STATISTICAL ANALYSIS OF CUPCAKE HEIGHT WHEN SUBSTITUTING FOR BUTTERMILK IN RED VELVET CUPCAKES

Evaluated by Kellie Chieu, Jennifer Godwin, and Ashley Saltzman
ST 370 Project
Fall 2014

Executive Summary
In our study we investigated how the two common substitutes for buttermilk (milk with vinegar and milk with lemon juice) affect the height of red velvet cupcakes while baking. Our response variable was the height of the baked cupcake five minutes after it comes out of the oven, measured in millimeters. Our two factors were lemon juice and vinegar. The levels of our amount factor was 0.5 tablespoons, 1 tablespoon, and 1.5 tablespoons.

To replicate our findings, we baked three separate batches of six cupcakes for each level of our factors. We used the random number generator from a calculator to randomize the order in which we made the batches. This was done in a completely randomized blocked design. In the raw text file for MatLab, the cupcakes using the same type and amount were grouped and labeled 1 - 18. In addition, lemon cupcakes were labeled 0 and vinegar was 1. However, the order in which the batches were made is outlined in the Excel sheet.

From the stats function, it was shown that lemon produced the greatest mean and therefore the taller cupcakes compared to vinegar. Furthermore, the one tablespoon had the least standard deviation and thus had the most consistent results when comparing just the means. On the other hand, the 0.5 tablespoon created the greatest standard deviation in cupcake height. The fitted interaction means of the data sets also agreed with these findings, and it showed that the 1 tablespoon had the least variation.

Using the mfit function, it was shown that the amount of the lemon or vinegar was more important in influencing the height of the cupcake than the type of substance. The mplot of our data showed interactions between all levels of the factors.

During our statistical analysis, we used boxplots to analyze the values and spread of the data. The boxplot analysis further supported our findings that .5 tablespoon had a much larger standard deviation than 1 tablespoon. The boxplot for the type of cupcake showed that the lemon juice cupcakes had a larger standard deviation overall than the vinegar cupcakes. From this, we can conclude that the level of 1 tablespoon was the most consistent for each factor.

In both of the models, all of the terms were statistically significant. These terms included the type of variable (lemon or vinegar) and the amount of the variable (0.5, 1, and 1.5 tablespoon). The correlation for our models were significantly low.

All of the data we collected indicated that the 0.5 tablespoon level had the most variation but the highest mean, while the 1 tablespoon level had the least variation and the smallest mean. When analyzing the data by type, all of our data suggested that lemon juice had the highest overall mean. Our initial hypothesis was that the cupcakes made with vinegar would be taller than the cupcakes made with lemon juice. Our findings oppose this hypothesis. Since our correlation was not very strong, further experiments can be performed to benefit this study.

Table of Contents
Reason for Our Experiment
Prior Expectations
Methods
  ● Description
  ● Recipe
Statistical Analysis
  ● Comparison of Means
  ● M-plot of Data
  ● B-plot of Data
  ● Anova Tables
Practical Implications of the Study
Additional Questions

Reason for Our Experiment
In this study, we expected to learn which ingredient would result in a taller cupcake: lemon juice or vinegar. We varied the levels of the factor to determine if differing amounts of lemon juice or vinegar would be comparable. This study could be of interest to someone who sells cupcakes. If a professional pastry chef wanted to make their cupcakes taller to appeal to the customer, they might consider our data when choosing ingredients.

Prior Expectations
Before performing our experiment, we predicted that the cupcakes made with vinegar in their batter would rise higher than the cupcakes made with lemon juice in their batter. Through our experiment and our statistical analysis, it appears that the cupcakes made with lemon juice rise higher overall, against our original hypothesis.

Methods
To bake the cupcakes, we used the same measuring cups and bowls for each batch. We washed our equipment between making each batch to prevent contamination of the ingredients. When measuring the ingredients, a flat knife was used to scrape off excess amounts. This ensured that all ingredients would be present in the same amounts except for the factor we were testing. Dry ingredients were mixed first, then wet ingredients were added to the mixture. In a separate bowl, we added the varying ingredient to the milk and stirred. This mixture stood for five minutes before we added it to the batter. A hand whisk was used to mix the batter into a smooth consistency. We poured the batter into the cupcake pan using the same 1 tablespoon measuring cup so each cupcake would have the same amount of batter before baking. The same oven was used to bake all batches at 350 degrees Fahrenheit for 15 minutes. The cupcakes cooled for exactly 5 minutes. We stuck a toothpick through the center of each cupcake and marked the height with a pencil. We then measured the toothpick using a digital caliper.

Cupcake Recipe
  ● 1¼ cup flour
- ¾ cup sugar
- ½ teaspoon baking soda
- ½ teaspoon salt
- ¼ teaspoon cocoa powder
- ¾ cup vegetable oil
- 1 egg
- ½ teaspoon vanilla extract
- ½ cup milk
- Variable
Raw Data

Statistical Analysis
Using the stats function, we analyzed the data by substitution used (cupcake.type for vinegar and lemon juice) and the amount of the substance put in each cupcake (cupcake.amount).

| Table 1 |
|---|---|
| Command: stats(Cupcake.height,Cupcake.type) |
| Output: Lemon Vinegar |
| N | 54.0000 | 54.0000 |
| Mean | 33.2222 | 31.7278 |
| Std. Dev. | 2.0971 | 1.2702 |
| Q1 | 31.7750 | 30.8500 |
Median  32.7500  31.5500  
Q3      35.4000  32.9000  
Min     28.5000  29.1000  
Max     37.2000  34.6000  
Range   8.7000   5.5000

From Table 1, it can be seen that the mean of lemon was larger than vinegar by 1.494. It can also be noted that lemon had a greater standard deviation than vinegar. While standard deviation difference was 0.8269, this was less significant than the mean difference.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Command:</td>
</tr>
<tr>
<td>stats(Cupcake.height,Cupcake.amount)</td>
</tr>
<tr>
<td>Output:</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Q1</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Q3</td>
</tr>
<tr>
<td>Min</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>Range</td>
</tr>
</tbody>
</table>

As shown in Table 2, the standard deviation of 1 tablespoon was the smallest while the 0.5 tablespoon had the greatest. The 0.5 tablespoon also produced the greatest mean of 33.7444.

Table 3
Command:
mfit(Cupcake.height,Cupcake.type,Cupcake.amount)
Output:
Overall Mean
32.475

Fitted Main Effect of Y variable, y, by X variable, x1
Source    N  Main Effect
Lemon     54  0.74722
Vinegar   54  -0.74722

Fitted Main Effect of Y variable, y, by X variable, x2
Source      N  Main Effect
0.5         36  1.26940
1           36  -0.89444
1.5         36  -0.89444

Table of 2-way x1 by x2 Interaction Effects

<table>
<thead>
<tr>
<th>x1</th>
<th>Lemon</th>
<th>Vinegar</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2</td>
<td>0.5</td>
<td>1.21940</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-0.12222</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>-1.09720</td>
</tr>
</tbody>
</table>

Comparing 0.74722 and 1.269, the greatest absolute values of the main effects from the x1 variable and x2 variable respectively, it can be deduced that the amount was more important than type. For the interaction, the largest magnitude was 1.21940. So the relative ranking is X1 > X2 > interaction in terms of importance in mean shifts.

The fitted interaction effects show the difference between the predicted mean when using ONLY the two main effects of X1 and X2 and the actual mean from the data. The table shows that the 0.5 tablespoon data had an interaction effect number of 1.21940, therefore it had the greatest difference from the predicted mean of X1 and X2 alone. The 1 tablespoon mean had the least difference from the predicted mean from the main effects only prediction, which suggests that it has the smallest interaction.

Command: M-plot for mplot(Cupcake.height,Cupcake.type,Cupcake.amount)
Graph 1 showed the mplot of our data without any parallel lines, which indicates an interaction exists between the levels of factors. The slopes of the .5 tablespoon line and the 1 tablespoon line are negative, while the slope of the 1.5 tablespoon line is positive. The positive slope indicates that the mean of cupcakes with 1.5 tablespoons of vinegar was higher than the mean of cupcakes with 1.5 tablespoons of lemon juice. The 1 tablespoon line and the 1.5 tablespoon line are very close together, which suggests that this difference is less important than the difference shown by the .5 tablespoon line.

**Command:** `bplot(cupcake.height, cupcake.type)`

**Output:**
From the box plot (Graph-2), we can see that the cupcakes made with lemon juice have a larger standard deviation than the cupcakes made with vinegar. The median of the data values for the lemon juice cupcakes is also higher, which is important because median is not affected by outliers.

Command: bplot(cupcake.height, cupcake.amount)
Output:
From the boxplot (Graph-3), we can see that the .5 tablespoon data set has a large spread and standard deviation, as well as a larger median. The data from the 1 tablespoon set has less variation and a smaller standard deviation, which shows it is more consistent.

Table 4
>> lm(cupcake)

Variable Names in Input Dataset
type
trial
amount
height

Enter class or model statement or type ex for examples

lm>> class type amount

Enter model statement

lm>> model height = type + amount + type*amount

Sequential Sums of Squares ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>1</td>
<td>60.3008</td>
<td>60.3008</td>
<td>47.5644</td>
<td>4.5622e-10</td>
</tr>
<tr>
<td>amount</td>
<td>2</td>
<td>91.8772</td>
<td>45.9386</td>
<td>36.2357</td>
<td>1.2903e-12</td>
</tr>
<tr>
<td>type*amount</td>
<td>2</td>
<td>97.4117</td>
<td>48.7058</td>
<td>38.4184</td>
<td>3.6582e-13</td>
</tr>
<tr>
<td>Error</td>
<td>102</td>
<td>129.3128</td>
<td>1.2678</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-square 0.65872
Standard Error 1.126

lm>> model log(height) = type + amount + type*amount

Sequential Sums of Squares ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>1</td>
<td>0.054315</td>
<td>0.054315</td>
<td>43.9663</td>
<td>1.6206e-09</td>
</tr>
<tr>
<td>amount</td>
<td>2</td>
<td>0.081573</td>
<td>0.040787</td>
<td>33.0158</td>
<td>8.7836e-12</td>
</tr>
<tr>
<td>type*amount</td>
<td>2</td>
<td>0.087949</td>
<td>0.043974</td>
<td>35.5961</td>
<td>1.8778e-12</td>
</tr>
<tr>
<td>Error</td>
<td>102</td>
<td>0.126010</td>
<td>0.001235</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-square 0.63982
Standard Error 0.035148

The ANOVA table above showed all factors having a p-value less than .05, which suggests all factors were statistically significant. The R-square value of the height model was .65872. This low value implies that the model equation is probably not a reliable indicator of cupcake height. The log height model decreased the R-square and the
standard error. Since the log function did not make our equation more accurate, the data showed a more linear relationship.

Command: lm>> rplot by type
Output:

This plot for the lemon and vinegar levels shows that they have about the same spread of residuals, which indicates that a linear fit is appropriate for both types. The residuals for lemon seem to concentrate more around \( y = 0 \) than the vinegar residuals.

Command: lm>> rplot by amount
Output:
The line of best fit that was provided by the lm function creates a model from the entire data set. The residual plot above shows the difference between the value of specific data points and the line of best fit that is going through the data. This plot shows that the data for 1 tablespoon has the smallest residual sum, while the 1.5 data has more negative residuals and the .5 tablespoon data has more positive residuals. It can also be noted that the majority of the points fall within two standard deviations of the mean. The 1.5 tablespoon level seems to have a few more data points outside the two standard deviation than the other two sets.

Practical Implications of the Study
If a specific cupcake height and consistency was desired this study could be used as a guide. It is recommended that using 1 tablespoon of whichever variable will produce consistently shorter cupcakes but with less deviation. Someone who was interested in baking the tallest cupcakes might consider using .5 tablespoons of either lemon juice or vinegar. However, someone manufacturing a larger amount of cupcakes would probably want their cupcakes to be of consistent size, so one customer would not receive a larger cupcake than another. If this was the case, it would be more practical to use 1 tablespoon of vinegar or lemon juice to keep their products consistent.

Additional Questions
Since our equation of best fit did not have a significantly high R-square, the equation was not a reliable indicator of cupcake height. This raises the question of what other factors could affect the height of the cupcake. In another experiment, we could test other ingredients and see if that made a difference in the R-square of the model. We could also make a comparison between the standard deviations of each factor level (ex: ½ tablespoon lemon compared to ½ tablespoon vinegar) to be more specific. We could also examine how cupcakes made with buttermilk or plain milk compared to the heights of the two substitutes tested in our study.