ST_518 Project

MA, HK, BA, JF

2022-11-19

Executive Summary

Text goes here. . . ..

Introduction

Experimental Design

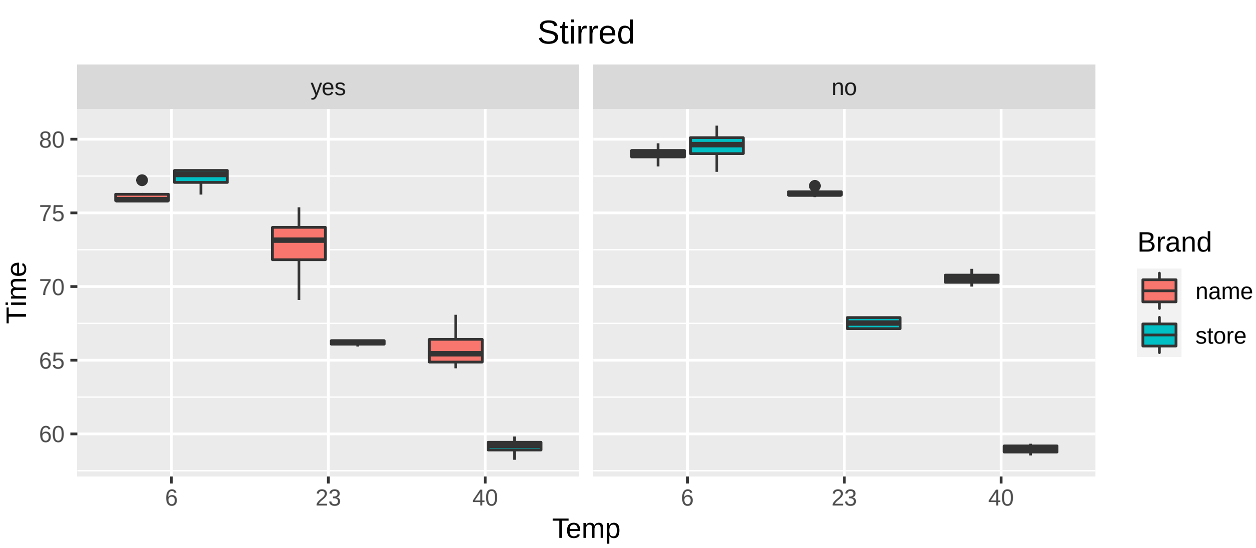
Description here

Exploratory Data Analysis

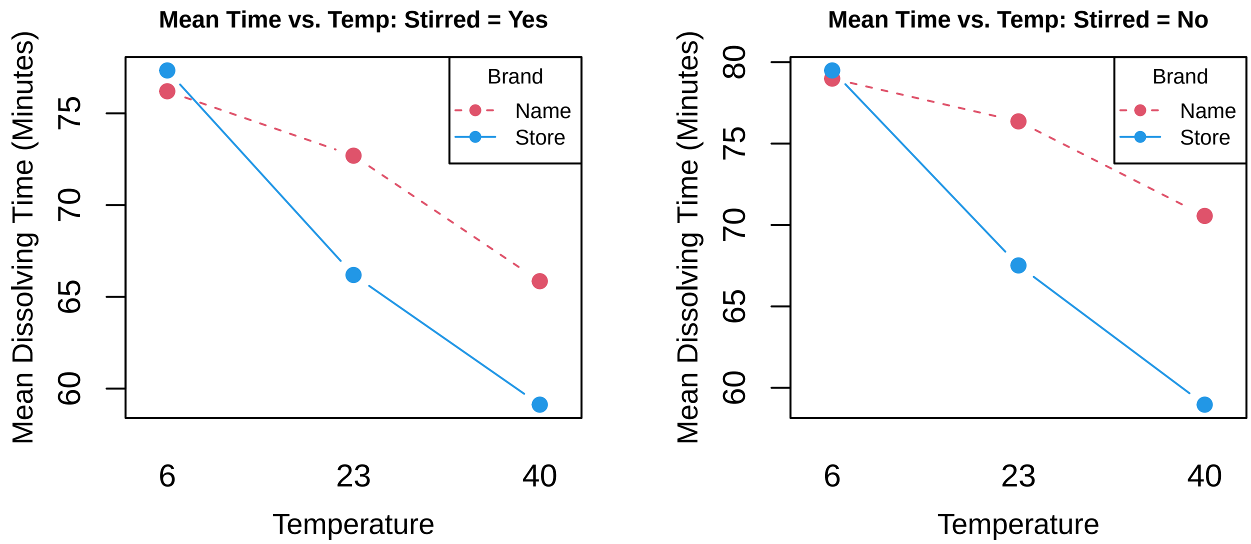
For this study we are presented with data from an 'Effervescent Experiment'. The data contained dissolving times of two different brands of cold medicine tablets that were obtained under various conditions. Those conditions included varying water temperatures (6°C, 23°C, 40°C) and the presence of stirring (magnetic stir bar at 350 rpm). This was a complete block design with stirring acting as the blocking effect. In all, the data contained 48 rows and 6 columns. The 6 columns include 3 explanatory variables (Brand, Temp, Stirred categorical factors), 2 response variables (Time and Org Time, both numerical) and 1 descriptor (sample order). Prior to starting any analysis, we will explore the data to gain an understanding of what to expect and to check for violations of any assumptions.

Brand	Temp	Stirred	25%	Mean	Median	75%	Var	n
name	6	yes	75.83358	76.20241	75.89223	76.26107	0.4593492	4
name	6	no	78.79910	78.99061	79.04435	79.23586	0.4146440	4
name	23	yes	71.82180	72.69145	73.14894	74.01859	6.9869087	4
name	23	no	76.20492	76.36351	76.27622	76.43481	0.1078134	4
name	40	yes	64.87321	65.85343	65.43863	66.41886	2.5499751	4
name	40	no	70.28754	70.55511	70.50947	70.77705	0.2544033	4
store	6	yes	77.06561	77.33703	77.60659	77.87801	0.5964884	4
store	6	no	79.01994	79.49240	79.63219	80.10465	1.6942517	4
store	23	yes	66.08831	66.19126	66.22629	66.32923	0.0411024	4
store	23	no	67.14393	67.51552	67.52360	67.89520	0.2060739	4
store	40	yes	58.90895	59.12529	59.21659	59.43293	0.4320148	4
store	40	no	58.76884	58.96347	58.99050	59.18513	0.1202191	4

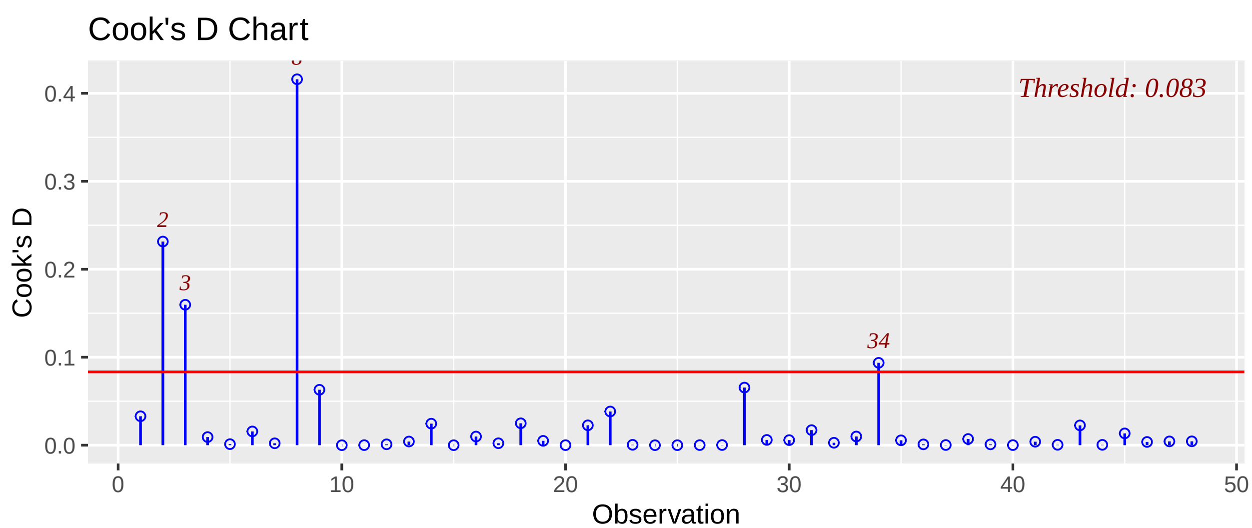
From the summary statistics table, we can see that each group has exactly 4 entries, so no imbalance concerns. The variance seems to jump by quite a large amount between the groups, so contrast analysis might be a concern due to the small sample size.



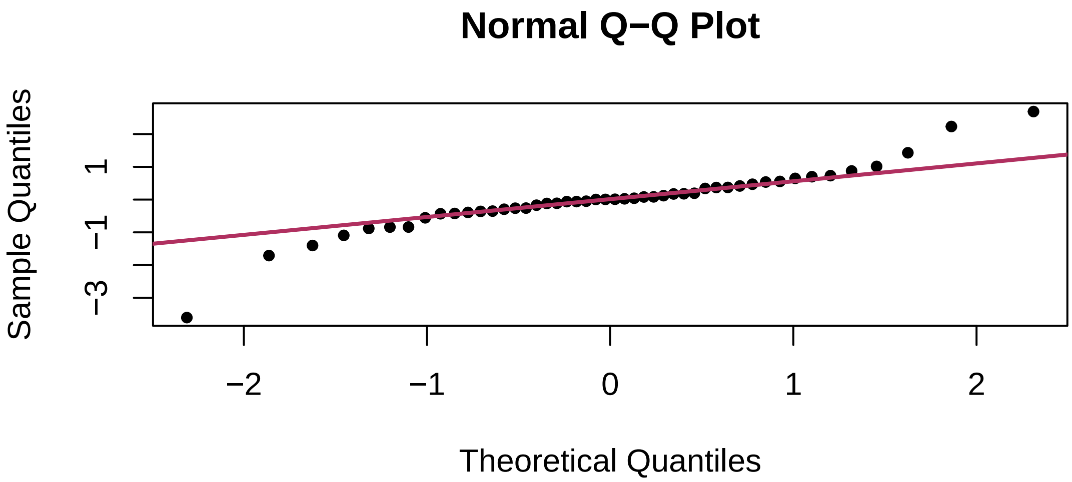
Immediately it can be seen that stirring seems to increase the variance of the name brand medicine. Also, an interaction effect between temperature and brand can be deduced if lines are drawn through the centers of the boxes. We can also see that temperature has an inverse effect on dissolving times whether stirring is present or not. Stirring might have an additive effect regardless of temperature.



The possible interaction between brand and temperature becomes even more noticeable in the preceding three factor interaction plots. Specifically the brand and temperature interaction can be seen with increasing temperature the store brand line has a more negative slope than the name brand line. In addition, there might be a slight three factor interaction between brand, temperature, and stirring as the name and store brand lines appear to be closer together in the stirred=yes plot than the stirred=no plot.



From the boxplots, we were able to see a small amount of outliers. To confirm if there are any of concern we plotted the Cook's Distance for each point based on a full linear model. Point 8 has a higher Cook's distance than the rest of the points which may require removal for analysis if it is suspected of causing issues in the analysis. This would have to be weighed against the risks cause by introducing imbalances.



Finally, we check the normality of the data. Here a QQ plot is generated for the full model residuals. The data seems to be indicative of heavy tails. This might pose a problem for some of our analyses.

Analysis and Results

Contrasts

contrast	estimate	SE	df	lower.CL	upper.CL
stirred	-2.413294	0.3102781	36	-3.042567	-1.784021
branding	5.338595	0.3102781	36	4.709322	5.967868

contrast	estimate	SE	df	lower.CL	upper.CL
stirredbrand	-3.720646	0.4387996	36	-4.610573	-2.8307197
stirredstore	-1.105942	0.4387996	36	-1.995869	-0.2160151

contrast	estimate	SE	df	lower.CL	upper.CL
temp6_23	7.315177	0.3800116	36	6.360952	8.269401
temp6_40	14.381286	0.3800116	36	13.427062	15.335510
temp23_40	7.066109	0.3800116	36	6.111885	8.020334

Conducting a linear contrast analysis on each of the explanatory variables reveals that there are significant differences between groups based on factors. In the first case, we contrasted the means of stirred versus not stirred. Here the difference in means is -2.41 with an upper 95% confidence limit of -3.04 and a lower 95% CI limit of -1.78. In other words, on average stirring medicine reduces dissolving time by between 3.04 and 1.78 minutes regardless of brand or temperature. When looking only at brand, name brand dissolving times were on average between 4.71 and 5.97 (95% CI) minutes slower than store brand. Since neither of the intervals contained zero we can conclude that there is a difference between brands and between the presence of stirring.

While significant for both store and name brands, stirring had more of an impact to dissolving times for name brand than it did for the store brand. Stirring reduced name brand dissolving times by 2.83 and 4.61 minutes whereas for the store brand that interval was 0.22 and 2 minutes.

A similar analysis was completed on the three levels of temperature……..

Model Development

Three models were developed and analyzed for this paper:

$Model\ 1 : Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\gamma\beta)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}$
$Model\ 2 : Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \epsilon_{ijkl}$
$Model\ 3 : Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijkl}$

Where α is brand effect, β is temperature effect, γ is stir effect. i, j, k are 1, 2, 1, 2, 3, and 1, 2, respectively. ϵ_{ijkl} is assumed to be normally distributed with a μ of 0 and a variance of σ^2_ϵ. μ is the overall mean and is an unknown value.

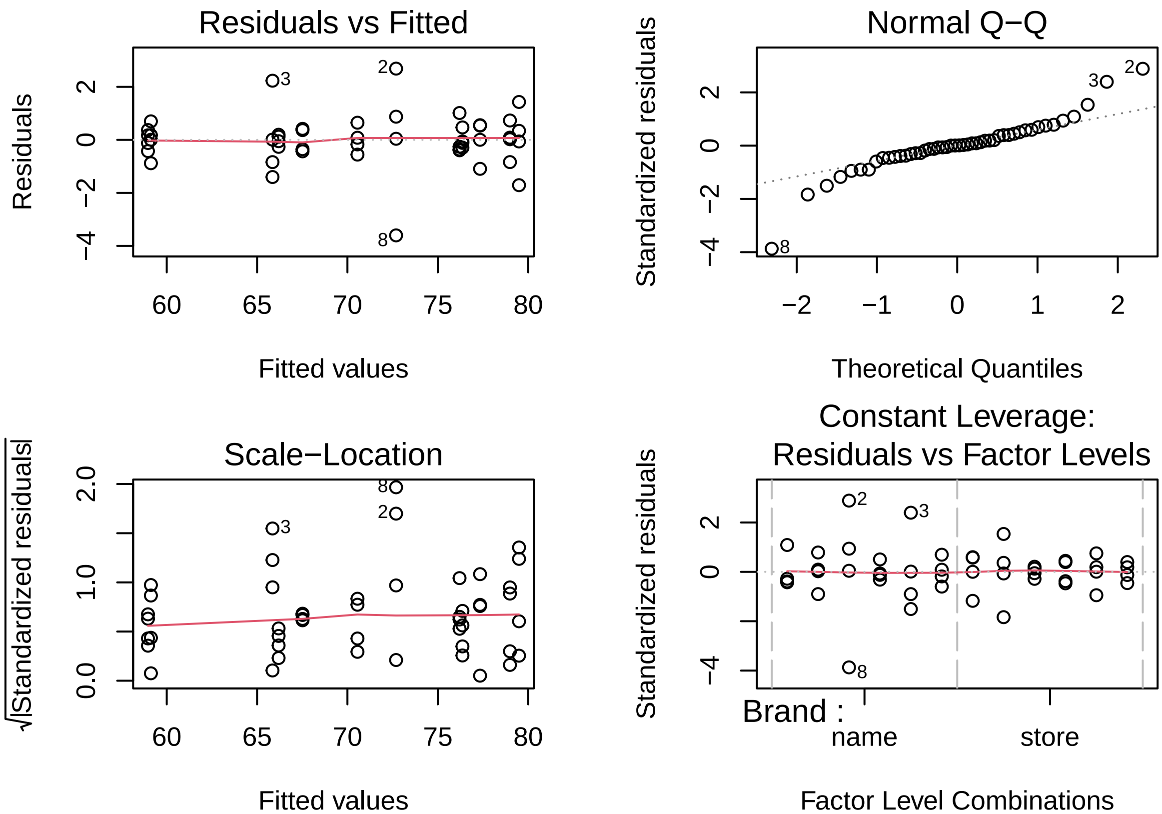
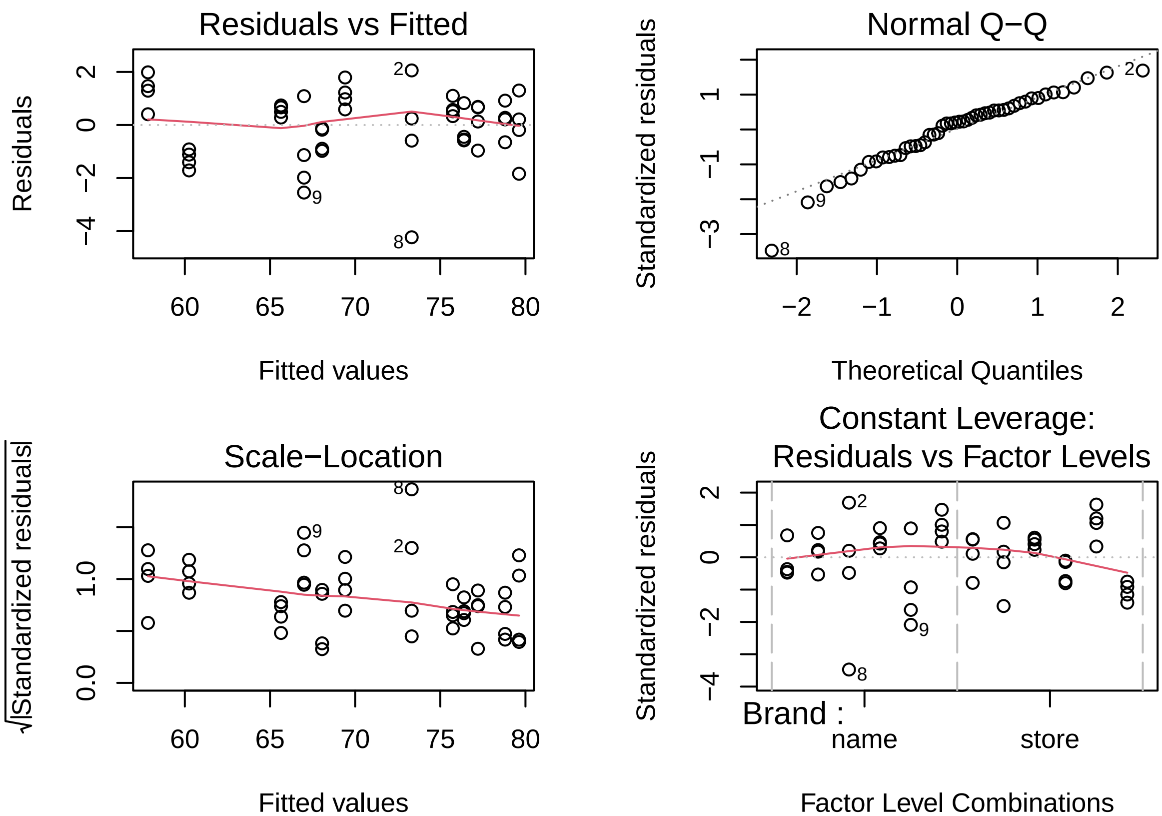


Table 5: Model 1 ANOVA Results

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Brand	1	342.007154	342.007154	296.0407972	0.0000000
Temp	2	1654.736551	827.368276	716.1685387	0.0000000
Stirred	1	69.887866	69.887866	60.4948148	0.0000000
Brand:Temp	2	231.851912	115.925956	100.3453058	0.0000000
Brand:Stirred	1	20.510041	20.510041	17.7534556	0.0001609
Temp:Stirred	2	0.124706	0.062353	0.0539727	0.9475345
Brand:Temp:Stirred	2	9.056126	4.528063	3.9194837	0.0288376
Residuals	36	41.589732	1.155270	NA	NA

Table 6: Model 2: ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Brand	1	342.00715	342.007154	196.71962	0e+00
Temp	2	1654.73655	827.368276	475.89522	0e+00
Stirred	1	69.88787	69.887866	40.19891	1e-07
Brand:Temp	2	231.85191	115.925956	66.67963	0e+00
Residuals	41	71.28061	1.738551	NA	NA



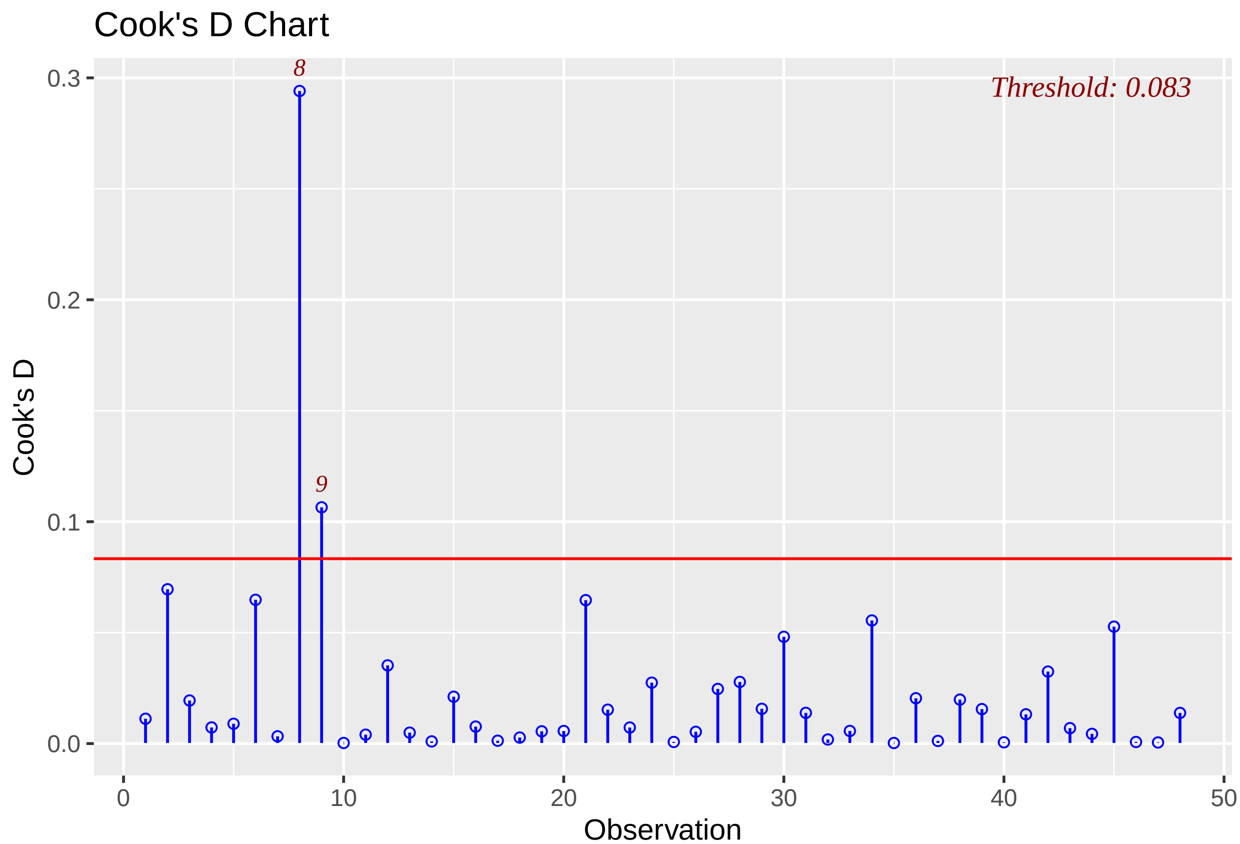
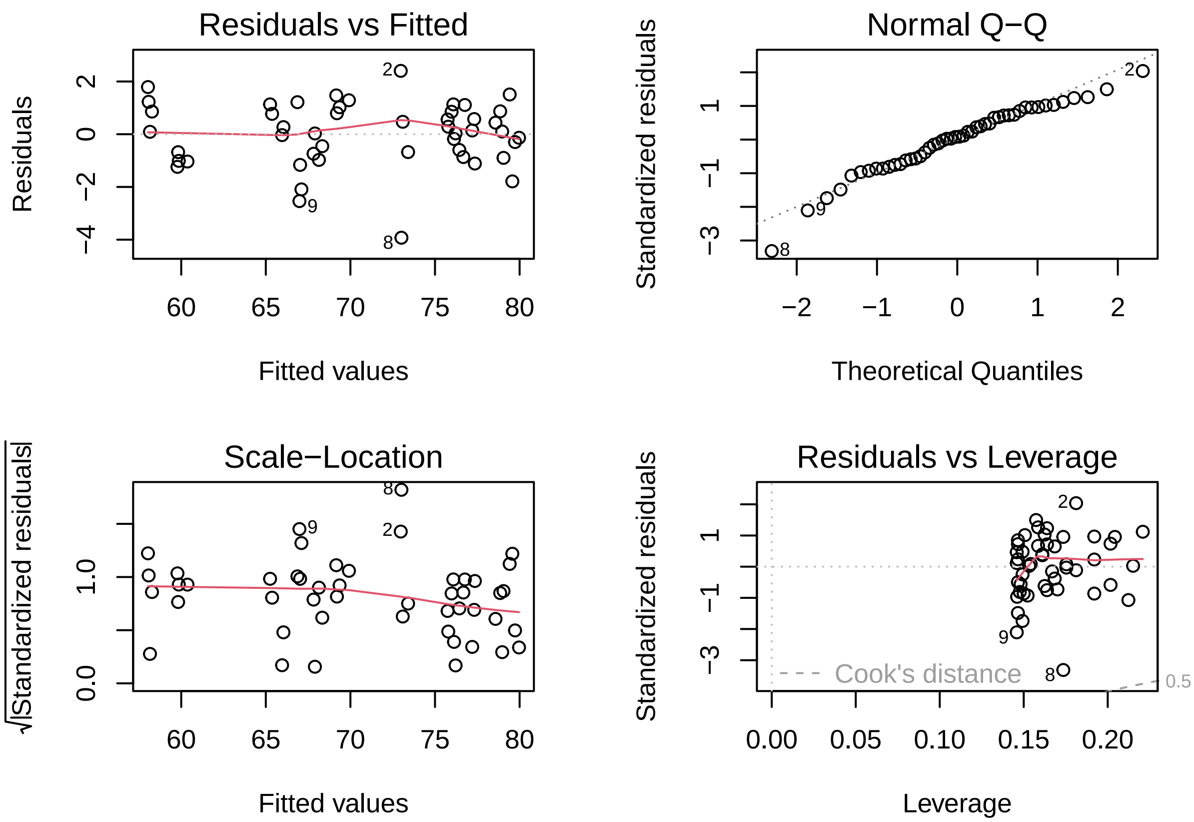


Table 7: Model 3: ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Brand	1	342.0071543	342.0071543	201.4523573	0.0000000
Temp	2	1654.7365514	827.3682757	487.3444529	0.0000000
Stirred	1	69.8878657	69.8878657	41.1660257	0.0000001
Order	1	0.9059095	0.9059095	0.5336076	0.4693512
Brand:Temp	2	234.3183134	117.1591567	69.0102180	0.0000000
Residuals	40	67.9082953	1.6977074	NA	NA

```
## Anova Table (Type III tests)
##
## Response: Time
##                Sum Sq Df    F value    Pr(>F)
## (Intercept) 22049.8  1 12987.9591 < 2.2e-16 ***
## Brand           3.6  1     2.1070   0.15442
## Temp          335.9  2    98.9237 3.275e-16 ***
## Stirred         6.2  1     3.6261   0.06409 .
## Order           3.4  1     1.9864   0.16645
## Brand:Temp    234.3  2    69.0102 1.076e-13 ***
## Residuals      67.9 40
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
0.5
Leverage

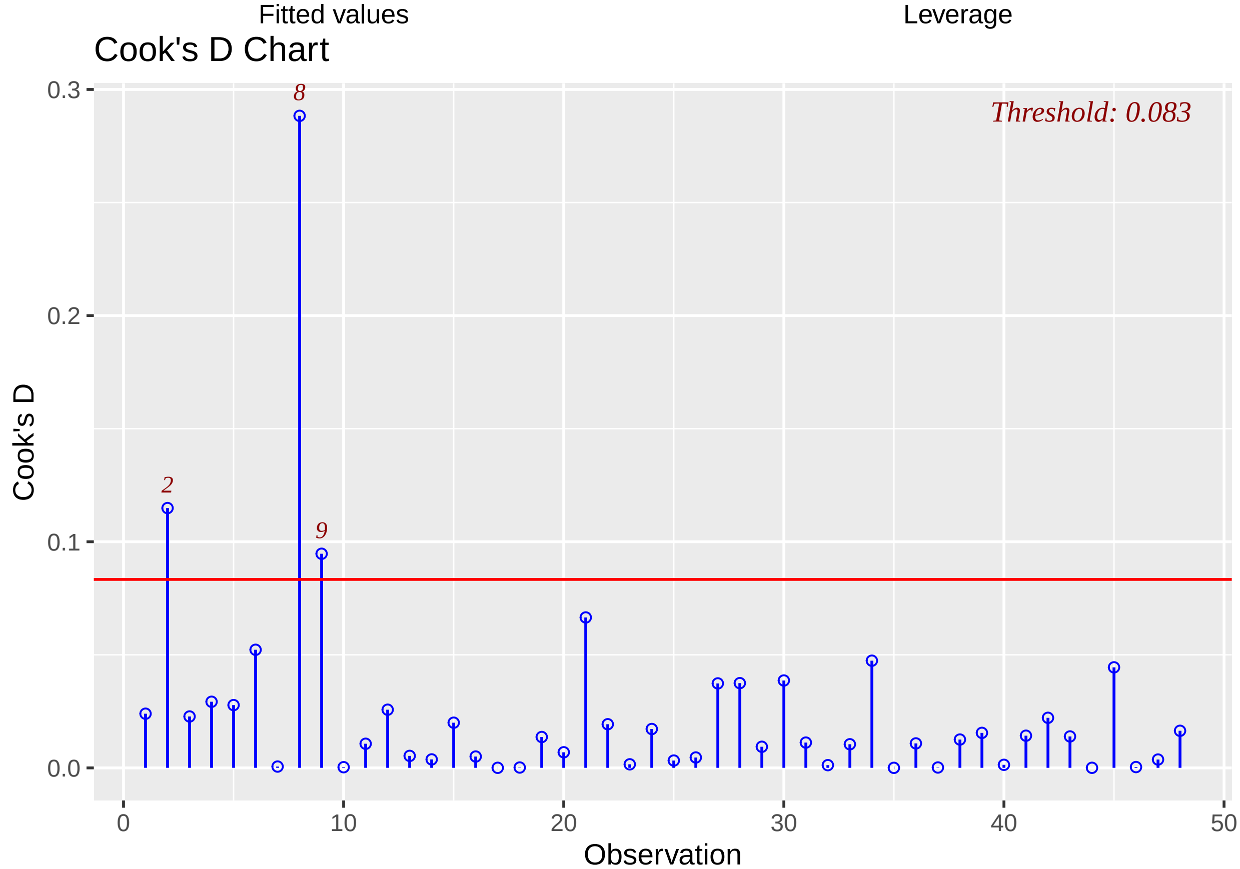
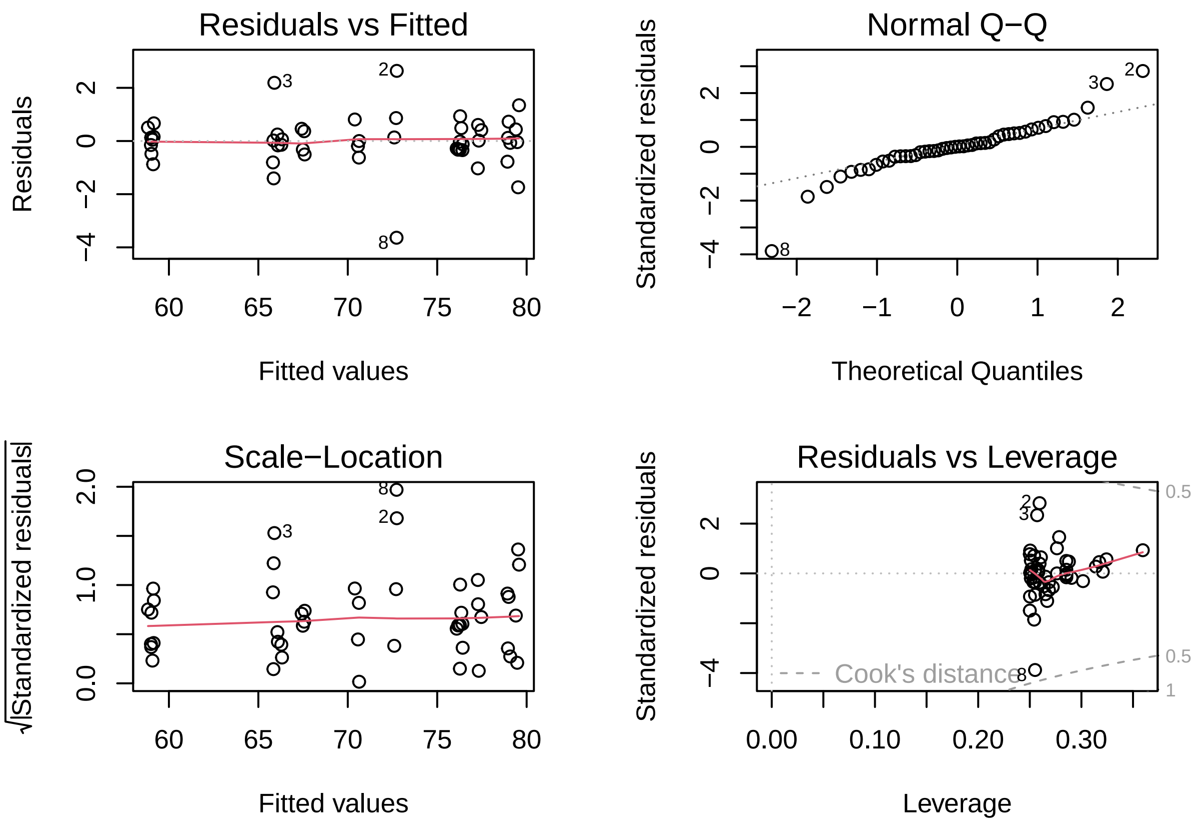

Cook's D Chart
Threshold: 0.083
Cook's D
Observation

Table 8: Model 4 ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Brand	1	342.0071543	342.0071543	289.5117212	0.0000000
Temp	2	1654.7365514	827.3682757	700.3736925	0.0000000
Stirred	1	69.8878657	69.8878657	59.1606229	0.0000000
Order	1	0.9059095	0.9059095	0.7668595	0.3871609
Brand:Temp	2	234.3183134	117.1591567	99.1761391	0.0000000
Brand:Stirred	1	17.2952414	17.2952414	14.6405566	0.0005144
Temp:Stirred	2	0.0420436	0.0210218	0.0177951	0.9823712
Brand:Temp:Stirred	2	9.2246692	4.6123346	3.9043772	0.0294693
Residuals	35	41.3463412	1.1813240	NA	NA

```
## Anova Table (Type III tests)
##
## Response: Time
##                    Sum Sq Df    F value    Pr(>F)
## (Intercept)       15086.6  1 12770.9606 < 2.2e-16 ***
## Brand                 2.4  1     1.9942  0.166731
## Temp                220.9  2    93.4777 9.153e-15 ***
## Stirred               9.9  1     8.3899  0.006465 **
## Order                 0.2  1     0.2060  0.652697
## Brand:Temp           69.5  2    29.3957 3.224e-08 ***
## Brand:Stirred         0.4  1     0.3458  0.560255
## Temp:Stirred          3.5  2     1.4816  0.241180
## Brand:Temp:Stirred    9.2  2     3.9044  0.029469 *
## Residuals            41.3 35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Model Selection

Results

Conclusion

Appendix: Code

```
library(tidyverse)
library(emmeans)
library(lme4)
library(lmerTest)
library(olsrr)
library(car)
df_eff <- read_csv('effervescence.csv', col_types = 'fffnnn')
df_stats <-
df_eff %>% group_by(Brand, Temp, Stirred) %>%
summarise('25%' = quantile(Time, probs = 0.25),
          'Mean' = mean(Time),
          'Median' = median(Time),
          '75%' = quantile(Time, probs = 0.75),
          'Var' = var(Time),
          'n' = n())

knitr::kable(df_stats)
df_eff %>% ggplot() + geom_boxplot(aes(fill = Brand, y = Time, x = Temp)) +
  facet_grid(cols = vars(Stirred)) + labs(title = "Stirred") + theme(
  plot.title = element_text(hjust = 0.5)
)
##3 factor interaction plot based on HW7 code
par(mfrow=c(1,2), mar = c(3.5,3.5,2,2))
with(df_eff%>%filter(Stirred=="yes"),interaction.plot(Temp,Brand,Time,
            type="b", pch=19, col=c(2,4), ylab="", xlab = "",
            main="Mean Time vs. Temp: Stirred = Yes",
            cex.main = 0.75, legend = FALSE))
legend("topright",
       title = "Brand",
       c("Name", "Store"),
       cex = 0.7,
       col = c("#DF536B","#2297E6"),
       pch = c(19,19), lty = c(2,1))
title(xlab = "Temperature", ylab = "Mean Dissolving Time (Minutes)", line = 2.25, cex.lab = 0.9)
#```

#```{r, echo=FALSE, message=FALSE, error=FALSE, fig.dim=c(6,3), dpi=250}
with(df_eff%>%filter(Stirred=="no"),interaction.plot(Temp,Brand,Time,
          type="b", pch=19, col=c(2,4), ylab="", xlab = "",
          main="Mean Time vs. Temp: Stirred = No",
          cex.main = 0.75, legend = FALSE))
legend("topright",
       title = "Brand",
       c("Name", "Store"),
       cex = 0.7,
       col = c("#DF536B","#2297E6"),
       pch = c(19,19), lty = c(2,1))
title(xlab = 'Temperature', ylab = "Mean Dissolving Time (Minutes)", line = 2.25, cex.lab = 0.9)
aov_eff <- aov(lm_eff <- lm(Time ~ Brand * Temp * Stirred, data = df_eff))

#cooksD_values <- cooks.distance(lm_eff)
```

```
#ggplot() + geom_col(aes(y = cooksD_values, x = 1:length(cooksD_values)), width = 0.025, col = 'red')
#    geom_point(aes(y = cooksD_values, x = 1:length(cooksD_values))) + xlab('Sample Points') + ylab("Co
#    geom_hline(yintercept = 0.25, lty = 2) + labs(title = "Cook's Distance for each sample point")
ols_plot_cooksd_chart(lm_eff)
qqnorm(lm_eff$resid, pch = 20)
qqline(lm_eff$resid, col = "maroon", lwd = 2)
means_eff <- emmeans(aov_eff, specs = c('Brand', 'Temp', 'Stirred'))
#summary(means_eff)
cont_str_brd <-
contrast(means_eff, list(stirred = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6, -1/6, -1/6, -1/6, -1/6, -1/6, -1/6),
                         branding = rep(c(1/6,-1/6), 6)
                         )
         )

cont_strbrd <-

contrast(means_eff, list(stirredbrand = c(1/3, 0,  1/3, 0, 1/3, 0, -1/3, 0, -1/3, 0, -1/3, 0),
                         stirredstore = c(0, 1/3, 0,  1/3, 0, 1/3, 0, -1/3, 0, -1/3, 0, -1/3)
                         )
         )
cont_temp <-
contrast(means_eff, list(temp6_23 = c(1/4, 1/4, -1/4, -1/4, 0, 0, 1/4, 1/4, -1/4, -1/4, 0, 0),
                         temp6_40 = c(1/4, 1/4, 0, 0, -1/4, -1/4, 1/4, 1/4, 0, 0, -1/4, -1/4),
                         temp23_40 = c(0, 0, 1/4, 1/4, -1/4, -1/4, 0, 0, 1/4, 1/4, -1/4, -1/4)
                         ), options=list(adjust="bonferroni"))
knitr::kable(confint(cont_str_brd))
knitr::kable(confint(cont_strbrd))
knitr::kable(confint(cont_temp))
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(aov_eff)
knitr::kable(summary(aov_eff)[[1]], 'simple', caption = 'Model 1 ANOVA Results')
#model with stirred as block effect without interaction
aov_block_eff <- aov(lm_block_eff <- lm(Time ~ Brand * Temp + Stirred, data = df_eff))
#summary(lm_block_eff)
knitr::kable(summary(aov_block_eff)[[1]], caption = "Model 2: ANOVA Table")
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(aov_block_eff)
ols_plot_cooksd_chart(lm_block_eff)
#added covariate Order model with stirred as block effect without interaction
aov_block_order_eff <- aov(lm_block_order_eff <- lm(Time ~ Brand * Temp + Stirred + Order, data = df_ef
#summary(lm_block_order_eff)
knitr::kable(summary(aov_block_order_eff)[[1]], caption = "Model 3: ANOVA Table")
Anova(aov_block_order_eff, type=3) # type 3 SS
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(lm_block_order_eff)
ols_plot_cooksd_chart(lm_block_order_eff)
#added covariate Order to model with 3 factor interaction
aov_three_order_eff <- aov(lm_three_order_eff <- lm(Time ~ Brand * Temp * Stirred + Order, data = df_ef
#summary(lm_three_order_eff)
knitr::kable(summary(aov_three_order_eff)[[1]], 'simple', caption = "Model 4 ANOVA Table")
Anova(aov_three_order_eff, type=3) # type 3 SS
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(lm_three_order_eff)
```