An Experimental Analysis on Baking: The Effect of Butter Temperature, Baking Temperature, and Time on Cookie Size

May 09, 2025

Assmaa Aziz, Katherine Cui, Nguyen Mai

1. Introduction

Cookies are a universally enjoyed dessert; however, achieving consistent baking results can be more difficult than expected. Many amateur bakers, including us, follow recipes closely, but outcomes often vary in subtle ways. These inconsistencies are frequently due to the effects and interactions of various factors relating to temperature and time. We hoped to better understand the effects of these various individual factors and how they work together in influencing cookie outcomes.

We designed an experiment that investigates how baking factors affect cookie size. Specifically, we focused on three factors: butter temperature (melted or room temperature), baking temperature (325°F or 375°F), and baking time (10, 12, or 15 minutes). We wanted to determine which set of conditions would produce the largest cookies. Choosing diameter as our response variable offered a clear and approachable way to quantify differences in baking outcomes from different sets of conditions.

Our experiment was structured to evaluate the individual effect of each factor and to examine how these factors interact. For instance, we were interested in whether using melted butter at a lower temperature for a longer duration would produce cookies that spread more than those baked under higher temperatures. Gaining insight into such combinations may help explain why some batches of cookies turn out differently despite following the same set of instructions.

The experiment was purposely designed to reflect typical home baking conditions. All ingredients, tools, and procedures were carefully selected to ensure the study could be easily replicated in any household kitchen, ensuring that our findings are directly applicable to everyday baking. By conducting a carefully randomized and replicated study, we hope to better understand the science behind baking and, therefore, provide practical tips that help bakers with more consistent results.

2. Methodology

Following a CR[3] format, this experiment's three factors included butter temperature, baking temperature, and baking time. A completely randomized design was chosen as a simple and straightforward approach for studying the effects of the three factors, and the random assignment mechanism ensured that confounding factors were in expectation balanced between groups. A complete block design was considered during the design process; however, since the experimental units were only solid objects (and not something more complex like human or animal subjects) and cookies were baked in a single session under similar conditions, a CR design sufficed. Furthermore, a Latin Square design was not well suited since we had more than one factor of interest, and a split plot design was unfitting since there are no factors that are difficult to change. Ultimately, a completely randomized factorial design proved ideal not only for practicality but also as it allowed full flexibility in treatment assignment. Since every cookie could be independently assigned to any of the treatment combinations, we could randomize the order of runs to minimize systematic bias from time or environmental factors. By doing so, we could observe not only how each condition affected the cookies but also how different combinations might lead to different results.

The levels within the three factors were melted/room temperature, 325/375°F, and 10/12/15 minutes. The levels were chosen from customary practices and values in various baking procedures of various cookie types and recipes. For example, 325°F and 375°F are extremely common cookie baking temperatures, and these standard levels were chosen to generalize the experiment results to typical baking activity. These three factors and levels combine to twelve (2 x 2 x 3) different treatment combinations.

The diameter of the baked cookies, measured in inches, was the response variable of our experiment. Cookie size varies heavily depending on factors such as butter temperature, baking temperature, and baking time, and diameter is an appropriate measure for cookie size as it is well-defined, easy to measure, and a definition of size. This response variable is of continuous numeric data type (double data type in R).

As each individual ball of cookie dough was the experimental unit, they were randomly assigned to treatments. The randomization procedure was carried out using R. Treatments were labelled from 1-12, and *sample()* was used for random sampling to assign treatments. The output vector [10, 11, 6, 4, 7, 2, 12, 10, 4, 7, 5, 2, 3, 6, 11, 8, 1, 5, 9, 3, 12, 9, 1, 8] detailed the randomly assigned treatment combination for a particular dough ball.

The results of our experiment are detailed in the following table below. Note that we completed two runs for our experiment to ensure the reliability and accuracy of our results; we wanted replication to prove that the results were consistent and not a one-time occurrence. Run 1 is displayed on the left, and Run 2 is displayed on the right. Our original data was measured in cm due to the particular caliper used in the experiment; however, the values have been converted to inches below as specified in our experiment.

Table A: Experiment Results
**Run 1 (left) & Run 2 (right)*

Cookie Number	Butter Temperature (R/M)	Baking Temperature (℉)	Time (minutes)	Diameter (inches)
1	Room	350	10	3.03
2	Room	350	12	3.31
3	Room	350	15	3.11
4	Room	375	10	2.95
5	Room	375	12	2.68
6	Room	375	15	2.8
7	Melted	350	10	2.87
8	Melted	350	12	3.07
9	Melted	350	15	3.07
10	Melted	375	10	2.64
11	Melted	375	12	2.44
12	Melted	375	15	2.32

Cookie Number	Butter Temperature (R/M)	Baking Temperature (℉)	Time (minutes)	Diameter (inches)
1	Room	350	10	3.39
2	Room	350	12	3.5
3	Room	350	15	4.06
4	Room	375	10	3.15
5	Room	375	12	3.19
6	Room	375	15	3.46
7	Melted	350	10	2.83
8	Melted	350	12	3.07
9	Melted	350	15	2.83
10	Melted	375	10	2.52
11	Melted	375	12	2.64
12	Melted	375	15	2.68

In principle, because the experiment was carried out in a home kitchen, there were undoubtedly shortcomings relating to variability. With a lack of regulated environment conditions and specialized equipment, it was not possible to ensure that each cookie was baked and measured under the exact same circumstances. For example, where the cookies were placed on the baking tray and how long the cookie dough sat might have affected how the cookies turned out. Measurement with a caliper is much more precise than a regular ruler but can still face human reading error especially on such a small scale. Combined with high variability, our relatively small sample size of 24 cookies may contribute to lower power as low variance and larger sample sizes is what is required for higher power. The rationale for not increasing sample size further lied in balancing time and energy constraints versus power; the power increase we would obtain from an extra run of cookies was not nearly significant enough in comparison to the extra effort it would require to bake a new batch. Our sample size promised an acceptable two runs with sufficient power for our particular experiment.

As the experiment was conducted, other factors came into play that affected variability. Although detailed procedures were closely and carefully followed to ensure limited variability, quite a bit of imprecision came with the experimental unit themselves. Each ball of cookie dough faced significant variability in terms of volume due to the human error that came with using kitchen cookie scoops. Even after scooping, placing each experimental unit onto baking paper led to fluctuations in landing and position which we then realized had some effect on the way the dough would eventually spread. The cookie dough also did not always spread in a perfect circular shape, leading to many instances of oblong cookies. To prevent leniency in definition, we clarified that diameter was measured by the longest ‘diameter’ for each cookie. However, this new definition comes with some level of vagueness in terms of accurate measurement especially with oddly shaped cookies. Alongside the inherent error that comes with measuring with a kitchen caliper without cross checking, our design suffered somewhat in terms of reliability in practice. Although there might have been a few setbacks in terms of reliability, overall validity remained good. Diameter, no matter what way it is measured, maps directly to size. This experiment also proved to be quite replicable when carried out, as the procedure was straightforward and concrete to be followed multiple times in different runs without much variation.

3. Results

Our project aimed to measure the effect of butter temperature, baking temperature, baking time, and their interactions on cookie diameter. Across all our observations, we found the minimum cookie size was 2.32 inches, the maximum size was 4.06 inches, and the average size was 2.98 inches. To summarize our data and to compare factors, we plotted box plots using the R library tidyverse. We observed that, on average, cookies made from room-temperature butter were larger than those made from melted butter (*Figure A*). On average, cookies baked at 350°F were larger than those baked at 375°F (*Figure B*). On average, cookies baked for 12 minutes were the largest, followed by those baked for 15 minutes (*Figure C*).

Figure A

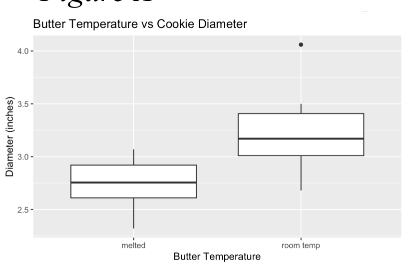


Figure B

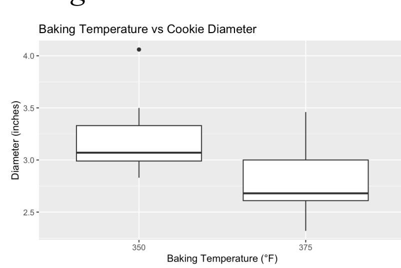
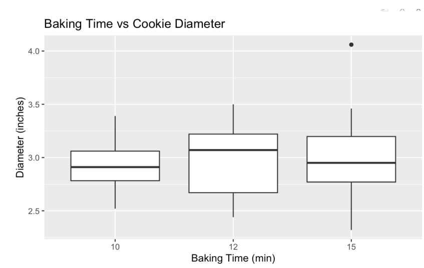


Figure C



To look for any possible two-way and three-way interactions, we plotted interaction plots using the R library tidyverse. We did not find any interaction between baking temperature and butter temperature (*Figure D*). The effects of these two factors are independent. We observed interactions between butter temperature and baking time and between baking temperature and baking time (*Figures F, G*). Longer baking times and room-temperature butter resulted in larger cookies (*Figure F*). Room-temperature butter led to larger cookies across all conditions (*Figure E*). The largest cookies occurred when using room-temperature butter, baking at 350°F, and baking for 15 minutes (*Figure E*).

Figure D

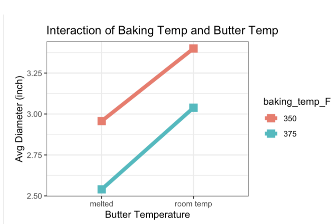


Figure F

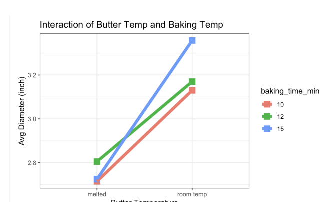


Figure E

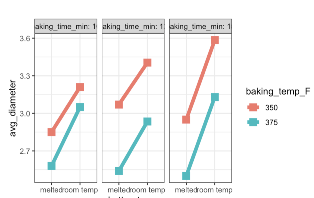
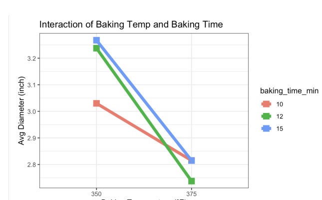


Figure G



ANOVA

The model for our CB[3] experiment can be laid out as:

$$\mathbf{y_{ijk}} = \boldsymbol{\mu} + \alpha_i + \beta_j + \lambda_k + (\alpha\beta)_{ij} + (\alpha\lambda)_{ik} + (\beta\lambda)_{jk} + (\alpha\beta\lambda)_{ijk} + \varepsilon_{ijk}$$

*In this model, i represents the level of butter temperature, j represents the level of baking temperature, and k represents the level of baking time. $\boldsymbol{\mu}$ is the overall mean, α_i is the main effect of butter temperature, β_j is the main effect of baking temperature, λ_k is the main effect of baking time, ε_{ijk} is the random error, and the rest of the terms are the respective interaction effects.

The results of the ANOVA model are summarized below.

Figure H: ANOVA Summary

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
butter_temp	1	1.3301	1.3301	15.594	0.00193
baking_temp_F	1	0.9087	0.9087	10.654	0.00678
baking_time_min	2	0.0566	0.0283	0.332	0.72412
butter_temp:baking_temp_F	1	0.0045	0.0045	0.053	0.82148
butter_temp:baking_time_min	2	0.0809	0.0405	0.474	0.63352
baking_temp_F:baking_time_min	2	0.0933	0.0466	0.547	0.59263
butter_temp:baking_temp_F:baking_time_min	2	0.0033	0.0017	0.019	0.98073
Residuals	12	1.0235	0.0853		

We observe that the F-statistics for butter and baking temperature are quite significant. This means that we can be assured that our choices to test for butter and baking temperature were reasonable.

In addition, we can calculate R-squared, a measure of the percentage of variation in the response that is explained by the model, using the formula $R^2 = SSR/SST$. With a relatively strong R-squared of around 0.71 (71%), we can conclude that our model fits the data well.

Randomization-Based Inference

To test whether the effects we saw were due to our treatments or due to chance, we performed randomization-based inference. Our null hypothesis was the following: Butter temperature, baking temperature, baking time, and their interactions have no effect on cookie size. The mean response is the same across groups, and any observed difference is due to chance. Our alternative hypothesis was the following: At least 1 factor or its interactions have an effect on cookie size. The mean response differs for at least one of our treatment groups.

We performed a simulation where the responses to each treatment were randomly sampled. From this simulated data, we made an ANOVA model and found simulated F statistics for each of our treatment effects. We repeated this simulation 500 times and found the following p-values for each treatment effect. Our resulting p-values are similar to the output of the ANOVA table, which suggests our simulation was done correctly.

Table B: P-values

Treatment Effect	P-value
Butter temperature	0.002
Baking temperature	0.006
Baking Time	0.706
Butter temperature * Baking temperature	0.832
Butter temperature * Baking time	0.63
Baking temperature * Baking time	0.592
Butter temperature * Baking temperature * Baking time	0.974

For the main effects of butter temperature and baking temperature, using an $\alpha = 0.05$, we reject the null hypothesis. From our test, we can conclude that butter temperature and baking temperature have an effect on cookie size. For our remaining main effect and our interaction effects, we fail to reject the null hypothesis. We can conclude that baking time and all the interactions between factors have no effect on our response: cookie size.

A limitation of our experiment is that, due to time constraints, we only have two replications and 24 observations in total. We used one batch of dough to control for dough variability, but this means the results of our analysis may not generalize for different dough batches or slight differences in ingredients, such as variation in brand.

4. Discussion/Conclusion

Our results revealed that butter temperature and baking temperature had the most substantial impact on cookie diameter. Cookies made with room-temperature butter and baked at a lower temperature generally resulted in larger sizes. In contrast, baking time, despite being varied across three levels, showed a less consistent effect on the outcome. Hence, we can say that while bake duration does play a role in the final product, it is not as critical as the butter's temperature or the oven's heat. Although we initially expected interactions between the variables, such as longer baking times offsetting the effects of melted butter, our results did not provide strong evidence of such interactions. Each factor mostly appeared to operate independently.

In conducting this experiment, we encountered several forms of variability. Differences in dough scooping, placement on baking trays, and final cookie shapes added unavoidable noise to the data. While these variations were minimized to the best of our ability, they reflect the kind of uncertainty present in most home baking environments. However, the overall design of the experiment allowed us to draw general conclusions with practical applications.

Given the results, we came up with a few ideas that could help us explore the topic in greater depth. We would be interested in testing different flavors of Betty Crocker cookie mixes (peanut butter, sugar cookie, birthday cake, or red velvet) since the dough composition in different flavors may interact differently with butter temperature and baking conditions. In addition, increasing the number of replicates would improve our ability to yield stronger statistical conclusions. Finally, although we chose to bake all the cookies in a shared kitchen to save time and keep conditions consistent, future versions of this experiment could explore how results may vary when each team member bakes individually in their own home kitchen.

We do recognize that our scooping method was prone to human error. A possible solution would be to use a cookie dough portioner or to weigh each dough ball, which could improve consistency and help reduce variation in portion size.

In conclusion, by carrying out a structured experiment and analyzing the results carefully, we are able to identify the most influential factors that contribute to larger cookies: using room temperature butter and baking at 350°F. These findings not only reinforce our understanding of baking factors but also demonstrate how statistical reasoning can be applied to everyday tasks in a practical and meaningful way.

Appendix
Lab Notes

cookies data

replication	butter temp	baking temp	time	diameter (in)	comments		replication	butter temp	baking temp	time	diameter (in)	comments
1	Room temperature	350°F	12 min	3.31			2	Melted	375°F	12 min	2.64	
1	Room temperature	350°F	15 min	3.11			2	Room temperature	350°F	12 min	3.5	
1	Room temperature	375°F	10 min	2.95			2	Room temperature	375°F	10 min	3.15	
1	Melted	350°F	15 min	3.07			2	Melted	350°F	10 min	2.83	
1	Room temperature	375°F	15 min	2.8			2	Room temperature	375°F	12 min	3.19	looks overdone, almost burnt edges
1	Melted	350°F	10 min	2.87			2	Room temperature	350°F	10 min	3.39	
1	Melted	350°F	12 min	3.07			2	Melted	350°F	15 min	2.83	
1	Melted	375°F	10 min	2.64	thicker than others		2	Melted	375°F	10 min	2.52	looks a bit uneven
1	Melted	375°F	15 min	2.32	tiny		2	Melted	375°F	15 min	2.68	
1	Room temperature	350°F	10 min	3.03			2	Room temperature	350°F	15 min	4.06	outlier, a lot bigger
1	Room temperature	375°F	12 min	2.68			2	Room temperature	375°F	15 min	3.46	
1	Melted	375°F	12 min	2.44			2	Melted	350°F	12 min	3.07	

R Code

project

Project Data

```
#Load in data
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.3     ✔ tidyr     1.3.1
✔ purrr     1.0.2
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to
become errors
```

```
replication = c(rep(1, 12), rep(2, 12))
butter_temp = rep(c(rep('room temp', 6), rep('melted', 6)), 2)
baking_temp = rep(c(rep(c(350, 375), each = 3)), 2)
baking_time = rep(c(10, 12, 15), 8)

responses_cm = c(7.7, 8.4, 7.9, 7.5, 6.8, 7.1, 7.3, 7.8, 7.8, 6.7, 6.2, 5.9,
                 8.6, 8.9, 10.3, 8, 8.1, 8.8, 7.2, 7.8, 7.2, 6.4, 6.7, 6.8)

responses_in = round(responses_cm/2.54, 2)

cookies = data.frame(cookie_id = 1:24, replication = replication,
                     butter_temp = butter_temp, baking_temp_F = baking_temp,
                     baking_time_min = baking_time,
                     diameter_inch = responses_in)
cookies$butter_temp = factor(cookies$butter_temp)
cookies$baking_temp_F = factor(cookies$baking_temp_F)
cookies$baking_time_min = factor(cookies$baking_time_min)
head(cookies, 3)
```

```
  cookie_id replication butter_temp baking_temp_F baking_time_min diameter_inch
1         1           1   room temp           350              10          3.03
2         2           1   room temp           350              12          3.31
3         3           1   room temp           350              15          3.11
```

Treatment Assignment

```
#Randomly assign treatments

set.seed(100)
treatments = c(rep(1:12, each = 1, times = 2))
treatment_assignment = sample(treatments, replace = FALSE)
treatment_assignment
```

```
[1] 10 11  6  4  7  2 12 10  4  7  5  2  3  6 11  8  1  5  9  3 12  9  1  8
```

```
#Each index in our vector represents one of our dough balls. The value at that
#index is the randomly assigned treatment combination.
```

Visualizations and Summary Statistics

```
summary(cookies)
```

```
   cookie_id      replication     butter_temp baking_temp_F baking_time_min
 Min.   : 1.00   Min.   :1.0   melted   :12   350:12        10:8
 1st Qu.: 6.75   1st Qu.:1.0   room temp:12   375:12        12:8
 Median :12.50   Median :1.5                                15:8
 Mean   :12.50   Mean   :1.5
 3rd Qu.:18.25   3rd Qu.:2.0
 Max.   :24.00   Max.   :2.0
 diameter_inch
 Min.   :2.320
 1st Qu.:2.680
 Median :2.990
 Mean   :2.984
 3rd Qu.:3.160
 Max.   :4.060
```

Main Effects

Butter Temperature

```
#Main Effects

#Figure A
butter_temp_vs_resp = ggplot(cookies, aes(x = butter_temp, y = diameter_inch)) +
  geom_boxplot() + labs(title = 'Butter Temperature vs Cookie Diameter',
                        x = 'Butter Temperature', y = 'Diameter (inches)')
#butter_temp_vs_resp
```

Baking Temperature

```
#Figure B

baking_temp_vs_resp = ggplot(cookies, aes(x = baking_temp_F, y = diameter_inch))
  geom_boxplot() + labs(title = 'Baking Temperature vs Cookie Diameter',
                        x = 'Baking Temperature (°F)', y = 'Diameter (inches)')
#baking_temp_vs_resp
```

Baking Time

```
#Figure C
baking_time_vs_resp = ggplot(cookies, aes(x = baking_time_min,
                                           y = diameter_inch)) + geom_boxplot() +
  labs(title = 'Baking Time vs Cookie Diameter', x = 'Baking Time (min)',
       y = 'Diameter (inches)')
#baking_time_vs_resp
```

Interaction Effects

2-Way: Butter Temp and Baking Temp

```
#Interaction plots
#Figure D
group_means <- cookies |>
  group_by(butter_temp, baking_temp_F) |>
  summarize(avg_diameter = mean(diameter_inch),
            n = n())
```

```
`summarise()` has grouped output by 'butter_temp'. You can override using the
`.groups` argument.
```

```
#group_means

plot1 = ggplot(group_means, aes(x = butter_temp,
                                y = avg_diameter,
                                color = baking_temp_F,
                                group = baking_temp_F)) +
  geom_point(shape = 15, size = 4) +
  geom_line(lwd = 2) +
  theme_bw() + labs(title = 'Interaction of Baking Temp and Butter Temp',
                    x = 'Butter Temperature', y = 'Avg Diameter (inch)')
```

2-Way: Butter Temp and Baking Time

```
#Interaction plots
#Figure F
```

```
group_means <- cookies |>
  group_by(butter_temp, baking_time_min) |>
  summarize(avg_diameter = mean(diameter_inch),
            n = n())
```

```
`summarise()` has grouped output by 'butter_temp'. You can override using the
`.groups` argument.
```

```
#group_means

plot2 = ggplot(group_means, aes(x = butter_temp,
                                y = avg_diameter,
                                color = baking_time_min,
                                group = baking_time_min)) +
  geom_point(shape = 15, size = 4) +
  geom_line(lwd = 2) +
  theme_bw() + labs(title = 'Interaction of Butter Temp and Baking Temp',
                    x = 'Butter Temperature', y = 'Avg Diameter (inch)')
```

2-Way: Baking Temp and Baking Time

```
#Interaction plots
#Figure G

group_means <- cookies |>
  group_by(baking_temp_F, baking_time_min) |>
  summarize(avg_diameter = mean(diameter_inch),
            n = n())
```

```
`summarise()` has grouped output by 'baking_temp_F'. You can override using the
`.groups` argument.
```

```
#group_means

plot3 = ggplot(group_means, aes(x = baking_temp_F,
                                y = avg_diameter,
                                color = baking_time_min,
                                group = baking_time_min)) +
  geom_point(shape = 15, size = 4) +
  geom_line(lwd = 2) +
  theme_bw() + labs(title = 'Interaction of Baking Temp and Baking Time',
                    x = 'Baking Temperature (°F)', y = 'Avg Diameter (inch)')
```

3-Way: Butter Temp, Baking Temp, Baking Time

```
##Interaction plots
#Figure E
```

```
group_means <- cookies |>
  group_by(butter_temp, baking_temp_F, baking_time_min) |>
  summarize(avg_diameter = mean(diameter_inch),
            n = n())
```

```
`summarise()` has grouped output by 'butter_temp', 'baking_temp_F'. You can
override using the `.groups` argument.
```

```
#group_means

plot4 = ggplot(group_means, aes(x = butter_temp,
                                y = avg_diameter,
                                color = baking_temp_F,
                                group = baking_temp_F)) +
  geom_point(shape = 15, size = 4) +
  geom_line(lwd = 2) +
  facet_wrap(vars(baking_time_min), labeller = label_both) +
  theme_bw()
```

ANOVA

```
#ANOVA
anova_model <- aov(diameter_inch ~ butter_temp*baking_temp_F*baking_time_min,
                   data = cookies)
anova_tbl = summary(anova_model)
anova_tbl
```

```
                                          Df Sum Sq Mean Sq F value  Pr(>F)
butter_temp                                1 1.3301  1.3301  15.594 0.00193 **
baking_temp_F                              1 0.9087  0.9087  10.654 0.00678 **
baking_time_min                            2 0.0566  0.0283   0.332 0.72412
butter_temp:baking_temp_F                  1 0.0045  0.0045   0.053 0.82148
butter_temp:baking_time_min                2 0.0809  0.0405   0.474 0.63352
baking_temp_F:baking_time_min              2 0.0933  0.0466   0.547 0.59263
butter_temp:baking_temp_F:baking_time_min  2 0.0033  0.0017   0.019 0.98073
Residuals                                 12 1.0235  0.0853
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
obs_f = anova_tbl[[1]]$`F value`[1:7]
#obs_f
```

Randomization-based Inference

```
#Randomization-based Inference
set.seed(100)
```

```r
shuffle_three_factor <- function(y, a, b, c) {
  Y <- sample(y) # source of randomness

  cookies <- data.frame(diameter_inch = Y,
                        butter_temp = a,
                        baking_temp_F = b,
                        baking_time_min = c)

  anova_model <- aov(diameter_inch ~ butter_temp*baking_temp_F*baking_time_min,
                     data = cookies)
  anova_table <- summary(anova_model)
  f_stats <- anova_table[[1]]$`F value`
  f_stats[1:7]
}


null_stats <- replicate(500, shuffle_three_factor(cookies$diameter_inch,
                                                  cookies$butter_temp,
                                                  cookies$baking_temp_F,
                                                  cookies$baking_time_min))



null_butter_temp = null_stats[1, ]
null_baking_temp = null_stats[2, ]
null_baking_time = null_stats[3, ]
null_butter_baking_temp = null_stats[4, ]
null_butter_baking_time = null_stats[5, ]
null_baking_temp_time = null_stats[6, ]
null_butter_baking_temp_time = null_stats[7, ]

p1 = mean(null_butter_temp > obs_f[1])
p2 = mean(null_baking_temp > obs_f[2])
p3 = mean(null_baking_time > obs_f[3])
p4 = mean(null_butter_baking_temp > obs_f[4])
p5 = mean(null_butter_baking_time > obs_f[5])
p6 = mean(null_baking_temp_time > obs_f[6])
p7 = mean(null_butter_baking_temp_time > obs_f[7])

df = data.frame(p_butter_temp = p1, p_baking_temp = p2, p_baking_time = p3,
                p_butter_baking_temp = p4, p_butter_baking_time = p5,
                p_baking_temp_time = p6, p_butter_baking_temp_time = p7)
df
```

```
  p_butter_temp p_baking_temp p_baking_time p_butter_baking_temp
1         0.002         0.006         0.706                0.832
  p_butter_baking_time p_baking_temp_time p_butter_baking_temp_time
1                 0.63              0.592                     0.974
```