**Stats 512 Final Project – PC**

**Background**

Our company makes intravenous solutions for medical use. These are produced primarily in flexible, plastic containers, but a legacy product still exists where the solution is filled into glass bottles. On the production line, the bottles are loaded onto a conveyor, washed and dried, and then are automatically filled with the desired solution. Immediately following this step, stoppers are loosely placed on the filled bottles. The next step is a critical process: a machine lifts the stopper from the bottle in order to draw a vacuum, then replaces and seals the stopper, covering it with an aluminum shell. Following this process, the bottles are steam sterilized, labeled and packaged for shipment.

**Introduction**

The vacuum sealing step in the process outlined above is the subject of this analysis. It is essential to obtain a minimum level of vacuum within the glass container in order to ensure a consistent, durable seal. The seal provides a barrier between the sterilized contents of the bottle and the outside environment. The manufacturing process must be controlled in order to consistently produce units with adequate vacuum. The specified minimum vacuum for finished product is 15 inches of mercury (in-Hg). The purpose of this experiment is to determine the optimal settings to ensure that the minimum vacuum is consistently achieved.

**Data Description**

The critical operating parameters within this step are the conveyor **Line Speed** (measured in bottles per minute, bpm) and **Machine Vacuum Setting** (measured in inches of mercury, in-Hg). These are the predictor variables. The response variable is the resulting **Bottle Vacuum Reading** (measured with a vacuum gauge equipped with a needle that can pierce the stopper in order to measure internal bottle vacuum level). This is a destructive test and so each bottle can only be measured one time. The experimental unit is a bottle.

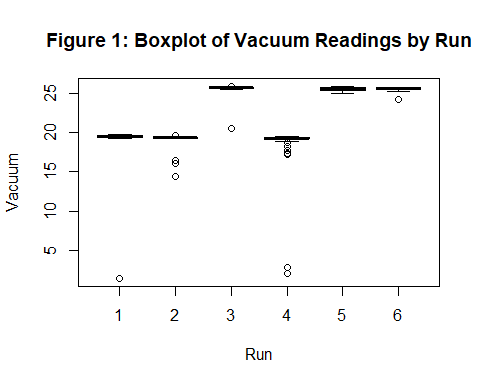
For Line Speed, there were 4 levels chosen and 2 levels of Vacuum Setting. The bottle Vacuum readings were taken 24 hours following steam sterilization. Both predictor and response variables are continuous. For each run, there were 48 samples measured. This information is summarized in Table 1.

**Summary Statistics**

The results of each run were averaged and given in Table 1, below:

**Table 1: Experiment Results**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Predictors** | | **Response** | | |
| Run | Line Speed | Vacuum Setting | # Vacuum Readings (n) | Average Vacuum | Standard Deviation |
| 1 | 60 | 21.0 | 48 | 18.75625 | 3.658735 |
| 2 | 55 | 20.6 | 48 | 19.15417 | 0.949347 |
| 3 | 55 | 27.4 | 48 | 25.55208 | 0.750883 |
| 4 | 85 | 20.6 | 48 | 18.42917 | 3.425621 |
| 5 | 85 | 27.4 | 48 | 25.53958 | 0.192121 |
| 6 | 100 | 27.4 | 48 | 25.55833 | 0.222972 |



The boxplot shown in Figure 1 illustrates the dispersion of the readings. Runs #1 and #4 had individual bottles with very low vacuum readings.

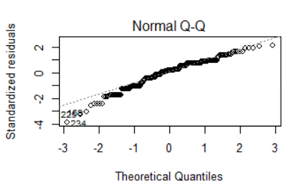
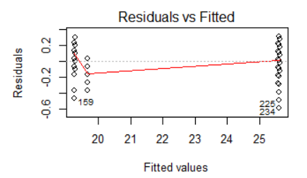
**Analysis**

The experiment was not a balanced design due to the fact that the production line was not available to perform all combinations of factors. A linear model was used to fit the data, then the results were checked for outliers.

The extreme outliers were removed and the model was fitted to the reduced dataset. The results were similar in that the machine vacuum setting was the only significant factor for the new model (Table 2). However the model assumptions were still not met, so another round of outlier removal was performed. This was repeated until a residuals Q-Q Plot showed reasonable linearity (Figure 2). The 4 models used to evaluate the data are given in Table 2. These included the all combinations of with or without outliers and with or without predictor variable interactions.

Finally, the predictive function was used to calculate the minimum machine vacuum setting required to achieve the specified Bottle Vacuum Minimum Reading. The prescribed machine vacuum setting was taken as the minimum prediction value required to achieve a minimum value of 15.0 in-Hg for the lower bound of the prediction interval of the Bottle Vacuum Reading. The prediction was performed for each model at the highest and lowest speed settings available (55 and 100 bottles per minute). These results are listed in table 4.

**Figure 2: Fitted vs Residuals and Normal Q-Q Plot of Data with Outliers Removed**



**Table 2: Significant Predictor variables**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Model Description | Significant Factors | P-Value |
| 1 | Raw Data fit with  Vac Setting \* Speed interaction | Vacuum Setting | < 0.001 |
| 2 | Raw Data  without interaction | Vacuum Setting | < 0.001 |
| 3 | Outliers Removed  With Vac Setting\*Speed interaction | Vacuum Setting | < 0.001 |
| 4 | Outliers Removed without interaction | Vacuum Setting | < 0.001 |

**Table 3: Coefficients of Fitted Model**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Model Description | Intercept | | Vac\_Setting | Speed | Vac\_Set\*Speed |
| 1 | Raw Data With  Vac Setting \* Speed interaction | 2.147221 | | 0.852339 | -0.065035 | 0.002392 |
| 2 | Raw Data  without interaction | -2.105798 | | 1.023231 | -0.004841 | n/a |
| 3 | Outliers Removed  With Vac Setting\*Speed interaction | -0.3983257 | 0.9538216 | | 0.0073200 | -0.0003328 |
| 4 | Outliers Removed without interaction | 0.1908541 | 0.9302840 | | -0.0011047 | n/a |

**Table 4: Prediction Values from Each Model**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Vacuum Setting | Speed | Lower bound of Predicted  Vacuum Reading |
| 1 | 20.9 | 55 | 14.95 |
| 1 | 21.0 | 55 | 15.05 |
| 1 | 21.6 | 100 | 14.99 |
| 1 | 21.7 | 100 | 15.10 |
| 2 | 21.0 | 55 | 14.94 |
| 2 | 21.1 | 55 | 15.04 |
| 2 | 21.2 | 100 | 14.91 |
| 2 | 21.3 | 100 | 15.01 |
| 3 | 16.3 | 55 | 14.95 |
| 3 | 16.4 | 55 | 15.05 |
| 3 | 16.2 | 100 | 14.93 |
| 3 | 16.3 | 100 | 15.02 |
| 4 | 16.3 | 55 | 14.99 |
| 4 | 16.4 | 55 | 15.09 |
| 4 | 16.3 | 100 | 14.94 |
| 4 | 16.4 | 100 | 15.03 |

**Results and Conclusions**

All models with or without outliers, identified the machine vacuum setting as the only critical parameter. The Speed setting was not significant at the α=0.05 level. Neither was the interaction.

From the prediction values table, the difference between Models 1 & 2 versus 3 & 4 can easily be seen. There is very little difference whether or not the predictor variables interaction is taken into account. This difference amounts to a few tenths in terms of the Machine Vacuum Setting parameter. However, the difference is quite large when including or excluding the outliers. The reasons for the outliers are due to aberrations in the capping/sealing process. This could include improper placement of the stopper on the bottle, mismatched diameters of the bottle opening and the stopper, improper fitting of the vacuum seal head to the shoulder of the bottle. All of these are mostly attributable to dimensional variation or distortions of the bottle and or stopper. Therefore, it is recommended to exclude the outliers and use models 3 or 4 to determine the minimum allowed Machine Vacuum Setting. The minimum prescribed would therefore be 16.4 in-Hg. This allows for any speed setting within the limits of the speed levels tested (55 – 100 bottles per minute).

The maximum value that the machine vacuum setting is capable of producing could change over time. This value slowly decreases due to vacuum pump wear and is a good indicator of when preventive maintenance is due. It is therefore recommended to service the vacuum pump whenever the pump is incapable of meeting this minimum value. From a practical approach, it would be better to service the vacuum pump when the pump could no longer pull vacuum to 18.5 in-Hg. This allows for normal variation so as to prevent inadvertently dropping below the minimum requirement.

**Future Study**

Additional tests should be conducted to determine and eliminate the causes of the outliers. One such test could be to rerun the same bottle and stopper combination to see if it would result in a similar low reading. This would confirm bottle and/or stopper dimensional issues. Exchanging the stopper and bottle could help pinpoint which component is the source. If the same bottle/stopper combination did not result in a low bottle vacuum reading after a second run, other variables should be explored.

**References**

1. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
2. John Fox and Sanford Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
3. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

**Appendix - R Code**

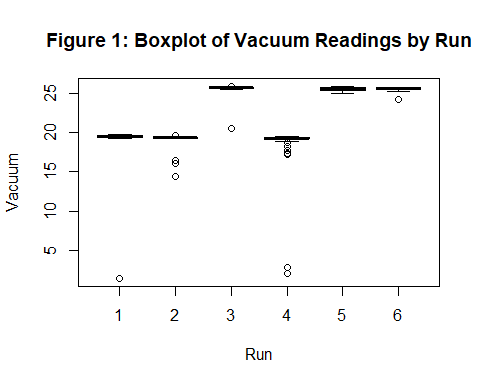
library(ggplot2)  
library(dplyr)  
library(car)  
library(emmeans)  
library(multcomp)  
Bottles <- read.csv("C:/Users/paulc/Documents/Documents/Stats CSU/Stats 512/Project/Experimental Results.csv", quote = "'")  
str(Bottles)

## 'data.frame': 288 obs. of 4 variables:  
## $ Run : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Vacuum : num 19.5 19.5 19.5 19.6 19.7 19.6 19.5 19.6 19.7 19.6 ...  
## $ Speed : int 60 60 60 60 60 60 60 60 60 60 ...  
## $ Vac\_Set: num 21 21 21 21 21 21 21 21 21 21 ...

Bottles$Run <- as.factor(Bottles$Run)  
  
SumStats <- summarise(group\_by(Bottles, Run),  
n = n(),  
mean = mean(Vacuum),  
sd = sd(Vacuum))  
str(SumStats)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 6 obs. of 4 variables:  
## $ Run : Factor w/ 6 levels "1","2","3","4",..: 1 2 3 4 5 6  
## $ n : int 48 48 48 48 48 48  
## $ mean: num 18.8 19.2 25.6 18.4 25.5 ...  
## $ sd : num 3.659 0.949 0.751 3.426 0.192 ...

library(ggplot2)  
boxplot(Vacuum ~ Run, data = Bottles, main = "Figure 1: Boxplot of Vacuum Readings by Run")



Model1 <- lm(Vacuum ~ Vac\_Set\*Speed, data = Bottles)  
Anova(Model1, type = 3)

## Anova Table (Type III tests)  
##   
## Response: Vacuum  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 1.05 1 0.2364 0.6272   
## Vac\_Set 102.13 1 22.9033 2.743e-06 \*\*\*  
## Speed 4.97 1 1.1149 0.2919   
## Vac\_Set:Speed 4.33 1 0.9701 0.3255   
## Residuals 1266.38 284   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(Model1)

##   
## Call:  
## lm(formula = Vacuum ~ Vac\_Set \* Speed, data = Bottles)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.7584 0.0477 0.1706 0.5416 1.1339   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.147221 4.416635 0.486 0.627   
## Vac\_Set 0.852339 0.178100 4.786 2.74e-06 \*\*\*  
## Speed -0.065035 0.061593 -1.056 0.292   
## Vac\_Set:Speed 0.002392 0.002429 0.985 0.325   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.112 on 284 degrees of freedom  
## Multiple R-squared: 0.7227, Adjusted R-squared: 0.7197   
## F-statistic: 246.7 on 3 and 284 DF, p-value: < 2.2e-16

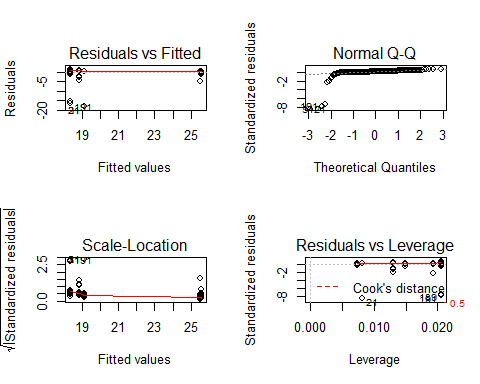
Model2 <- lm(Vacuum ~ Vac\_Set + Speed, data = Bottles)  
Anova(Model2, type = 3)

## Anova Table (Type III tests)  
##   
## Response: Vacuum  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 22.94 1 5.1458 0.02405 \*   
## Vac\_Set 2887.73 1 647.6737 < 2e-16 \*\*\*  
## Speed 1.77 1 0.3981 0.52858   
## Residuals 1270.71 285   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(Model2)

##   
## Call:  
## lm(formula = Vacuum ~ Vac\_Set + Speed, data = Bottles)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.6916 0.0355 0.2534 0.6084 0.9388   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.105798 0.928301 -2.268 0.0241 \*   
## Vac\_Set 1.023231 0.040206 25.449 <2e-16 \*\*\*  
## Speed -0.004841 0.007673 -0.631 0.5286   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.112 on 285 degrees of freedom  
## Multiple R-squared: 0.7217, Adjusted R-squared: 0.7198   
## F-statistic: 369.6 on 2 and 285 DF, p-value: < 2.2e-16

par(mfrow = c(2, 2))  
plot(Model1)

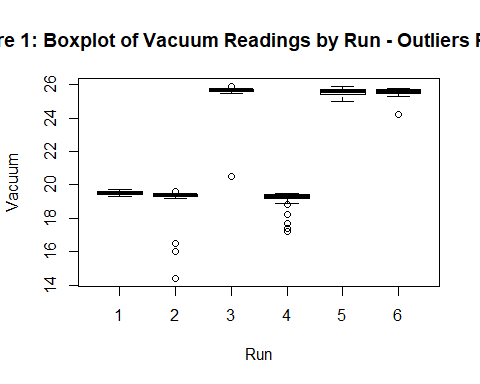


outlierTest(Model1)

## rstudent unadjusted p-value Bonferroni p  
## 21 -9.740231 1.6026e-19 4.6155e-17  
## 31 -9.740231 1.6026e-19 4.6155e-17  
## 191 -8.828427 1.1315e-16 3.2588e-14  
## 189 -8.288726 4.6566e-15 1.3411e-12

Bottles\_out <- Bottles[-c(21,31,191,189),]

boxplot(Vacuum ~ Run, data = Bottles\_out, main = "Figure 1: Boxplot of Vacuum Readings by Run - Outliers Removed")



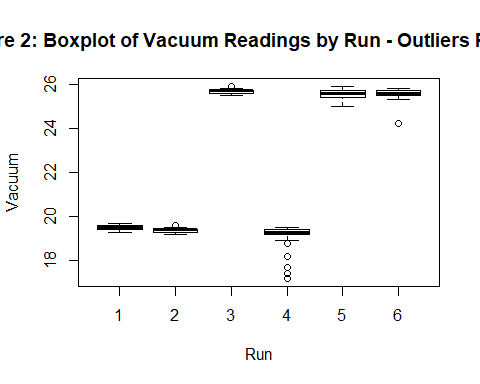
ModelA <- lm(Vacuum ~ Vac\_Set\*Speed, data = Bottles\_out)  
Anova(ModelA, type = 3)

## Anova Table (Type III tests)  
##   
## Response: Vacuum  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 0.001 1 0.0031 0.9558   
## Vac\_Set 121.073 1 383.3200 <2e-16 \*\*\*  
## Speed 0.012 1 0.0381 0.8454   
## Vac\_Set:Speed 0.011 1 0.0342 0.8534   
## Residuals 88.439 280   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

outlierTest(ModelA)

## rstudent unadjusted p-value Bonferroni p  
## 119 -10.774227 7.3229e-23 2.0797e-20  
## 65 -9.855430 7.4239e-20 2.1084e-17  
## 68 -5.976601 6.9392e-09 1.9707e-06  
## 86 -4.935516 1.3770e-06 3.9107e-04

Bottles\_out2 <- Bottles\_out[-c(63,66,84,117),]  
boxplot(Vacuum ~ Run, data = Bottles\_out2, main = "Figure 2: Boxplot of Vacuum Readings by Run - Outliers Removed")



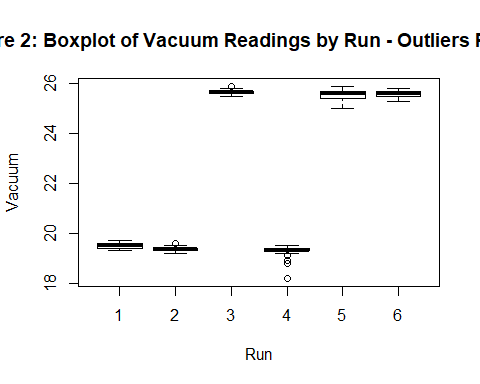
ModelB <- lm(Vacuum ~ Vac\_Set\*Speed, data = Bottles\_out2)  
Anova(ModelB, type = 3)

## Anova Table (Type III tests)  
##   
## Response: Vacuum  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 0.222 1 2.8501 0.09250 .   
## Vac\_Set 110.665 1 1418.8977 < 2e-16 \*\*\*  
## Speed 0.344 1 4.4137 0.03656 \*   
## Vac\_Set:Speed 0.222 1 2.8401 0.09307 .   
## Residuals 21.526 276   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

outlierTest(ModelB)

## rstudent unadjusted p-value Bonferroni p  
## 148 -7.573802 5.5303e-13 1.5485e-10  
## 147 -6.643053 1.6402e-10 4.5924e-08  
## 149 -6.643053 1.6402e-10 4.5924e-08  
## 184 -5.339720 1.9516e-07 5.4646e-05  
## 278 -5.028242 8.9394e-07 2.5030e-04

Bottles\_out3 <- Bottles\_out2[-c(141, 142, 143, 178, 270),]  
boxplot(Vacuum ~ Run, data = Bottles\_out3, main = "Figure 2: Boxplot of Vacuum Readings by Run - Outliers Removed")



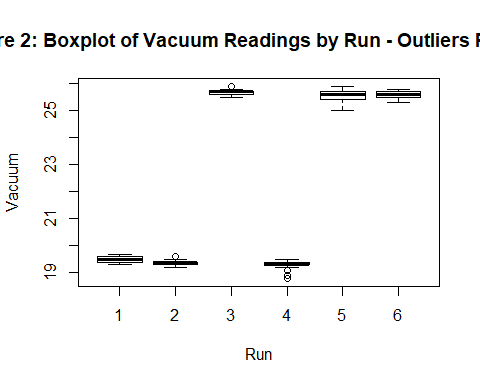
ModelC <- lm(Vacuum ~ Vac\_Set\*Speed, data = Bottles\_out3)  
Anova(ModelC, type = 3)

## Anova Table (Type III tests)  
##   
## Response: Vacuum  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 0.008 1 0.2827 0.5954   
## Vac\_Set 118.155 1 4439.0348 <2e-16 \*\*\*  
## Speed 0.014 1 0.5341 0.4655   
## Vac\_Set:Speed 0.028 1 1.0332 0.3103   
## Residuals 7.213 271   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

outlierTest(ModelC)

## rstudent unadjusted p-value Bonferroni p  
## 165 -7.192807 6.2561e-12 1.7204e-09

Bottles\_out3 <- Bottles\_out3[-c(156),]  
boxplot(Vacuum ~ Run, data = Bottles\_out3, main = "Figure 2: Boxplot of Vacuum Readings by Run - Outliers Removed")



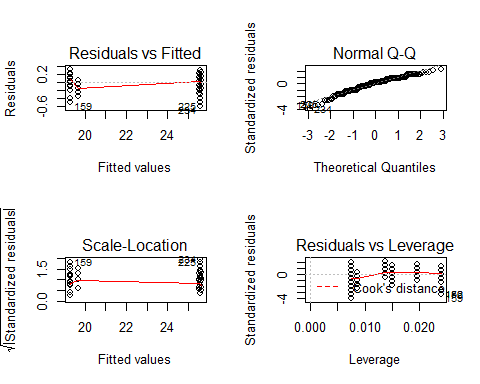
Model3 <- lm(Vacuum ~ Vac\_Set \* Speed, data = Bottles\_out3)  
Anova(Model3, type = 3)

## Anova Table (Type III tests)  
##   
## Response: Vacuum  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 0.033 1 1.4931 0.22280   
## Vac\_Set 119.279 1 5320.2724 < 2e-16 \*\*\*  
## Speed 0.057 1 2.5388 0.11225   
## Vac\_Set:Speed 0.077 1 3.4135 0.06576 .   
## Residuals 6.053 270   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(Model3)

##   
## Call:  
## lm(formula = Vacuum ~ Vac\_Set \* Speed, data = Bottles\_out3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.58345 -0.08157 0.02408 0.11655 0.32408   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.3983257 0.3259839 -1.222 0.2228   
## Vac\_Set 0.9538216 0.0130768 72.940 <2e-16 \*\*\*  
## Speed 0.0073200 0.0045941 1.593 0.1122   
## Vac\_Set:Speed -0.0003328 0.0001801 -1.848 0.0658 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1497 on 270 degrees of freedom  
## Multiple R-squared: 0.9977, Adjusted R-squared: 0.9977   
## F-statistic: 3.892e+04 on 3 and 270 DF, p-value: < 2.2e-16

par(mfrow = c(2, 2))  
plot(Model3)



Model4 <- lm(Vacuum ~ Vac\_Set + Speed, data = Bottles\_out3)  
Anova(Model4, type = 3)

## Anova Table (Type III tests)  
##   
## Response: Vacuum  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 0.18 1 7.8984 0.005309 \*\*   
## Vac\_Set 2230.42 1 98606.3496 < 2.2e-16 \*\*\*  
## Speed 0.09 1 3.8719 0.050121 .   
## Residuals 6.13 271   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(Model4)

##   
## Call:  
## lm(formula = Vacuum ~ Vac\_Set + Speed, data = Bottles\_out3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.58673 -0.07016 0.02984 0.10606 0.31327   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.1908541 0.0679095 2.810 0.00531 \*\*   
## Vac\_Set 0.9302840 0.0029625 314.016 < 2e-16 \*\*\*  
## Speed -0.0011047 0.0005614 -1.968 0.05012 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1504 on 271 degrees of freedom  
## Multiple R-squared: 0.9977, Adjusted R-squared: 0.9976   
## F-statistic: 5.787e+04 on 2 and 271 DF, p-value: < 2.2e-16

NewData1 <- data.frame(Vac\_Set = 21.0, Speed = 55)  
  
predict(Model1, NewData1, interval = "prediction")

## fit lwr upr  
## 1 19.23235 15.05192 23.41278

NewData2 <- data.frame(Vac\_Set = 21.1, Speed = 55)  
  
predict(Model2, NewData2, interval = "prediction")

## fit lwr upr  
## 1 19.2181 15.04471 23.3915

NewData3 <- data.frame(Vac\_Set = 16.4, Speed = 55)  
  
predict(Model3, NewData3, interval = "prediction")

## fit lwr upr  
## 1 15.34675 15.04545 15.64804

NewData4 <- data.frame(Vac\_Set = 16.4, Speed = 55)  
  
predict(Model4, NewData4, interval = "prediction")

## fit lwr upr  
## 1 15.38675 15.08717 15.68633