2022

Brewing the Best Cup

A 2^3 FACTORIAL STUDY

COREY ZHANG

EXECUTIVE SUMMARY

Background: Like many other coffee lovers, my boyfriend and I find ourselves in a constant pursuit for the best cup of coffee. We've conducted dozens of "homegrown" experiments over the years, where we've manipulated multiple factors several times in order to perfect our coffee making skills. The three most common factors we've manipulated have been the temperature of the water used, the type of water used, and the form of coffee purchased to try to enhance our coffee. However, given the random and undocumented approaches so far, we're left not knowing which factors actually relate to coffee taste, and which factors are not relevant.

Purpose: The purpose of this experiment was to test the effect of water components and type of coffee purchased on coffee taste ratings. I hypothesized that "Coffee Type" and "Water Temperature" would have a significant main effect on coffee taste, and that "Water Type" would not significantly affect our "Coffee Taste" ratings. I had no apriori hypotheses regarding any potential interaction effects between these three variables on "Coffee Taste" ratings, and so, those analyses are considered exploratory in nature.

Procedure: Using a 2^3 factorial design, I examined the main effects of and interactions between the three factors of interest. In each of the 24 experimental runs, my boyfriend and I rated a cup of morning coffee on a scale from 1-very bad to 10-very good. I was interested in examining each of our responses individually, as well as our average coffee ratings. Specifically, I tested the effects of water temperature (cold versus room), water type (filtered versus faucet), and type of coffee (whole bean versus ground) on three ratings of coffee taste.

Main Conclusions: Findings from this experiment support my hypothesis that "Coffee Type" had a significant effect on "Coffee Taste", such that the use of whole bean coffee lead to greater ratings of coffee taste than did pre-ground coffee. Additionally, some support for the effect of “Water Type” and “Water Temperature” on “Coffee Taste” was found, such that the use of cold, filtered water may lead to better ratings of coffee taste. No significant interactions between these three factors were observed. Taken together, these findings suggest that whole bean coffee may lead to better tasting coffee, and should be used to optimize coffee taste. Future experiments should aim to replicate these findings in bigger samples, and explore other potential factors that can lead to the best cup of coffee.

Background:

An article posted in USA Today suggests that over 80% of Americans drink coffee[1]. Any one in the coffee business knows that several factors go into making the perfect cup of coffee. In fact, almost everyone in my life above the age of 16 is an avid coffee drinker. And each of these individuals has their own preferences for their cup of coffee—some like it iced, while others like it served hot and immediately. Some like it bitter, some like it on the sweeter side. It is no surprise then that my boyfriend and I are also coffee lovers—there’s not much we look forward to more than a good cup of coffee.

And just like all of our family and friends, my boyfriend and I have become pretty big coffee snobs. We are both particular in how we prefer to have our coffee, and will go to great lengths to ensure a great cup of coffee. In a never-ending pursuit to make the “best” cup of coffee at home, we frequently play around with different procedures to find a way to make the perfect cup. Sometimes we create combinations that lead to coffee that tastes too bitter or not flavorful at all, and other times we find that one approach leads to the best coffee we’ve ever made. However, because we typically change multiple factors each time, it is hard to know which factors contribute to better tasting coffee. Thus, I decided to design an experiment with the hopes of identifying which factors actually have an effect on our coffee preferences

Method

The Design Plan:

As mentioned above, there are dozens of potential factors that may contribute to overall coffee quality. For this project, I decided to isolate three potentially-relevant factors that I believe to be related to coffee quality—type of water used, temperature of water used, and whether whole bean or pre-ground coffee was used. I chose these three factors because they were the most convenient factors to manipulate in the context of this project. Additionally, these were three salient factors that my boyfriend and I have typically played around with in previous attempts to enhance our coffee making abilities. Each of these factors has two levels:

Type of water: In order to test the effects of type of water, I will manipulate the use of faucet water or filtered water.

Temperature of water. For temperature of water used, I will manipulate the use of cold versus room-temperature water. For the temperature factor, the water was either set on the kitchen counter for exactly two hours (room temperature condition), or the water was refrigerated for exactly two hours (cold temperature condition).

Coffee type. Finally, I will also manipulate use the use of whole bean and pre-ground coffee. For the whole bean coffee conditions, I ground the coffee immediately before the coffee is brewed. Folgers 100% Columbian coffee will be used for both conditions to control for the effect of brand name coffee quality.

Response variables. I will have a total of three response variables. My boyfriend and I will both independently rated the taste of coffee and provide a "Coffee Taste" rating using a scale of 1-very bad to 10-very good. Additionally, I will look at the average of our Coffee Taste ratings as a third response variable.

Given that I have three factors with two levels each, this is a 2^3 factorial design experiment. Using the JMP Classical, Full Factorial Design option, I generated a replicated experimental design consisting of 24 randomized runs depicted in Figure 1.

The Experimental Procedure:

For 24 days, I made a pot of coffee every morning. For all runs, in order to reduce potential NOISE, I used the same coffee maker and filter, and controlled the amount of water (6 cups) and coffee (6 tbsp.) used. I also used the same brand of coffee in all 24 trials. I poured the same amount of coffee (1 cup) into the same mug 5 minutes after the coffee is completely done brewing.

Both my boyfriend and I both provided ratings every day for each of the 24 runs. In order to reduce the likelihood of other flavors interfering with the taste rating, we both provided these ratings before eating or drinking anything else. We used a form to rate the coffee tastes as 1-very bad to 10-very good. We completed these ratings separately each day. My boyfriend was blinded to the run condition; however, I was not blinded since I was the one making the coffee. After all 24 runs, I inputted all of the responses and created an average rating score.

Data Quality Issues: Following data collection, when preparing to run my analyses, I realized that there was a couple of nuisance factors that I should have documented and controlled for. The main issue was that the exact time of day we tasted the coffee varied. Although I tried to keep it constant throughout the experimental runs, the exact time of day was slightly variable in each run. Additionally, sometimes one of us had brushed our teeth before provided the ratings, which may have affected the taste of coffee.

Data Analytic Plan:

I used the JMP software to analyze the data from this experiment. I used the proposed procedure outlined in Table 6.8 from the to analyze this 2^3 factorial design. First, in order to see the effects of each of my three factors, I estimated the three factor effects. This step helps me identify the direction and strength of the relationship between each of the factors and the ratings of coffee quality. Next, given that this is a replicated design, I fit the full model (Analyze→Fit Model→Macros→Full Factorial). Analyses of variance were used to assess test for the significance of main effects and interactions. Next, I refined the model by removing nonsignificant variables from the full model. I then checked for model adequacy by analyzing the residuals of the model. Once a refined and adequate model was specified, I then interpreted the results. This process was conducted three times: once with my boyfriend's ratings as the response variable, the next using my own ratings, and finally, using the average of our ratings as the response variable. I hypothesized that "Coffee Type" and "Water Temperature" would have a significant main effect on coffee taste, and that "Water Type" would not significantly affect our "Coffee Taste" ratings. I had no apriori hypotheses regarding any potential interaction effects between these three variables on "Coffee Taste" ratings.

Results:

Boyfriend's Ratings:

I first analyzed the effects of each of the three factors on coffee taste ratings using JMP's "Fit Y by X". Results from this analysis showed a significant effect of "Coffee Type" on my boyfriends "Coffee Taste" ($F = 59.68$, $p < 0.0001$), with Whole Bean coffee having a significantly higher mean than Pre-Ground coffee (9.0 and 6.58, respectively). Results indicated no significant effect of "Water Type" ($F = 0.98$, $p = 0.33$) or "Water Temperature" ($F = 1.67$, $p = 0.21$) on boyfriend's "Coffee Taste" Rating. Next, I included all the factors in a Full Factorial, which provided estimates for each of the three factors and the three 2-way interactions. The ActualxPredicted Plot showed a good amount of overlap in scores, with a R2 of 0.89, indicating a good fitting model. I examined the Effect Summary and saw that “Coffee Type” ($p<0.001$), “Water Temp” ($p=0.005$), and “Water Type” ($p=0.02$) showed significant effects on Boyfriend’s “Coffee Taste” Rating. All four-interaction terms showed non-significant effects. I then reduced the model by removing non-significant terms in the model, which resulted in the loss of the all three 2-way interaction terms. The Effect Summary showed a significant effect of “Coffee Type” ($p<0.001$), “Water Temp” ($p=0.007$), and “Water Type” ($p=0.03$). The Residual by Row Plot shows a pretty random pattern of data points, and the new R2 of 0.84, suggests that this reduced model accounts for 84% of the variability in the Coffee Taste responses, and an overall good fitting model. All relevant JMP output can be found in Figure 2.

Own Ratings:

I followed the same procedure as above for my own Coffee Ratings. I analyzed the effects of each of the three factors on coffee taste ratings using JMP’s “Fit Y by X”. Results from this analysis showed a similar pattern of results. There was a significant effect of “Coffee Type” on my own “Coffee Taste” ($F = 320.40$, $p < 0.001$), with Whole Bean coffee having a significantly higher mean than Pre-Ground coffee (9.83 and 6.75, respectively). Results indicated no significant effect of “Water Type” ($F = 0.02$, $p = 0.90$) or “Water Temperature” ($F = 0.14$, $p = 0.71$) on my own “Coffee Taste” Rating. Next, I included all the factors in a Full Factorial, which provided estimates for each of the three factors and the three 2-way interactions. The ActualxPredicted Plot showed a good amount of overlap in scores, with a R2 of 0.96, indicating a good fitting model. I examined the Effect Summary and saw that only “Coffee Type” ($p<0.001$) was significant on my own rating of “Coffee Taste” ($p<0.001$). All other factors showed non-significant effects. I then reduced the model by removing non-significant terms in the model, starting with the least significant factor. In total, I removed the following terms: the 3-way interaction, "WaterType*CoffeeType" ($p=0.62$), "WaterTypexWaterTemp" ($p=0.14$), "WaterType" ($p=0.62$), "WaterTemp"($p=0.14$), "CoffeeType*WaterTemp" ($p=0.14$) and "WaterTemp" ($p=0.15$). This final Effect Summary then only consisted of a significant effect of "Coffee Type" on my own "Coffee Rating" ($p<0.001$). The new R2 of 0.94 suggests that this reduced model accounts for 94% of the variability in the Coffee Taste responses, and an overall good fitting model. All relevant JMP output can be found on Figure 3.

Average Ratings:

In order to have a more complete and parsimonious model, I created an average score of my and my boyfriend's ratings of "Coffee Taste". Similar to the above analyses, there was a significant effect of "Coffee Type" on the average "Coffee Taste" ($F = 164.10$, $p < 0.001$), with Whole Bean coffee having a significantly higher mean than Pre-Ground coffee (9.42 and 6.67, respectively). Results indicated no significant effect of "Water Type" ($F = 0.16$, $p = 0.69$) or "Water Temperature" ($F = 0.66$, $p = 0.43$) on my own "Coffee Taste" Rating. Next, I included all the factors in a Full Factorial, which provided estimates for each of the three factors and the three 2-way interactions. The ActualxPredicted Plot showed a good amount of overlap in scores, with a R2 of 0.94, indicating a good fitting model. I examined the Effect Summary and saw that "Coffee Type" ($p<0.001$) and "WaterTemp" ($p=0.01$) has a significant effect on average ratings of "Coffee Taste", while the interaction terms between WaterType*WaterTemp ($p=0.07$) and CoffeeType*WaterTemp ($p=0.07$) were approaching significance. All other factors showed non-significant effects. I then reduced the model by removing non-significant terms in the model, starting with the least significant factor. In total, I removed the following terms: the 3-way interactions ($p=1.0$), the "WaterTypexCoffeeType" term ($p=0.62$), "WaterTypexWaterTemp" ($p=0.06$), "WaterType" ($p=0.20$). The final reduced model consisted of the main effects of "WaterTemp"(p=0.02) and "CoffeeType" (p=<0.001). The new R2 of 0.91 suggests that this reduced model accounts for 91% of the variability in the Coffee Taste responses, and an overall good fitting model. All relevant JMP output can be found on Figure 4.

Discussion

This experiment aimed to examine multiple factors that influence coffee taste. Specifically, I assessed the effect of water temperature, water type, and type of coffee used on coffee taste ratings. This was conducted using a screening design of 24 randomized runs, and utilized ratings provided by my boyfriend and me. Results from this study provided strong support that "Coffee Type" had a significant influence on my, my boyfriend's and our average ratings of "Coffee Taste". Findings showed that, holding all other variables constant, using whole bean coffee was related to significantly greater ratings of coffee taste than when using pre-ground coffee. In relation to my boyfriend's ratings, it seems like there was also a significant main effect of "Water Temperature", with cold water resulting in statistically significant greater ratings of taste, and a significant main effect of "Water Type", with a preference for filtered water over faucet water in relation to "Coffee Taste". The effect of "Water Temperature" on "Coffee Taste" persisted when examining these effects with average ratings.

Overall, the findings from this study provide helpful information regarding what factors affect our coffee taste ratings. Specifically, it is useful to know that, regardless of water temperature and type, using whole bean coffee grounds substantially improves coffee taste, according to my boyfriend and me. This is helpful information as we can now be sure to use whole bean coffee moving forward. More investigation into the effects of the characteristics of water used and potential other factors is needed to determine other changes that can be made to enhance our coffee taste.

Limitations and Future Directions:

Of course, the results from this study should be considered in light of the study's limitation. First, one major limitation is that I was not blinded to the experimental run condition, since I was the one preparing coffee. While I tried to be objective as possible, I believe my ratings would have been more accurate had I been blinded to the experimental conditions. A second limitation relates to my response variable. We used a one-item survey of coffee tasting of one to ten. Looking at the raw data, I noticed that both raters only used only 6 to 10 on the scale. It would be helpful to have a finer range of ratings that can be used. In addition to an overall "Coffee Taste" rating, it could be useful to include other components or interpretations of coffee taste (e.g. aroma, likelihood to have a second cup) to provide more information. Third, only three relevant factors were manipulated in this experiment. Similarly, only two raters were utilized. Future experiments should aim to include additional factors that may be relevant to enhancing coffee taste and include a larger sample of independent raters.

Lessons Learned:

When applying these techniques to a real experiment, I had a newfound appreciation for JMP's Screening Design option to design randomized experimental runs. I was thinking of how to do this by hand, and became easily discouraged. However, it is clear how using this design approach can be used when designing a real experiment. Another striking lesson I learned through this process is the importance of precision when conducting experiments. While I tried to think of all potential confounding factors (e.g. mug type, amount of time before serving coffee), I inevitably failed to consider many potential factors. Because I did not expect any additional NOISE factors that were controllable, I did not think document any of these factors. As a result, I wasn't able to using blocking techniques to help account for NOISE factors. Through doing a real experiment, I realize the importance of systematically measuring these factor so that blocking can be used to minimize the variability transmitted to the response ratings. Similarly, a considerable amount of time should be used when thinking about how to measure variables of interest, which ultimately relates to how the data analysis will be conducted and the interpretation of these results.

References:

Fernau, K., & Republic, T. A. (2013, April 10). Coffee grinds fuel for the nation. Retrieved from https://www.usatoday.com/story/money/business/2013/04/09/coffee-mania/2069335/

Figure 1. *Screening Experimental Design*

Project.jmp

Project.jmp
Design 2x2x2 Factorial
Model
Evaluate Design
DOE Dialog

Columns (7/0)
Pattern
WaterType
CoffeeType
WaterTemp
Boyfriend's Rating
Own Rating
AverageRatings

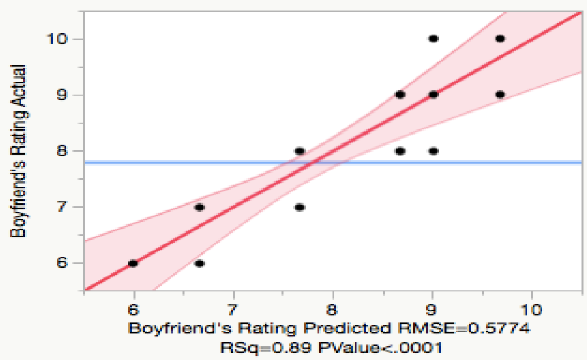
Rows
All rows 24
Selected 0
Excluded 0
Hidden 0
Labelled 0

	Pattern	WaterType	CoffeeType	WaterTemp	Boyfriend's Rating	Own Rating	AverageRatings
1	211	Filtered	Ground	Room	6	7	6.5
2	111	Faucet	Ground	Room	6	7	6.5
3	111	Faucet	Ground	Room	6	7	6.5
4	221	Filtered	Whole Bean	Room	8	9	8.5
5	122	Faucet	Whole Bean	Cold	8	9	8.5
6	112	Faucet	Ground	Cold	7	7	7
7	222	Filtered	Whole Bean	Cold	9	10	9.5
8	212	Filtered	Ground	Cold	8	7	7.5
9	111	Faucet	Ground	Room	6	6	6
10	112	Faucet	Ground	Cold	6	7	6.5
11	212	Filtered	Ground	Cold	7	7	7
12	212	Filtered	Ground	Cold	8	7	7.5
13	122	Faucet	Whole Bean	Cold	9	10	9.5
14	121	Faucet	Whole Bean	Room	8	10	9
15	221	Filtered	Whole Bean	Room	10	10	10
16	112	Faucet	Ground	Cold	7	7	7
17	222	Filtered	Whole Bean	Cold	10	10	10
18	122	Faucet	Whole Bean	Cold	9	10	9.5
19	211	Filtered	Ground	Room	6	6	6
20	121	Faucet	Whole Bean	Room	9	10	9.5
21	222	Filtered	Whole Bean	Cold	10	10	10
22	211	Filtered	Ground	Room	6	6	6
23	121	Faucet	Whole Bean	Room	9	10	9.5
24	221	Filtered	Whole Bean	Room	9	10	9.5

Figure 2. *Model Results for Boyfriend's Ratings*

Original

Actual by Predicted Plot



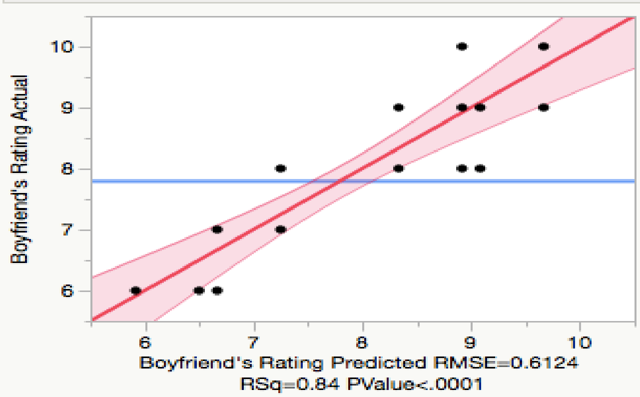
Effect Summary

Source	LogWorth	PValue
CoffeeType	7.713	0.00000
WaterTemp	2.237	0.00579
WaterType	1.604	0.02490
WaterType*WaterTemp	1.017	0.09616
CoffeeType*WaterTemp	1.017	0.09616
WaterType*CoffeeType	0.138	0.72829
WaterType*CoffeeType*WaterTemp	0.138	0.72829

Remove Add Edit FDR

Reduced:

Actual by Predicted Plot



Effect Summary

Source	LogWorth	PValue
CoffeeType	8.254	0.00000
WaterTemp	2.150	0.00708
WaterType	1.520	0.03018

Remove Add Edit Undo FDR

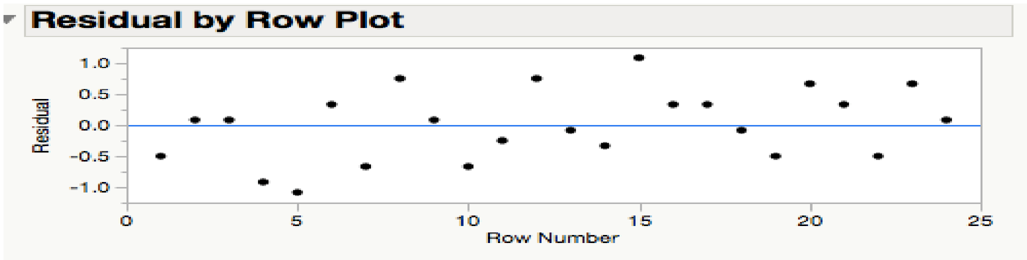
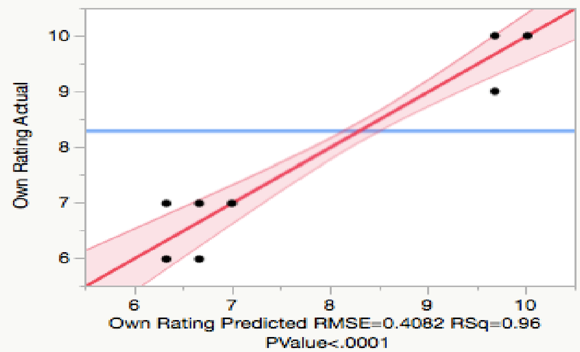

Residual by Row Plot
Residual
1.0
0.5
0.0
-0.5
-1.0
0
5
10
15
20
25
Row Number

*Figure 3. Model Results for **Own Ratings***

Original Model

Actual by Predicted Plot



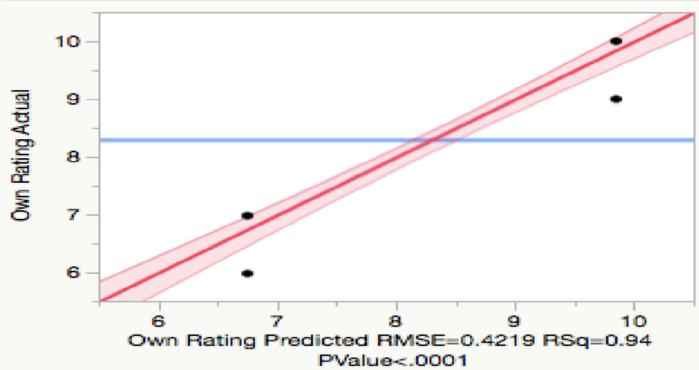
Effect Summary

Source	LogWorth	PValue
CoffeeType	11.499	0.00000
WaterTemp	0.815	0.15309
WaterType*WaterTemp	0.815	0.15309
CoffeeType*WaterTemp	0.815	0.15309
WaterType	0.205	0.62388 ^
WaterType*CoffeeType	0.205	0.62388
WaterType*CoffeeType*WaterTemp	0.205	0.62388

Remove Add Edit FDR ('^' denotes effects with containing effects above them)

Reduced Model

Actual by Predicted Plot



Effect Summary

Source	LogWorth	PValue
CoffeeType	13.874	0.00000

Remove Add Edit Undo FDR

Figure 3. Model Results for Average of Both Ratings