Cookie Data Project Template

Garrett Peters

1. Description

The goal of this project is to determine what variables go into what makes the best tasting cookie. Since the definition of what a good tasting cookie is subjective, we decided to measure this by taking the diameter of the cookie. We determined that the smaller the diameter, the more the cookie rose, and therefore was fluffier and more delicious. This study would be optimal for anyone who is trying to figure out how to make their homemade cookies taste better.

The population of this study is all cookies Prof. Michelle Paret will bake in her oven. The sample that will be looking at is 24 cookies that were made using the same oven, on the same day. Our sample has 8 total combinations between our 3 key explanatory variables. This is because each of our variables has two levels. This means that we measured two types of fats, two amounts of flour, and two different lengths of time for the dough to chill before baking. The fourth explanatory variable in this study is how the cookies were baked. The 24 cookies were cooked in 4 batches of 6 cookies in each batch. Since there are 8 possible combinations and 24 total cookies, this would mean that each combination was replicated 3 times. This is a randomized experiment therefore we can determine which combination gives us our best cookie.

1.1 Research Questions

Question 1) What explanatory variables are significant, and affect the taste of the cookie the best? Are there variables that it doesn't matter which type you use?

Question 2) What is the optimal blend of flour quantity, chilling duration, and choice of fat to achieve the best tasting cookie?

1.2 Variables

Variable	Type	Description	Levels and Ranges
Fat	Categorical	Type of fat used inside dough	2 levels – Butter and Margarine
Flour	Numeric	Amount of flour measured in cups used	2 levels – 1 cup and 1.75 cups
ChillTime	Numeric	Duration in minutes inside refrigerator before baking	2 levels – 360 mins and 135 mins
Diameter	Numeric	Longest distance from one side of cookie to the other	5.4 – 10 centimeters
Batch	Categorical	6 cookies cooked at 4 separate times	A,B,C,D

2. Exploratory Data Analysis

Table 1: five-number summary, mean, standard deviation, and sample size (n)

min <dbl>	Q1 <dbl>	median <dbl>	Q3 <dbl>	max <dbl>	mean <dbl>	sd <dbl>	n <int>	missing <int>
5.4	5.675	7.15	8.9	10	7.4	1.823041	24	0

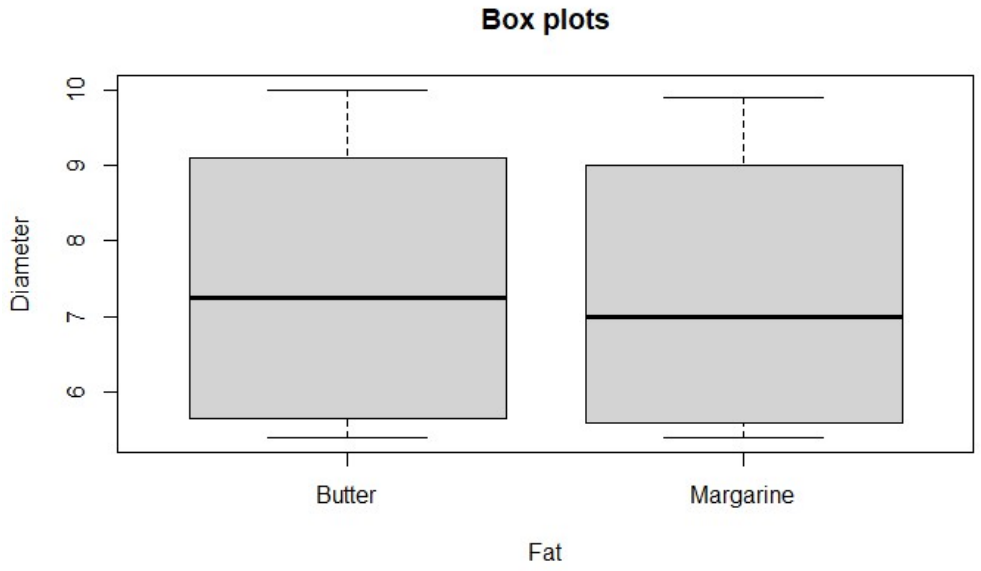
This table shows us that there are 24 total cookies, and the smallest cookie diameter is 5.4 centimeters, and the max is 10 centimeters. The median cookie diameter is 7.15 centimeters, but the average is +*slightly bigger at 7.4 centimeters. The standard number of centimeters away from the mean diameter is 1.82 centimeters.

Table 2: Histogram of cookie Diameter

This histogram shows the distribution of 24 cookies. Here you can see that this is close to a non-symmetric bi-modal distribution, this means that there is something going on that caused the cookies to be so drastic from each other. You can also tell this because this graph does not have any cookies that fit our mean. There is no bar in the 7-8 range.

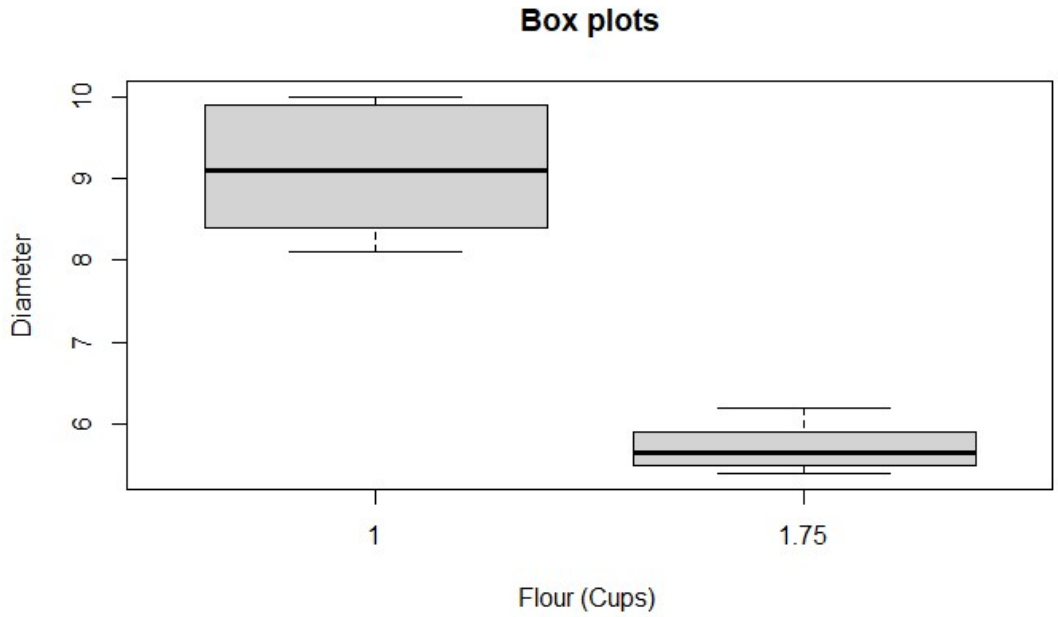
There are 10 cookies with a diameter between 5 and 6 centimeters, 2 cookies between 6 and 7 centimeters, 6 cookies in the 8-to-9-centimeter range, and 6 cookies with a diameter between 9 and 10 centimeters.

Table 3: Box plots of cookie diameter group by type of fat



Here you can see that there are two box plots showing the distribution of cookie diameter grouped by the type of fat being used. Looking at it there does not seem to be a significant difference in the look of both box plots. Both have very similar medians and both slightly right skewed. You can tell they are right skewed because both medians are slightly closer to both the first quartile and minimum value.

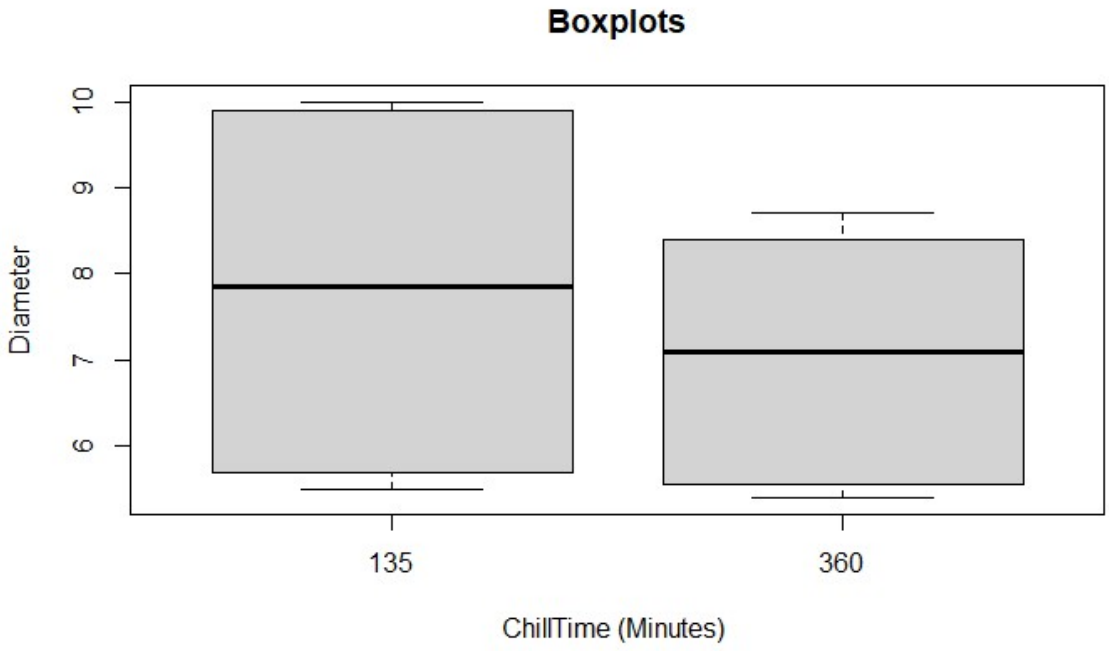
Table 4: Box plots of cookie diameter grouped by amount of flour



Here we have two box plots that show the distribution of cookie diameter grouped by the amount of flour used. We see here that there seems to be a significant difference between these two types. The median diameter of a cookie that was baked using 1 cup of flour is slightly above 9, whereas the median diameter of a cookie that was baked using 1.75 cups of flour is slightly below 6.

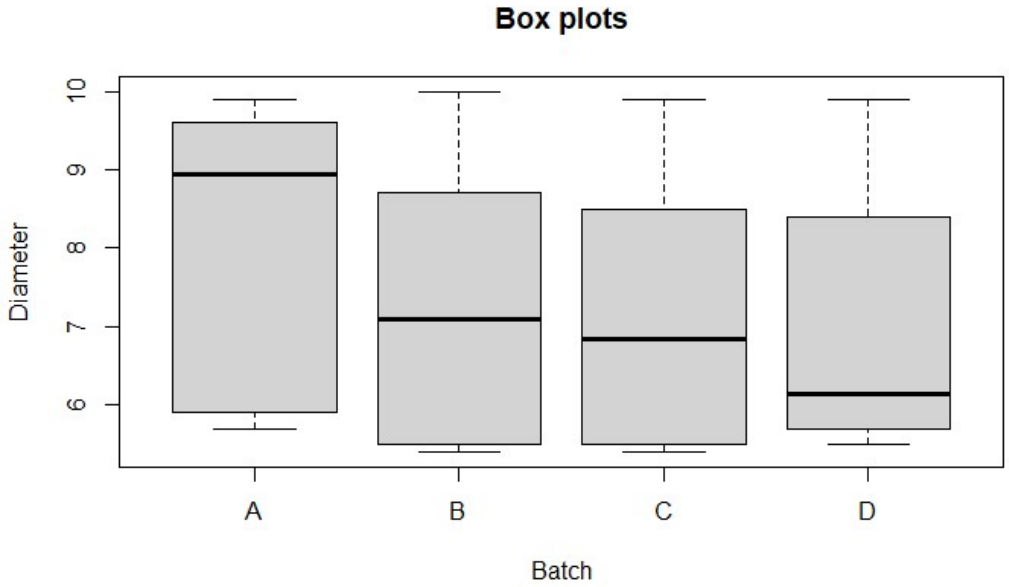
The box plots do not share any values. The range of cookie diameter of 1 cup also seems to be a lot bigger than the range for 1.75 cups of flour.

Table 5: Box plots of cookie diameter group by different chill times



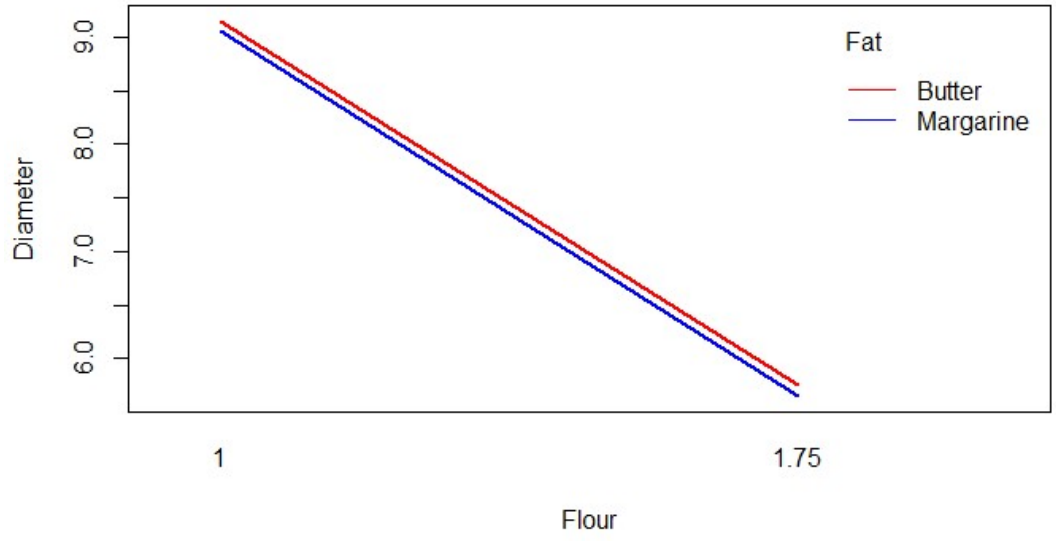
Here we have two box plots that show the distribution of cookie diameter grouped by the amount of time the dough was chilled. The two times have pretty similar medians, and both are pretty normally distributed. The only thing I see that is worth noting here is that the longer the dough was chilled the more consistent the diameter. You can see this by looking at the end tails of the two boxplots. The dough chilled for 360 minutes has a tighter box plot.

Table 6: Box plots of cookie diameter group by batch



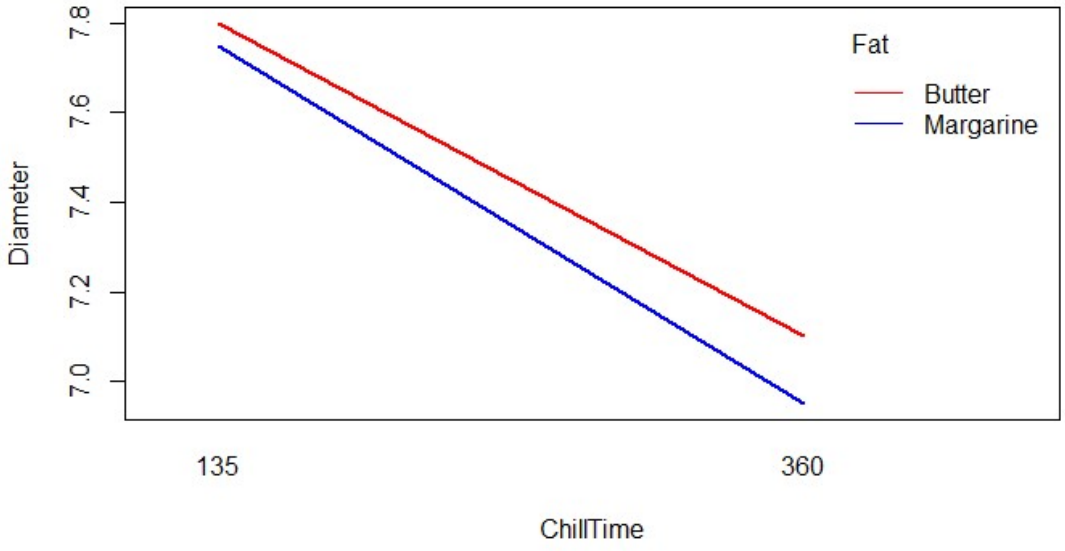
Here we see four box plots showing the cookie diameter distribution grouped by which batch they were baked. This is interesting because even though each batch has a very similar range and spread, the medians are all pretty different. Each batch on paper has the same 6 cookies, so all these box plots should look the same. However, there is a pretty big difference between Batch A and Batch D. This tells me that even though the oven temp is said to be constant for all four batches, it might be advantages to wait to bake the cookies later, maybe

Table 7: Interaction plot of Diameter by Fat and Flour



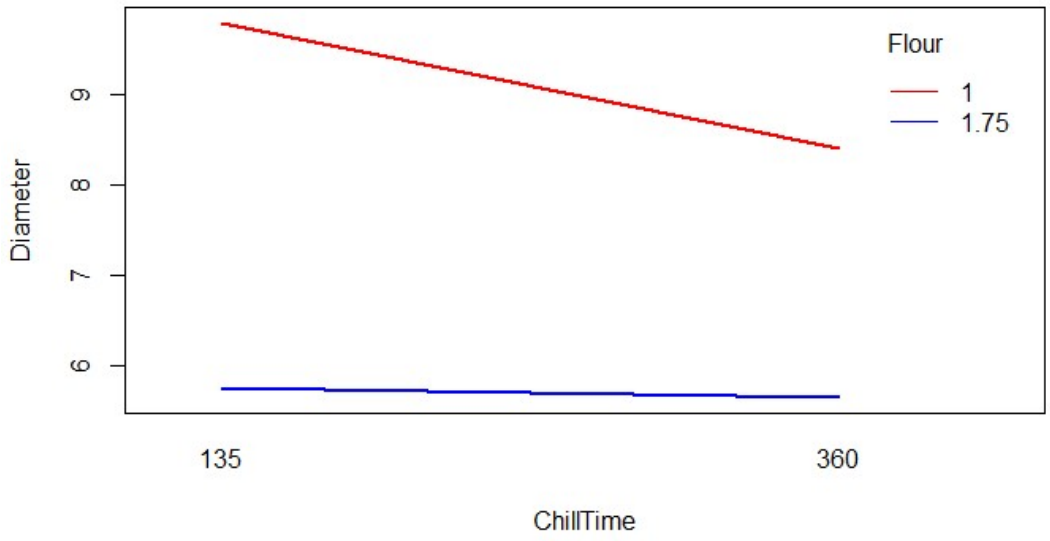
Here we have a plot of cookie diameter looking for an interaction between the type of fat used, and the amount of flour used. Butter is being shown by the red line, and margarine is being shown by the blue line, since these two lines do not cross in this plot, it tells us that these two variables are not related and do not affect each other. This tells us that no matter the type of fat, cookie diameter is still going to be similar for each amount of flour used. Although in for both amount of flour, margarine seems to produce the smaller diameter cookie. This plot once again confirms with table 4 that 1.75 cups of flour produce a smaller cookie diameter.

Table 8: Interaction plot of Diameter by Fat and ChillTime



Here we have a plot of cookie diameter looking for an interaction between the type of fat used, and the amount of time the dough was chilled. Butter is shown by the red line, and margarine is shown by the blue line, since these two lines do not cross in this plot, it tells us that these two variables are not related and do not affect each other. This tells us that no matter the type of fat used, the diameter of cookie is going to be similar for each duration of chill time. However, margarine produces the smaller diameter cookie for duration. This table also confirms with table 5 that the 360-minute duration to chill produces a smaller diameter cookie.

Table 9: Interaction plot of Diameter by Flour and ChillTime



Here we have a plot of cookie diameter looking for an interaction between duration of chill time, and the amount of flour used. 1 cup flour is being shown by the red line, and 1.75 cups of flour is being shown by the blue line, since these two lines do not cross in this plot, it tells us that these two variables are not related and do not affect each other. This plot confirms with table 4 that 1.75 cups of flour produce the smallest diameter cookies, and that generally chill time does not have much difference for 1.75 cups of flour. It does also tell us that if you do use 1 cup of flour, there is a pretty big difference in diameter, and favors chilling the dough longer.

3. Statistical Analysis

Analysis of Variance (ANOVA) was used to analyze the chocolate chip cookie data to evaluate the effect of fat, flour, and chill time on cookie diameter. The analysis included all two-way interaction terms, in addition to the batch variable since six cookies were baked at a time in a single batch. See the Appendix for the ANOVA results, which were conducted using Minitab Statistical Software.

To reduce the ANOVA model and identify the significant factors, Akaike's Information Criterion (AIC) was used. Per Table 11, the AIC model reduction revealed a significant interaction between the flour and chill time variables. Therefore, these variables should not be assessed independently. Rather, both variables should be evaluated simultaneously in terms of their effect on cookie diameter.

None of the other two-way interaction terms, nor the fat and batch variables were significant, and thus are not shown in the final model provided in Table 11. Therefore, there is not sufficient evidence of a difference in population mean diameter for cookies made using butter versus margarine. The batch variable was also not statistically significant, and thus was excluded from the final model shown in Table 11.

Variable	*P-value*
Flour	0.000
Chill Time	0.000
Flour*Chill Time	0.000

Table 10: Final model summary for cookie diameter ANOVA

Per a residuals analysis, all model assumptions were satisfied for the final model, including normality, constant variance, and independence. See the Appendix for details. The residuals analysis also revealed there were no extreme outliers that needed to be investigated.

The adjusted R-squared for the analysis, shown in the Appendix, was extremely strong at 98.24%. Thus, flour and chill time account for nearly all the variability in cookie diameter. Also, the lack-of-fit p-value is not significant, which indicates the recommended ANOVA model is valid.

To then determine which flour amount and chill time minimizes cookie diameter, Tukey multiple comparisons were used with an alpha level of 0.05, as shown in Table 12. The comparisons analysis reveals that for 1.75 cups of flour, there is no significant difference in population mean diameter for dough chilled for either 135 or 360 minutes, with predicted cookie diameters of 5.75 and 5.65 cm, respectively.

Flour*ChillTime	N	Mean	Grouping		
1.00 135	6	9.80	A		
1.00 360	6	8.40		B	
1.75 135	6	5.75			C
1.75 360	6	5.65			C

Table 11: Tukey comparison results to determine optimal factor settings

However, if only 1 cup of flour is used, then the cookie diameters are significantly larger (i.e., less desirable) and there is a significant difference depending on chill time. Specifically, if the dough is chilled for 360 minutes, the predicted diameter is 8.4 cm. Or, if the dough is chilled for 135 minutes, the predicted cookie diameter is the largest at 9.8 cm, per Figure 13. This concludes the statistical analysis for the cookie diameter study.

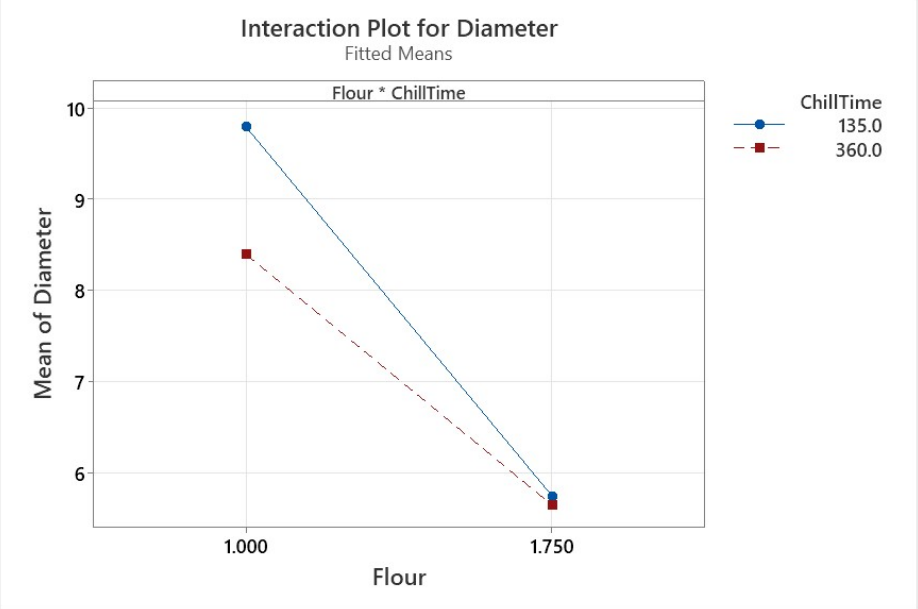


Figure 13: Predicted cookie diameters for flour and chill time

4. Recommendations

Question 1) What explanatory variables are significant, and affect the taste of the cookie the best? Are there variables that it doesn't matter which type you use?

- The explanatory variables that are significant are ChillTime and Flour per table 10, we determined based on table 7 and table 8 that fat is not significant.

Question 2) What is the optimal blend of flour quantity, chilling duration, and choice of fat to achieve the best tasting cookie?

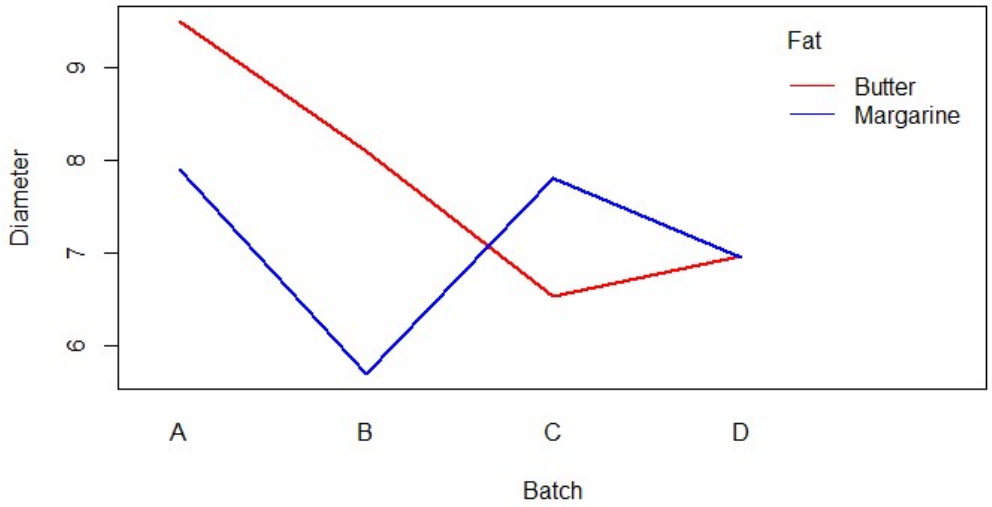
-By looking at table 11, it shows us that for the optimal cookie to use 1.75 cups of flour, when using 1.75 cups of flour the difference between type of fat and ChillTime is insignificant.

5. Resources

For resources for general linear models: https://online.stat.psu.edu/stat504/lesson/6/6.1
For resources for Tukey comparisons: https://online.stat.psu.edu/stat500/book/export/html/610
For more information on r studio: https://online.stat.psu.edu/statprogram/tutorials/statistical-software/r

6. Additional Considerations

Table 12: Interaction Plot of Diameter by Fat and Batch



Here we have a plot of cookie diameter looking for an interaction between the type of fat used, and the batch number. The Butter is being shown by the red line, and margarine is being shown by the blue line, these lines do cross multiple times throughout the graph. This tells us that these two variables affect each other in some sort of way.

The batch variable refers to what order the cookies were placed into the oven, A being first, D being last. We were not told where batches were placed while the other batches were being cooked. This graph shows me that cookies baked with margarine produced overall smaller cookies than with butter, and to get the smallest cookie, to let the dough sit and settle in open air for the amount of time they are supposed to be baked, and then place them into the oven to bake.

If you let dough that is used with butter, sit and settle for two bake times, they will actually start produce smaller cookies compared margarine cookies that have settled for the same amount of time. Lastly, after 3 batches, margarine and butter actually produce almost identical diameters.