Margarita Mayhem
Stats 263 Final Project
Jessica Houghton and Sasha Yousefi

Introduction

Imagine a 87 degree summer day in Palo Alto. Sun is glistening, everyone is outside, the smell of sunscreen and barbeques wafting in the air. What could possibly make this day any better? There is only one answer - a margarita made to statistical perfection.

In this project, we used a $2^{(k-p)}$ fractional design to investigate the most important variables in margarita making. We varied four factors - tequila brand, lime juice freshness, garnish, and sweetener. Since a full factorial experiment would have required 16 different iterations of margaritas, we decided to go with a $2^{(4-1)}$ experiment, which saved us in cost, time, and participant fatigue. We coordinated our tasting experiment along with our annual friendsgiving, where each of our fourteen willing participants tried and ranked each iteration of margarita on a scale of {very bad, bad, mediocre, good, very good}.

While none of our results were statistically significant, we learned noticeable trends in the data. For our main effects, more expensive tequila and salt as a garnish were linked to higher ratings. Agave as a sweetener led to a slightly higher result, but not quite as noticeable of a trend as the former two. Lastly, freshly squeezed lime juice slightly decreased ratings from store bought lime juice. While the interaction between tequila:sweetener and tequila:garnish were negligible, there appeared to be a noticeable interaction between tequila:lime juice. In the tequila:lime juice interaction, the type of lime juice did not affect preference when paired with cheaper tequila, but people seemed to prefer store bought lime juice when paired with more expensive tequila.

If we were to redo the experiment, we would change a couple of design decisions - we would give a larger volume of margarita to taste, would have participants spit out the beverage after each taste, reduce the salt to beverage ratio, and would have a wider range of experienced tequila drinkers. A larger volume of beverage would allow the participants to make more informed decisions on quality with multiple sips. However, with larger beverages, alcohol is more likely to alter the quality of our ratings. To remedy this, we would ask participants not to consume their beverage and for them to spit it out after each taste. Additionally, we received notes that the garnish to beverage ratio was too high. This is likely since our taster cups were so small and we chose to garnish the entire rim (see Figure 8 in the appendix) . We would decrease this ratio so the participants' decision is based more on the total sum of margarita ingredients rather than whether they preferred salt or Tajin. Lastly, many of our participants were quite new to tequila and had only tried it a handful of times before. In order to get a better distribution of margarita quality, we would have more experienced participants in the mix.

Methods

Experiment design:

In order to find the most important factors in a margarita, we used a fractional factorial design. There were 4 different factors of particular interest in making the best margarita: tequila (Hornitos vs Esponda), lime juice (store bought vs fresh squeezed), sweetener (simple syrup vs agave), and garnish (Tajin vs salt) on the rim. Each of these factors had two levels, the first of which was the theoretical

“worse” ingredient for margarita making (cheaper tequila, store bought lime juice, and non-traditional sweetener and garnish), and the second of which is typically used in a traditional margarita. A full factorial experiment design with these four factors would give $2^4 = 16$ margarita recipes. In order to be more realistic with our margarita-making time and resources, we decided a $2^{(4-1)}$ fractional factorial design would be optimal for efficiency while still exploring the main effects and some of the interaction terms. This design led to $2^{(4-1)} = 8$ different margarita recipes.

When designing our factorial experiment, we chose a resolution of IV in order to have a balance between minimum expense and minimum confounding interactions. This resolution specifies that no main effects are confounded with each other, no main effects are confounded with any two factor interactions, and some two factor interactions are confounded with each other. We chose I=TLSG (T=tequila, L=lime, S=sweetener, G=garnish). This choice gives the following aliasing patterns found in Figure 1 where each main effect is aliased with a three level interaction and two level interactions are aliased with each other. This was ideal for our study as we hypothesized that lower level effects would be more significant than the interactions, making it acceptable for main effects to be aliased with likely small three level interactions. However, the downside of this design is that each two level interaction is aliased with another, making it difficult to determine the true effect between two variables. The full factorial design table with the chosen permutations for each margarita can be found in Figure 7 in the Appendix.

```
tequila = lime:sweetener:garnish
lime = tequila:sweetener:garnish
sweetener = tequila:lime:garnish
garnish = tequila:lime:sweetener
tequila:lime = sweetener:garnish
tequila:sweetener = lime:garnish
tequila:garnish = lime:sweetener
```

Figure 1: Aliasing pattern of the margarita experiment

Margarita Recipe:

We followed this margarita recipe to measure out proportions of each of our ingredients. The recipe for one margarita calls for 1 oz of Cointreau, 1 oz lime juice, 2 oz tequila, and ⅓ oz sweetener. Each margarita was made in a batch size with the tequila, lime, sweetener, and garnish varying according Figure 7 in the Appendix. Although the type of ingredient changed per margarita, the proportions of each ingredient constant for each bach. 1 oz was poured into 14 small cups (one per participant) labeled with the margarita recipe number and topped with the appropriate garnish.

Participants:

As everyone has different preferences and dislikes when it comes to a good margarita, we decided to sample as many participants as possible within a reasonable budget. In total, 14 participants between the ages 22-24 sampled each of the 8 margaritas. Each participant was tested separately without the influence of others with the following procedure. First, a random number n between 1-8 was drawn and the participant tasted the nth margarita cup, containing 1 oz of that recipe. After tasting, the participant rated the margarita with one of the following choices: {“very bad”, “bad”, “ok”, “good”, “very good”}

each of which were stored as {-2,-1,0,1,2}. Then, the participant cleanses their palette by eating a bite of bread and drinking a sip of water. These steps are repeated 7 more times, one for each margarita (sampled in a random order).

Results

After running our fractional factorial, the results are a tad disappointing given that none of the main effects or interactions are significant. The multiple R-squared is 0.02652 which shows that the ratings for each margarita are not well explained by the varying ingredients. It is also important to note that 8 interactions were not defined given the aliasing structure shown in Figure 1. Although none of our main effects or interactions are significant, it is still quite interesting to explore how each factor affects the rating of the margarita. An ordering by magnitude for each effect can be found in Figure 3. It seems that the tequila, tequila and lime interaction, and garnish are the most important factors in the model with the highest effect size. This is also evident in Figure 4 which shows a normal QQ Plot. Tequila, tequila and lime, and garnish all lie the furthest away from the reference line meaning that they (visually) have more extreme values than would be expected than if they came from a normal distribution.

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.35714 -1.00000 -0.07143  0.85714  2.14286

Coefficients: (8 not defined because of singularities)
                                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                                           -0.14286    0.30858  -0.463    0.644
tequilaEspolon                                         0.39286    0.43640   0.900    0.370
limeFresh Squeezed                                     0.07143    0.30858   0.231    0.817
sweetenerAgave                                         0.07143    0.30858   0.231    0.817
garnishSalt                                            0.21429    0.30858   0.694    0.489
tequilaEspolon:limeFresh Squeezed                     -0.32143    0.43640  -0.737    0.463
tequilaEspolon:sweetenerAgave                          0.03571    0.43640   0.082    0.935
tequilaEspolon:garnishSalt                            -0.03571    0.43640  -0.082    0.935
limeFresh Squeezed:sweetenerAgave                           NA         NA      NA       NA
limeFresh Squeezed:garnishSalt                              NA         NA      NA       NA
sweetenerAgave:garnishSalt                                  NA         NA      NA       NA
tequilaEspolon:limeFresh Squeezed:sweetenerAgave            NA         NA      NA       NA
tequilaEspolon:limeFresh Squeezed:garnishSalt               NA         NA      NA       NA
tequilaEspolon:sweetenerAgave:garnishSalt                   NA         NA      NA       NA
limeFresh Squeezed:sweetenerAgave:garnishSalt               NA         NA      NA       NA
tequilaEspolon:limeFresh Squeezed:sweetenerAgave:garnishSalt NA        NA      NA       NA

Residual standard error: 1.155 on 104 degrees of freedom
Multiple R-squared:  0.02652,   Adjusted R-squared:  -0.039
F-statistic: 0.4047 on 7 and 104 DF,  p-value: 0.8975
```

Figure 2: Linear Model Coefficients

```
tequilaEspolon:garnishSalt
                -0.07142857
tequilaEspolon:sweetenerAgave
                 0.07142857
             sweetenerAgave
                 0.14285714
          limeFresh Squeezed
                 0.14285714
                 garnishSalt
                 0.42857143
tequilaEspolon:limeFresh Squeezed
                -0.64285714
               tequilaEspolon
                 0.78571429
```

Figure 3: Main effects and interactions sorted by magnitude

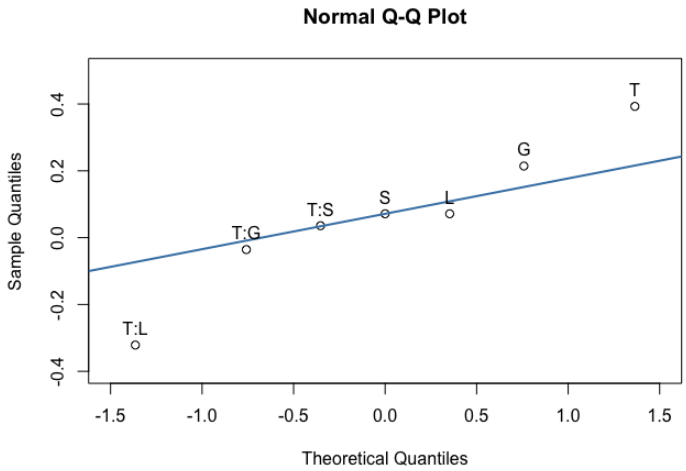

Figure 4: Normal QQPlot of the model's coefficients

A graph of the main effects and interactions can also be informative even though the effect sizes are not statistically significant. Looking at the graphs below one can see that Esponada tequila, store bought lime juice, agave, and salt all lead to slightly better (although not statistically significant) ratings. Similarly, the interaction plots are interesting to look at in order to analyze how two different variables might relate to the ratings of the margaritas. In particular, it looks like store bought lime leads to better ratings when paired with Esponada tequila, but the lime juice does not matter as much with Hornitos tequila. Both sweetener and garnish have similar effects on both types of tequila, although agave and salt had very small, insignificantly better ratings. However, it is important to note that these interactions are aliased with sweetener:garnish, lime:garnish, and lime:sweetener. Given our experimental design, it is impossible to completely disentangle these aliased interactions without further runs to complete a full factorial experiment.

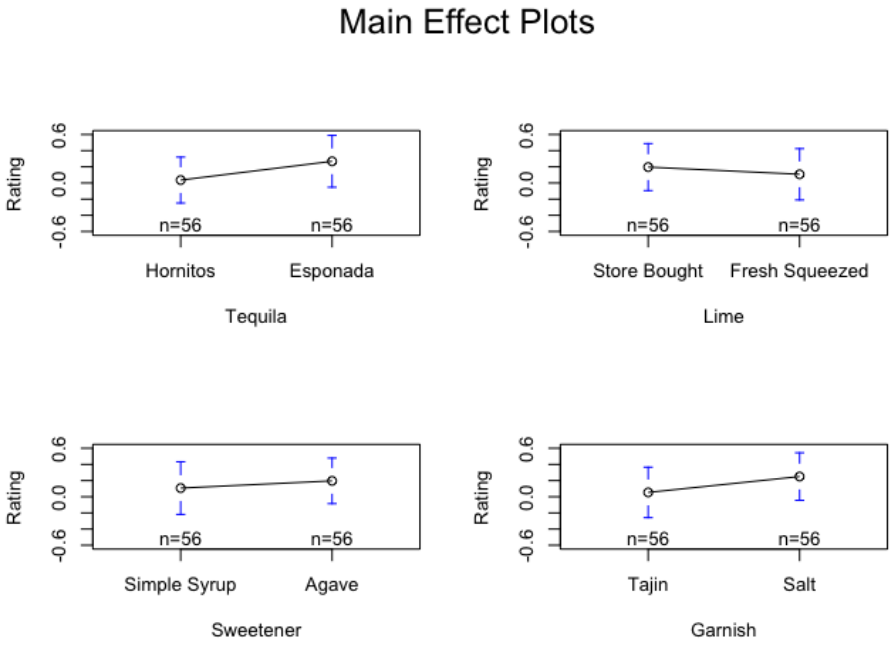

Figure 5: Main effect plots of each ingredient of interest

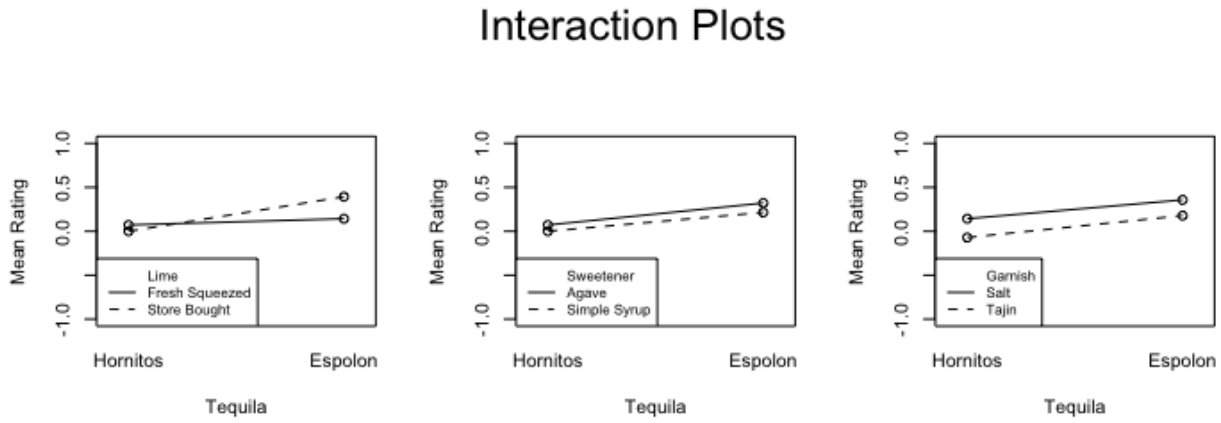

Figure 6: Interaction plots between the measured 2-level interactions. The plots for sweetener:garnish, lime:garnish, and lime:sweetener are not shown as they are aliased with each of the interactions above.

Discussion

Although our experiment did not yield significant results, we still learned about some of the most important factors that go into making the best margarita. Before this experiment, we hypothesized that a combination of more expensive tequila, fresh squeezed lime, agave, and salt would be the best combination as they are the typical ingredients in a classic margarita. Most of this was true through our experiment, with the exception that store bought lime juice led to better ratings than fresh squeezed lime juice. Our theory is that the store bought lime juice might have a sweeter taste compared to an acidic taste of fresh squeezed lime which could be more appealing to our group of participants that are relatively young and inexperienced tequila drinkers. The most important two level interaction in our study was tequila:lime juice which was aliased with sweetener:garnish. This interaction showed that store bought lime juice was better with nicer tequila, but the lime juice does not matter as much with the cheaper tequila. The other two interactions between tequila:sweetener and tequila:garnish seemed to not have any effect on each other. However, each of these two level interactions were aliased with other two level interactions, making the effects of each difficult to distinguish without a full factorial experiment.

As mentioned in the introduction, there are a few changes to the design of the experiment that are worth exploring. One of the items of feedback from our participants was that the amount of margarita in each cup was not enough to sufficiently rate the margarita. Thus, it might be beneficial to have >1 oz per cup, so long as the participants spit out the margarita after each tasting in order to avoid having too much alcohol in their system which may compromise their future ratings. Additionally, we noticed that the garnish was likely overpowering, as the cups held little volume and the garnish covered the entire rim. This likely affected the results as the salt or Tajin were overpowering and may have masked the effects of other ingredients. To remedy this, we could include a much thinner layer of garnish. Lastly, including more experienced tequila drinkers in the study would likely give us a better measure of quality.

All in all, we had a wonderful time conducting this experiment. Not only did it teach us the ins and outs of how to design a $2^{(k-p)}$ factorial experiment, but it gave us a super fun activity to do with friends during our friendsgiving party. Everyone thoroughly enjoyed the project, and we're happy to have a bit more insight on what exactly makes the perfect margarita. We hope to implement these changes in the future to see how our results improve!

Appendix

	I	T	L	S	G	TL	TS	TG	LS	LG	SG	TLS	TLG	TSG	LSG	TLSG
(1)	+					+	+	+	+	+	+					+
tl	+	+	+			+					+			+	+	+
ts	+	+		+			+			+			+		+	+
tg	+	+			+			+	+			+			+	+
ls	+		+	+				+	+				+	+		+
lg	+		+		+		+			+		+		+		+
sg	+			+	+	+					+	+	+			+
tlsg	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
t	+	+							+	+	+	+	+	+		
l	+		+				+	+			+	+	+		+	
s	+			+		+		+		+		+		+	+	
g	+				+	+	+		+				+	+	+	
tls	+	+	+	+		+	+		+			+				
tlg	+	+	+		+	+		+		+			+			
tsg	+	+		+	+		+	+			+			+		
lsg	+		+	+	+				+	+	+				+	

Figure 7: Chosen margarita recipe permutations for the $2^{(4-1)}$ experiment. Empty blocks represent negative (-) values. Each row represents an observation and each column represents an effect of interest. The rows highlighted in green represent the permutations of margarita that were done in this experiment. The white highlighted rows indicate the permutations not done in this experiment, but that would have be done in a complete 2^4 factorial design.

Figure 8: Picture of the margarita experiment before tasting. Each cluster of cups has 14 cups (one per person) and represents a single margarita recipe. The bread at the top of the table was used to palette cleans in between tastings.