number of orders shipped on the given day that arrived late to the customers.
2. **Emp\_absent** : Number of employees who were on leave on the given day.
3. **Total\_ship** : Total number of orders that were shipped on the given day
4. **Ship\_carrier** : The predominant carrier on the given day.

Dependent and independent variables in our dataset:

* In our data set, late\_ship is the dependent variable as the number of orders shipped on a day that arrives late is dependent on different factors like the total number of staff available to ship the orders on that day, the total number of orders that were shipped on that day and the carrier for those orders on that day.
* Emp\_absent, Total\_ship and Ship\_carrier are the independent variables.
* Ship\_carrier is a nominal independent variable with FedEx and UPS as the carrier companies. This is a binary variable with 0 for FedEx and 1 for UPS. The cases in this variable are evenly split among all the orders, which means that there are 60 cases for FedEx and 60 cases for UPS.



**Complete listing of the data used for our project**

|  |  |  |  |
| --- | --- | --- | --- |
| **Late\_Ship** | **Emp\_Absent** | **Total\_Ship** | **Ship\_Carrier** |
| 2 | 7 | 30 | 0 |
| 5 | 8 | 25 | 1 |
| 6 | 8 | 50 | 0 |
| 4 | 2 | 33 | 0 |
| 5 | 3 | 45 | 1 |
| 6 | 9 | 26 | 1 |
| 4 | 10 | 27 | 0 |
| 8 | 11 | 27 | 0 |
| 11 | 2 | 43 | 1 |
| 25 | 9 | 45 | 1 |
| 13 | 1 | 40 | 0 |
| 25 | 8 | 47 | 1 |
| 9 | 11 | 31 | 0 |
| 8 | 3 | 28 | 0 |
| 2 | 7 | 35 | 1 |
| 12 | 9 | 26 | 1 |
| 3 | 5 | 30 | 0 |
| 8 | 9 | 37 | 1 |
| 18 | 7 | 28 | 0 |
| 6 | 6 | 29 | 0 |
| 4 | 2 | 25 | 1 |
| 3 | 9 | 26 | 0 |
| 8 | 11 | 28 | 1 |
| 7 | 1 | 36 | 1 |
| 13 | 7 | 41 | 1 |
| 5 | 2 | 28 | 0 |
| 12 | 7 | 49 | 1 |
| 4 | 10 | 33 | 0 |
| 3 | 4 | 31 | 0 |
| 8 | 4 | 39 | 1 |
| 21 | 10 | 50 | 1 |
| 7 | 9 | 28 | 0 |
| 6 | 1 | 35 | 1 |
| 12 | 3 | 41 | 0 |
| 3 | 9 | 35 | 1 |
| 8 | 11 | 38 | 0 |
| 9 | 1 | 45 | 1 |
| 14 | 5 | 39 | 0 |
| 16 | 7 | 49 | 1 |
| 16 | 3 | 41 | 1 |
| 7 | 2 | 34 | 0 |
| 8 | 5 | 37 | 1 |
| 11 | 8 | 47 | 0 |
| 18 | 11 | 50 | 1 |
| 13 | 3 | 43 | 1 |
| 17 | 9 | 38 | 0 |
| 9 | 7 | 33 | 0 |
| 5 | 9 | 38 | 0 |
| 7 | 2 | 37 | 1 |
| 6 | 5 | 46 | 1 |
| 21 | 2 | 49 | 0 |
| 18 | 4 | 44 | 0 |
| 5 | 6 | 29 | 0 |
| 4 | 5 | 25 | 1 |
| 19 | 8 | 38 | 0 |
| 17 | 9 | 39 | 0 |
| 17 | 11 | 40 | 1 |
| 9 | 10 | 41 | 1 |
| 5 | 9 | 42 | 1 |
| 6 | 4 | 43 | 0 |
| 8 | 2 | 50 | 0 |
| 2 | 9 | 33 | 1 |
| 12 | 1 | 45 | 0 |
| 3 | 8 | 26 | 0 |
| 8 | 11 | 35 | 1 |
| 18 | 2 | 38 | 1 |
| 16 | 5 | 45 | 0 |
| 16 | 2 | 39 | 0 |
| 7 | 4 | 49 | 1 |
| 8 | 6 | 50 | 1 |
| 11 | 5 | 43 | 0 |
| 18 | 8 | 38 | 1 |
| 7 | 2 | 46 | 0 |
| 6 | 9 | 49 | 0 |
| 21 | 11 | 44 | 1 |
| 18 | 1 | 29 | 1 |
| 5 | 7 | 25 | 0 |
| 4 | 2 | 38 | 1 |
| 19 | 7 | 36 | 0 |
| 3 | 2 | 41 | 0 |
| 8 | 3 | 28 | 1 |
| 7 | 9 | 49 | 0 |
| 13 | 10 | 25 | 1 |
| 5 | 5 | 50 | 1 |
| 12 | 8 | 33 | 1 |
| 7 | 11 | 45 | 0 |
| 6 | 3 | 26 | 1 |
| 21 | 9 | 45 | 0 |
| 18 | 7 | 40 | 0 |
| 5 | 9 | 47 | 1 |
| 4 | 11 | 41 | 1 |
| 19 | 1 | 46 | 0 |
| 5 | 5 | 48 | 1 |
| 12 | 5 | 44 | 0 |
| 4 | 8 | 39 | 1 |
| 4 | 11 | 49 | 0 |
| 5 | 3 | 50 | 1 |
| 6 | 9 | 45 | 0 |
| 4 | 2 | 41 | 1 |
| 8 | 9 | 42 | 1 |
| 11 | 11 | 26 | 0 |
| 25 | 1 | 35 | 1 |
| 6 | 5 | 38 | 0 |
| 12 | 8 | 45 | 1 |
| 3 | 11 | 39 | 1 |
| 8 | 3 | 49 | 0 |
| 3 | 9 | 43 | 0 |
| 8 | 7 | 38 | 0 |
| 21 | 9 | 33 | 1 |
| 10 | 7 | 40 | 1 |
| 8 | 4 | 35 | 0 |
| 2 | 6 | 39 | 0 |
| 12 | 5 | 42 | 0 |
| 19 | 8 | 47 | 1 |
| 5 | 2 | 47 | 0 |
| 12 | 3 | 39 | 0 |
| 11 | 9 | 38 | 1 |
| 18 | 7 | 41 | 1 |
| 7 | 9 | 49 | 1 |
| 6 | 11 | 50 | 0 |

**Regression analysis for the given model combinations**

**Preprocessing:**

**Code:**

rm(list=ls())  
library(rio)  
QMB\_ <- read\_excel(“QMB .xlxs”) colnames(QMB\_)=tolower(make.names(colnames(QMB\_)))  
str(QMB\_) ***## ship carrier -- o for fedex and 1 for UPS***

**Output:**

## 'data.frame': 120 obs. of 4 variables:  
## $ late\_ship : num 2 5 6 4 5 6 4 8 11 25 ...  
## $ emp\_absent : num 7 8 8 2 3 9 10 11 2 9 ...  
## $ total\_ship : num 30 25 50 33 45 26 27 27 43 45 ...  
## $ ship\_carrier: num 0 1 0 0 1 1 0 0 1 1 ...

**Code:**

QMB\_$ship\_carrier<-as.factor(QMB\_$ship\_carrier)  
str(QMB\_)

**Output:**

## 'data.frame': 120 obs. of 4 variables:  
## $ late\_ship : num 2 5 6 4 5 6 4 8 11 25 ...  
## $ emp\_absent : num 7 8 8 2 3 9 10 11 2 9 ...  
## $ total\_ship : num 30 25 50 33 45 26 27 27 43 45 ...  
## $ ship\_carrier: Factor w/ 2 levels "0","1": 1 2 1 1 2 2 1 1 2 2 ...

**Code:**

table(QMB\_$ship\_carrier)

attach(QMB\_)

**Output:**  
## 0 1   
## 60 60

**Interpretation:** The above output shows that the fourth variable – ship\_carrier is split evenly with 60 values for Fedex and 60 values for UPS.

***## Simple Regression model of the form (y, X1)***

**Code:**  
out1<- lm(late\_ship~emp\_absent, data= QMB\_)  
summary(out1)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ emp\_absent, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.789 -4.727 -1.777 3.245 15.301   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.68738 1.18848 8.151 4.33e-13 \*\*\*  
## emp\_absent 0.01125 0.16787 0.067 0.947   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.91 on 118 degrees of freedom  
## Multiple R-squared: 3.804e-05, Adjusted R-squared: -0.008436   
## F-statistic: 0.004489 on 1 and 118 DF, p-value: 0.9467

**Interpretation**: The above summary shows that the Intercept coefficient is 9.68738 and the slope is 0.01125. The independent variable in this scenario is emp\_absent and the dependent variable is late\_ship. As the p-value is 0.947 which is very high than 0.05, we cannot reject the null hypothesis as the value is insignificant.

***## Simple Regression model of the form (y, X2)***

**Code:**

out2<- lm(late\_ship~total\_ship, data= QMB\_)  
summary(out2)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ total\_ship, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.792 -4.559 -1.519 3.663 15.908   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.97282 2.73482 1.087 0.2792   
## total\_ship 0.17485 0.06916 2.528 0.0128 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.756 on 118 degrees of freedom  
## Multiple R-squared: 0.05139, Adjusted R-squared: 0.04335   
## F-statistic: 6.392 on 1 and 118 DF, p-value: 0.01278

**Interpretation**: The above summary shows that the Intercept coefficient is 2.97282 and the slope is 0.0117485. The independent variable in this scenario is total\_ship and the dependent variable is late\_ship. As the p-value is 0.0128, less than 0.05, we can reject the null hypothesis and accept the alternate as the value is very significant.

***## Simple Regression model of the form (y, X3)***

**Code:**

out3<- lm(late\_ship~ship\_carrier, data= QMB\_)  
summary(out3)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ ship\_carrier, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.300 -4.300 -2.217 2.783 14.700   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.2167 0.7598 12.131 <2e-16 \*\*\*  
## ship\_carrier1 1.0833 1.0745 1.008 0.315   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.885 on 118 degrees of freedom  
## Multiple R-squared: 0.008542, Adjusted R-squared: 0.0001394   
## F-statistic: 1.017 on 1 and 118 DF, p-value: 0.3154

**Interpretation**: The above summary shows that the Intercept coefficient is 9.2167 and the slope is 1.0833. The independent variable in this scenario is ship\_carrier and the dependent variable is late\_ship. As the p-value is 0.3154 which is very high than 0.05, we cannot reject the null hypothesis as the value is insignificant.

***## Multiple Regression model of the form (y, X1, X2)***

**Code:**  
out4<- lm(late\_ship~emp\_absent+total\_ship, data= QMB\_)  
summary(out4)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ emp\_absent + total\_ship, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.782 -4.636 -1.564 3.565 16.085   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.73707 2.98796 0.916 0.3615   
## emp\_absent 0.03290 0.16439 0.200 0.8417   
## total\_ship 0.17557 0.06953 2.525 0.0129 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.78 on 117 degrees of freedom  
## Multiple R-squared: 0.05171, Adjusted R-squared: 0.0355   
## F-statistic: 3.19 on 2 and 117 DF, p-value: 0.04477

**Interpretation**: From the above summary, we can see that the Intercept coefficient is 2.73707 and the slopes are β1=0.03290 and β2=0.17557. The independent variables in this model are emp\_absent and total\_ship and the dependent variable is late\_ship. As the p-value for total\_ship is 0.0129 which is less than 0.05, we can reject the null hypothesis and accept the alternate, which means this variable is significant.

***## Multiple Regression model of the form (y, X1, X3)***

**Code:**

out5<- lm(late\_ship~emp\_absent+ship\_carrier, data= QMB\_)  
summary(out5)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ emp\_absent + ship\_carrier, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.310 -4.298 -2.214 2.797 14.721   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.193074 1.286781 7.144 8.32e-11 \*\*\*  
## emp\_absent 0.003826 0.168027 0.023 0.982   
## ship\_carrier1 1.082249 1.080087 1.002 0.318   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.91 on 117 degrees of freedom  
## Multiple R-squared: 0.008546, Adjusted R-squared: -0.008402   
## F-statistic: 0.5042 on 2 and 117 DF, p-value: 0.6053

**Interpretation**: From the above summary, we can see that the Intercept coefficient is 9.193074 and the slopes are β1=0.003826 and β2=1.082249. The independent variables in this model are emp\_absent and ship\_carrier and the dependent variable is late\_ship. As the p-value for both independent variables is less than 0.05, we cannot reject the null hypothesis, which means these variables are insignificant.

***## Multiple Regression model of the form (y, X2, X3)***

**Code:**

out6<- lm(late\_ship~total\_ship+ship\_carrier, data= QMB\_)  
summary(out6)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ total\_ship + ship\_carrier, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.574 -4.294 -1.606 3.921 15.426   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.63792 2.76298 0.955 0.3417   
## total\_ship 0.17140 0.06933 2.472 0.0149 \*  
## ship\_carrier1 0.93765 1.05356 0.890 0.3753   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.762 on 117 degrees of freedom  
## Multiple R-squared: 0.05777, Adjusted R-squared: 0.04166   
## F-statistic: 3.586 on 2 and 117 DF, p-value: 0.03078

**Interpretation**: From the above summary, we can see that the Intercept coefficient is 2.63792 and the slopes are β1=0.17140 and β2= 0.93765. The independent variables in this model are total\_ship and ship\_carrier and the dependent variable is late\_ship. As the p-value for total\_ship is 0.0149 which is less than 0.05, we can reject the null hypothesis and accept the alternate, which means this variable is significant.

***## Full Multiple Regression model of the form (y, X1, X2, X3)***

**Code:**  
out7<- lm(late\_ship~emp\_absent+total\_ship+ship\_carrier, data= QMB\_)  
summary(out7)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ emp\_absent + total\_ship + ship\_carrier,   
## data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.586 -4.266 -1.694 3.818 15.570   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.45383 3.00825 0.816 0.4163   
## emp\_absent 0.02609 0.16474 0.158 0.8744   
## total\_ship 0.17200 0.06972 2.467 0.0151 \*  
## ship\_carrier1 0.92974 1.05916 0.878 0.3819   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.786 on 116 degrees of freedom  
## Multiple R-squared: 0.05797, Adjusted R-squared: 0.03361   
## F-statistic: 2.379 on 3 and 116 DF, p-value: 0.07329

**Interpretation**: From the above summary, we can see that the Intercept coefficient is 2.45383 and the slopes are β1=0.02609, β2= 0.17200 and β3= 0.92974. The independent variables in this model are emp\_absent, total\_ship and ship\_carrier and the dependent variable is late\_ship. As the p value for total\_ship is 0.0151 which is less than 0.05, we can reject the null hypothesis and accept the alternate, which means this variable is significant.

***##Multiple regression model using an interaction term of the form (y, X1, X2, X1X2)***

**Code:**

out8<- lm(late\_ship~emp\_absent+total\_ship+emp\_absent\*total\_ship, data= QMB\_)  
summary(out8)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ emp\_absent + total\_ship + emp\_absent \*   
## total\_ship, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.974 -4.756 -1.488 3.809 15.943   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 4.335244 6.561976 0.661 0.510  
## emp\_absent -0.203687 0.879595 -0.232 0.817  
## total\_ship 0.134641 0.164973 0.816 0.416  
## emp\_absent:total\_ship 0.006089 0.022234 0.274 0.785  
##   
## Residual standard error: 5.803 on 116 degrees of freedom  
## Multiple R-squared: 0.05232, Adjusted R-squared: 0.02782   
## F-statistic: 2.135 on 3 and 116 DF, p-value: 0.09959

**Interpretation**: From the above summary, we can see that the Intercept coefficient is 4.335244 and the slopes are β1=-0.203687, β2= 0.134641, and β3= 0.006089. The independent variables in this model are emp\_absent, total\_ship, and the interaction between emp\_absent and total\_ship. The dependent variable is late\_ship. As the p values for all the variables are greater than 0.05, we cannot reject the null hypothesis, which means the variables are insignificant.

***##simple regression model using squared terms of the form (y, X1, X12)***

**Code:**

***## regression of squared values***  
QMB\_$emp\_abs2 <- QMB\_$emp\_absent^2  
QMB\_$total\_ship2<- QMB\_$total\_ship^2  
  
out9<- lm(late\_ship~emp\_absent+emp\_abs2, data = QMB\_)  
summary(out9)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ emp\_absent + emp\_abs2, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.735 -4.482 -1.735 3.099 15.438   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.54295 2.03842 5.172 9.68e-07 \*\*\*  
## emp\_absent -0.38555 0.78511 -0.491 0.624   
## emp\_abs2 0.03287 0.06352 0.517 0.606   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.929 on 117 degrees of freedom  
## Multiple R-squared: 0.002321, Adjusted R-squared: -0.01473   
## F-statistic: 0.1361 on 2 and 117 DF, p-value: 0.8729

**Interpretation**: From the above summary, we can see that the Intercept coefficient is 10.54295 and the slopes are β1=-0.38555 and β2= 0.03287. The independent variable in this model is emp\_absent and the dependent variable is late\_ship. We have a squared term of emp\_absent which is an independent variable here. As the p values for all the variables are greater than 0.05, we cannot reject the null hypothesis, which means the variables are insignificant.

***##simple regression model using squared terms of the form (y, X2, X22)***

**Code:**

out10<- lm(late\_ship~total\_ship+total\_ship2,data = QMB\_)  
summary(out10)

**Output:**

##   
## Call:  
## lm(formula = late\_ship ~ total\_ship + total\_ship2, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.578 -4.456 -1.801 3.706 15.201   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -15.75934 13.21769 -1.192 0.2356   
## total\_ship 1.21075 0.71858 1.685 0.0947 .  
## total\_ship2 -0.01373 0.00948 -1.448 0.1502   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.73 on 117 degrees of freedom  
## Multiple R-squared: 0.06809, Adjusted R-squared: 0.05216   
## F-statistic: 4.275 on 2 and 117 DF, p-value: 0.01615

**Interpretation**: From the above summary, we can see that the Intercept coefficient is -15.75934 and the slopes are β1=1.21075 and β2= -0.01373. The independent variable in this model is total\_ship and squared value of total\_ship. The dependent variable is late\_ship. As the p-value is 0.01615 which is less than 0.05, so we are enough confident to reject the null hypothesis and it is significant for the regression model.

***##Determining which fit is the best fit***

**Code:**

library(stargazer)

**Output:**

##   
## Please cite as:

## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. <https://CRAN.R-project.org/package=stargazer>

**Code:**

stargazer(out1,out2,out3,out4,out5,out6,out7,out8,out9,out10, type= "html", out="Project1.htm")

**Output:**

Calendar

Description automatically generated



***## best model found is out10***  
**Interpretation**:

From the above comparison sheet for all the regression model output, we can say that the model out10 is the best model as its R square value is best and its adjusted R2 value is the highest. It says that if the IV’s in the out10 regression model explains the dependent variable best. Also, the p-value is less than 0.05 for both IV’s, so we are enough confident to reject the null hypothesis and it is significant for the regression model.

***##Assessment of whether our model violates any of the common regression assumptions for linearity, independence of errors, normality of errors, and equality of error variances.***

***## Checking LINE assumptions***

**Code:**  
par(mfrow=c(2,2))  
*#Linearity*  
plot(QMB\_$late\_ship, out10$fitted.values,  
 pch=16, main="Late Shipment Vs Fitted Values")  
abline(0,1,lwd=3,col="red")  
*#Normality*  
qqnorm(out10$residuals,pch=16,  
 main="Normality of Residuals")  
qqline(out10$residuals,lwd=3,col="red")  
hist(out10$residuals, col = "cadetblue3", probability = TRUE,  
 main="Histogram of Residuals")  
curve(dnorm(x,0,sd(out10$residuals)),  
 from=min(out10$residuals),  
 to=max(out10$residuals),  
 lwd=3,col="brown4",add=TRUE)  
*#Equality of Variances*  
plot(out10$fitted.values,rstandard(out10),pch=16,  
 main="Equality of Variance- Standardized Residuals")  
abline(0,0,col="red",lwd=3)

par(mfrow=c(1,1))

**Output:**

**Chart

Description automatically generated**

**Interpretation**: For the best fit model that is out10, considering the LINE assumptions and from the graphs obtained above, we can deduce the following statements:

The data in our model is not linear as the data points are spread throughout the space and there are very few points near the abline.

The data in our model satisfies Normality assumption as almost all the data points lie on the abline in the Normality of Residuals graph, with very few outliers. We can see that data is right skewed.

The data satisfies the Equality assumption as the data points are almost mirroring each other above and below the abline in the Equality of Variance graph. Also, it is heteroscedastic.

**##Written interpretations of the model’s slope and intercept coefficients.**

**Code:**

summary(out10)

**Output:**

## Call:  
## lm(formula = late\_ship ~ total\_ship + total\_ship2, data = QMB\_)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.578 -4.456 -1.801 3.706 15.201   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -15.75934 13.21769 -1.192 0.2356   
## total\_ship 1.21075 0.71858 1.685 0.0947 .  
## total\_ship2 -0.01373 0.00948 -1.448 0.1502   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.73 on 117 degrees of freedom  
## Multiple R-squared: 0.06809, Adjusted R-squared: 0.05216   
## F-statistic: 4.275 on 2 and 117 DF, p-value: 0.01615

**Interpretation**:

Model Output: Y= -15.75 +1.210(total\_ship)-0.013 (total\_ship2)

The intercept coefficient is -15.75934 and the slopes are 1.21075 and -0.01373. As the beta coefficient for total\_ship is positive, it means that a change in this variable will affect the output variable directly or positively. This means, if the total shipments on a day increase, the number of late arrived shipments also increases. As the coefficient of total\_ship2 is negative, the change in this variable will affect the outcome variable inversely.

***##The two types of prediction confidence intervals resulting from the independent variable values you choose***

**Code:**

***## Two types of prediction confidence interval***  
late\_ship\_predict<- data.frame(total\_ship = 39, total\_ship2=1521)  
predict(out10,late\_ship\_predict,interval = "predict")

**Output:**

## fit lwr upr  
## 1 10.57833 -0.8672893 22.02395

**Code:**

predict(out10,late\_ship\_predict,interval = "confidence")

**Output:**

## fit lwr upr  
## 1 10.57833 9.084852 12.07181

**Interpretation**: We can see that the fitted late orders for a given day when the total no of orders are 39 is 10.57833. Hence, we can say that if the total number of orders placed on a day is 39, then there could be approximately 11 orders which may arrive late. On a larger population, for 39 total orders, it could lie between a lower limit of 9.084852 and an upper limit of 12.07181 at a 95% confidence interval. For prediction, we have the lower limit as -0.8672893 and the upper limit as 22.02395 which is a wider range.