Analyzing Optimal combination of Milk Type, Temperature and Frothing to Boost Coffee Tasting Experience via Factorial Experiment

Tianxiao Ma

16 April, 2021

I. Description

With the growth of the vegan market, almond milk is becoming a trend as a milk substitute.However, taste is the golden criterion for determining whether almond milk can substitute the position of milk among coffee drinkers.This study is designed around the coffee taste, with the objective of comparing the effects of various combinations of milk type, milk frothing, and coffee temperature on taste. Three null hypotheses were made based on the most interested factor - milk type, including substituting almond milk doesn't affect rating while holding other factors constant; frothing almond milk has no impact on the taste, regardless of coffee temperature; the combination of frothed almond milk in hot coffee doesn't interact with taste compared to cold coffee without frothing cow milk. Ultimately, the study aims to discover the optimal combinations to enhance coffee tasting experience. The experimenter examined 8 combinations of experimental conditions in each of the 2 replicates to analyze the impact of various combinations of milk and temperature conditions on the rating of coffee taste.

This replicated 2^3 factorial experiment has one quantitative factor - coffee temperature, and two qualitative factors - milk type and foaming status.Details about factors used in the report and respective levels are displayed in Table I. Cups of coffee, which served as the experimental units, were subjected to various treatments. To avoid Systemic bias, the experiment restricted milk brands to Natrel 2% cow milk and Almond Breeze unsweetened original almond milk. Frothing time is controlled at 15 seconds. All 16 experimental runs took place at home. This ensures a feasibly consistent room temperature, milk pour method, coffee-to-water ratio, coffee brand(Ground Tim Hortons original blend), brewing method, and amount of coffee(100ml) and milk(50ml). Apparatus and example pictures of coffee under various conditions in this experiment have been introduced in Figures 1 below.

Table 1: Factors and respective levels

Factors	Level I	Level II
Milk Type(M)	Cow Milk(-1)	Almond Milk(+1)
Frothing(F)	Not froth(-1)	Froth(+1)
Coffee temperature(T)	10°C (-1)	50°C (+1)

Subjectivity greatly affects taste rating, resulting in unavoidable bias in the response. Section IV delves the subject further into the shortcomings as well as potential improvements. Given the different predilections amongst people, the author instructed participants to use water as a baseline (5 on a scale of 1-10) to minimize noises caused by unclear rating standards. Then participants were asked to give rates from three aspects: smoothness, blandness and aftertaste based on assigned coffee(See Appendix I for more details on rating standard).

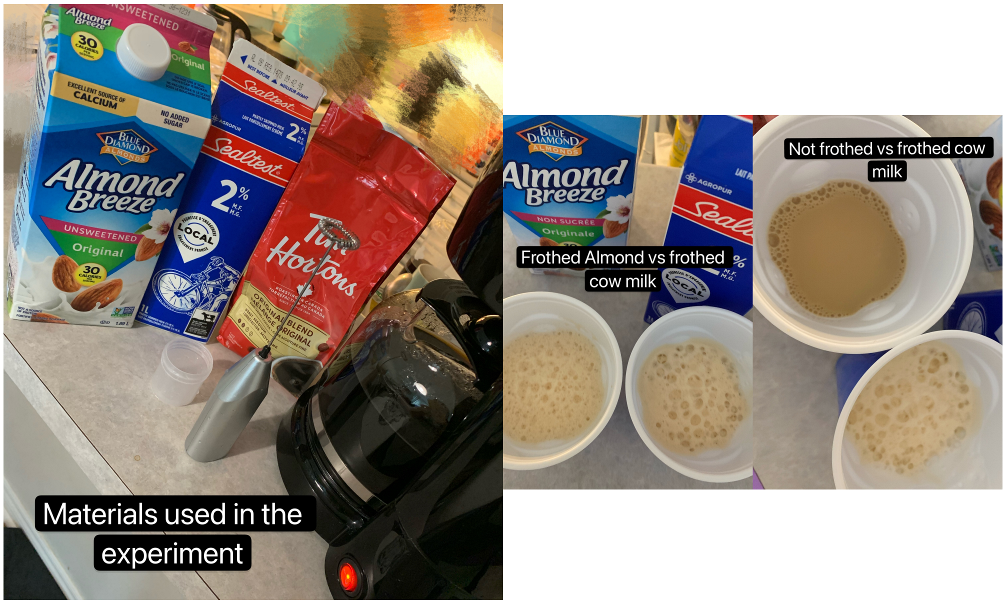

Figure 1: Example pictures of the experiment

The experiment took place on April 10th, 2021.Sixteen neighbors (7 households, including 1 household with 5 members, 1 three-people households and 3 two-people households and 2 single households) participated in the replicated experiment. All participants were acknowledged with relevant information before providing available time slots two days prior to the experiment. The experimenter brewed 16 cups of black coffee in the morning,then chilled half of them while keeping the other half warm. Experimental runs were completely randomized by an online number generator between 1-16(CalculatorSoup, n.d.) and thus prevented any selection bias in runs.After treatment, the experimenter delivered sufficient cups of coffee to assigned neighbors in predetermined time slots.Neighbors were asked to send ratings back within 20 mins based on smoothness,blandness and aftertaste,either through snapchat or call.In the following sections, the average rating of the three aspects are used as the response to perform further analysis.

II. Data Analysis

Table 2 introduces the results from running all 16 records in 2 replications. “run1” and “run2” represent the randomized run order.The variable “average_taste” records the response taste averaged over two genuinely replicated runs. Furthermore, taste records from the first and second runs were saved in “taste 1” and “taste 2,” respectively, while “diff_taste” recorded the differences between them.

Table 2: Design matrix of coffee taste design

run1	run2	M	F	T	taste1	taste2	diff_taste	average_taste
6	13	-1	-1	-1	5.67	7.00	-1.33	6.335
2	7	1	-1	-1	6.00	7.00	-1.00	6.500
3	15	-1	1	-1	6.00	5.67	0.33	5.835
14	10	1	1	-1	6.67	7.33	-0.66	7.000
8	4	-1	-1	1	2.00	7.00	-5.00	4.500
16	9	1	-1	1	6.00	5.00	1.00	5.500
5	12	-1	1	1	8.33	9.00	-0.67	8.665
11	1	1	1	1	10.00	8.67	1.33	9.335

Cube plots and Interaction plots

Figure 1: Cube plot for coffee taste

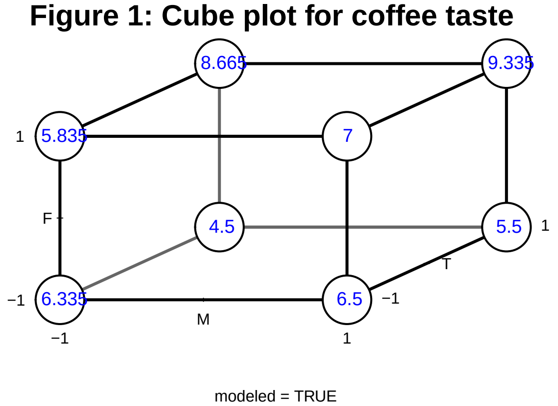

Figure 2: Interaction between Milk type(M) and Frothing(F)

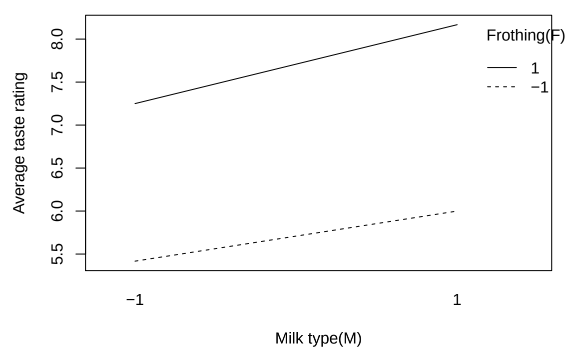

Figure 3: Interaction between Milk type(M) and Temperature(T)

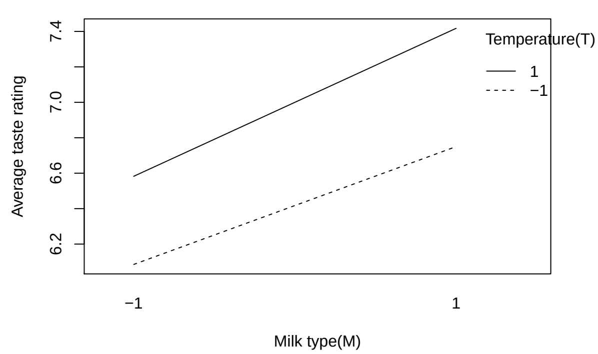

Figure 4: Interaction between Frothing(F) and Temperature(T)

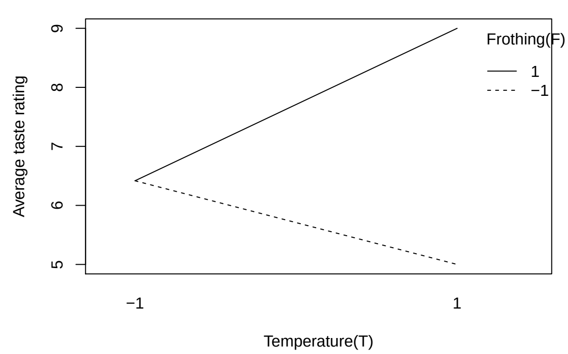

Figure 2 depicts a cube plot that intuitively visualizes the main effects and interactions. Calculations of each factorial effects were performed based on the cube plot in the next section. To directly visualize observed interactions, figure 3, 4 and 5 sequential display interactions between milk type and frothing, milk type and temperature, frothing and temperature.

Notice how lines in Figure 2 and 3 are approximately parallel, implying that the two-factor interaction between milk type and temperature, as well as the interaction between milk type and frothing have weak influences on coffee taste. On the other hand, Figure 4 depicts the shifting relationship between temperature and taste, with the direction of change dependent on frothing, and the intersection of two lines on the left indicating temperature-frothing interaction. When we use frothed milk, there is a positive relationship between temperature and taste, while for not using frothed milk the relationship turns to be negative.Determination of the interaction's significance was conducted in regression analysis parts.

Main effects and Interaction Effects

Table 3: Main Effects and Interactions

Factor	Effects
M	0.7500
F	2.0000
T	0.5825
MT	0.0850
MF	0.1675
TF	1.9900
MTF	-0.3325

Table III obtains manually calculated main effects and interactions from the cube plot, with the direction and magnitude appearing to be consistent with interaction plots.The procedures for calculating are as follows:

$$Main : Milk = (6.5 + 7 + 9.335 + 5.5 - 6.335 - 5.835 - 8.665 - 4.5)/4$$

$$Main : Frothing = (5.835 + 7 + 8.665 + 9.335 - 6.335 - 6.5 - 4.5 - 5.5)/4$$

$$Main : Temperature = (8.665 + 4.5 + 9.335 + 5.5 - 5.835 - 7 - 6.335 - 6.5)/4$$

$$Interation : MT = [[(9.335 + 5.5 - 8.665 - 4.5)/2] - [(7 + 6.5 - 5.835 - 6.335)/2]]/2$$

$$Interation : MF = [[(9.335 + 7 - 5.5 - 6.5)/2] - [(8.665 + 5.835 - 4.5 - 6.335)/2]/2$$

$$Interation : TF = [[(8.665 + 9.335 - 4.5 - 5.5)/2] - [(5.875 + 7 - 6.335 - 6.5)/2]]/2$$

$$Interation : MTF = [[(9.335 + 4.5 - 8.665 - 5.5)/2] - [(7 + 6.335 - 5.835 - 6.5)/2]/2$$

From the magnitude of each effect, we observe that the main effect of frothing, as well as the interaction between temperature and frothing has the largest effect on coffee taste. However, further investigation of significance needs to be analyzed using hypothesis tests in regression analysis.

Estimated variance & standard error of factorial effects

Suppose that the variance of each measurement is σ^2, then the estimated variance at each set of conditions is $s^2 = difference^2/2$.

The average of these single-degree-of-freedom estimates yields an

$$s^2 = [(-1.33^2)/2 + (1^2)/2 + (0.33^2)/2 + (-0.66^2)/2 + (-5^2)/2 + (1^2)/2 + (-0.67^2)/2 + (1.33^2)/2]/8 = 1.9707$$

with 8-degree-of-freedom pooled estimate.

$$Var(effect) = (1/8 + 1/8)s^2 = 1.9707/4 = 0.492675$$

The standard error of any factorial effect is

$$se(effect) = \sqrt{0.492675} = 0.702$$

When interpreting the results, the standard error of any factorial gives good intuition of to what extent an effect should be considered significant. In the following analysis, we consider effects more than twice the standard errors(1.4 points on a 1-10 scale) significant.

The multivariate linear regression analysis

Let $taste_i$ become the taste rating from the i_{th} run, with chosen coded factor names and levels as showed in Table 1. The multivariate linear model of this 2^3 factorial design is:

$$taste_i = \beta_0 + \beta_1 M_i + \beta_2 T_i + \beta_3 F_i + \beta_4 M_i T_i + \beta_5 M_i F_i + \beta_6 T_i F_i + \beta_7 M_i T_i F_i + \epsilon_i$$

where MT, TF, and MTF are two-factor interactions, MTF represents the three-factor interaction. In the following analysis, the factorial estimates should be twice the estimated least square coefficients in regression.

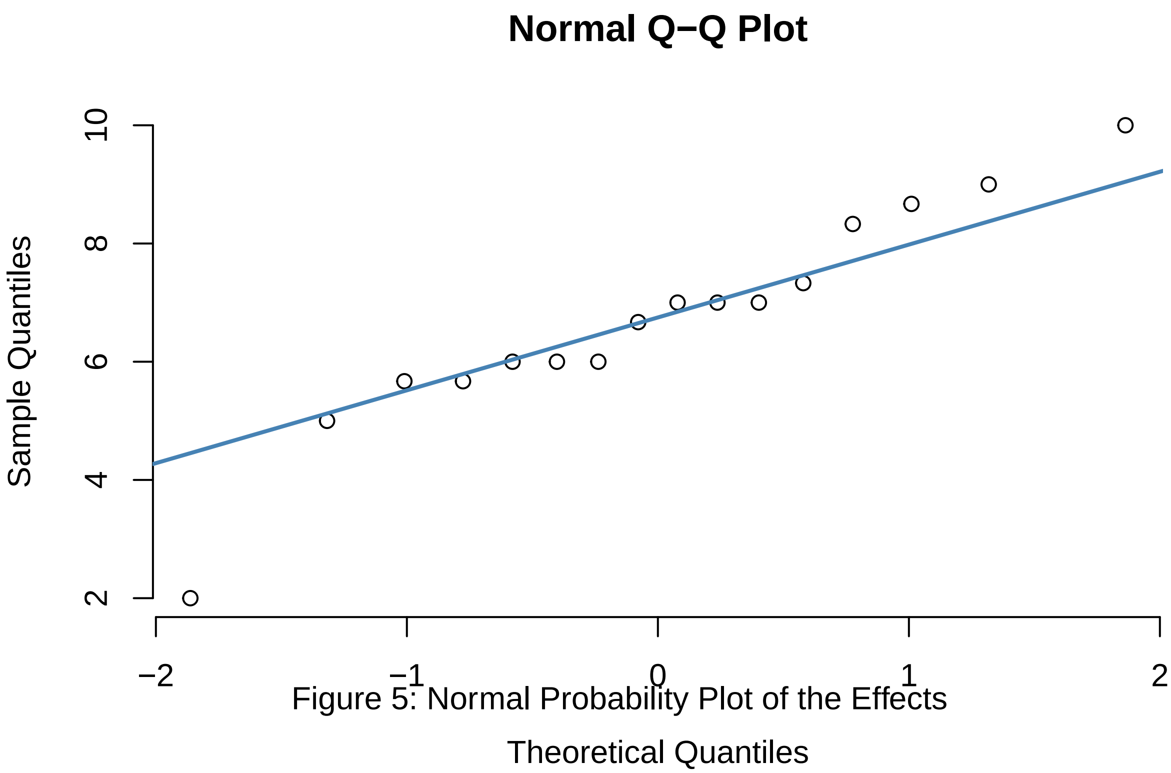

Figure 5: Normal Probability Plot of the Effects

Normality check and constant variance check were performed before regression analysis. Outliers are found in both tails of the normal Q-Q plot, indicating a left-skewed trend with heavy tails. However, because the sample size is small, we are unable to rule out such outliers. Given that the majority of the dots are clustered around a straight line, it is reasonable to assume that taste, as a response variable, is normally distributed and has constant variance for each experimental combination. Furthermore, because taste ratings were assigned to 16 different participants in the experimental runs, the independent assumption should be met. Thus we can proceed with the regression analysis with confidence. The regression yielded the following least squared estimates:

```
##               Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)  6.70875  0.3509541 19.1157500 5.810607e-08
## M            0.37500  0.3509541  1.0685159 3.164663e-01
## T            0.29125  0.3509541  0.8298807 4.306722e-01
## F            1.00000  0.3509541  2.8493758 2.149969e-02
## M:T          0.04250  0.3509541  0.1210985 9.065994e-01
## M:F          0.08375  0.3509541  0.2386352 8.173877e-01
## T:F          1.00000  0.3509541  2.8493758 2.149969e-02
## M:T:F       -0.16625  0.3509541 -0.4737087 6.483657e-01
```

Table 4: 95% confidence interval for each effect

	X2.5..	X97.5..
(Intercept)	11.798897	15.036103
M	-0.868603	2.368603
T	-1.036103	2.201103
F	0.381397	3.618603
M:T	-1.533603	1.703603
M:F	-1.451103	1.786103
T:F	0.381397	3.618603
M:T:F	-1.951103	1.286103

The estimated factorial effects are the twofold of least squares coefficients, and the intercept $\beta_0 = 6.70875$ is the average taste score in the runs. The estimated factorial effects are consistent with the manually calculated result stored in Table III, as expected. The equation had an adjusted R square value of 0.431, suggesting that, in the experiment, 43.1 % of the variability in taste could be explained by main effects frothing, interaction between coffee temperature and frothing, as well as the sample mean, assuming all other effects were indeed negligible. This analysis resulted in the model regression equation focused on factorial effects

$$\hat{taste}_i = 6.70875 + 0.75M_i + 0.5825T_i + 2F_i + 0.085M_iT_i + 0.1675M_iF_i + 2T_iF_i - 0.3325M_iT_iF_i$$

After obtaining p-values and confidence intervals for factorial effects adapted from the multivariate linear regression model, we can take hypothesis tests to examine the significance of each effect. Given the null hypothesis for each parameter is that

$$H_0 : \beta = 0$$

While the alternative hypothesis is

$$H_a \neq 0$$

We focus on the most interested $\beta_1, \beta_4, \beta_5, \beta_7$ which involves milk type first. If the p-value is smaller than the significance level, 0.05, then we have statistically strong evidence to reject the null hypothesis. Meanwhile, the 95% confidence interval for each effect could also help with determining significant effects by checking the exclusion of zero.

At the significance level $\alpha = 0.05$, unfortunately, none of the interested effects related to milk type turns to statistically impact the taste of coffee. The main effects milk type failed to obtained statistical significance on taste with P-value largely exceeds the 0.05 significance level(p-value 0.32). The 95% confidence interval for the not significant main effect of milk type is (-0.87, 2.37), which includes 0 as well as all the other confidence intervals for effects involved milk type. As a result, we failed to reject the null hypotheses for main effect milk type, that is, we have no evidence that the mean taste rating is different when using almond milk compared to cow milk, regardless of other factors.The only significant main effects was frothing, as well as the two-factor interaction between frothing and coffee temperature. Their p-value is smaller than significance level(identical p-value for both effects: 0.0215), as well as the 95% confidence interval excludes zero(identical 95% CI: [0.381397,3.618603]). While the two-way interaction between temperature and frothing has 95% confidence interval (-1.46,4.96) excludes 0.

However, even the interaction between frothing and temperature is statistically significant,each of their effect cannot be interpreted separately because of the large TF interaction as shown in Figure IV, as lines crossed on the left. The temperature effect lowers 0.2 units from taste when not frothing the milk, and 4 units higher when the milk is frothed. Noted that the only negative effect is the three-factor interaction among all effects, but only to a marginal extent. Practically, as taste was designed on a scale of 1-10, a change of almost 14% of the scale(as suggested by the estimated standard error of any factorial effect) could be considered substantial.The effect of milk over the ranges studied is to increase the coffee taste by 0.75 points, regardless of the tested level of frothing and coffee temperature.

Besides, the effects of temperature and frothing cannot be interpreted separately because of the large positive TF interaction as shown in Figure 4 as lines crossed in the left. The temperature effect shows a negative 0.2 points when the milk is not frothed but 4 points higher when the milk is frothed. Results adapted from the confidence interval of TF interaction excludes zero, which indicates the interaction between temperature and frothing is statistically significant.

III: Conclusions

This factorial experiment with 2 replications successfully investigated the effect of milk type, frothing, and coffee temperature on the taste of brewed coffee. Furthermore, it was carried out with control methods in place to minimize the effect of background variables, which was obvious due to the presence of significant effects.

Based on this experiment, we failed to reject all null hypotheses proposed around milk type, as substituting milk with almond milk don't impact the coffee taste significantly, no matter holding temperature and frothing constant or interacting with other factors. Practically, as taste was designed on a scale of 1-10, a change of almost 14% of the scale(as suggested by the estimated standard error of any factorial effect) could be considered substantial.The most interested main effect of milk type suggests that, using almond milk will improve the rating by 0.75 points compared to using cow milk on average, regardless of temperature and frothing. The effect is not statistically significant, but still gives good intuition that almond milk might be used more in the future as the market expands, if we consider a 7.5% increase significant. The only statistically significant interaction is between coffee temperature and frothing, with a shifting relationship between temperature and taste depending on frothing. When we use frothed milk, there is a positive relationship between temperature and taste, while for not using frothed milk the relationship turns to be negative.

The experiment showed that, on average, the expected taste rating would: Increase by 2 points on a taste scale of 1-10 when using frothed milk in 50°C coffee compared to 10°C coffee, regardless of milk type, while using non-frothed milk will lower 0.2 points of rating in hot coffee compared to cold ones.

Eventually, the experiment provides theoretical optimized results for coffee tasters: Using frothed cow milk in hot coffee will vastly enhance the coffee tasting experience. On the other hand, the experiment also provides the combination to avoid: almond milk in frothed hot coffee. The combination may marginally disappoint the drinking experience by around 0.33 points on a 1-10 scale.

IV: Discussion and Limitations

The experiment is restricted geographically and timely under the setting of pandemic, further studies should consider allocating more participants and replicate the experiment more times. As for the experiment design itself, the largest source of bias comes from the response variable, taste. 16 participants have very different tastes as well as unique opinions towards coffee, not limited to dairy options, temperature and whether to froth the milk, but also bean types, brewing methods etc. To obtain a relatively fair response, the experimenter provided a rating standard (use water as a baseline of 5, rate on smoothness, blandness and aftertaste). However, noisy unobservables can not be eliminated such as participants' mood, weather, specific preference on dairy/temperature/frothing and so on. Future studies should also refine the considerably subjective response to a more specific aspect or could be run by involving multiple experienced coffee taste testers in the experimentation process. This could help relate, to some capacity, the effect of each factor on taste.

Furthermore, future studies should consider including additional controls, such as background variables, to reduce noise. It may incur a higher cost and necessitate the inclusion of more effects in the experiment, but it effectively eliminates experimental noise.

Due to the small sample size, this experiment failed to exclude outliers that influence the accuracy of factorial effect estimates as variance increases. Additional runs should be performed to validate the validity of the regression models, as the ones outlined in this report could only be said with certainty of the two replicated experiments, but not necessarily of any larger counterparts. This could also reduce the possibility of outliers for invalid model assumptions such as the violation of normality, non-constant variance. Future studies can consider more design methods such as blocked factorial design to improve the accuracy of factorial estimates.

References

Furey, Edward “Random Number Generator”; CalculatorSoup, https://www.calculatorsoup.com - Online Calculators

Appendix

I. Coffee Rating Standard

Coffee taste scores revolve around three elements: Smoothness, Blandness, and Aftertaste. All ratings are given on a scale of 1 to 10, with 1 being the worst and 10 being the best.

Smoothness focuses on the mouthfeel, a higher rating of smoothness indicates that the coffee has a velvety smoothness, which provides a higher level of experience of the moist and smoothness in the mouth.

Blandness describes the taste of coffee, a low rating indicates the coffee is tasteless, watery and pale. A high score indicates the coffee has rich and thick taste.

Aftertaste refers to the flavour that lingers in the mouth after the coffee has been swallowed. A higher aftertaste rating indicates that the enjoyable taste is retained in the mouth for the appropriate amount of time.