**Tastiest Lemonade**

Courtney Mohs, Kate Eichstaedt, Matt Wendel, Cole Darling, and Sheehan Breyne-Grace

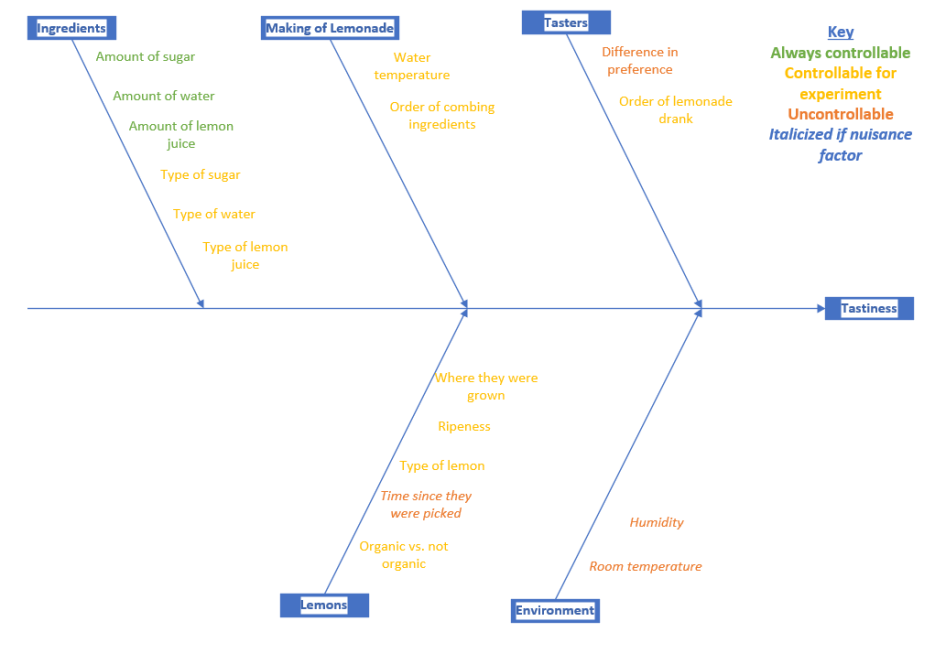
Statistics 424

April 15, 2024

# Introduction

On a hot summer day as you walk through the neighborhood there may be a familiar sighting of a lemonade stand 50 cents per cup, maybe even signs that include ‘best lemonade in the world’. However, the lemonade stand around the corner also said best lemonade, but one is sweeter and the other more sour. The team wanted to investigate what makes the tastiest cup of lemonade. We will look at the optimal combination of lemon juice, sugar, and water to make the tastiest lemonade recipe. To determine which lemonade recipe leads to the tastiest lemonade, a taste test will be performed. In the taste test, tasters will be given random samples of the various lemonade

recipes and then will be asked to rank sweetness, sourness, and overall tastiness on a scale of 0-20 and also summing all categories for a ranking of 0-60. All tasters will be given informed consent before the experiment is conducted and will be able to discontinue the experiment if they desire. The group wants to answer the question if differing the amounts and type of sugar, lemons and water make a difference in an individual's ability to detect differences in sour and sweetness of the lemonade as well as being to detect the optimal combination of ingredients to make the tastiest lemonade in the neighborhood.



**Figure 1**: Fishbone Diagram of Factors Affecting Tastiness of Lemonade

# Methods

Using the cause and effect fishbone diagram along with our research question, we were able to identify several variables. We identified our main response variable to be the taste of the lemonade, which we further broke down into sweetness, sourness, and tastiness which we could them sum together to be an overall tastiness response variable. We identified the three most important factors to be the three main elements that give lemonade its taste – the type of lemon juice, the amount of lemon juice, and the amount of sugar. Additionally, the difference in ratings between tasters was identified as a nuisance factor which we addressed by blocking by taster.

Based off the fishbone diagram, it was clear that multiple factors impact the taste of lemonade, so it was decided it would be best to use a factorial design to test multiple factors. In order to test three factors while also maintaining a feasible number trial to conduct, we elected to use a 23 design. By testing 3 factors at 2 levels each, in addition to 2 repeated measures and 3 tasters, we had created 48 data points. If we tested more factors or at more than 2 levels per factor, the experiment would have become too large for the scope of this project.

We applied our model (Figure 2) to each of our three measurements, sweetness, tastiness, and sourness, and to the combined score.

**Figure 2**: Statistical Model and Variable Description

A screenshot of a computer

Description automatically generated

We used an ANOVA specifically for factorial design to determine if the means of the different recipes differ from each other. We did this for each response variable. This is because a high score of 20 may not be advantageous for the tastiest lemonade. We assumed an average of 10 for sweetness and sourness is the desired level. However, a 20 on the overall tastiness scale is the more desirable ranking. After the ANOVA test determined that at least one of the means differ from each other, and then we used Fisher’s LSD to determine which means of each recipe differed from each other. We collected data in Excel and used R for all of the analysis.

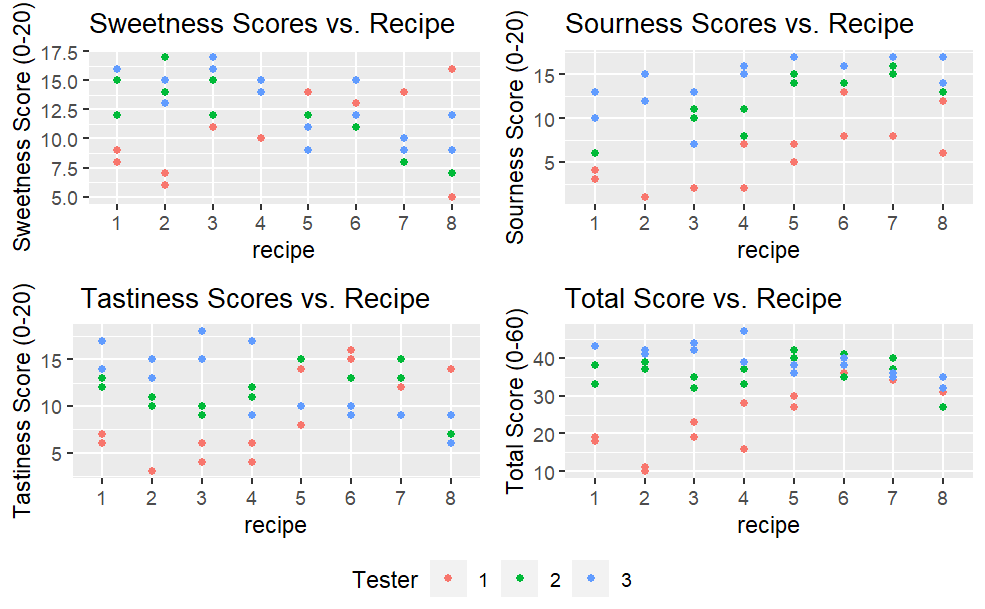
For the response variable we asked the tasters to rank sweetness, sourness, and overall tastiness on a scale of 0-20. We also summed each individual score to get a score between 0 and 60. There were three tasters (blocked by taster), and each person tasted each recipe twice. To make sure that our data was unbiased, it was a blind taste test, so the tasters did not know the recipe they were receiving. During testing, each taster was given a cup that only the experimenter knew the recipe for, and then the taster was asked to rank the sweetness, sourness, and overall tastiness of the sample. Additionally, the order that each person tasted the lemonade in was random. In order to prevent the tasters from influencing one another’s ranks, each taster was asked to record their scores privately. This was then repeated for all the samples. In between each sample, the tasters ate a plain cracker to serve as a palette cleanser. Additionally, each taster tried each lemonade twice to increase the sample size.

# Discussion and Analysis

# Recipe’s Tested

|  |  |  |  |
| --- | --- | --- | --- |
| Recipe | Lemon Juice Type | Amount of Sugar (cups) | Amount of Lemon Juice (cups) |
| 1 | Fresh | 1/3 | ¼ |
| 2 | Concentrate | 1/3 | ¼ |
| 3 | Fresh | ½ | ¼ |
| 4 | Concentrate | ½ | ¼ |
| 5 | Fresh | ½ | ½ |
| 6 | Concentrate | ½ | ½ |
| 7 | Fresh | 1/3 | ½ |
| 8 | Concentrate | 1/3 | ½ |

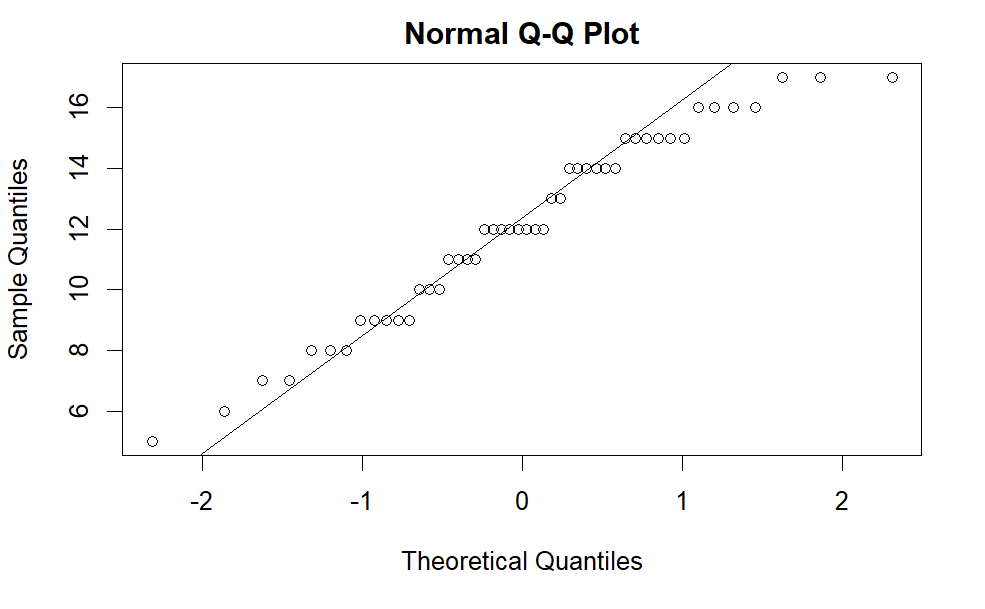
# Summary Data Plots



The above plots, in general, show that the scores of each tester appear to differ from one another. More specifically, the two score replicates within each tester appear to be grouped more closely than the scores of other testers. For example, recipe 3 on the Total Score versus Recipe plot shows that tester 3 scored the recipe around 40-45, tester 2 scored near 20, and tester three fell near the middle of the other testers scores. The replicates within each tester appear to all vary by less than 5, but the scores of tester 3 and 1 differ by ~20. This phenomenon indicated that blocking by tester within our factorial design might be necessary in order to offset the impact of variance associated with the innate “taste” bias of each of our testers. There also appears to be discrepancies in the variance across recipes which is another limitation we will analyze further in our ANOVA’s below.

# ANOVA - Sweetness

Analysis of Variance Table  
  
Response: Sweetness  
 Df Sum Sq Mean Sq F value Pr(>F)   
lemon\_type 2 39.035 19.517 2.4234 0.10220   
sugar\_amount 1 16.044 16.044 1.9922 0.16625   
lemon\_amount 1 51.338 51.338 6.3744 0.01587 \*  
Tester 2 39.125 19.562 2.4290 0.10169   
lemon\_type:sugar\_amount 1 1.174 1.174 0.1457 0.70478   
lemon\_type:lemon\_amount 1 9.389 9.389 1.1658 0.28708   
sugar\_amount:lemon\_amount 1 0.667 0.667 0.0828 0.77513   
Residuals 38 306.042 8.054   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1



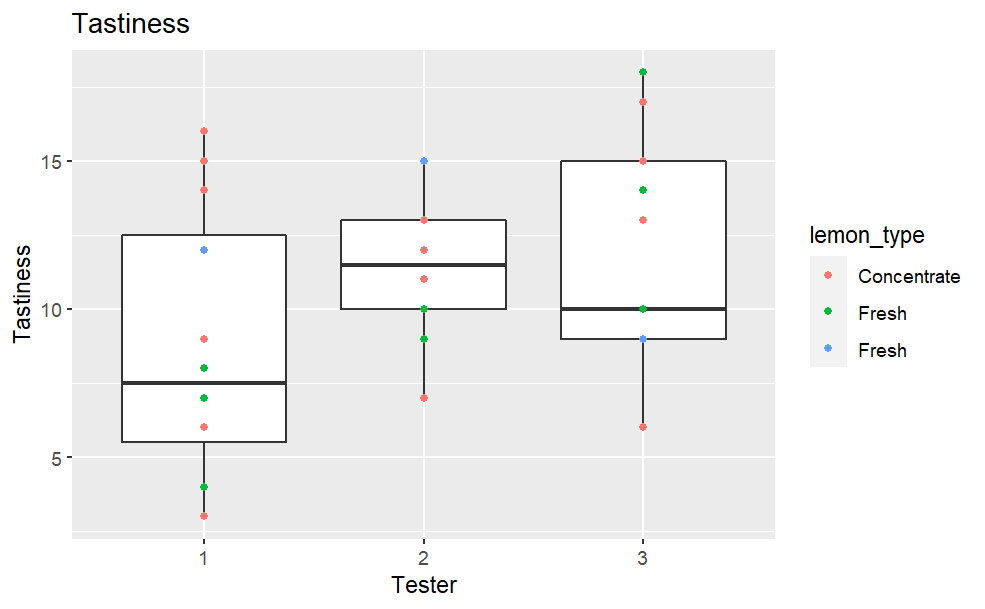
### ANOVA - Sourness

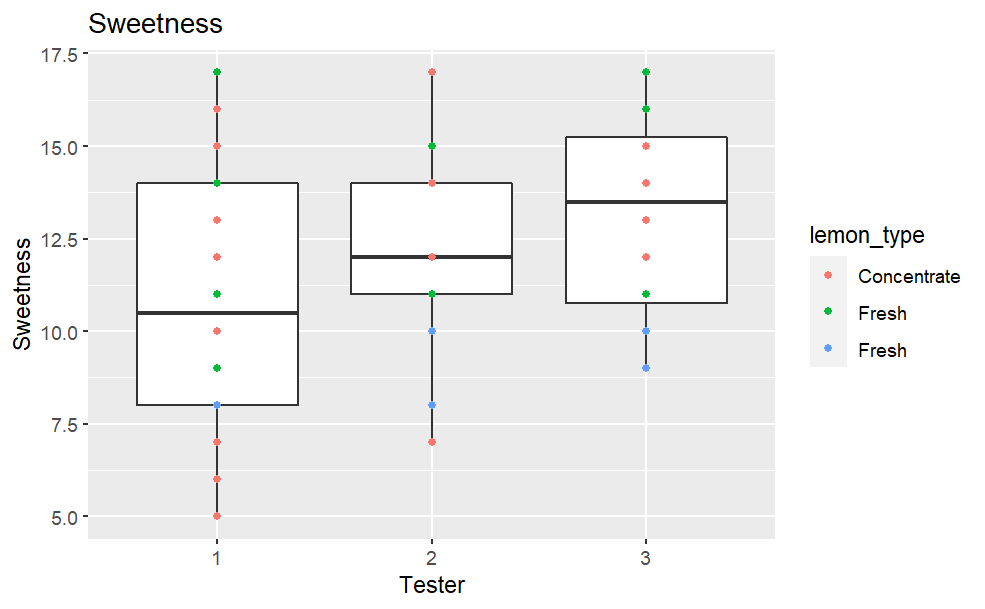
Analysis of Variance Table  
  
Response: Sourness  
 Df Sum Sq Mean Sq F value Pr(>F)   
lemon\_type 2 144.20 72.101 11.3902 0.0001332 \*\*\*  
sugar\_amount 1 20.54 20.544 3.2455 0.0795565 .   
lemon\_amount 1 156.82 156.817 24.7734 1.425e-05 \*\*\*  
Tester 2 609.29 304.646 48.1270 3.848e-11 \*\*\*  
lemon\_type:sugar\_amount 1 4.69 4.694 0.7416 0.3945461   
lemon\_type:lemon\_amount 1 2.72 2.722 0.4300 0.5159165   
sugar\_amount:lemon\_amount 1 0.17 0.167 0.0263 0.8719581   
Residuals 38 240.54 6.330   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

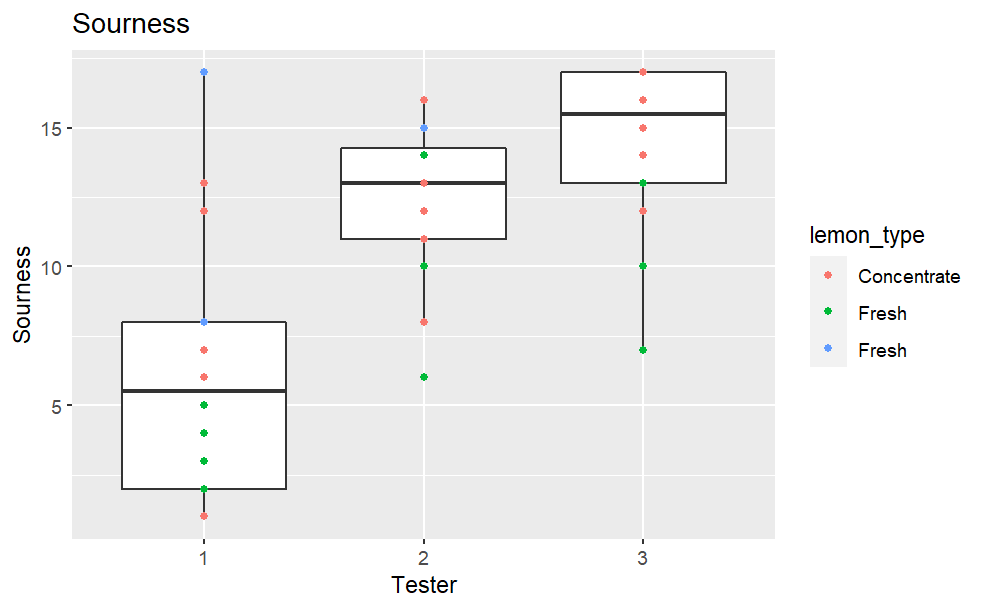
### ANOVA - Tastiness

Analysis of Variance Table  
  
Response: Tastiness  
 Df Sum Sq Mean Sq F value Pr(>F)   
lemon\_type 2 24.76 12.382 0.8904 0.41890   
sugar\_amount 1 12.47 12.469 0.8966 0.34967   
lemon\_amount 1 6.02 6.017 0.4326 0.51466   
Tester 2 95.54 47.771 3.4351 0.04253 \*  
lemon\_type:sugar\_amount 1 20.25 20.250 1.4561 0.23501   
lemon\_type:lemon\_amount 1 1.13 1.125 0.0809 0.77763   
sugar\_amount:lemon\_amount 1 12.04 12.042 0.8659 0.35797   
Residuals 38 528.46 13.907   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

### Block Variance







# Discussion of ANOVA results

As our plot of Sourness by block indicates, the variance across tester may not be consistent for this ANOVA test (although it does seem to be consistent for the Sweetness and Tastiness tests). This violates one of the major assumptions that we need to make to draw inferences from this model. Because the block with higher variance also has the lowest values, a logarithmic transformation is unlikely to fix the problem. We attempted to fix this problem below, but for now we will continue our discussion of this data while acknowledging that this violation may affect our inferences.

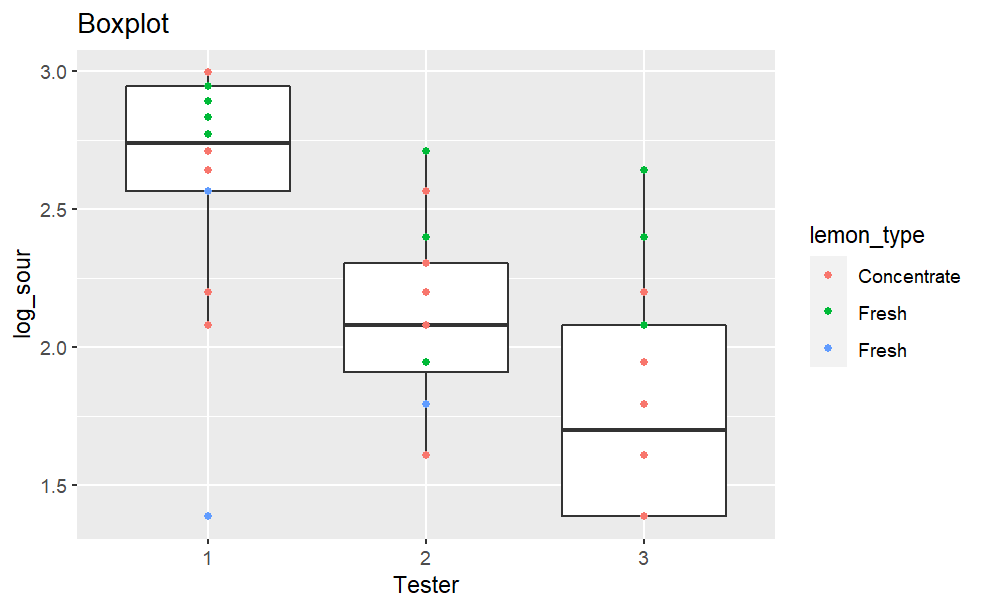
Our assumption of normality for the residuals has not been violated, as one can tell from our QQ-plot for sweetness. The data follows similar patterns for the other two measures, there is code in the appendix that can show this. Our residuals vs fitted values also do not follow any suspicious patterns, so we can safely assume normal distribution of our error terms.

Each ANOVA test shows strong evidence for a block effect, which indicates that each taster has a different taste for lemonade. The only other statistically significant hypothesis is the effect of the amount of lemon on Sourness score. This indicates, not surprisingly, that the amount of lemon juice has an effect on the tasters’ perception of a recipe’s sourness.

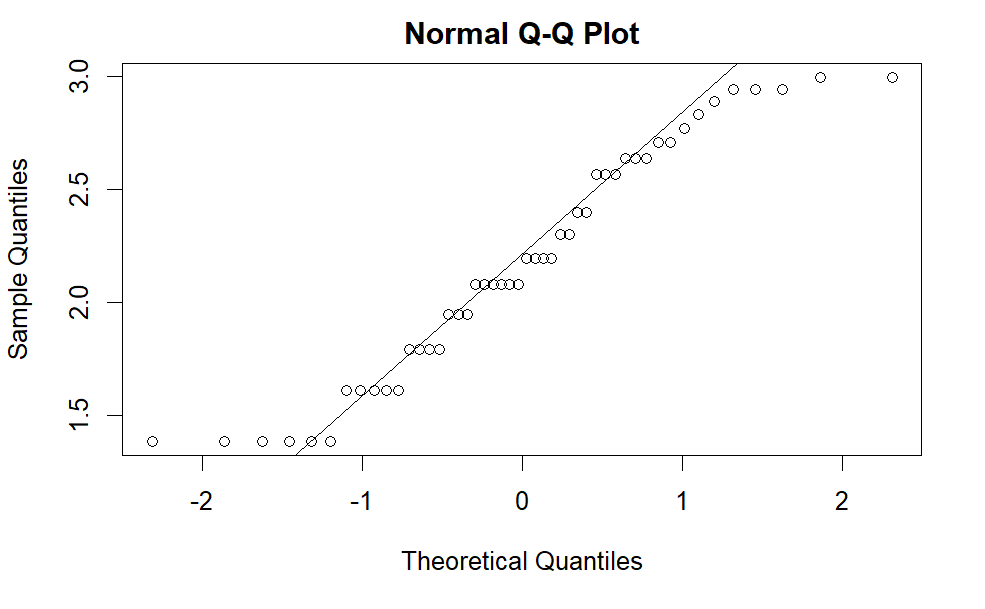
# Variance regularized Sourness results

Seeing as our equal variance across blocks assumption has been violated, we attempted to account for this by changing our scale from 1-20 to 20-1 and taking the log of the values. This transformation fixed the issue with our assumption, but we would need to do more advanced and involved testing to know if this model remains valid after transformation. The transformation leaves us with an ANOVA that does not display the same ‘lemon amount’ effect that our base data showed. Therefore, we should be skeptical of the validity of this hypothesis.

### Sourness ANOVA - block variance regularized



Analysis of Variance Table  
  
Response: log\_sour  
 Df Sum Sq Mean Sq F value Pr(>F)   
lemon\_type 2 1.9499 0.97496 14.3967 2.218e-05 \*\*\*  
sugar\_amount 1 0.2775 0.27752 4.0980 0.0500 .   
lemon\_amount 1 1.6865 1.68649 24.9034 1.368e-05 \*\*\*  
Tester 2 5.7593 2.87965 42.5222 2.017e-10 \*\*\*  
lemon\_type:sugar\_amount 1 0.0286 0.02859 0.4222 0.5198   
lemon\_type:lemon\_amount 1 0.0773 0.07734 1.1420 0.2920   
sugar\_amount:lemon\_amount 1 0.0086 0.00863 0.1274 0.7231   
Residuals 38 2.5734 0.06772   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

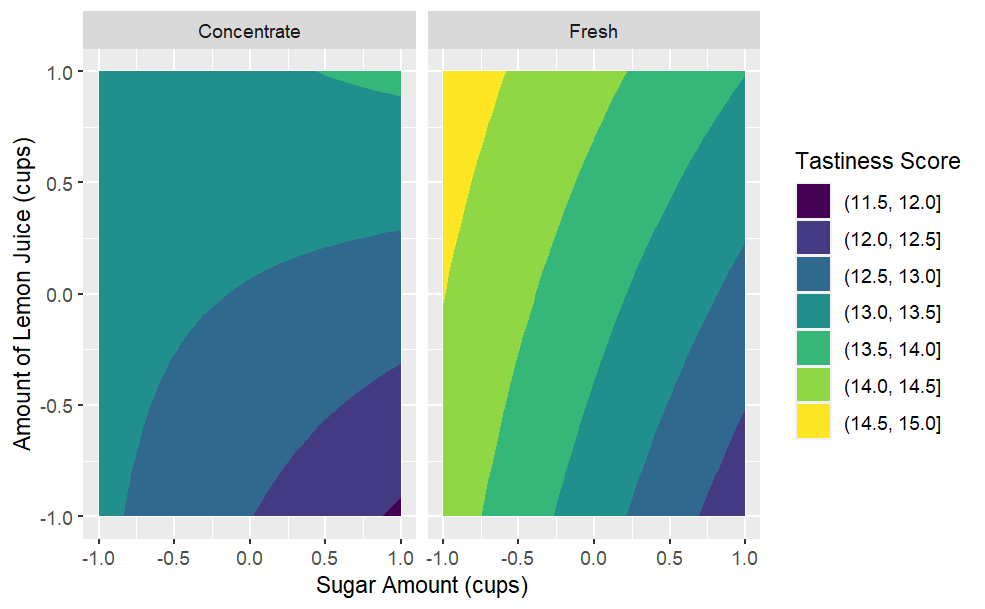


### Further Analysis of Tastiness Scores

### Tastiness Model Call: lm(formula = Tastiness ~ lemon\_type \* sugar\_amount \* lemon\_amount + Tester, data = d\_coded)

### \*d\_coded refers to a dataframe with fitted/coded values for each variable

### Response Surface Plots and the Ideal Cup of Lemonade



Discussion of surface response plots and the ideal cup of lemonade:

The plot above is derived from predictions of our model for Tastiness scores applied to a fitted/coded version of our original data table. The plot indicates that, when concentrate lemon juice was used, the testers are predicted to find the lemonade most tasty when the values of sugar and lemon juice are both high (with 1/2 cup of lemon juice and 1/2 cup of sugar). When freshly squeezed lemon juice was used, the plot predicts that testers will score the recipe more highly if high amounts of lemon juice and low amounts of sugar are used (with 1/2 cup of lemon juice and 1/3 cup of sugar). With both concentrate and fresh lemon juice, recipes with high sugar amounts (1/2 cup) and low amounts of lemon (¼ cup) are predicted to have the lowest tester scores. The ideal predicted cup would include fresh lemon juice as its plot reaches the higher predicted scores (up to 14.5-15) than the concentrate plot. Then, to get the highest predicted tastiness scores of the Fresh contour plot, the ideal lemon juice amount would be ~1/2 cup and the sugar amount would be 1/3 cups. While the plot does provide interesting insights, the model summary in our code revealed that the effect of the tester is the only one that is significant in predicting Tastiness scores meaning that these observations are also statistically insignificant.

But, Ignoring the insignificance, the tastiest cup of lemonade would include: freshly squeezed lemon juice, 1/3 cup of sugar,1/2 cup of lemon juice which corresponds to recipe 7.

# Conclusion

To conclude, this experiment was performed to determine the tastiest lemonade recipe by varying several factors of the recipe. To determine this, we used a 23 design, where amount of lemon juice, amount of sugar, and type of lemon juice were varied accordingly. This created 8 recipes for the three tasters to try. The tasters tried each recipe twice using blocking to try and avoid any nuisance factors that may have arisen from taster bias. Each tester then gave the recipe a rating for sourness, sweetness, and overall tastiness. The data was then analyzed in several ways, including graphical analyses and ANOVA tests. Although not all the data collected showed significance, we did find that there are significant differences between taster's ratings and that the amount of lemon juice influences the tasters rating of sourness. By using a response surface plot and ignoring insignificance, we found that the tastiest recipe would be comprised of fresh lemon juice, 1/3 cup sugar, and ½ cup lemonade. There are certain limitations with this study, including the number of tasters, and taster bias. For future tests, more tasters should be used. In addition, the tasters should have their palates trained for about a week to try and eliminate taste biases.

