# ST_518 Project

MA, HK, BA, JF

2022-11-24

## Executive Summary

Text goes here. . . ..

### Introduction

For this paper we have been presented with data gathered on the dissolving cold medicine in water. The dataset contains dissolving characteristics of different cold medicine brands done under various environmental conditions. The goal of this paper is to answer the following questions:

- Are the dissolving characteristics different between brands?
- Does temperature of the water influence dissolving characteristics? If so, is there an interaction effect between brand and temperature?
- Does stirring influence dissolving times and is there an interaction with the other two effects?

### Experimental Design

Data used for this study was obtained from an experiment called the 'Effervescent Experiment'. The experiment was conducted on two different brands of cold medicine, 'Name' and 'Store'. Each brand was dissolved in water at three different but evenly spaced temperatures, 6°C, 23°C, and 40°C. Data was gathered on each combination of brand and temperature using 4 replications. The experiment was then repeated with the introduction of stirring as a blocking factor. Samples were stirred using a magnetic stirrer spinning at 350 rpm. The data was then tabulated and made available for this analysis.

## Exploratory Data Analysis

### Summary

In total, the provided dataset contains 48 rows and 5 columns. The 5 columns include 3 explanatory categorical variables (Brand, Temp in °C, and Stirring), a single continuous response variable (Time, in minutes) and one descriptor (order). Prior to analysis, the data will be explored to gain a better understanding of what to expect and, more importantly, check for any potential violations of analytical assumptions.

From the summary statistics table (see Appendix 1, Table 1), we can see that each group has exactly 4 entries, eliminating concerns with respect to design imbalance. Constructing a means table for the data without taking into account stirring. We can see that there does appear to be a disparity between the mean dissolving times of store brand cold medications when compared to name brand.

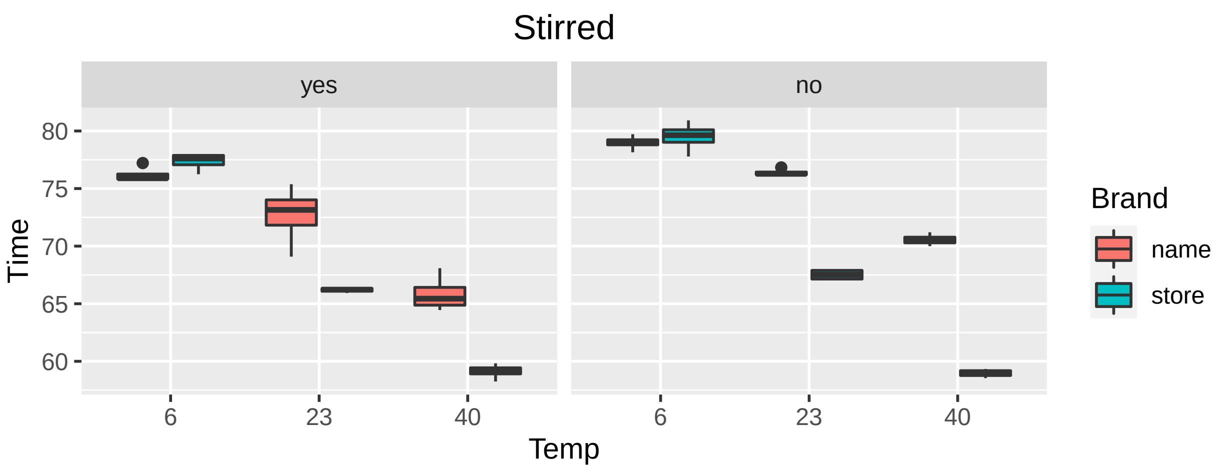
| Temp | name | store | TempMean |
|---|---:|---:|---:|
| 6 | 77.59651 | 78.41471 | 78.00561 |
| 23 | 74.52748 | 66.85339 | 70.69044 |
| 40 | 68.20427 | 59.04438 | 63.62433 |
| BrandMean | 73.44276 | 68.10416 | 70.77346 |

When inspecting the marginal means of store versus name brand, we find that store brand dissolves in less time that name brand. This disparity becomes more pronounced as the effect of temperature is introduced. It was observed that increasing has a more dramatic effect on store brand medicine than name brand. Store brand medicine drops from 78.42 to 59.04 ($\Delta$ of -19.38) seconds across a temperature change from 6°C to 6°C. Whereas, name brand medicine only drops from 77.60 to 68.20 ($\Delta$ of -9.40 seconds) across the same temperature interval.

Next, consider the range in variability at each factor and level. The range in values of the variance is 6.9458; an interesting result considering nine of the twelve observable variances fall within a range of 0.04 and 1.69. The variability between observations of name brand cold medicines are elevated, especially in cases when the observation was stirred at 23°C and 40°C. Variance of non-stirred observations, particularly within temperature values of 23°C and 40°C, are noticeably lower than their stirred counterparts within the same brand. The variance seems to jump by a large amount between the groups.

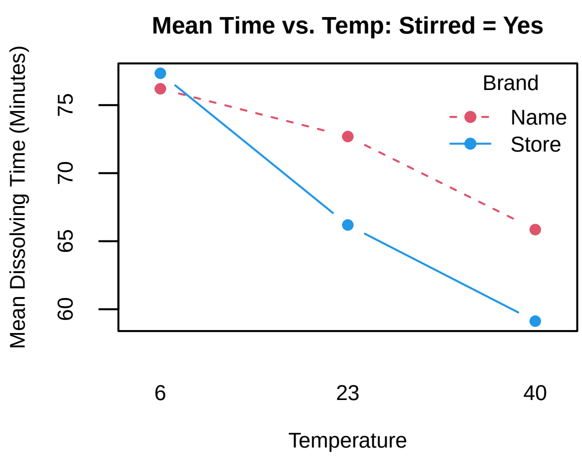
When plotting the data in a boxplot, there is a noticeable increase in the variance of the name brand medicine in the stirred block versus the non stirred. Additionally, stirring produced a decrease in the mean differences of brand within each temperature grouping. Earlier, we had introduce an insight from the summary statistics output indicating an inverse relationship between temperature and dissolve time–as temperature increases dissolve time decreases. The boxplot reinforces this idea.

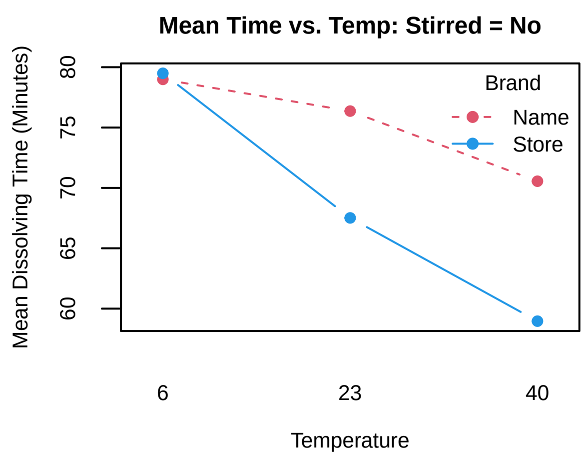
It is simply conjecture at this point, however we noticed that observations of dissolve time while stirring the water seems to have increased the name brand variability, while not stirring the water seems to have increased the store brand variability. Perhaps worth looking into the blocking effects of stirred on variability at various temperatures.




**Interactions**

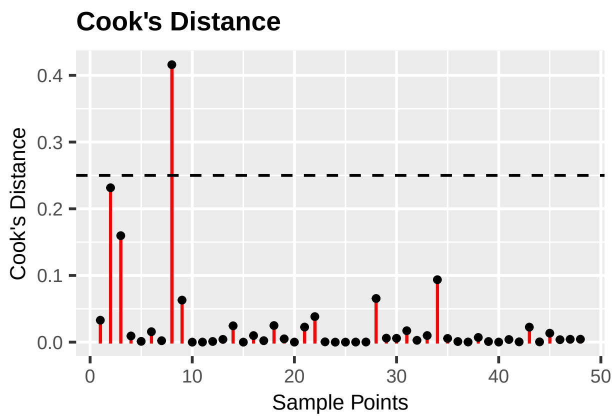
The possible interaction between brand and temperature becomes even more noticeable in the preceding three-factor interaction plots. Specifically, the brand and temperature interaction can be seen when the temperature increases. The slope for the store brand has a more pronounced negative slope than the slope of the name brand. 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.

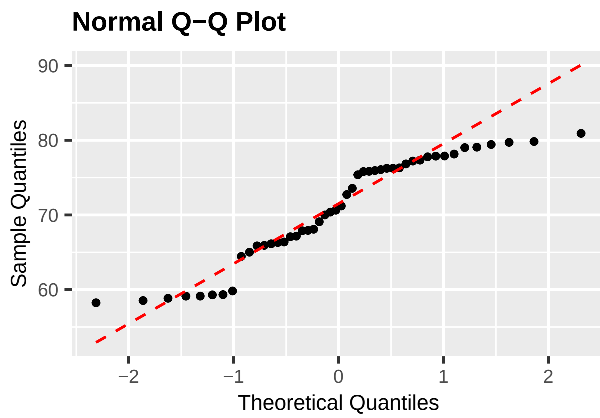


### Assumptions and Violations

In reference to the boxplots, we were able to see a small number of outliers. To confirm if there is any 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 caused by introducing imbalances.




Finally, we check the normality of the data. Here a Q-Q plot is generated for the full model residuals. The data appears to suffer from heavy tails, multimodality and/or gaps in data between the left tail and the center. Since downstream analysis hinges on the assumption that our data is normally distributed, these issues may pose a problem.

## Analysis and Results

### Model Development

The following 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 $\alpha$ is brand effect, $\beta$ is temperature effect, $\gamma$ is stir effect. i, j, k are (1, 2), (1,2,3), and (1,2), respectively. $\epsilon_{ijkl}$ is assumed to be normally distributed with a $\mu_\epsilon$ of 0 and a variance of $\sigma^2_\epsilon$. $\mu$ is the overall mean and is an unknown value.

Mixed Effects models:
$Model\ 4 : Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \epsilon_{ijk}$
Where both temperature and brand are random.
$Model\ 5 : Y_{ijk} = \mu + A_i + \beta_j + \gamma_k + (A\beta)_{ij} + \epsilon_{ijk}$
Where temperature is fixed and brand is random.
$Model\ 6 : Y_{ijk} = \mu + \alpha_i + B_j + \gamma_k + (\alpha B)_{ij} + \epsilon_{ijk}$
Where temperature is random and brand is fixed.

With order as a factor:
$Model\ 7 : Y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \nu_l + (\alpha\beta)_{ij} + \epsilon_{ijklm}$
Similar to model 2, but with an introduced effect, $\nu$, to represent order.

$Model\ 8 : Y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \nu_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\gamma\beta)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijklm}$
Similar to model 1, but with an introduced effect, $\nu$, to represent order.

### Model Selection

| Model | Root MSE | $R^2$ | *adj* $R^2$ | AIC | BIC |
|---|---|---|---|---|---|
| 1 | 1.074835 | 0.9824498 | 0.9770873 | 155.3374101 | 179.6630232 |
| 2 | 1.3185414 | 0.9699208 | 0.965519 | 171.1984162 | 186.1680243 |
| 3 | 2.655107 | 0.8720833 | 0.860184 | 236.6806137 | 247.9078198 |
| 7 | 1.302961 | 0.9713439 | 0.966329 | 170.8720441 | 187.7128532 |
| 8 | 1.0868873 | 0.9825525 | 0.9765706 | 157.0556795 | 183.2524936 |

### Contrasts

Conducting a linear contrast analysis on each of the explanatory variables reveals that there are significant differences between groups based on factors, see Appendix 1 Table 2, 3, 4, and 5 for full results.

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 for the three levels of temperature. Completing a contrast analysis using a Bonferroni correction we found that in pairwise cases each level was significantly different from the other. The 95% confidence limits were (6.25, 8.38), (13.32, 15.44), and (6.00, 8.13) for the pair wise comparisons of $6°C\ vs\ 23°C$, $6°C\ vs\ 40°C$, and $23°C\ vs\ 40°C$, respectively. Zero did not fall in any of those ranges. When comparing individual levels versus the remainder of the group, $23°C$ was found not to be significantly different from the rest of the levels. That confidence interval ranged from -1.04 to 0.80 minutes of dissolving time. Due to that, we do not have enough evidence to say $23°C$ is different from either $6°C$ or $40°C$.

### Random Effects Analysis

Expanding Table 20 from Appendix 1 to include portion of variance due to effect, the following table is produced:

| Cov Parm | Estimate | Portion |
|---|---|---|
| Brand | 9.420 | 0.132 |
| Temp | 44.465 | 0.623 |
| Brand*Temp | 14.071 | 0.197 |
| Residual | 3.361 | 0.047 |

Temperature is responsible for the largest portion of total variance. $\sigma^2_\beta$, at value of 44.465, explains 62.3% of the total variance ($\hat{\sigma}^2_{total}$ being 71.32).

The overall dissolving meantime, $\hat{Y}_{...}$, was found to be 70.773. The standard error of that value, $SE(\mu)$ was calculated to be 4.684748. From that we find the 95% confidence limit is ± 3.705803.

$\widehat{CV}$ was found to be 0.119

## Conclusion

## Appendix I: Analysis Tables and Figures

Table 4: Data Summary Table

| Brand | Temp | Stirred | Min | 25% | Mean | Median | 75% | Max | Range | Var | n |
|---|---|---|---|---|---|---|---|---|---|---|---|
| name | 6 | yes | 75.80973 | 75.83358 | 76.20241 | 75.89223 | 76.26107 | 77.21547 | 1.4057377 | 0.4593492 | 4 |
| name | 6 | no | 78.15246 | 78.79910 | 78.99061 | 79.04435 | 79.23586 | 79.72130 | 1.5688327 | 0.4146440 | 4 |
| name | 23 | yes | 69.08937 | 71.82180 | 72.69145 | 73.14894 | 74.01859 | 75.37855 | 6.2891789 | 6.9869087 | 4 |
| name | 23 | no | 76.06895 | 76.20492 | 76.36351 | 76.27622 | 76.43481 | 76.83265 | 0.7636940 | 0.1078134 | 4 |
| name | 40 | yes | 64.45156 | 64.87321 | 65.85343 | 65.43863 | 66.41886 | 68.08492 | 3.6333543 | 2.5499751 | 4 |
| name | 40 | no | 69.99943 | 70.28754 | 70.55511 | 70.50947 | 70.77705 | 71.20207 | 1.2026434 | 0.2544033 | 4 |
| store | 6 | yes | 76.24402 | 77.06561 | 77.33703 | 77.60659 | 77.87801 | 77.89089 | 1.6468708 | 0.5964884 | 4 |
| store | 6 | no | 77.78345 | 79.01994 | 79.49240 | 79.63219 | 80.10465 | 80.92176 | 3.1383169 | 1.6942517 | 4 |
| store | 23 | yes | 65.92809 | 66.08831 | 66.19126 | 66.22629 | 66.32923 | 66.38436 | 0.4562787 | 0.0411024 | 4 |
| store | 23 | no | 67.08353 | 67.14393 | 67.51552 | 67.52360 | 67.89520 | 67.93138 | 0.8478521 | 0.2060739 | 4 |
| store | 40 | yes | 58.24407 | 58.90895 | 59.12529 | 59.21659 | 59.43293 | 59.82388 | 1.5798100 | 0.4320148 | 4 |
| store | 40 | no | 58.53920 | 58.76884 | 58.96347 | 58.99050 | 59.18513 | 59.33370 | 0.7945066 | 0.1202191 | 4 |

Table 5: Least Squares Means

| Brand | Temp | Stirred | emmean | SE | df | lower.CL | upper.CL |
|---|---|---|---|---|---|---|---|
| name | 6 | yes | 76.20241 | 0.5374175 | 36 | 75.11248 | 77.29235 |
| store | 6 | yes | 77.33703 | 0.5374175 | 36 | 76.24709 | 78.42696 |
| name | 23 | yes | 72.69145 | 0.5374175 | 36 | 71.60152 | 73.78138 |
| store | 23 | yes | 66.19126 | 0.5374175 | 36 | 65.10132 | 67.28119 |
| name | 40 | yes | 65.85343 | 0.5374175 | 36 | 64.76350 | 66.94337 |
| store | 40 | yes | 59.12529 | 0.5374175 | 36 | 58.03535 | 60.21522 |
| name | 6 | no | 78.99061 | 0.5374175 | 36 | 77.90068 | 80.08055 |
| store | 6 | no | 79.49240 | 0.5374175 | 36 | 78.40247 | 80.58233 |
| name | 23 | no | 76.36351 | 0.5374175 | 36 | 75.27358 | 77.45344 |
| store | 23 | no | 67.51552 | 0.5374175 | 36 | 66.42559 | 68.60546 |
| name | 40 | no | 70.55511 | 0.5374175 | 36 | 69.46518 | 71.64505 |
| store | 40 | no | 58.96347 | 0.5374175 | 36 | 57.87354 | 60.05341 |

Table 6: Contrast Stirred and Brand

| 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 |

Table 7: Contrast Stirred versus Brand

| 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 |

Table 8: Contrast Temperatures

| contrast | estimate | SE | df | lower.CL | upper.CL |
|---|---|---|---|---|---|
| temp6__23 | 7.3151767 | 0.3800116 | 36 | 6.254195 | 8.3761584 |
| temp6__40 | 14.3812861 | 0.3800116 | 36 | 13.320305 | 15.4422678 |
| temp23__40 | 7.0661094 | 0.3800116 | 36 | 6.005128 | 8.1270910 |
| temp6__rest | 10.8482314 | 0.3290997 | 36 | 9.929394 | 11.7670685 |
| temp23__rest | -0.1245337 | 0.3290997 | 36 | -1.043371 | 0.7943034 |
| temp40__rest | -10.7236978 | 0.3290997 | 36 | -11.642535 | -9.8048607 |

Table 9: Model 1: ANOVA Table

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Brand | 1 | 342.007 | 342.007 | 296.041 | 0.000 |
| Temp | 2 | 1654.737 | 827.368 | 716.169 | 0.000 |
| Stirred | 1 | 69.888 | 69.888 | 60.495 | 0.000 |
| Brand:Temp | 2 | 231.852 | 115.926 | 100.345 | 0.000 |
| Brand:Stirred | 1 | 20.510 | 20.510 | 17.753 | 0.000 |
| Temp:Stirred | 2 | 0.125 | 0.062 | 0.054 | 0.948 |
| Brand:Temp:Stirred | 2 | 9.056 | 4.528 | 3.919 | 0.029 |
| Residuals | 36 | 41.590 | 1.155 | NA | NA |

Table 10: Model 1: Type III ANOVA Table

| | Sum Sq | Df | F value | Pr(>F) |
|---|---|---|---|---|
| (Intercept) | 23227.231 | 1 | 20105.451 | 0.000 |
| Brand | 2.575 | 1 | 2.229 | 0.144 |
| Temp | 221.582 | 2 | 95.901 | 0.000 |
| Stirred | 15.548 | 1 | 13.458 | 0.001 |
| Brand:Temp | 80.110 | 2 | 34.672 | 0.000 |
| Brand:Stirred | 0.400 | 1 | 0.347 | 0.560 |
| Temp:Stirred | 3.668 | 2 | 1.588 | 0.218 |
| Brand:Temp:Stirred | 9.056 | 2 | 3.919 | 0.029 |
| Residuals | 41.590 | 36 | NA | NA |

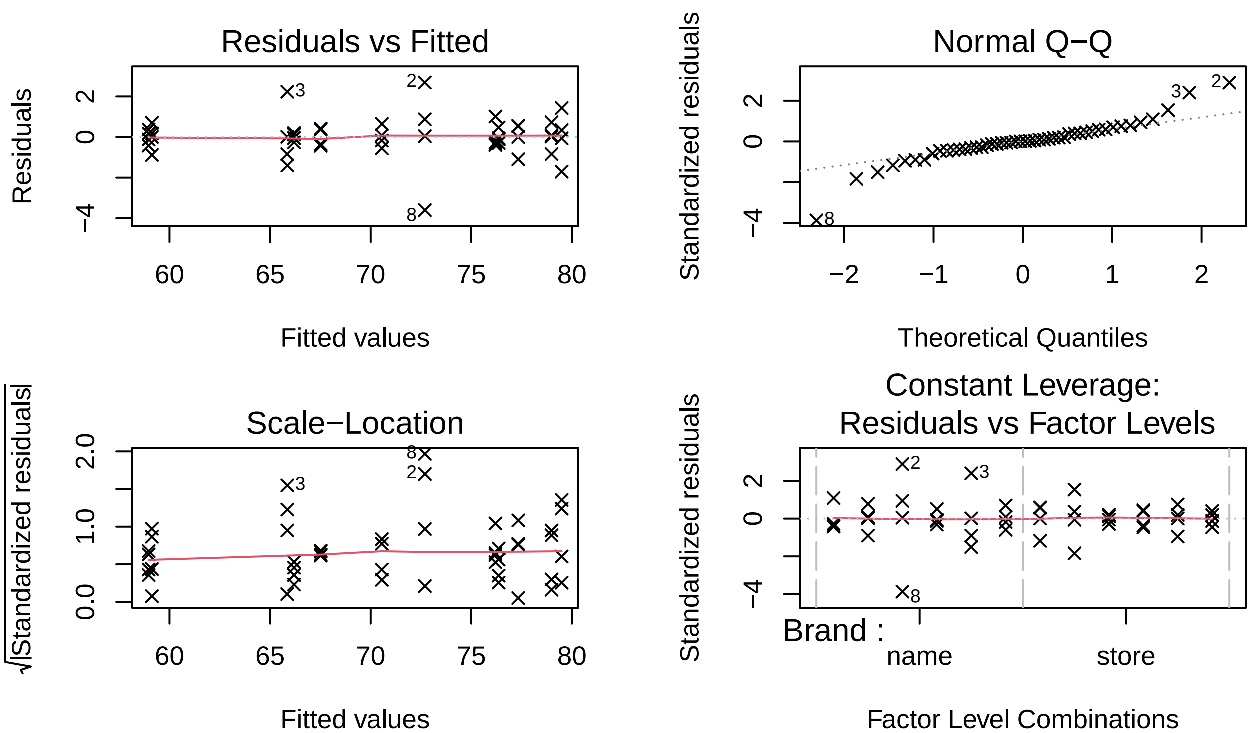


Table 11: Model 2: ANOVA Table

| | 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 12: Model 2: Type III ANOVA Table

| | Sum Sq | Df | F value | Pr(>F) |
|---|---|---|---|---|
| (Intercept) | 40014.251 | 1 | 23015.858 | 0.000 |
| Brand | 2.678 | 1 | 1.540 | 0.222 |
| Temp | 366.976 | 2 | 105.541 | 0.000 |
| Stirred | 69.888 | 1 | 40.199 | 0.000 |
| Brand:Temp | 231.852 | 2 | 66.680 | 0.000 |
| Residuals | 71.281 | 41 | NA | NA |

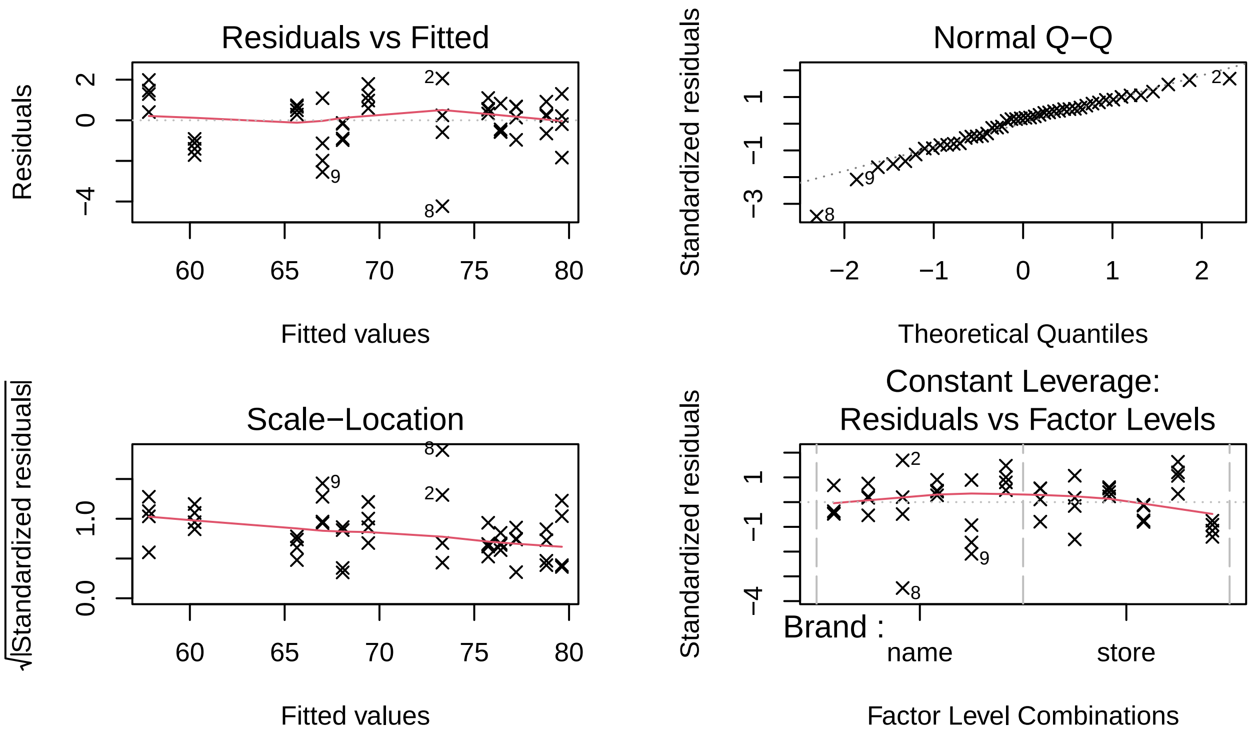


Table 13: Model 3: ANOVA Table

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Brand | 1 | 342.00715 | 342.007154 | 48.514451 | 0.0000000 |
| Temp | 2 | 1654.73655 | 827.368276 | 117.363970 | 0.0000000 |
| Stirred | 1 | 69.88787 | 69.887866 | 9.913744 | 0.0029802 |
| Residuals | 43 | 303.13252 | 7.049593 | NA | NA |

Table 14: Model 3: Type III ANOVA Table

| | Sum Sq | Df | F value | Pr(>F) |
|---|---|---|---|---|
| (Intercept) | 60625.967 | 1 | 8599.924 | 0.000 |
| Brand | 342.007 | 1 | 48.514 | 0.000 |
| Temp | 1654.737 | 2 | 117.364 | 0.000 |
| Stirred | 69.888 | 1 | 9.914 | 0.003 |
| Residuals | 303.133 | 43 | NA | NA |

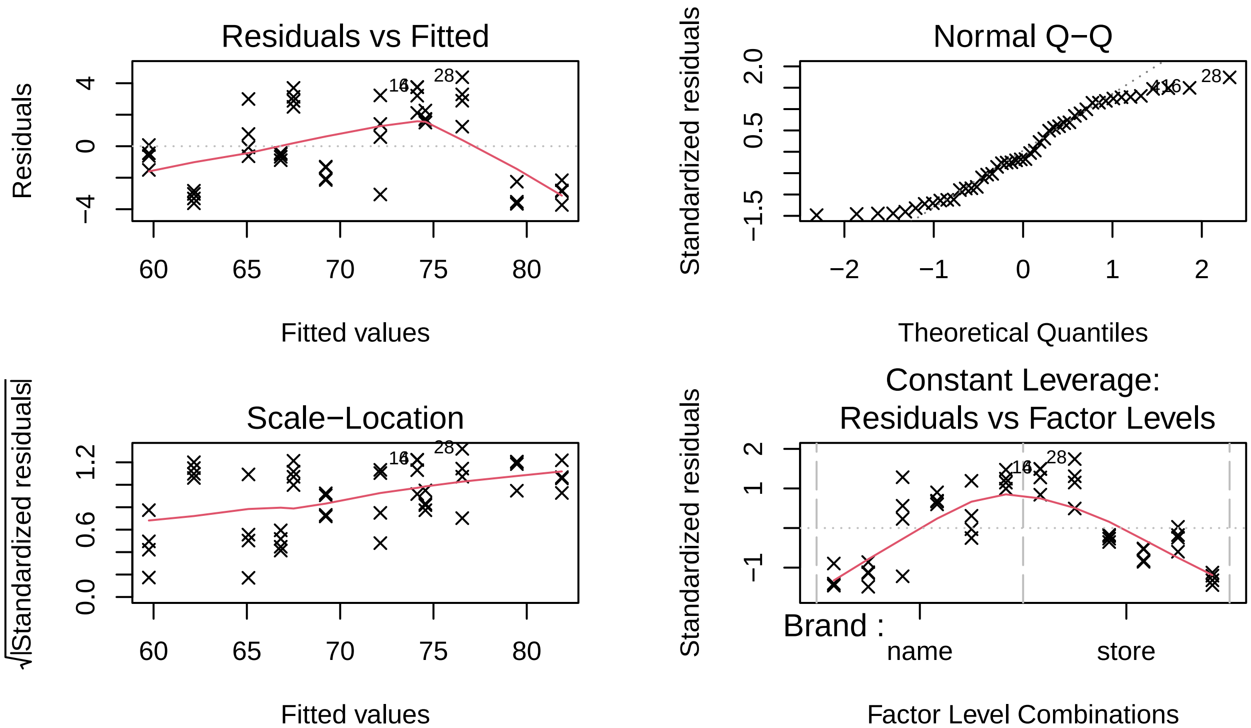


Table 15: Mixed Effects Models

| Source | Df | Sum Sq | Mean Sq | Error Term | Error Df | F Score | Pr>F |
|---|---|---|---|---|---|---|---|
| Brand | 1 | 342.0072 | 342.007154 | MSAB | 2 | 2.950221 | 0.2280 |
| Temp | 2 | 1654.7366 | 827.368276 | MSAB | 2 | 7.137041 | 0.1229 |
| Brand*Temp | 2 | 231.8519 | 115.925956 | MSE | 42 | 34.491507 | 0.0000 |
| Residual | 42 | 141.1685 | 3.361154 | NA | NA | NA | NA |

Table 16: Model4: Brand is Fixed and Temperature is Random

| Brand | Mean | Effect |
|---|---|---|
| name | 73.443 | 0.000 |
| store | 68.104 | -5.339 |

| Cov Parm | Estimate |
|---|---|
| Temp | 44.465 |
| Brand*Temp | 14.071 |
| Residual | 3.361 |

Table 18: Model5: Brand is Random and Temperature is Fixed

| Temp | Mean | Effect |
|---|---|---|
| 6 | 78.006 | 0.000 |
| 23 | 70.690 | -7.315 |
| 40 | 63.624 | -14.381 |

| Cov Parm | Estimate |
|---|---|
| Brand | 9.42 |
| Brand*Temp | 14.071 |
| Residual | 3.361 |

Table 20: Model6: Brand and Temperature are Random

| Cov Parm | Estimate |
|---|---|
| Brand | 9.42 |
| Temp | 44.465 |
| Brand*Temp | 14.071 |
| Residual | 3.361 |

Table 21: Model 7: 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 |

Table 22: Model 7: Type III ANOVA Table

| | Sum Sq | Df | F value | Pr(>F) |
|---|---|---|---|---|
| (Intercept) | 22049.754 | 1 | 12987.959 | 0.000 |
| Brand | 3.577 | 1 | 2.107 | 0.154 |
| Temp | 335.887 | 2 | 98.924 | 0.000 |
| Stirred | 6.156 | 1 | 3.626 | 0.064 |
| Order | 3.372 | 1 | 1.986 | 0.166 |
| Brand:Temp | 234.318 | 2 | 69.010 | 0.000 |
| Residuals | 67.908 | 40 | NA | NA |

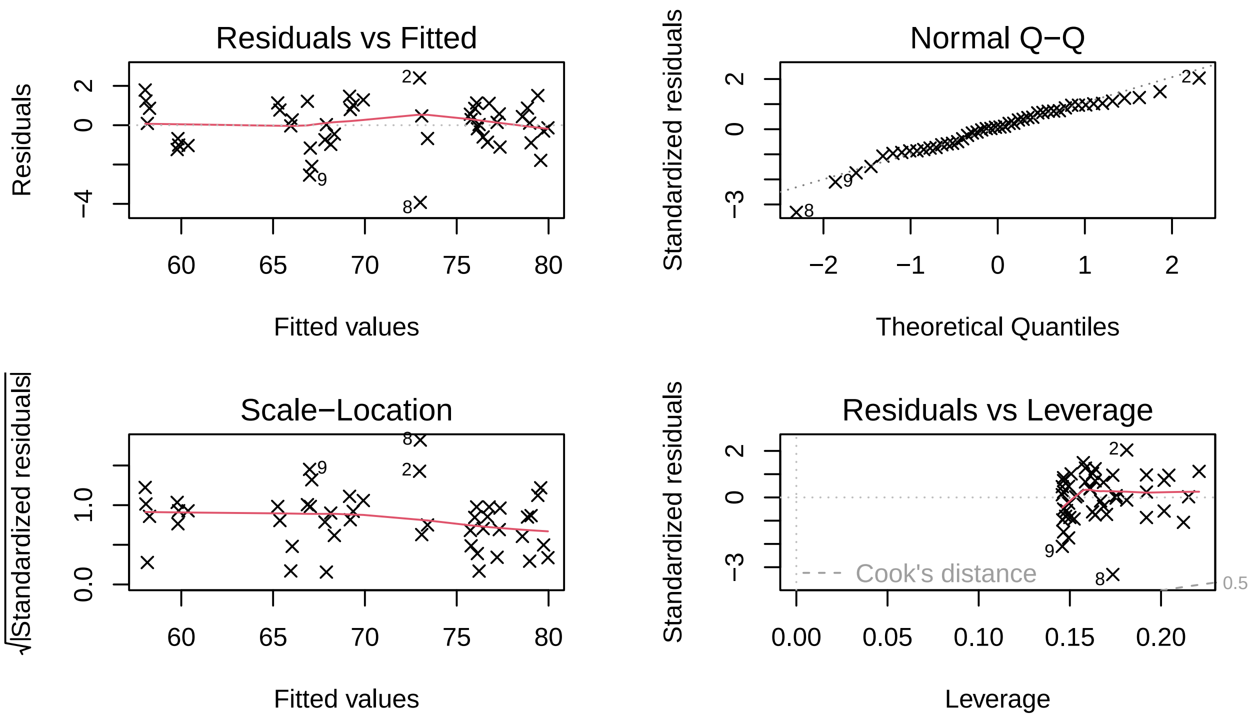
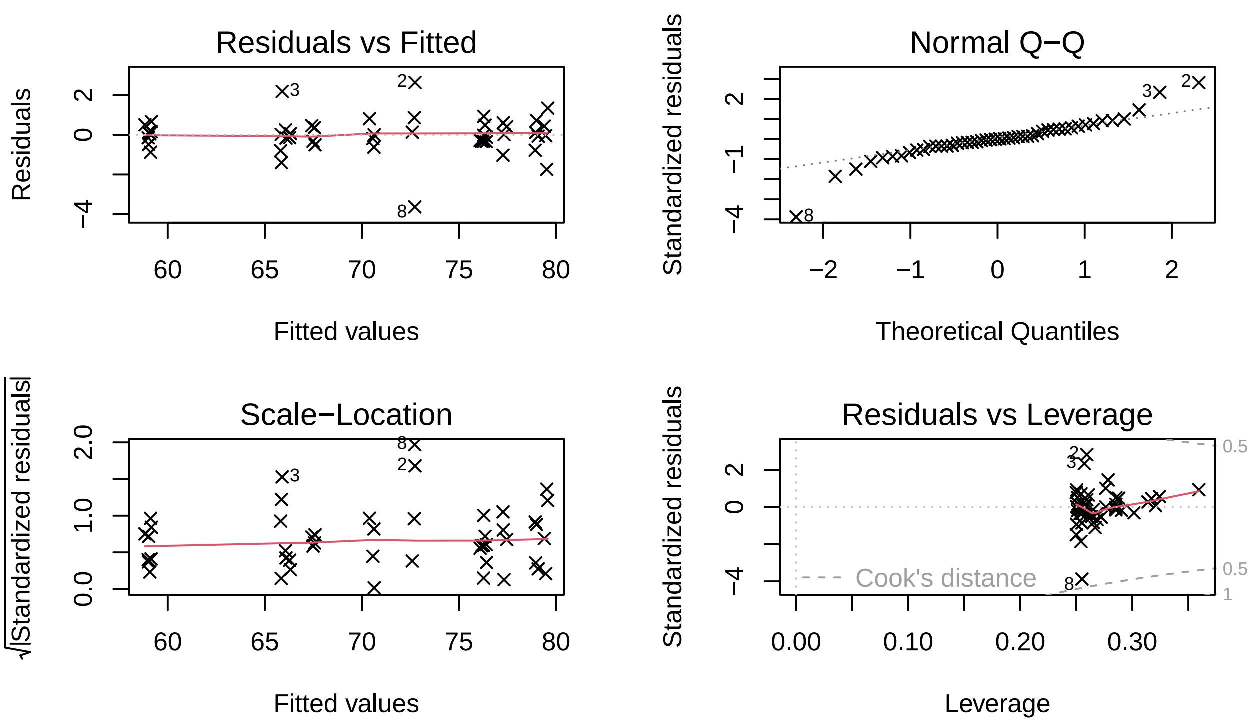


Table 23: Model 8: ANOVA Table

| | 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 |

Table 24: Model 8: Type III ANOVA Table

| | Sum Sq | Df | F value | Pr(>F) |
|---|---|---|---|---|
| (Intercept) | 15086.643 | 1 | 12770.961 | 0.000 |
| Brand | 2.356 | 1 | 1.994 | 0.167 |
| Temp | 220.855 | 2 | 93.478 | 0.000 |
| Stirred | 9.911 | 1 | 8.390 | 0.006 |
| Order | 0.243 | 1 | 0.206 | 0.653 |
| Brand:Temp | 69.452 | 2 | 29.396 | 0.000 |
| Brand:Stirred | 0.409 | 1 | 0.346 | 0.560 |
| Temp:Stirred | 3.500 | 2 | 1.482 | 0.241 |
| Brand:Temp:Stirred | 9.225 | 2 | 3.904 | 0.029 |
| Residuals | 41.346 | 35 | NA | NA |



## Appendix: Code

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

means_table <- df_eff %>% group_by(Brand, Temp) %>% summarise('Mean' = mean(Time))
means_table <- means_table %>% pivot_wider(names_from = Brand, values_from = Mean)
means_table$TempMean <- rowMeans(means_table[,2:3])
means_table <- cbind('Temp' = c('6', '23', '40', 'BrandMean'),rbind(means_table[,2:4], colMeans(means_ta
knitr::kable(means_table)
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=20, col=c(2,4), ylab="", xlab = "",
            main="Mean Time vs. Temp: Stirred = Yes",
            cex.main = 0.75, cex.axis = 0.7, legend = FALSE))
legend("topright",
       title = "Brand",
       c("Name", "Store"),
       bty = "n",
       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 = 0.7)
#```

#```{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=20, col=c(2,4), ylab="", xlab = "",
          main="Mean Time vs. Temp: Stirred = No",
          cex.main = 0.75, cex.axis = 0.7, legend = FALSE))
legend("topright",
```

```r
       title = "Brand",
       c("Name", "Store"),
       bty = "n",
       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 = 0.7)
#model1
aov_eff <- aov(lm_eff <- lm(Time ~ Brand * Temp * Stirred, data = df_eff))

cooksD_values <- cooks.distance(lm_eff)

CD_plot <- 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)), shape = 20) +
  xlab('Sample Points') + ylab("Cook's Distance") +
  geom_hline(yintercept = 0.25, lty = 2) +
  labs(title = "Cook's Distance") +
  theme(
      plot.title = element_text(size = 10, face = "bold"),
      axis.text = element_text(size = 7),
      axis.title = element_text(size = 8)
)

#CD_plot <- ols_plot_cooksd_chart(lm_eff)
qqplot1 <- ggplot(df_eff, aes(sample = Time)) +
  stat_qq(shape = 20) +
  stat_qq_line(linetype = "dashed", col = 'red') +
  labs(x = "Theoretical Quantiles",
       y = "Sample Quantiles",
       title = "Normal Q-Q Plot") +
  theme(
      plot.title = element_text(size = 10, face = "bold"),
      axis.text = element_text(size = 7),
      axis.title = element_text(size = 8)
)

plot_grid(CD_plot, qqplot1)
#qqnorm(df_eff$Time, pch = 20)
#qqline(df_eff$Time, col = "maroon", lwd = 2)
#model with stirred as block effect without interaction
#model2
aov_block_eff <- aov(lm_block_eff <- lm(Time ~ Brand * Temp + Stirred, data = df_eff))
#model3
aov_block_eff_noint <- aov(lm_block_eff <- lm(Time ~ Brand + Temp + Stirred, data = df_eff))
#added covariate Order model with stirred as block effect without interaction
#model7
aov_block_order_eff <- aov(lm_block_order_eff <- lm(Time ~ Brand * Temp + Stirred + Order, data = df_ef
#model8
aov_three_order_eff <- aov(lm_three_order_eff <- lm(Time ~ Brand * Temp * Stirred + Order, data = df_ef
RMSE_function <- function(df_aov){
```

```r
    r_mse <- sqrt(sum(df_aov$residuals^2)/df_aov$df)

    r_s <- 1 - tail(summary(df_aov)[1][[1]][[2]], n = 1) / sum(summary(df_aov)[1][[1]][2])

    a_r_s <- 1 - (1 - r_s)*(nrow(df_aov$model) - 1)/(df_aov$df)

    aic_ <- AIC(df_aov)

    bic_ <- BIC(df_aov)

    output_stats <- c(
        r_mse,
        r_s,
        a_r_s,
        aic_,
        bic_
    )
    return(output_stats)
}


model1 <- RMSE_function(aov_eff)
model2 <- RMSE_function(aov_block_eff)
model3 <- RMSE_function(aov_block_eff_noint)
model7 <- RMSE_function(aov_block_order_eff)
model8 <- RMSE_function(aov_three_order_eff)
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),
                         temp6_rest = c(1/4, 1/4, -1/8, -1/8, -1/8, -1/8, 1/4, 1/4, -1/8, -1/8, -1/8, -
                         temp23_rest = c(-1/8, -1/8, 1/4, 1/4, -1/8, -1/8, -1/8, -1/8, 1/4, 1/4, -1/8, -
                         temp40_rest = c(-1/8, -1/8, -1/8, -1/8, 1/4, 1/4, -1/8, -1/8, -1/8, -1/8, 1/4,
                         ), options=list(adjust="bonferroni")
         )
#knitr::kable(confint(cont_str_brd))
#knitr::kable(confint(cont_strbrd))
#knitr::kable(confint(cont_temp))
lm_eff_me <- lm(Time ~ Brand * Temp, data = df_eff)
```

```
aov_eff_me <- aov(lm_eff_me)
anova_eff_me <- anova(lm_eff_me)

me_table <- as_tibble(summary(aov_eff_me)[[1]][,1:3])
me_table <- cbind('Source' = c('Brand', 'Temp', 'Brand*Temp', 'Residual'),
                  me_table,
                  'Error Term' = c("MSAB", "MSAB", "MSE", "NA"))

MSA <-  me_table[1, 4]
MSB <-  me_table[2, 4]
MSAB <- me_table[3, 4]
MSE <-  round(me_table[4, 4],3)

a <- length(levels(df_eff$Brand))
b <- length(levels(df_eff$Temp))
n <- nrow(df_eff)/(a*b)

sigma_ab <- round((MSAB - MSE) / n,3)
sigma_a <- round((MSA - MSAB) / (b * n), 3)
sigma_b <- round((MSB - MSAB) / (a*n), 3)

f_scores <- c(MSA/MSAB,MSB/MSAB,MSAB/MSE, NA)

error_dof <- c(rep(tail(me_table$Df, n = 2)[1], 2), tail(me_table$Df, n = 1)[1], NA)

me_table['Error Df'] <- error_dof

me_table["F Score"] <- f_scores
f_test <- round(1 - pf(me_table['F Score'][[1]],
                       me_table['Df'][[1]],
                       me_table['Error Df'][[1]]),
                4)

me_table['Pr>F'] <- f_test

se_mu <- sqrt((MSA+MSB-MSAB)/(a*b*n))

dof_app <- (MSA + MSB - MSAB)^2 / (MSA^2/(a-1) + MSB^2/(b-1) + MSAB^2 / ((a-1)*(b-1)))

CV_hat <- sqrt(sum(c(sigma_a, sigma_b,sigma_ab, MSE)))/mean(df_eff$Time)

re_table <- as_tibble(cbind(c('Brand', 'Temp' , 'Brand*Temp', 'Residual'), round(c(sigma_a, sigma_b,sig
colnames(re_table) <- c('Cov Parm', 'Estimate')
re_table$Estimate <- as.numeric(re_table$Estimate)
re_table$Portion <- round(as.numeric(re_table$Estimate) / sum(as.numeric(re_table$Estimate)),3)
knitr::kable(re_table)
knitr::kable(df_stats, caption = "Data Summary Table")
knitr::kable(summary(means_eff), caption = "Least Squares Means")
knitr::kable(confint(cont_str_brd), caption = "Contrast Stirred and Brand")
knitr::kable(confint(cont_strbrd), caption = "Contrast Stirred versus Brand")
knitr::kable(confint(cont_temp), caption = "Contrast Temperatures")
knitr::kable(summary(aov_eff)[[1]], 'simple', digits = 3, caption = 'Model 1: ANOVA Table')
knitr::kable(Anova(aov_eff, type=3), 'simple', digits = 3, caption = 'Model 1: Type III ANOVA Table') #
```

```
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(aov_eff, pch = 4)
knitr::kable(summary(aov_block_eff)[[1]],
             'simple' , caption = "Model 2: ANOVA Table")
knitr::kable(Anova(aov_block_eff, type=3),
             'simple', digits = 3, caption = 'Model 2: Type III ANOVA Table')
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(aov_block_eff, pch = 4)
knitr::kable(summary(aov_block_eff_noint)[[1]],
             'simple', caption = "Model 3: ANOVA Table")
knitr::kable(Anova(aov_block_eff_noint, type=3),
             'simple', digits = 3, caption = 'Model 3: Type III ANOVA Table')
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(aov_block_eff_noint, pch = 4)

name_mean <- df_eff %>% filter(Brand == 'name') %>% select(Time) %>% unlist() %>% mean()
brand_fe <- df_eff %>% group_by(Brand) %>% summarise('Mean' = mean(Time), 'Effect' = Mean - name_mean)

name_mean <- df_eff %>% filter(Temp == '6') %>% select(Time) %>% unlist() %>% mean()
temp_fe <- df_eff %>% group_by(Temp) %>% summarise('Mean' = mean(Time), 'Effect' = Mean - name_mean)
knitr::kable(me_table, caption = 'Mixed Effects Models')

knitr::kable(brand_fe, digits = 3,
             caption = 'Model4: Brand is Fixed and Temperature is Random')

knitr::kable(cbind(c('Temp', 'Brand*Temp', 'Residual'),
                   c(sigma_b, sigma_ab, MSE)), col.names = c('Cov Parm', 'Estimate'), caption = "")
knitr::kable(temp_fe, caption = 'Model5: Brand is Random and Temperature is Fixed',
             digits = 3)
knitr::kable(cbind(c('Brand', 'Brand*Temp', 'Residual'),
                   c(sigma_a, sigma_ab, MSE)), col.names = c('Cov Parm', 'Estimate'),
             caption = "")
knitr::kable(cbind(c('Brand', 'Temp' , 'Brand*Temp', 'Residual'),
                   c(sigma_a, sigma_b,sigma_ab, MSE)), col.names = c('Cov Parm', 'Estimate'),
             caption = 'Model6: Brand and Temperature are Random')
knitr::kable(summary(aov_block_order_eff)[[1]],
             'simple', caption = "Model 7: ANOVA Table")
knitr::kable(Anova(aov_block_order_eff, type=3), 'simple',
             digits = 3, caption = 'Model 7: Type III ANOVA Table')
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(aov_block_order_eff, pch = 4)
knitr::kable(summary(aov_three_order_eff)[[1]],
             'simple', caption = "Model 8: ANOVA Table")
knitr::kable(Anova(aov_three_order_eff, type=3),
             'simple', digits = 3, caption = 'Model 8: Type III ANOVA Table')
par(mfrow=c(2,2), mar = c(5,5,2,2))
plot(lm_three_order_eff, pch = 4)
```