

Statistics 850

A 2_{IV}^{7-2} Experiment

Coleman Breen

April 30, 2021

Summary

An unreplicated 2_{IV}^{7-2} quarter-fractional experiment was conducted to model the contribution of seven variables on the quality scores of frozen margaritas. Subsequent analysis found three main-effects to be significant under three hypothesis testing methods: assuming higher order interactions are negligible, Lenth's method, and Dong's method. One two-factor interaction was also found significant by all three of the methods.

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.

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 thoroughly 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. 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 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. It is used to distill tequila. Many margarita recipes call for agave syrup to add sweetness to the tart base of lime juice or lime concentrate. Some even use it as a sugar substitute in coffee or tea. 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 a 2^{7-2}_{IV} design, discussed more thoroughly in the following section.

Table 1: Factor levels of seven experimental variables.

Factor	Level I	Level II	Abbreviation
Base Flavor	Lime	Strawberry	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 (Cheap)	Cointreau (Expensive)	E
Tequila Brand	Altos (Cheap)	Casamigos (Expensive)	F
Tequila Color	Blanco	Reposado	G

Experimental Design

The following sections cover the design of the experiment:

Quarter-fractional Design Matrix

Our budget is 32 runs, which will be collected in four blocks of eight. Blocking is discussed more in the following subsection. With seven variables and 32 runs budgeted, we use a 2^{7-2}_{IV} 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 ; CG and EF . These six two-factor interactions in plain English, are any two of Lime Juice, Triple Sec, Tequila Brand, Tequila Color. Now there are $\binom{7}{2} - 6 = 21 - 6 = 15$ estimable two-factor interactions.

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 cover nearly all of the common ingredients in frozen margaritas. 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. Crucially, the order of the blocks is randomized as well as the runs within each block.

Protocol

The following protocol is used to create eight runs in one block. The recipe is found in table 3. 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 in case the reader would like to enjoy

Equipment

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

	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.

	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	1 cup
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

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. Because of the time it takes

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 nectar (if called for) 7. Salt 8. 12 ice cubes

Blend for two minutes on medium setting. Pour mixture into highball glass and place back in the freezer. Wash out pitcher and dry with towel. Repeat the aforementioned steps for the seven additional 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. The taster 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 then think the second drink is much better. 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. We will discuss this in depth in the model diagnostics section.

Analysis of Data

Full Model

We start by fitting a model with a global mean, 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 global mean, $A/2$ is half of the A main effect, $\dots, \epsilon \sim \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown. Let x_1 be 1 if A is in its high level and -1 if A is in its low level. Define x_2, \dots, x_7 similarly for the B, \dots, G factors. The $Block_k$ variable with $k \in \{1, 2, 3\}$ is 1 if the corresponding run is in block k and zero otherwise (i.e. runs in block 4 have zeros for each of the three blocking variables).

Then the model can be expressed as follows.

$$\begin{aligned} \mathbb{E}[Y] = & \eta + \frac{A}{2}x_1 + \frac{B}{2}x_2 + \frac{C}{2}x_3 + \frac{D}{2}x_4 + \frac{E}{2}x_5 + \frac{F}{2}x_6 + \frac{G}{2}x_7 \\ & + \frac{AB}{2}x_1x_2 + \frac{AC}{2}x_1x_3 + \frac{AD}{2}x_2x_4 + \frac{AE}{2}x_1x_5 + \frac{AF}{2}x_1x_6 + \frac{AG}{2}x_1x_7 + \frac{BC}{2}x_2x_3 \\ & + \frac{BD}{2}x_2x_4 + \frac{BE}{2}x_2x_5 + \frac{BF}{2}x_2x_6 + \frac{BG}{2}x_2x_7 + \frac{CD}{2}x_3x_4 + \frac{DE}{2}x_3x_5 + \frac{DF}{2}x_4x_5 + \frac{DG}{2}x_4x_7 \\ & + Block_1 + Block_2 + Block_3 + \epsilon \end{aligned}$$

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 is

Table 5: Estimated main and non-confounded two-factor interactions.

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

good news because it 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 zero. If this is

We cannot estimate σ^2 because our experiment is unreplicated. This prevents us from using traditional t-tests for the null hypothesis that an effect is zero. Lenth $\theta_1 = A, \theta_2 = B, \dots, \theta_{22} = DG$. The null hypotheses are $H_0^1 : \theta_1 = 0; \dots; H_0^{22} : \theta_{22} = 0$. The alternative hypotheses are $H_0^1 : \theta_1 \neq 0; \dots; H_0^{22} : \theta_{22} \neq 0$. Set the size of the test to $\alpha = 0.1$.

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 .

Lenth's Method

Lenth's method defines the following:

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

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

$$\nu = g/3 \text{ and } \gamma = 0.5(1 - (1 - \alpha)^{1/g})$$

The critical value for the test that rejects the null at level $\alpha = 0.1$ is $t_{\nu, \gamma} \times PSE$. Calculation

Critical value $c_{Lenth's} = 1.46$

Dong's Method

Dong's method is a less conservative improvement to Lenth's method. It 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.

Loh's Method

Adapt Siyu's code to work for this.

Summary of Significant Factors and interpretation

Check for errors fluctuating over run order? Or total score

Restricted Model

We now consider a restricted model

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

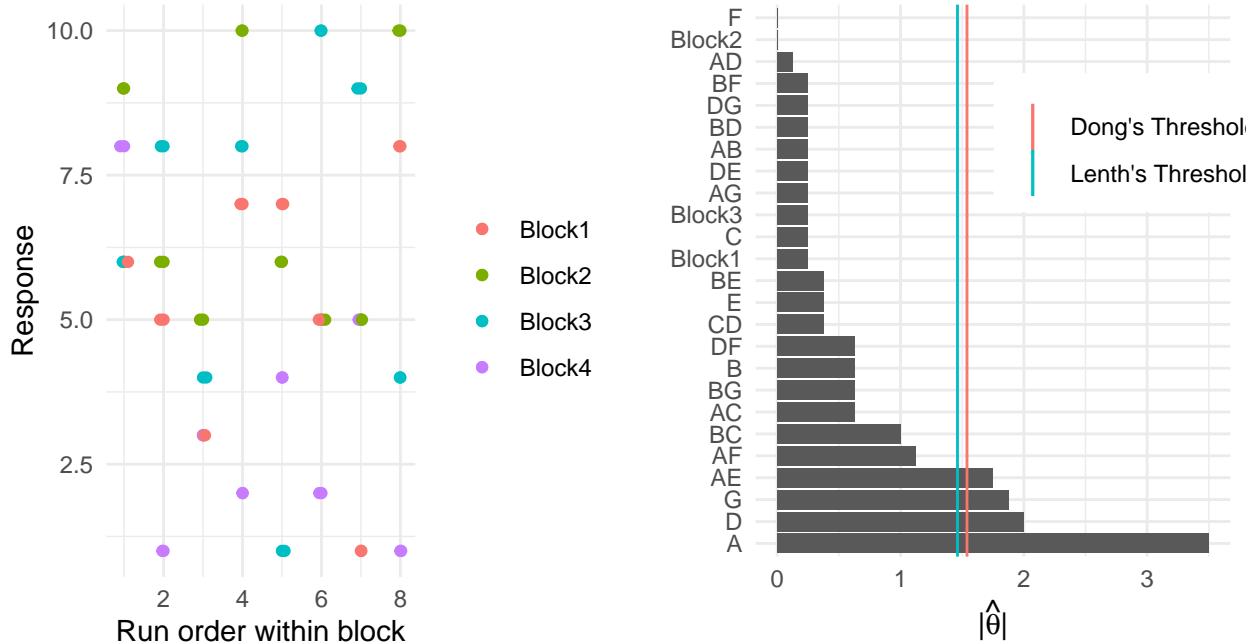


Figure 1: A, D, G, and AE effects are significant by both methods.

Because the contrasts are orthogonal, we can simply pluck the estimates of the regression coefficients from table 3 and divide by two:

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

Reduced Model Diagnostics

There are three fundamental assumptions made in our linear regression model. First, Slight departure from normality as seen in the QQ normal plot in figure 3.

Box-Cox

Perhaps there exists a transformation on Y that can accomplish a few things: (i) simplify the model by only including main-effects, (ii) make residuals more normally distributed, and (iii) improve the overall fit and interpretability of the model. We will experiment with a Box-Cox transform on the response, Y .

See slide 182

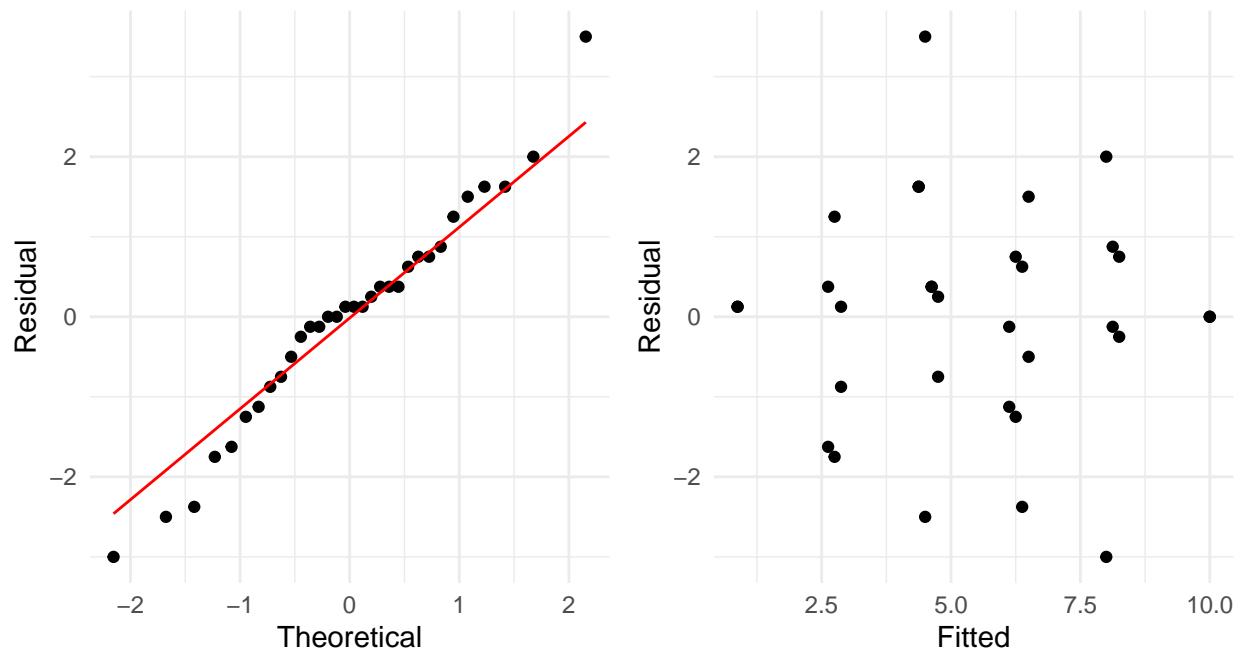
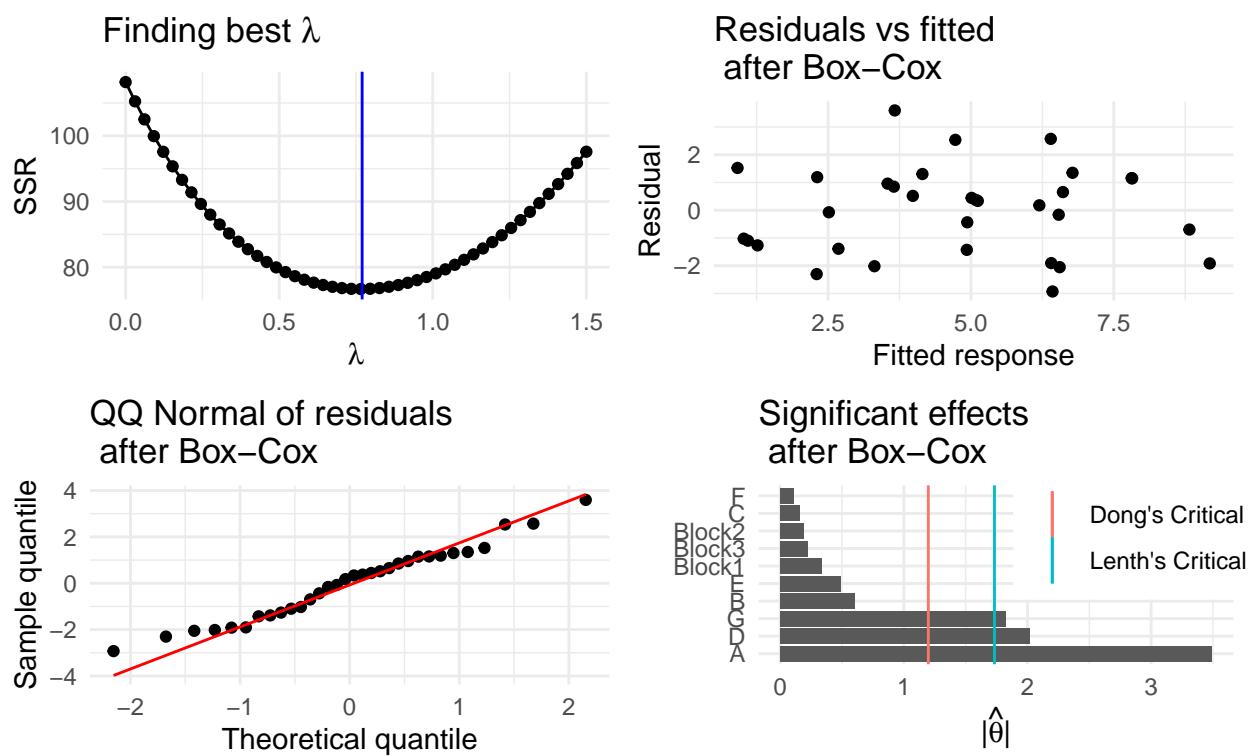


Figure 2: Some caption about residuals



Follow-up Study

After the initial analysis, there are eight runs left in the budget. In our primary analysis, we found just one significant two-factor interaction. By using a fractional design, we confounded six total two factor interactions, including a few of interest: $FG = 'TequilaBrand : TequilaColor'$; $EF = 'TripleSecBrand : TequilaBrand'$; and $EG = 'TripleSecBrand : TequilaColor'$.

Budget of one eighth additional runs. Could turn our quarter-fractional design into a three-eights experiment as laid out in Wu and Hamada (cite).

View this as a pilot to a larger follow-up experiment. Data cannot be incorporated into main data set. Why?

```
## $sig.effects
## named numeric(0)
##
## $critical
## [1] 13.3594
```

We find just one significant main effect, A, based on Dong's method. In a restricted parameter space,

Limitations

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

Tasters

First, the use of multiple, randomly-selected tasters could have been used and a summary statistic (average or median) could have been used. This was considered early on, but was difficult to coordinate given the ongoing pandemic.

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

Batch Sizes When making “frozen” margaritas in single-serving batches, the drinks are more of a slushy consistency than a true frozen margarita. This would have wasted some 85% of the raw materials and would have increased the cost of the experiment six-fold.

Conclusion

Three models were considered: an unrestricted model with seven main effects and 15 two-factor interactions; a reduced model with three main effects and one two-factor interaction; and a Box-Cox transformed model which had no significant effects.

Appendix

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

Recipes

Randomization of Blocks and Runs

Code

Helper functions

Hypothesis testing

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

Analysis and plotting of “follow-up” dataset

References

- [1] Box, Hunter, and Hunter.
- [2] Wu and Hamada. Experiments: Planning, Analysis, and Parameter Design Optimization. 2nd Edition.
- [3] Mead, Gilmour, and Mead. Statistical Principles for the Design of Experiments: Applications to Real Experiments.

Supplemental



