

What Makes a Good Frozen Margarita?

A 2_{IV}^{7-2} Experiment to Understand the Variables Effecting Taste

Statistics 850 Final Project

Coleman Breen

April 30, 2021

Summary

An unreplicated quarter-fractional experiment was conducted to model the contribution of seven variables on the taste scores of frozen margaritas. The merits and limitations of the 2_{IV}^{7-2} design, blocking scheme, and experimental protocol are discussed in depth. Subsequent analysis found three main-effects to be significant under two hypothesis testing frameworks: Lenth's method and Dong's method. One two-factor interaction was also found significant by both methods. The addition of frozen strawberries generally improved taste scores whereas the addition of agave syrup generally decreased them. Blanco tequila instead of reposado is preferred. A Box-Cox transform was used to (slightly) strengthen model assumptions. Box-Cox transformation did not identify any additional significant effects. An unreplicated 2^3 follow-up study was conducted to estimate two-factor interactions that were confounded in the main experiment. This follow-up study did not identify any significant effects.

Introduction

Background

A margarita in its simplest form is a cocktail comprised of tequila, triple sec, and lime juice. Its frozen cousin is made by blending all of these ingredients with ice. More complex recipes call for additional fruits—strawberries, pomegranates, or peaches. They can also call for small quantities of raw agave syrup (from which tequila is distilled), orange juice, or freshly-squeezed lime juice. We want to learn which factors contribute to an overall better (conversely worse) taste. On a personal note, frozen margaritas are a family favorite and I want to make the best possible drinks for my family and friends as pandemic protocols are lifted and we can get together again.

Factors

We consider seven factors that comprise a frozen margarita. Three of these factors—tequila brand, tequila color, and triple sec brand—are “substitution” variables. We use either type A or type B for this variable. The other five variables—frozen strawberries, orange juice, lime juice, and agave—are “inclusion” variables. These are either included or excluded for a given run. The seven variables are discussed more below.

Tequila Brand: We consider two tequila manufacturers that are representative examples for medium-priced and higher-priced tequilas. Does more expensive tequila contribute to a higher taste score? The first company Altos is a middle-priced tequila that is commonly used in mixed drinks. In Table 1 we call it “cheap” but that is just in relation to the second brand we consider. The second brand in our experiment is Casmaigos, founded in 2013 by George Clooney and two of his friends. After just four years, the actor [sold the company](#) for one billion US dollars. This tequila is commonly served on its own and is known for its complex flavors.

Tequila Color: Tequila is distilled from agave, a palm-like shrub that grows in Mexico. Like whiskey, the differentiation of tequila comes from the aging process. There are three types of tequila: blanco (white), reposado (rested), and anejo (vintage). Because blanco is not aged, it has strong flavors that go well in mixed drinks. Reposado is aged in oak barrels for two to 12 months. This aging cuts some of the sharpness and gives the tequila a subtler flavor. Anejo is aged even longer and is generally consumed on its own. For this reason, we only consider blanco and reposado tequilas. This factor is called tequila color because blanco tequila is clear and reposado is light brown.

Triple Sec: Triple sec is a sweet orange-flavored liqueur that is used in margaritas, long island iced teas, and cosmopolitans. The gold-standard triple sec is made by a company called Cointreau. Cointreau is relatively expensive, so we are interested in seeing if Cointreau produces a different taste score than that of a cheaper triple sec (in our experiment we use De Kuyper).

Frozen Strawberries: Frozen strawberries are included in many recipes, and are by far the most popular added fruit. Margaritas with strawberries are sweeter and fruitier. We consider two levels: margaritas with and without frozen strawberries.

Orange Juice: Orange juice (without pulp) is a natural compliment to the lime flavoring of margaritas, and the orange flavoring of triple sec. Does the addition of orange juice change the flavor score? The presence or absence of orange juice is another factor.

Lime Juice: In addition to frozen lime concentrate, recipes occasionally call for a small quantity of freshly-squeezed lime juice to bring out the citrus flavors. Our experiment includes drinks with and without added lime juice.

Agave: Agave syrup comes is a sugary substance produced by agave plants. Some people use it sweeten tea or coffee, but it is most famous for being used to distill tequila. Many margarita recipes call for agave syrup to add sweetness to the tart lime base. The final factor is whether or not to add agave to the margarita.

The seven factors and their two levels are found in the following table. For ease of notation, the first seven capital letters of the Roman alphabet are used as abbreviations. The first five variables— A through D —are the so-called starting variables and are in standard order. F and G are generated using Box, Hunter, and Hunter’s generators for a 2^7-2 design, discussed more thoroughly in the following section.

Experimental Design

The following sections cover the design of the experiment. First the design matrix determining the factor levels for 32 runs is introduced. Blocking is discussed in the following subsection. Finally, the experimental protocol and some practical

Table 1: Factor levels of seven experimental variables. Quantities are given to make one run (one glass) at a time.

Factor	Level I	Level II	Abbreviation
Strawberry	None	2 oz	A
Orange Juice	None	1 oz	B
Lime Juice	None	1.5 oz	C
Agave Nectar	None	2 tbsp	D
Triple Sec Brand	De Kuyper	Cointreau	E
Tequila Brand	Altos	Casamigos	F
Tequila Color	Blanco	Reposado	G

considerations are presented.

Quarter-Fractional Design

Our budget is 32 runs, which will be collected in four blocks of eight. With seven variables and 32 runs budgeted, we use a 2^{7-2} design. The quarter-fractional design is an “aggressive” experiment; we will be able to estimate main effects as well as some second-order interaction effects assuming third-order and higher effects are negligible. We use the generators from Box, Hunter, and Hunter: $F = ABCD$ and $G = ABDE$. Then our identifying relation is $I = ABCDF = ABDEG = CEFG$. The resolution four alias, $CEFG$, gives us the six confounded two-factor interactions: CE and FG ; CF and EG ; and CG and EF . These six two-factor interactions in plain English, are any two of lime juice, triple sec brand, tequila brand, and tequila color. Now there are $\binom{7}{2} - 6 = 21 - 6 = 15$ estimable two-factor interactions. While it is a disadvantage to lose some two-factor interactions, overall this design is very efficient.

The proposed experiment is unreplicated, which naturally involves tradeoffs. On the one hand, the unreplicated design with a pre-determined budget allows us to include seven variables with only 32 runs. These seven factors include nearly all of the common ingredients in frozen margaritas, so we are covering a lot of ground. On the other hand, we lose the ability to estimate the variance of the model residuals. This makes significance testing a bit trickier, but we have tools to make good inferences.

Blocking

Thirty-two runs are spread over four blocks, where each block is run on a different day. Blocking is necessary for two reasons. First and foremost, we have just one taster who could not do an adequate job judging 32 different margaritas. Anything past, say, the 15th drink would get a ten out of ten. Second, equipment limits make batches of eight. The two blocking variables are generated from CE and CF since these are already confounded by the quarter-fractional design. Of course this implicitly confounds $CE * CF = EF$, but this two-factor interaction was also confounded by the fractional design. Therefore, we do not lose any more of the two-factor interactions besides the six already lost. The runs and blocks are summarized in the table below. The design matrix, with run numbers, factor levels, and blocking is found in table 2. Crucially, the order of the blocks is randomized as well as the runs within each block (table 3). The randomization scheme is found in the appendix, and consists of randomizing the block order, then for each block randomizing the run order. Both steps are done using R’s `sample()` function with a seed for reproducability.

Budget

The budget is outlined in the appendix, but the total cost to repeat this experiment is approximately \$280. Because I had some of the ingredients beforehand, this total for me was closer to \$200. This is a lot of money but intentional. I wanted there to be real costs involved, because then I would have “skin in the game.” I had to be very careful with my runs and experimental design. I wanted to simulate a real-life situation as much as possible; it needed to sting to mess up. I found that this made me clearer in my thinking and more thoughtful in creating and refining the protocol.

Pre-experiment

Before running the main experiment, I made a batch of four frozen margaritas to see how my protocol could be improved. This helped me overcome four problems. First, I was able to get in the “groove” of making margaritas quickly without sacrificing quality. Second, I improved the single-drink recipe. Recipes are typically given to yield one pitcher (eight cups

Table 2: 32 Runs in standard order with blocks.

Run	A	B	C	D	E	F	G	Block	Run	A	B	C	D	E	F	G	Block
1	1	1	1	1	1	1	1	B4	17	1	1	1	1	-1	1	-1	B2
2	-1	1	1	1	1	-1	-1	B3	18	-1	1	1	1	-1	-1	1	B1
3	1	-1	1	1	1	-1	-1	B3	19	1	-1	1	1	-1	-1	1	B1
4	-1	-1	1	1	1	1	1	B4	20	-1	-1	1	1	-1	1	-1	B2
5	1	1	-1	1	1	-1	1	B2	21	1	1	-1	1	-1	-1	-1	B4
6	-1	1	-1	1	1	1	-1	B1	22	-1	1	-1	1	-1	1	1	B3
7	1	-1	-1	1	1	1	-1	B1	23	1	-1	-1	1	-1	1	1	B3
8	-1	-1	-1	1	1	-1	1	B2	24	-1	-1	-1	1	-1	-1	-1	B4
9	1	1	1	-1	1	-1	-1	B3	25	1	1	1	-1	-1	-1	1	B1
10	-1	1	1	-1	1	1	1	B4	26	-1	1	1	-1	-1	1	-1	B2
11	1	-1	1	-1	1	1	1	B4	27	1	-1	1	-1	-1	1	-1	B2
12	-1	-1	1	-1	1	-1	-1	B3	28	-1	-1	1	-1	-1	-1	1	B1
13	1	1	-1	-1	1	1	-1	B1	29	1	1	-1	-1	-1	1	1	B3
14	-1	1	-1	-1	1	-1	1	B2	30	-1	1	-1	-1	-1	-1	-1	B4
15	1	-1	-1	-1	1	-1	1	B2	31	1	-1	-1	-1	-1	-1	-1	B4
16	-1	-1	-1	-1	1	1	-1	B1	32	-1	-1	-1	-1	-1	1	1	B3

Table 3: Blocks and runs randomized. B2 corresponds to block 2. Run order is given from left to right.

	1	2	3	4	5	6	7	8
B2	17	26	20	27	15	5	14	8
B4	24	1	10	11	4	31	30	21
B3	32	22	23	9	3	29	2	12
B1	19	6	18	28	16	25	13	7

Table 4: Recipe used for all runs. * One cup of ice is 16 cubes, but 12 were used in the experiment. This is discussed in Pre-Experiment. ** We round down to one-eighth tspn for ease of measuring.

	Pitcher		Drink (1/6 pitcher)	
	Ounces	Cups/shots	Ounces	Cups/shots
Frozen Limeade	12 oz	1.5 cups	2 oz	1/4 cup
Frozen Strawberries	12 oz	1.5 cups	2 oz	1/4 cup
Tequila	9 oz	6 shots	1.5 oz	1 shot
Triple Sec	4.5 oz	3 shots	3/4 oz	1/2 shot
Ice Cubes	48 oz	6 cups	16 oz	(12 cubes) *
Lime juice	1.5 oz	1 shot	1/4 oz	1/2 tbsp
Orange Juice	1 oz	2 tbsp	1/6 oz	1 tspn
Agave Nectar	1 oz	2 tbsp	1/6 oz	1 tspn
Salt	n/a	1 tspn	n/a	(1/8 tspn) **

of output). Scaling down a recipe by dividing all quantities by six works well for all of the ingredients except for ice. One cup of ice is equivalent to 16 cubes using my freezer, but this much ice waters down the drink. As a result the flavor of the drink was not strong enough to make sound judgments. Because of this, I reduced the amount of ice used to just 12 cubes. The third problem I encountered was that the drinks' textures were runnier than expected. To prevent the early drinks from melting while I make the later drinks, I decided to place the glasses in the freezer beforehand, and store the drinks in the freezer before presentation. Finally, I could not store four bottles of tequila and two bottles of triple sec in the freezer, as I planned for in the project proposal. Instead, I stored them in the refrigerator. Obviously this is a compromise and plays a part in the runnier-than-expected textures. All of the relevant changes are reflected in the following protocol.

Protocol

The following protocol is used to create eight runs in one block. The recipe is found in table 4. Measurements for a full pitcher and a single glass are presented. Our experiment used the “Drink” column so as to not waste raw ingredients. The full pitcher recipe is provided because that is a more common and useful recipe for others.

Equipment: A standard set of measuring cups is used to measure each ingredient. A Ninja Pro 1100 watt blender is used for all runs. Eight identical high-ball glasses are used to serve drinks.

Preparation: All tequila and triple sec is placed in the refrigerator 24 hours before the drinks are made. All high-ball glasses are placed in the freezer one hour prior to the start of the experiment. This helps keep the drinks close to freezing.

Creating the mixture: The following ingredients (in order) are added to the blender.

1. Limeade concentrate (and frozen strawberries if called for in run)
2. Tequila
3. Triple sec
4. Orange juice (if called for)
5. Lime juice (if called for)
6. Agave syrup (if called for)
7. Salt
8. Ice

Blend for two minutes on medium setting. Pour mixture into highball glass and place glass back in the freezer. Wash out pitcher and dry with towel. Repeat the aforementioned steps for the subsequent runs.

Data Collection

The eight identical glasses are presented in the (random) order in which they were made. Drinks are placed on top of blank note cards along with a glass of water (palette cleanser) and pen to write scores. One taster is used for all blocks. The taster is allowed twenty minutes to sample all eight drinks. The taster is allowed to try drinks as many times as needed and scores each drink from one to ten with one being “bad” and ten being “good”. The taster sampling all drinks at once removes bias in scores because it allows the taster to calibrate their scores. For example, if drinks were presented one at a time, the taster may initially think that the first drink is an eight, but after subsequent drinks the taster may realize that the first drink *should* have been scored as a five. Presenting all eight drinks simultaneously remedies this issue. However, scores of the drinks will not be independent. As we will discuss in later sections, the lack of independence for the response does not appear to translate to a lack of independence in the error terms.

Analysis of Data

Exploratory Data Analysis

Raw data are presented in the appendix. The histogram in figure 1 shows that the center of the distribution is around five and fairly Gaussian. The distribution is relatively symmetrical and does not exhibit any notable skew. The right panel in figure 1 shows the taste score as a function of block (represented by color), and run order (x-axis). Runs in block four appear enriched in the bottom half of the plot. That is, among the low-scoring runs, block four is over-represented. However, there does not appear to be a strong enough pattern to cause concern. We will be mindful of block effects in the rest of the analysis section.

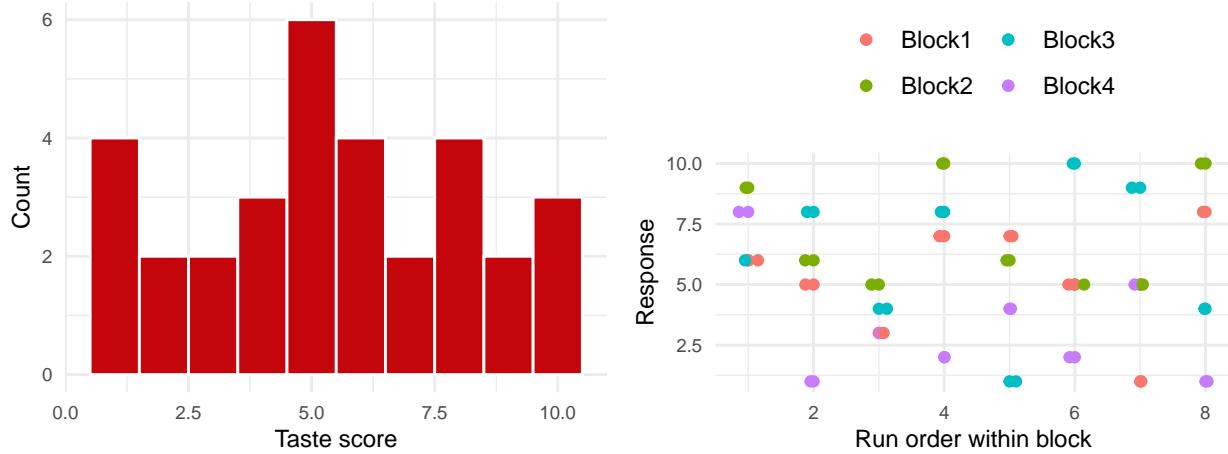


Figure 1: (L) Distribution of response variable. (R) Run order does not appear to bias data.

Table 5: Estimated main effects; non-confounded two-factor interactions; block effects; and intercept.

Effect	Estimate	Effect	Estimate	Effect	Estimate	Effect	Estimate
A	3.50	AB	-0.25	BD	-0.25	DG	0.25
B	-0.62	AC	-0.63	BE	0.37	Block1	-0.25
C	-0.25	AD	0.12	BF	0.25	Block2	0.00
D	-2.00	AE	1.75	BG	-0.62	Block3	-0.25
E	-0.37	AF	1.12	CD	-0.38	Intercept	5.50
F	0.00	AG	0.25	DE	-0.25		
G	-1.88	BC	1.00	DF	-0.62		

Full Model

We start by fitting a model with an intercept, all seven main effects, three blocking variables (four blocks can be coded as three variables), and 15 two-factor interactions. These 15 two-factor interactions, as discussed previously, are not confounded with any other two-factor interactions. In the usual notation, with η being the intercept, $A/2$ is half of the A main effect and the corresponding regression coefficient can be written as $\frac{A}{2}$. We assume the error term $\epsilon \sim \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown. Because our experiment is unreplicated σ^2 is un-estimable. Let x_1 be -1 if A is at Level I, and 1 if A is at Level II (see table 1). Define x_2, x_3, x_4, x_5, x_6 , and x_7 in the same manner for the B, C, D, E, F , and G factors respectively. The $Block_k$ variable with $k \in \{1, 2, 3\}$ is 1 if the corresponding run is in block k and zero otherwise. Runs in block 4 have zeros for each of the three blocking variables. The corresponding regression estimate for the three block variables are δ_1, δ_2 , and δ_3 respectively. Then the model can be expressed as follows.

$$\begin{aligned} \mathbb{E}[Y] = & \eta + \frac{A}{2}x_1 + \frac{B}{2}x_2 + \frac{C}{2}x_3 + \frac{D}{2}x_4 + \frac{E}{2}x_5 + \frac{F}{2}x_6 + \frac{G}{2}x_7 \\ & + \frac{AB}{2}x_1x_2 + \frac{AC}{2}x_1x_3 + \frac{AD}{2}x_1x_4 + \frac{AE}{2}x_1x_5 + \frac{AF}{2}x_1x_6 + \frac{AG}{2}x_1x_7 + \frac{BC}{2}x_2x_3 \\ & + \frac{BD}{2}x_2x_4 + \frac{BE}{2}x_2x_5 + \frac{BF}{2}x_2x_6 + \frac{BG}{2}x_2x_7 + \frac{CD}{2}x_3x_4 + \frac{DE}{2}x_3x_5 + \frac{DF}{2}x_3x_6 + \frac{DG}{2}x_3x_7 \\ & + \delta_1 Block_1 + \delta_2 Block_2 + \delta_3 Block_3 + \epsilon \quad (1) \end{aligned}$$

Estimates of the main and non-confounded two-factor interactions are in Table 5. Three things stand out to me. First, the magnitude of the strawberry (A) main effect is larger than all the others, and nearly twice as large as the second biggest (D aka agave syrup) effect. Second, the block variables are close to zero. This means that the block effects are, at the very least, not detrimental to the experiment. Finally, the estimate of the F main effect–tequila brand–is exactly zero. We explore the significance or lack thereof in the next section, but for now it surprises me that tequila brand does not impact the taste.

Before hypothesis testing we need to define some variables. Let $\theta_1 = A, \theta_2 = B, \dots, \theta_{22} = DG$. We consider 22 simultaneous null hypotheses: $H_0^1 : \theta_1 = 0; \dots; H_0^{22} : \theta_{22} = 0$. Let the size of each test be $\alpha = 0.1$. Then a significant effect is one in which the confidence interval does not contain zero. Equivalently, an estimated effect magnitude larger than some cutoff is deemed significant. We cannot use Bonferroni simultaneous confidence intervals because we are unable to obtain a reliable estimate of σ^2 (and hence cannot construct intervals based on the Student's t distribution).

Note on Using Higher-Order Interactions to Estimate σ^2

Seven of the 31 total degrees of freedom are allocated to estimate main effects. Another 15 are used to estimate two-factor interactions. Three more are used on estimating block effects. If we assume three-factor and higher interactions are negligible, then we have six degrees of freedom left to "estimate" σ^2 . However, as discussed in Mead's **Statistical Principles for the Design of Experiments**, some of the error degrees of freedom correspond to confounded two-factor interactions and so are not even reliable for estimating error variance (page 347). This limitation was known when the experiment was designed and can be overcome using Lenth's and Dong's methods for testing for significant effects.

Lenth's Method

Lenth's and Dong's methods allow us to perform significance testing one effects without trying to estimate σ^2 . Their Lenth's method defines the following:

$$s_0 = 1.5 \times \text{median}|\hat{\theta}_i|$$

$$PSE = 1.5 \times \text{median}\{|\hat{\theta}_i| : |\hat{\theta}_i| < 2.5s_0\}$$

$$\nu = g/3 \text{ and } \gamma = 0.5(1 - (1 - \alpha)^{1/g}) \text{ where } g \text{ is number of effects.}$$

The critical value for the test that rejects the null at level $\alpha = 0.1$ is $t_{\nu, \gamma} \times PSE$ is approximately 1.46. Implementation in code is relegated to the appendix.

Dong's Method

Dong's method is an improvement to Lenth's method. It uses the same g, ν from Lenth's method. Briefly, it requires two sequential calculations, written up by Professor Loh in slide 179 of the class notes based on a paper by Dong.:

$$m_1 = \#\{i : |\hat{\theta}_i| < 2.5s_0\} \quad \text{and} \quad s_1^2 = \frac{1}{m_1} \sum_{|\hat{\theta}_i| \leq 2.5s_0} \hat{\theta}_i^2$$

$$m_2 = \#\{i : |\hat{\theta}_i| < 2.5s_2\} \quad \text{and} \quad s_2^2 = \frac{1}{m_2} \sum_{|\hat{\theta}_i| \leq 2.5s_1} \hat{\theta}_i^2$$

Then $\hat{\theta}_i$ is significant if $\hat{\theta}_i > t_{m_1, \gamma} \times s_2$. Dong's method gives a critical value of 1.54, which classifies A, D, G , and AE as significant as well. The functional code can be found in the appendix.

Summary of Significant Factors and Interpretation

The three significant main effects have an important physical interpretation. All else equal, inclusion of frozen strawberries corresponds to an expected 3.5 point increase in taste score. Including agave syrup is associated with an expected decrease of one point. Likewise, margaritas with reposado are on average two points lower than those with blanco tequila. Finally, strawberry margaritas made with Cointreau are on average about 1.5 points better than non-strawberry margaritas made with De Kuyper. We now turn our attention to a restricted model with these four variables and subsequent model diagnostics.

Restricted Model

Having found four significant variables, we now consider a restricted model with extraneous variables left out. Consider the model below, and the corresponding regression coefficients.

$$\mathbb{E}[Y] = \eta + \frac{A}{2}x_1 + \frac{D}{2}x_4 + \frac{G}{2}x_7 + \frac{AE}{2}x_1x_5 + \epsilon \quad (2)$$

$$\mathbb{E}[Y] = 5.4375 + 1.75x_1 - x_4 - 0.9375x_7 + 0.875x_1x_5 + \epsilon$$

Restricted Model Diagnostics

There are three fundamental assumptions made in our linear regression model. First, that the variables satisfy an additive linear model. Second, we assume that the error terms are normally distributed with mean zero and constant variance. Finally, we assume that error terms are independent. The left plot of figure 2 shows that residuals follow a normal distribution fairly well. While there are slight deviations from the normal in the tails, overall the residual plot does not indicate a departure from normality. The right panel of figure 2 indicates that residuals are centered at zero. The variance of these residuals seems to be smaller at the extremes of the fitted value. This may indicate that our error terms do *not* have constant variance. While the pattern is not extreme enough for me to completely write off model (2), we may be able to improve on these diagnostics by transforming Y in the next section.

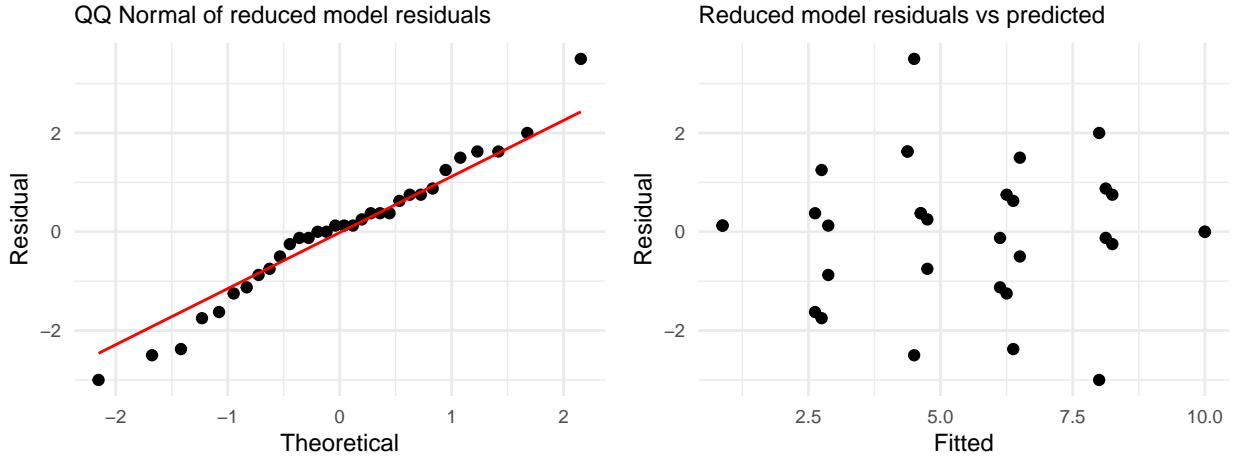


Figure 2: Model diagnostic plots for the reduced model (2). (L) Residuals from reduced model do not depart from normal distribution. (R) Residuals may have slightly non-constant variance, but are centered at zero.

Box-Cox Transformation

Initial model diagnostics do not suggest large departures from linear model assumptions. However, we may be able to improve upon our restricted model by transforming the taste scores. Using Professor Loh's description, assume that $\exists \lambda$ such that

$$y^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda \hat{y}^{\lambda-1}} & \text{if } \lambda \neq 0 \\ \hat{y} \log(y) & \text{otherwise} \end{cases}$$

are independent, normally distributed, same, variance, and satisfies an additive model (slide 181). Iterating over a range of potential λ values, we find the maximum likelihood estimate for $\lambda \approx 0.77$, which we round down to a tidy three-quarters. Then the Box-Cox transformed model is below with estimated regression coefficients.

$$\mathbb{E}[Y^{\frac{3}{4}}] = \eta + \frac{A}{2}x_1 + \frac{B}{2}x_2 + \frac{C}{2}x_3 + \frac{D}{2}x_4 + \frac{E}{2}x_5 + \frac{F}{2}x_6 + \frac{G}{2}x_7 + \delta_1 \text{Block}_1 + \delta_2 \text{Block}_2 + \delta_3 \text{Block}_3 + \epsilon \quad (3)$$

$$\mathbb{E}[Y^{\frac{3}{4}}] = 4.87 + 1.75x_1 - 0.3x_2 + -0.7x_3 - x_4 + -0.25x_5 - 0.01x_6 - 0.9x_7 - 0.2\text{Block}_1 - 0.1\text{Block}_2 - 0.1\text{Block}_3 + \epsilon$$

Repeating the process described in the previous section, we use Lenth's and Dong's methods and find A , D , and G main effects to be significant. This is visualized in the top right panel of figure 3. We use these three variables to define the Box-Cox reduced model and its least-square estimates for the regression coefficients.

$$\mathbb{E}[Y^{\frac{3}{4}}] = \eta + \frac{A}{2}x_1 + \frac{D}{2}x_4 + \frac{G}{2}x_7 + \epsilon \quad (4)$$

$$\mathbb{E}[Y^{\frac{3}{4}}] = 4.8 + 1.7x_1 - x_4 - 0.9x_7 + \epsilon$$

After Box-Cox transformation, we did not identify any new main effects. Conversely we did not rule any significant effects from the restricted model to be insignificant after performing the Box-Cox transformation. Between the full Box-Cox (3) and reduced Box-Cox model (4), we prefer the reduced as it is sparser and easier to interpret. Diagnostic plots for (4) are found in figure 3. It appears that error terms are well-behaved: they are normally distributed, centered at zero, and have constant variance.

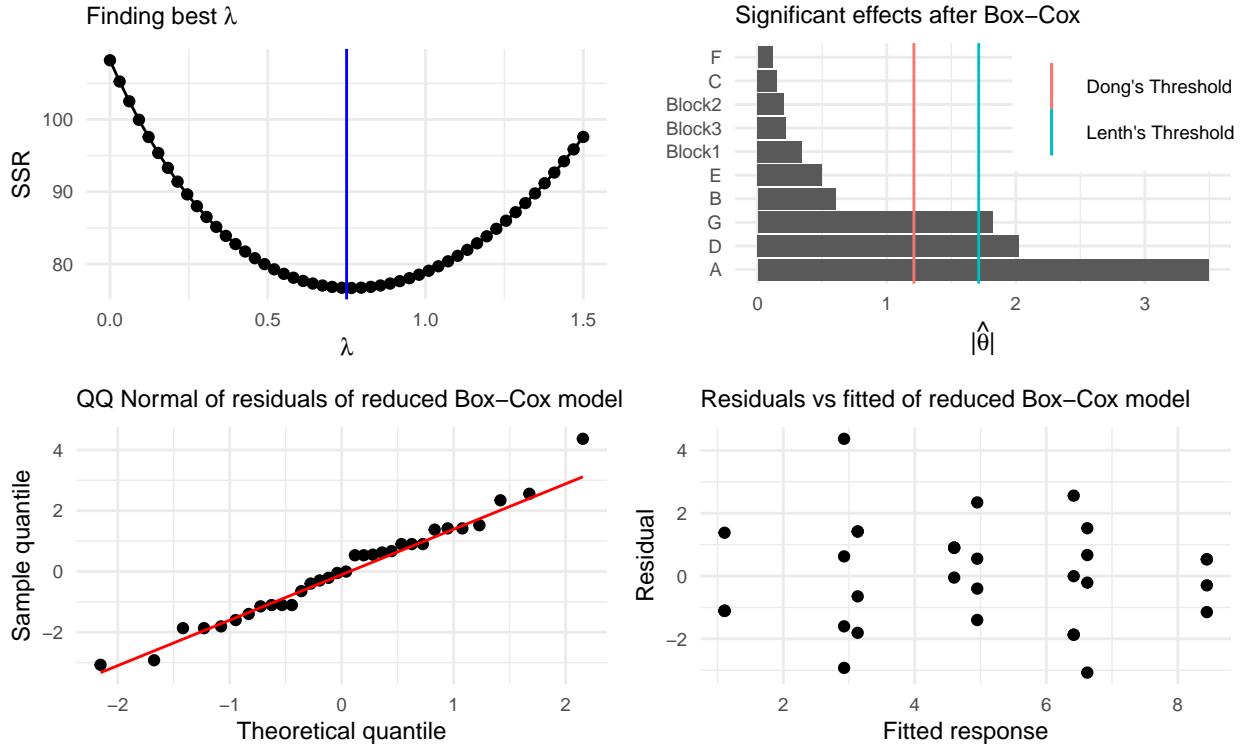


Figure 3: (TL) The optimal Box-Cox parameter is found by minimizing the model sum of square residuals. (TR) Three effects are found to be significant after transformation. (BL) QQ Normal plot does not suggest departure from normality assumption. (BR) Residuals are centered near zero and have mostly constant variance.

Final Model

The full-model (1) is unnecessarily bloated and has too many non-significant terms. The full Box-Cox model (3) is relatively parsimonious, but also includes numerous non-significant terms. The reduced Box-Cox model (4) is lean in parameters and does the best with model diagnostics, but is also less interpretable because it requires an exponential transformation on Y . The reduced model (2) is explainable, simple, and does well enough in model diagnostics. For this reason, we use this as our final model.

Follow-up Study: Unreplicated 2^3 Experiment to Examine Two-Factor Interactions.

After conducting 32 runs and analyzing the data we have enough tequila to make two more drinks for each color-brand combination: two Casamigos Blanco, two Altos Blanco, two Casamigos Reposado, and two Altos Reposado. There is enough of each triple sec to make at least four drinks. The other ingredients are relatively cheap and can be replenished. In short, our limiting factor is tequila.

Why not replicate? There are two reasons why I do not want to replicate the experiment. Most obviously, it would require doubling the financial and (experimental) costs. Second, replication does not get us anything I need or want. Replication would only allow us to get an estimate of σ^2 . It would allow us to build more robust simultaneous confidence intervals (e.g. Bonferroni or Scheffe). This is not needed. I feel confident in the classification of three main effects and one two-factor interaction being significant. There were effects that were “on the fence.” That is to say, no variables were *almost* significant. Because of this, I am not too concerned with estimating σ^2 and getting more robust confidence intervals.

Why not the other half-fraction? An additional 32 runs, as in the last paragraph, would require doubling the budget. Additionally, it would not get us much additional information for the time and money.

Why not adding a one-eighth fraction? Mead discusses irregular fractions for 2^n structures in section 14.5, including the three-eights design. It would be possible to turn our one-quarter-replicate into a three-eights-replicate by doing a

“partial fold over”. This would require an additional 16 runs and would not get us much in terms of estimations since some of the aliased two-factor interactions were used to block.

Follow-Up Design and Analysis

By using a fractional design, we confounded six total two factor interactions, including a few of interest: $FG = \text{TequilaBrand}:\text{TequilaColor}$; $EF = \text{TripleSecBrand}:\text{TequilaBrand}$; and $EG = \text{TripleSecBrand}:\text{TequilaColor}$. Before conducting the experiment, I assume the three factor interaction EFG is negligible. The model is

$$\mathbb{E}[Y^*] = \eta^* + \frac{E}{2} + \frac{F}{2} + \frac{G}{2} + \frac{EF}{2} + \frac{EG}{2} + \frac{FG}{2} + \frac{EFG}{2} + \epsilon.$$

By assuming the three-factor interaction is zero in order to estimate σ^2 . Of course, this estimate will be poor because it is calculated using only one degree of freedom. The EFG estimate is 1.5. Then $SD(\text{effect}) = \sqrt{\text{Var}(\text{effect})} = \sqrt{(EFG)^2} = 0.75$.

We continue to use size $\alpha = 0.1$. Then a Bonferroni-adjusted p-value is $0.1/6$ and the critical t -statistic is found to be $t_{1,0.1/6} \approx 19$. Then the confidence intervals take the form $\text{Estimate} \pm 19 * 1.5$. Clearly an interval this wide does not return any significant effects. Because this method is inferior in many respects, we also try Lenth’s and Dong’s methods with intervals of the form $\text{Estimate} \pm 15$ and $\text{Estimate} \pm 6$ respectively. Unsurprisingly, these methods do not return any significant effects either. In other words, there is not sufficient statistical evidence that the EF , FG , or EG interactions are significant in this follow-up study. This means that the two-factor interactions among tequila brand, tequila color, and triple sec brand are not significant.

The optimal λ for the Box-Cox transformation is -0.55 . However, ever after transformation, no main effects are deemed significant. At first glance we may worry that the G main effect is not significant in this analysis while it was in the main analysis. However, this is likely due to the within-batch lack of independence. The scorer was comparing strawberry–no lime juice–no orange juice–no agave margaritas to *each other* rather than a representation of the entire parameter space. In other words, each of the four blocks in the main experiment had a decent representation of each of the seven factor levels while this follow-up does not.

The main takeaway from this follow-up is that the confounded two-factor interactions from the main study are likely insignificant. Therefore, we do not need to concern ourselves with estimating them alongside the already significant variables in the reduced model.

Limitations

There are several ways to improve the design and execution of this experiment.

Taster(s): The use of multiple, randomly-selected tasters could have been used and a summary statistic (average or median) could have been used as the taste score. This was considered early on, but was difficult to coordinate given the shelter-in-place restrictions.

Replication: Replicating this experiment would allow us to estimate σ^2 which would help during model fitting and diagnostics.

Run Sizes: When making “frozen” margaritas in single-serving runs, the drinks are more of a slushy consistency than a true frozen margarita. Since this texture is consistent within and across blocks, it is not a cause for concern. However, a more faithful run would have involved making an entire pitcher at once which would have a better texture. This would have wasted some 85% of the raw materials and would have increased the cost of the experiment six-fold.

Conclusion

In our attempt to understand what contributes to a high-quality margarita, we designed and executed a 2^{7-2}_{IV} unreplicated experiment, allowing us to estimate all main-effects and most two-factor interactions. After considering four models (full, restricted, Box-Cox, and Box-Cox restricted), we presented the restricted model as the final model.

$$\mathbb{E}[Y] = 5.4375 + 1.75x_1 - x_4 - 0.9375x_7 + 0.875x_1x_5 + \epsilon \quad (2)$$

This model, in crude terms, suggests that i) adding strawberries improves taste scores; ii) agave syrup decreases taste scores; iii) blanco tequila is better than reposado; and iv) frozen strawberries with Cointreau is better than no strawberries and De Kupyer. A small follow-up study revealed that the two-factor interactions confounded in the main experiment are insignificant.

References

- [1] Box, Hunter, and Hunter. Statistics for Experimenters: Design, Innovation, and Discovery, 2nd Edition. 2005.
- [2] Dong. On the Identification of Active Contrasts in Unreplicated Fractional Factorials. 1993.
- [3] Lenth. Quick and Easy Analysis of Unreplicated Factorials. 1989.
- [4] Mead, Gilmour, and Mead. Statistical Principles for the Design of Experiments: Applications to Real Experiments. 2012.
- [5] Wu and Hamada. Experiments: Planning, Analysis, and Parameter Design Optimization, 2nd Edition. 2000.

Table 6: Approximate budget.

Ingredient	Quantity	Price
Casamigos blanco	750 ml	47.95
Casamigos reposado	750 ml	52.95
Altos blanco	None	29.95**
Altos reposado	None	29.95
De Kuyper Triple sec	750 ml	7.95
Cointreau (Expensive)	750 ml	39.95
Minutemaid Frozen lime concentrate	10 cans	25
Harris teeter frozen strawberries	5 tubs	25
Orange Juice (pulp-free)	Two 12 oz bottles	5
Limes	15	5
Agave syrup	One bottle	5
Morton fine sea salt	One container	5

Appendix

Supplemental information, including code, data, and pictures are found in the following sections.

Budget

Alcohol was purchased through the North Carolina Alcoholic Beverage Control (ABC) system. The price list is found on their [website](#).

Code

Helper functions

```

knitr:::opts_chunk$set(echo = FALSE)
knitr:::opts_chunk$set(eval = TRUE)
knitr:::opts_chunk$set(fig.height = 2.5)

fs <- 9 # font size for titles in ggplot

library(tidyverse) # for data wrangling
library(kableExtra) # for table outputs
library(wiscR) # some plotting functions
library(gridExtra)
library(latex2exp)
add_effect <- function(X, effect) {
  # X is a matrix (likely made by 'generate_initial_design')
  # effect is an effect represented by string of letters (e.g. "AB" or "ADE")
  # OR effect is a an integer
  factors <- colnames(X)

  if (typeof(effect) == "double"){

    combos <- t(combn(factors, 2))
    all_effects <- apply(combos, 1, paste0, collapse = "")
  } else {
    all_effects <- effect
  }
}

```

```

for (string in all_effects){
  v <- strsplit(string, split = "")[[1]]
  X <- cbind(X, apply(X[, v], 1, prod))
  colnames(X)[ncol(X)] <- string
}
return(X)
}

get_confounded <- function(I){
  # I is identifying relation (I = "ABCD")
  # Returns all confounded variables as data.frame

  I.split <- (strsplit(I, "")[[1]]

  # total number of combinations is  $2^n$ .
  # Subtract one since don't care about n choose 0
  # Subtract another 1 since don't need n choose n
  N <- 2^(length(I.split)) - 2
  df <- data.frame(lower.confounded = rep("", N),
                    higher.confounded = rep("", N),
                    stringsAsFactors = F)

  ix <- 1
  for (i in 1:(length(I.split)-1)) {
    combos <- apply(t(combn(I.split, i)), 1, paste0, collapse = "")
    for (string in combos) {
      tmp <- c(string, multiply_effects(c(I, string)))
      df[ix, ] <- tmp[order(nchar(tmp), tmp)]

      ix <- ix + 1
    }
  }
  return(distinct(df))
}

get_all_confounded <- function(multiple_Ids){
  N <- length(multiple_Ids)
  con.df <- get_confounded(multiple_Ids[1])

  for (i in 1:N){
    combos <- t(combn(multiple_Ids, i))
    for (j in 1:nrow(combos)){
      vec <- ifelse(i %in% c(1,N), combos[j, ], multiply_effects(combos[j, ]))
      tmp.df <- get_confounded(vec)

      con.df <- rbind(con.df, tmp.df)
    }
  }
  return(con.df %>% distinct())
}

generate_initial_design <- function(p){
  # Creates an X matrix with p columns. Each column is composed of 1s and -1s.
  # The patterns of + and - are different in each column. For example:
  # Col 1: + - + - + - ...
  # Col 2: + + - - + + ...
  # p is the number of factors and is only required (accepted) input.

  runs <- 2^p
  X <- matrix(rep(0, runs*p), nrow = runs)
}

```

```

colnames(X) <- LETTERS[seq( from = 1, to = p )] # assign column names from alphabet

for (jj in 1:p){
  pattern_multiplier <- 2^(jj - 1)
  string_multiplier <- 2^(p-jj)

  ix <- c(rep(1, pattern_multiplier), rep(-1, pattern_multiplier))
  X[,jj] <- rep(ix, string_multiplier)
}

return(X)
}

multiply_effects <- function(str_vec) {
  # helper function, multiply letter strings
  # E.g. multiply_effects("ABC", "CDE") --> "ABDE"
  N <- length(str_vec)
  while (N > 2) {
    str_vec <- c(multiply_effects(str_vec[1:2]), str_vec[3:N])
    N <- length(str_vec)
  }
  counts <- table(strsplit(paste0(str_vec[1],str_vec[2]), ""))
  # If the number of occurrences is odd, that letter stays (was not canceled)
  reduced <- paste0(names(counts[counts %% 2 == 1]), collapse="")
  return(reduced)
}

```

Experimental design

```

# 2^(7-2) means 5 variables in standard order + 2 "generated" using
# Box, Hunter, Hunter
X <- generate_initial_design(5)

generator1 <- "ABCD" # = F -> I = ABCDF
generator2 <- "ABDE" # = G -> I = ABDEG

X <- add_effect(X, generator1)
X <- add_effect(X, generator2)

colnames(X)[6:7] <- c("F", "G")

block1 <- "CE"
block2 <- "CF"

X <- add_effect(X, block1)
X <- add_effect(X, block2)

colnames(X)[8:9] <- c("BlockX", "BlockY")

# Change block variables from BlockX BlockY to 4 named blocks
df.design <- X %>%
  as.data.frame(row.names = F) %>%
  mutate(tmp = paste0(BlockX, BlockY)) %>%
  mutate(Block=recode_factor(tmp, '-1-1'="B1", '-11'="B2", '1-1'="B3", '11'="B4")) %>%
  select(-c(BlockX, BlockY, tmp)) %>%
  mutate(Run = 1:32) %>%
  relocate(Run)

```

Table 7: 32 Runs in standard order with blocks.

Run	A	B	C	D	E	F	G	Block	Run	A	B	C	D	E	F	G	Block
1	1	1	1	1	1	1	1	B4	17	1	1	1	1	-1	1	-1	B2
2	-1	1	1	1	1	-1	-1	B3	18	-1	1	1	1	-1	-1	1	B1
3	1	-1	1	1	1	-1	-1	B3	19	1	-1	1	1	-1	-1	1	B1
4	-1	-1	1	1	1	1	1	B4	20	-1	-1	1	1	-1	1	-1	B2
5	1	1	-1	1	1	-1	1	B2	21	1	1	-1	1	-1	-1	-1	B4
6	-1	1	-1	1	1	1	-1	B1	22	-1	1	-1	1	-1	1	1	B3
7	1	-1	-1	1	1	1	-1	B1	23	1	-1	-1	1	-1	1	1	B3
8	-1	-1	-1	1	1	-1	1	B2	24	-1	-1	-1	1	-1	-1	-1	B4
9	1	1	1	-1	1	-1	-1	B3	25	1	1	1	-1	-1	-1	1	B1
10	-1	1	1	-1	1	1	1	B4	26	-1	1	1	-1	-1	1	-1	B2
11	1	-1	1	-1	1	1	1	B4	27	1	-1	1	-1	-1	1	-1	B2
12	-1	-1	1	-1	1	-1	-1	B3	28	-1	-1	1	-1	-1	-1	1	B1
13	1	1	-1	-1	1	1	-1	B1	29	1	1	-1	-1	-1	1	1	B3
14	-1	1	-1	-1	1	-1	1	B2	30	-1	1	-1	-1	-1	-1	-1	B4
15	1	-1	-1	-1	1	-1	1	B2	31	1	-1	-1	-1	-1	-1	-1	B4
16	-1	-1	-1	-1	1	1	-1	B1	32	-1	-1	-1	-1	-1	1	1	B3

```
# Print table
t1 <- df.design[1:16, ]
t2 <- df.design[17:32,]
knitr::kable(list(t1, t2), booktabs = T, valign = 't', row.names = F,
            caption = '32 Runs in standard order with blocks.') %>%
  kable_classic()
```

```
set.seed(919)
block.order <- sample(unique(df.design$Block))

R <- matrix(nrow = 4, ncol = 8)
for (b in 1:4){
  R[b,] <- df.design %>%
    filter(Block == block.order[b]) %>%
    pull(Run) %>%
    sample()
}

row.names(R) <- block.order
colnames(R) <- 1:8

R %>%
  kable(valign = 'c', caption = 'Blocks and runs randomized. B2 corresponds to block 2. Run order is given fr
  kable_classic() %>%
  kable_styling(bootstrap_options = "striped")
```

Lenth's and Dong's methods

```
Lenths_method <- function(effects, alpha) {
  #effects: A vector or (preferably) a named list
  #alpha: level of test
```

Table 8: Blocks and runs randomized. B2 corresponds to block 2. Run order is given from left to right.

	1	2	3	4	5	6	7	8
B2	17	26	20	27	15	5	14	8
B4	24	1	10	11	4	31	30	21
B3	32	22	23	9	3	29	2	12
B1	19	6	18	28	16	25	13	7

```

abs_effects <- abs(effects)
g <- length(abs_effects)

s0 <- 1.5 * median(abs_effects)
PSE <- 1.5 * median(abs_effects[abs_effects < 2.5 * s0])

nu <- g / 3
gamma <- 0.5 * (1 - (1 - alpha)^(1/g))

critical <- qt(p = gamma, df = nu, lower.tail = F) * PSE
ix <- abs_effects > critical

return(list(sig.effects=effects[ix], critical=critical) )
}

# Dataset from slide 178 as gutcheck
# demo <- c(A=23, B=-5, C=1.5, AB=1.5, AC=10, BC=0, ABC=0.5)
# Lenthis_method(demo, 0.1)

Dongs_method <- function(effects, alpha) {
  g <- length(effects)
  s <- 1.5 * median(abs(effects))

  for (i in 1:2){
    ix <- abs(effects) <= 2.5*s
    m <- sum(ix)

    s_sq <- sum(effects[ix]^2) / m
    s <- sqrt(s_sq)
  }

  gamma = 0.5 * (1 - (1 - alpha)^(1/g))
  critical <- qt(p=gamma, df=m, lower.tail = F) * s
  sig_ix <- abs(effects) > critical
  return(list(sig.effects=effects[sig_ix], critical=critical) )
}

# demo <- c(A=23, B=-5, C=1.5, AB=1.5, AC=10, BC=0, ABC=0.5)
# Dongs_method(demo, 0.1)

```

Analysis and plotting of main dataset

```

tmp <- read.csv("../data/data.csv", row.names = 1)
colnames(tmp)[1:7] <- LETTERS[1:7]

df <- merge(x = tmp, y = df.design, by = LETTERS[1:7], all = TRUE, sort = FALSE)
contrasts(df$Block) <- matrix(c(1,0,0, 0,1,0, 0,0,1, 0,0,0), nrow = 4, byrow = T)

```

```

estimable_effects <- c("AB", "AC", "AD", "AE", "AF", "AG",
                      "BC", "BD", "BE", "BF", "BG",
                      "CD", "DE", "DF", "DG") # 15 estimable 2fi

for (ff in estimable_effects){
  df <- add_effect(df, ff)
}

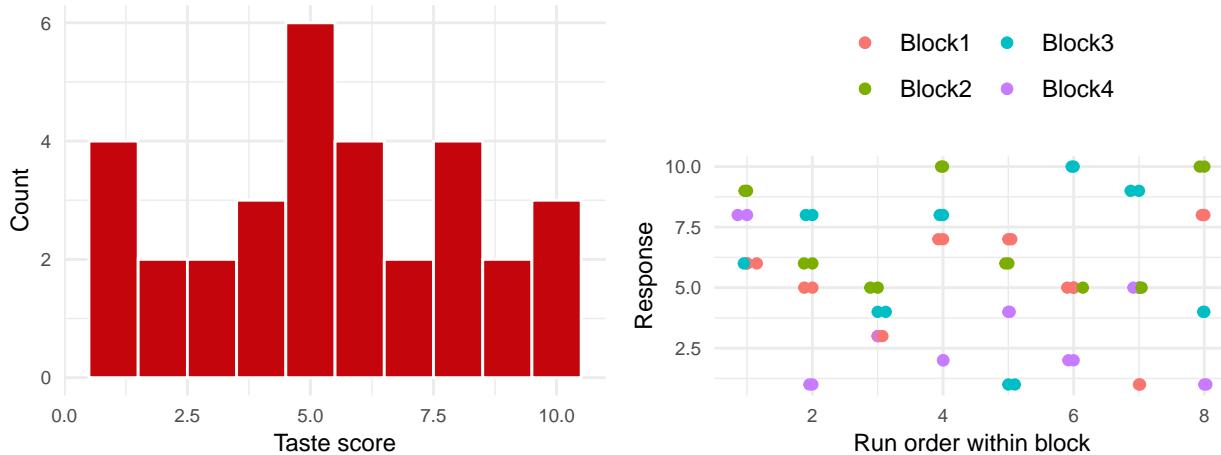
v <- c(LETTERS[1:7], estimable_effects, "Block", "Y")
df <- df %>% select(v)
order.df <- df$Y[R] %>% matrix(nrow=4, byrow = F) %>% t() %>% as.data.frame()
colnames(order.df) <- paste0("Block", c(4,2,3,1))
order.df$Order <- 1:8

p1 <- order.df %>%
  pivot_longer(cols = -Order, values_to = "Response", names_to = "Block") %>%
  ggplot(aes(x = Order, y = Response, color = Block)) +
  geom_point() +
  theme_minimal() +
  xlab("Run order within block") +
  geom_jitter(width = 0.15, height=0) +
  theme(legend.title = element_blank(),
        legend.text = element_text(size = fs),
        legend.position = "top",
        legend.background = element_rect(fill = "white", color="white")) +
  guides(color=guide_legend(ncol = 2,nrow=2, byrow = F)) +
  theme(text=element_text(size=fs))

p2 <- df %>%
  ggplot(aes(x = Y)) +
  geom_histogram(bins = 10, fill = "#C5050C", color="white") +
  theme(text=element_text(size=fs)) +
  theme_minimal() +
  xlab("Taste score") +
  ylab("Count") +
  theme(text=element_text(size=fs))

grid.arrange(grobs=list(p2,p1), ncol=2)

```



```

model.full <- lm(Y ~ . , data = df)
effects <- model.full$coefficients[-1]*2

```

Table 9: Estimated main effects; non-confounded two-factor interactions; block effects; and intercept.

Effect	Estimate	Effect	Estimate	Effect	Estimate	Effect	Estimate
A	3.50	AB	-0.25	BD	-0.25	DG	0.25
B	-0.62	AC	-0.63	BE	0.37	Block1	-0.25
C	-0.25	AD	0.12	BF	0.25	Block2	0.00
D	-2.00	AE	1.75	BG	-0.62	Block3	-0.25
E	-0.37	AF	1.12	CD	-0.38	Intercept	5.50
F	0.00	AG	0.25	DE	-0.25		
G	-1.88	BC	1.00	DF	-0.62		

```

tmp <- data.frame(Effect = c(names(effects), "Intercept"),
                  Estimate = c(round(unname(effects), 2), model.full$coefficients[1]))

t1 <- tmp[1:7, ]
t2 <- tmp[8:14,]
t3 <- tmp[15:21, ]
t4 <- tmp[22:26, ]

knitr::kable(list(t1,t2,t3,t4), booktabs = T, valign = 't', row.names = F,
            caption = 'Estimated main effects; non-confounded two-factor interactions; block effects; and intercept')
  
```



```

lenth <- Lenghts_method(effects, 0.10)
dong <- Dongs_method(effects, 0.10)
effects.df <- data.frame(Effect = names(effects),
                         Estimate = effects)

effects.df$Effect <- factor(effects.df$Effect, levels = effects.df$Effect[order(abs(effects.df$Estimate)), dec=1])

p2 <- effects.df %>%
  ggplot(aes(y = Effect, x = abs(Estimate))) +
  geom_col() +
  geom_vline(aes(xintercept = lenth$critical, color = "Lenth's Threshold")) +
  geom_vline(aes(xintercept = dong$critical, color = "Dong's Threshold")) +
  theme_minimal() +
  xlab(TeX("$| \hat{\theta} |")) +
  ylab("") +
  guides(fill=guide_legend(ncol=2)) +
  theme(legend.position = "top",
        legend.background = element_rect(fill = "white", color = "white"),
        legend.title = element_blank()) +
  theme(text=element_text(size=fs))

# Subset a data frame for box-cox
df.input <- df %>% select(c(A,B, C, D,E,F,G,Block, Y))
output <- get_best_lambda(df.input, 0, 1.5)

# Data wrangling for plotting
tmp <- data.frame(lambda = output$lambda.range, SSR = output$SSR)
p1 <- tmp %>%
  ggplot(aes(x = lambda, y = SSR)) +
  geom_point() +
  geom_line() +
  
```

```

theme_minimal() +
xlab(expression(lambda)) +
ylab("SSR") +
geom_vline(xintercept = 0.75, color = "blue") +
ggtitle(TeX("Finding best $\\lambda$"))+
theme(text=element_text(size=fs),
      plot.title = element_text(size = fs))

df.bc <- df.input %>%
  mutate(Y = transform_BoxCox(.Y, 0.75))
model.BC <- lm(Y ~ ., data = df.bc)
# anova(model.BC)

# [-1] drops the intercept
effects.BC <- 2*model.BC$coefficients[-1]
lenth.BC <- Lenghts_method(effects.BC, 0.1)
dong.BC <- Dongs_method(effects.BC, 0.1)

mod.BC.red <- lm(Y ~ A+D+G, data = df.bc)

p2 <- data.frame(fitted = mod.BC.red$fitted.values, residual=mod.BC.red$residuals) %>%
  ggplot(aes(x=fitted, y=residual)) +
  geom_point() +
  theme_minimal() +
  xlab("Fitted response") +
  ylab("Residual") +
  ggtitle("Residuals vs fitted of reduced Box-Cox model")+
  theme(text = element_text(size = fs)) +
  theme(plot.title = element_text(size = fs))

p3 <- wiscR::qqplot(mod.BC.red$residuals) + theme_minimal() +
  xlab("Theoretical quantile") +
  ylab("Sample quantile") +
  ggtitle("QQ Normal of residuals of reduced Box-Cox model") +
  theme(plot.title = element_text(size = fs))+ 
  theme(text = element_text(size = fs))

effects.BC.df <- data.frame(Effect = names(effects.BC),
                           Estimate = effects.BC)
effects.BC.df$Effect <- factor(effects.BC.df$Effect, levels = effects.BC.df$Effect[
  order(abs(effects.BC.df$Estimate), decreasing = T)])

p4 <- effects.BC.df %>%
  ggplot(aes(y = Effect, x = abs(Estimate))) +
  geom_col() +
  geom_vline(aes(xintercept = lenth.BC$critical, color = "Lenth's Threshold")) +
  geom_vline(aes(xintercept = dong.BC$critical, color = "Dong's Threshold")) +
  theme_minimal() +
  xlab(TeX("$| \hat{\theta} |$")) +
  ylab("") +
  theme(legend.position = c(0.8, 0.75),
        legend.text = element_text(size = 7),
        legend.background = element_rect(fill = "white", color = "white"),
        legend.title = element_blank()) +
  ggtitle("Significant effects after Box-Cox")

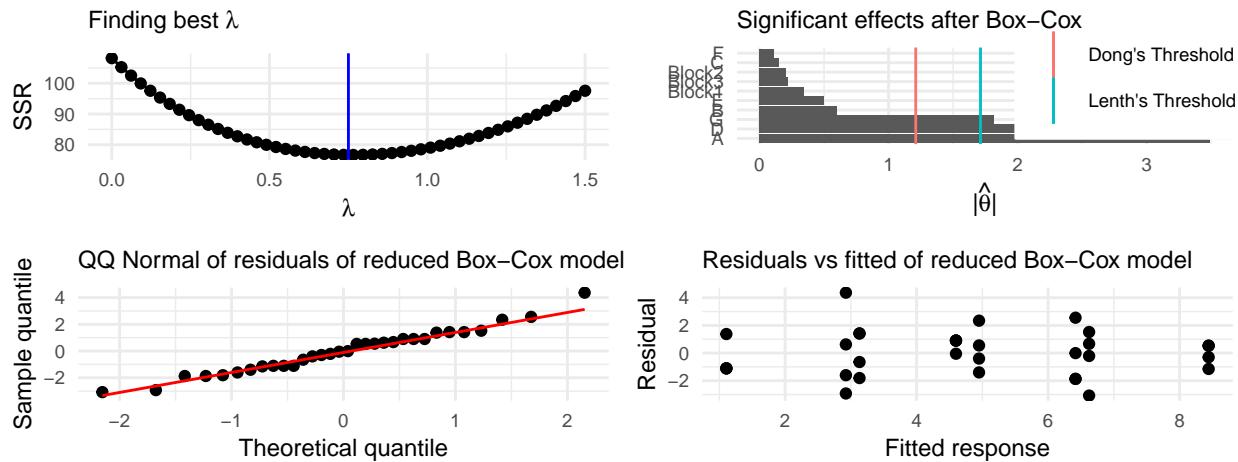
```

```

theme(text = element_text(size = fs),
      plot.title = element_text(size = fs))

grid.arrange(grobs=list(p1,p4, p3,p2), ncol = 2)

```



Analysis and plotting of “follow-up” dataset

```

X2 <- generate_initial_design(3) %>% as.data.frame()
row.names(X2) <- 1:8
colnames(X2) <- c("E", "F", "G")

set.seed(919)
ix <- sample(1:8)
X2 <- X2[ix, ]

write.csv(X2, "../design-matrices/design-followup.csv")
df.2 <- read.csv("../data/data-followup.csv", row.names = 1)

mod.2 <- lm(Y ~ .^3, data=df.2)

eff.2 <- 2 * mod.2$coefficients[-1]

bon.p <- 0.1 / 6
t <- qt(bon.p, df = 1, lower.tail = F)
sd <- (mod.2$coefficients[8]*2) %>% unname()

half.interval <- t*sd

lenth.2 <- Lenghts_method(eff.2, 0.1)
dong.2 <- Dongs_method(eff.2, 0.1)

out.2 <- get_best_lambda(df.2, lower = -1, upper = 0)
mod.2.BC <- lm(Y ~ ., out.2$df.transformed)
eff.2.BC <- mod.2.BC$coefficients[-1]*2

lenth.2.BC <- Lenghts_method(eff.2.BC, 0.1)
dong.2.BC <- Dongs_method(eff.2.BC, 0.1)

```

Table 10: Data for main experiment.

Data

Table 11: Data fro follow-up experiment.

	E	F	G	Y
3	1	-1	1	2
5	1	1	-1	6
4	-1	-1	1	2
8	-1	-1	-1	4
1	1	1	1	5
7	1	-1	-1	10
2	-1	1	1	2
6	-1	1	-1	3

Pictures



