Homework #1.1-Example C

C. Student

9/2/2022

Exploring Double Stuf vs. Regular Oreos®

For as long as people have sold products, goods, and services to others, there have also been claims of false advertising[1]. False advertising occurs when a company's advertisement contains untrue or misleading information about their product[2]. Subway® has been brought up several times for false advertising issues: once for their "footlong" sandwiches being allegedly less than 12 inches long[3] and more recently for their tuna salad alledgedly not containing any actual tuna[4].

Subway is not the only food brand to face such charges. Oreos® (under the Nabisco brand owned by Mondelēz International) have also faced this issue with regards to whether Double Stuf Oreos have twice the créme filling as regular Oreos. One article describes how one math teacher investigated this question and found that Double Stuf oreos had only around 1.86 times the amount of filling as Regular oreos[5].

Dr. Hatfield worked with an introductory statistics class to investigate whether double stuf oreos have twice the créme filling as regular oreos. They used a proper process to select a grocery store at random in Tempe, AZ and then to select 9 packages of regular and 9 packages of Double stuf from the shelves. They then split these 9 packages into 9 pairs (one of each type of cookie) to measure the créme filling masses (in grams) for each cookie.

I have been given one of these data sets (Oreo 3) to explore in context of the statistical research question "Do Double Stuf Oreos have double the créme filling mass as Regular Oreos?"

Data Exploration

A great place to begin is with a data visualization. Figure 1 shows side-by-side outlier box plots for créme filling mass by type of cookie. A couple of things that I quickly noticed was the presence of five potential outliers under the 1.5 IQR rule (marked by the solid black dots). Three of these outliers appear to be still somewhat close to the rest of the data. Two cookies appear to have a créme filling mass that is much more similar to the other type of cookie.

I reached out to Dr. Hatfield about the issue and he confirmed that there was a data entry error and that the cookie with row index 13 should be a "Regular" cookie and the cookie with row index 39 should be a "Double Stuf" cookie. I have corrected the data file and regenerated the box plot (see Figure 2).

Figure 3 provides a glimpse at modal clumps of cookies based on their filling masses. There appear to be two modal clumps (where the shadows are heaviest) around 3-3.5 grams and then around 7 grams. These

---

[1] https://www.truthinadvertising.org/timeline/

[2] https://dcba.lacounty.gov/portfolio/false-advertising/

[3] https://www.forbes.com/sites/ericgoldman/2016/02/29/why-the-subway-footlong-lawsuits-fell-short/?sh=571252669f25

[4] https://www.nbcnews.com/news/us-news/lawsuit-claims-subway-s-tuna-sandwiches-contain-no-tuna-n1256181

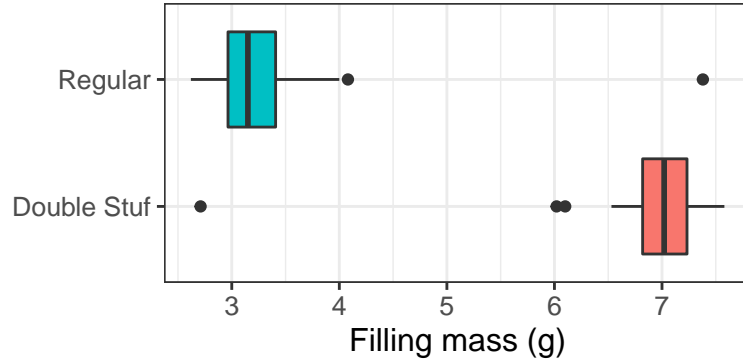[5] https://www.huffpost.com/entry/creme-oreo_n_3767983

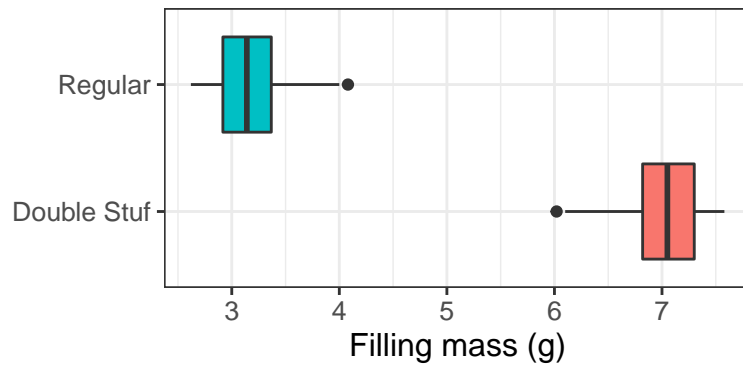Figure 1: Box plots of Filling Mass by Type

Figure 2: (Corrected) Box plots of Filling Mass by Type
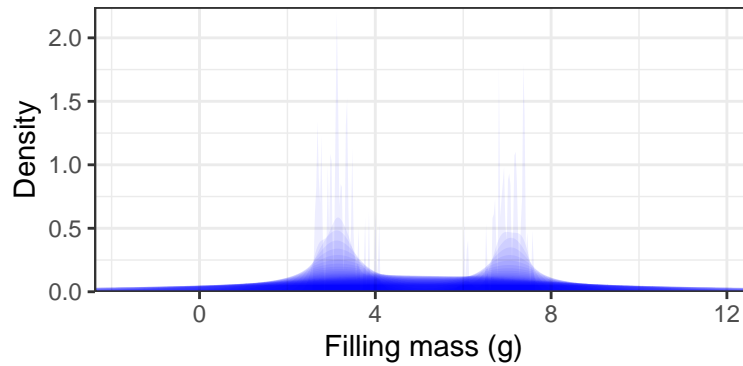

Figure 3: Shadowgram of Filling Masses

modal clumps appear to line up with the midpoints of the Regular and Double Stuf oreos, respectively, from Figure 2.

I also wanted to look at a histogram to get a better sense of the créme filling masses. Figure 4 provides us with side-by-side histograms for the two types of cookies. We can see a match between the centers of the modal clumps in the shadowgram (Figure 3) and the location of the highest bars in Figure 4. Further, we can get a sense of the spread of the filling masses. The Double Stuf cookies appear to go from a little under 6 grams to over 7.5 grams; the Regular cookies vary from a bit under 2.5 grams to just over 4 grams.
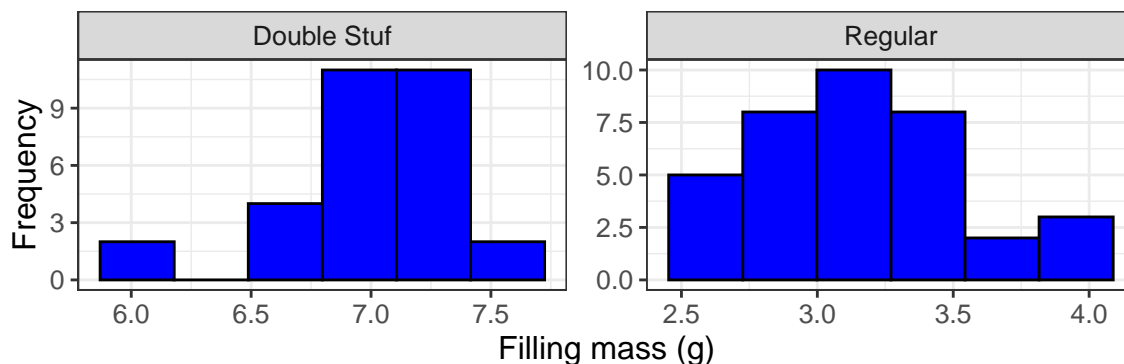

Figure 4: Histograms of Filling Mass by Type of Cookie

In addition to using data visualizations, I'm also going to use some descriptive statistics to help me get a better idea of what's going on in the data. Table 1 provides the various values of descriptive statistics for filling mass, broken up by the type of cookie. We can quickly see that there are 36 regular Oreos and 30 Double Stuf Oreos.

Table 1: Summary Statistics for Oreo Filling Masses

	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
Double Stuf	30	6.02	6.820	7.05	7.300	7.58	0.341	7.010	0.370	-0.888	0.558
Regular	36	2.62	2.918	3.14	3.368	4.08	0.334	3.179	0.376	0.543	-0.331

The *sample minimum* tells us the least amount of créme filling observed in each type of cookie; Regular had a cookie with 2.62 grams of filling while Double Stuf had a cookie with 6.02 grams. On the opposite end of things, the *sample maximum* tells us the greatest observed amount of créme filling. For Regular cookies, the heaviest (by filling) had 4.08 grams of filling and Double Stuf had 7.58 grams. While using the extrema is not necessarily the best answering the reserach question, I think that looking at their relative sizes can help us start to get a sense of what possible answers we could have. The ratio of Double Stuf to Regular for values of the *sample minimum* is 2.3; for values of the *sample maximum* we get 1.86. From these two ratios, I see that Double Stuf Oreos could come out on either end of the double question (more or less than double).

The *first quartile* (Q1) tells us the créme filling mass which cuts off the lightest 25% of our data. While this mass does not have to be observed, approximately 9 Regular cookies and 8 Double Stuf cookies will have less than 2.92 grams and 6.82 grams, respectively. The *third quartile* (Q3) tells us the créme filling mass which cuts off the heaviest 25% of our data. While this mass does not have to be observed, approximately 9 Regular cookies and 8 Double Stuf cookies will have more than 3.37 grams of filling and 7.3 grams, respectively.

If we want to estimate the location parameter for the distribution of créme filling masses, the *sample median* is a statistic which can provide an estimate. The *sample median* provides the mass of créme filling where half of the sample has less than this value and half the sample has at least this much créme filling. In other words, the value of the *sample median* is the center of our observations. For Regular cookies, this is 3.14 grams; for Double Stuf, 7.05 grams. The ratio of Double Stuf to Regular is 2.25. This suggests that Double Stuf Oreos might have at least twice the amount of filling than Regular Oreos. However, knowing the values

of the *sample median* only paints part of the story. I know that not every cookie is the same, thus we need a measure for the amount of variation. The *median absolute deviation (MAD)* provides the amount of créme filling that half of all cookies are within of the value of *sample median*. In other words, half of the Regular cookies are within 0.33 grams of 3.14 grams. For Double Stuf cookies, half are no more than 0.34 above or below 7.05 grams.

In addition to the *sample median*, we can use the *sample arithmetic mean (SAM)* to measure not only how well the two collections amassed créme filling, but to estimate the performance of the underlying process to fill the cookies with créme. The values of the *SAM* are 3.18 grams/cookie and 7.01 grams/cookie for Regular and Double Stuf, respectively. Thus, we can say that the collection of Regular Oreos accumulated 3.18 times as much créme filling as individual cookies. Similarly for Double Stuf cookies. The ratio of the values of the *SAM* of Double Stuf to Regular is 2.2; again, suggesting that Double Stuf Oreos might have more than double the stuf.

Just as the *MAD* provides some sense of variation, the *sample arithmetic variance (SAV)* does too. (As does the *sample arithmetic standard deviation–SASD*.). For Regular Oreos, the value of the *SAV* is 0.1417 sq. grams; for Double Stuf 0.1366 sq. grams. These values point out that the cookies vary from one another in terms of créme filling masses, with them having almost identical levels of pairwise deviation. The *SASD* can tell us about where accumulation of créme filling picks up and slows down. For Regular cookies, we see an uptick in accumulation at approximately 2.8 grams $(SAM(\mathcal{R}) * 1 \text{ cookie} - SASD(\mathcal{R}))$ and a slow down in accumulation near 3.56 grams $(SAM(\mathcal{R}) * 1 \text{ cookie} + SASD(\mathcal{R}))$.

The last two statistics in Table 1 are *sample skewness* and *sample excess kurtosis*. Both of these statistics provide us with more information about how the underlying filling process of the cookies behaves. The negative value of *sample skewness* for Double Stuf cookies indicates that the process appears to create cookies that have less créme filling than most Double Stuf cookies have. The postive value for Regular Oreos indicates the opposite: the filling process for Regular Oreos will fill some cookies with more créme filling than typical. For *sample excess kurtosis*, the positive value tells us that the filling process for Double Stuf Oreos will create more outliers than I would anticipate if the process behaved in a Gaussian way. The negative value for Regular Oreos suggests that that process creates fewer outliers than if that process also behaved in a Guassian way.

Now that I have a better understanding of the data, I believe that I can attempt to answer the statistical research question.

Statistical Inference

"Do Double Stuf Oreos have double the créme filling mass as Regular Oreos?" is a two-sample test of location. The null hypothesis would be that Double Stuf Oreos do have twice the amount of créme filling mass as Regular Oreos. The alternative hypothesis would be that they do not. We can express these two hypotheses as the following models:

$$H_0 : \frac{\mu_D}{\mu_R} = 2$$
$$H_1 : \frac{\mu_D}{\mu_R} \neq 2$$

While a two sample test of the ratio of location parameters does exist, I'm going to choose to use a log transformation to make models fit the form most often seen for this type of problem:

$$H_0 : \log(\mu_D) - \log(\mu_R) = log(2)$$
$$H_1 : \log(\mu_D) - \log(\mu_R) \neq log(2)$$

This will require applying the log transformation to the filling masses.

If I want to use Welch's t test, I will need to make sure that my data are 1) follow a Gaussian distribution, and 2) are independent. Figure 5 provides QQ plots for both Regular and Double Stuf Oreos when compared against a Gaussian distribution. There are just a couple of points (less than 10% of the data) beyond the envelope bounds for the Double Stuf cookies. I'm going to bank on the t test's relative robustness to slight departures from Gaussian assumption. I will note that after the log transformation, Regular Oreos have 0.3039 as the value of *sample skewness* while Double Stuf has -1.0384. The positive and negative values of *sample skewness* do have me worried that Welch's t test will produce questionable results.

To assess the issue of independence, knowing the order in which measurements were taken is important. There is no explicit column given for the measurement order. I have spoken to Dr. Hatfield and he confirmed that the data appear in the file in measurement order. Thus, we can look at the index plot in Figure 6. I do not see any patterns in the measurements of filling mass that suggest a violation of independence. However, I will note that Discovery UK's How It's Made does suggest that there could be a lack indpendence in the actual production process[6].
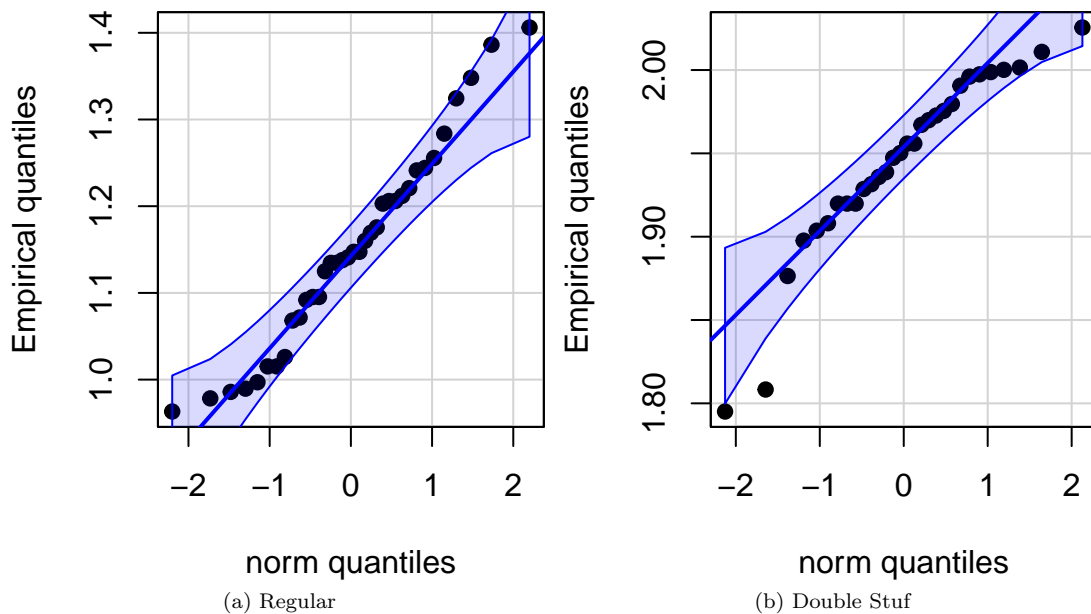

(a) Regular   (b) Double Stuf

Figure 5: QQ Plots for Log Filling Mass

Parametric Shortcut

Given all of this, I will proceed, with caution, to a Welch's t test using the log transformed masses and setting the null hypothesis value to log(2). For this test, I will use an Unusualness Threshold of 3% (confidence level of 97%) as asked.

I find that $t(51.46) \approx 4.72$, with a p-value of < 0.0001. Given that this is less than my Unusualness Threshold, I will declare that I have an unusual event under the null hypothesis and reject the null hypothesis. We would only get this value of the t statistic less than 1/100th of a percent of the time we repeat the study.

I found an interval estimate for the ratio using a method which will capture the true ratio of créme filling masses 97% of the time. For my samples, I found a transformed 97% confidence interval of (0.75, 0.84) log grams/gram. The un-transformed confidence interval is (2.11, 2.33).

I will decide to reject the null hypothesis and act as if Double Stuf Oreos to Regular Oreo créme filling ratio is something other than 2. In particular, we have some evidence in this study to suggest that the ratio is

---

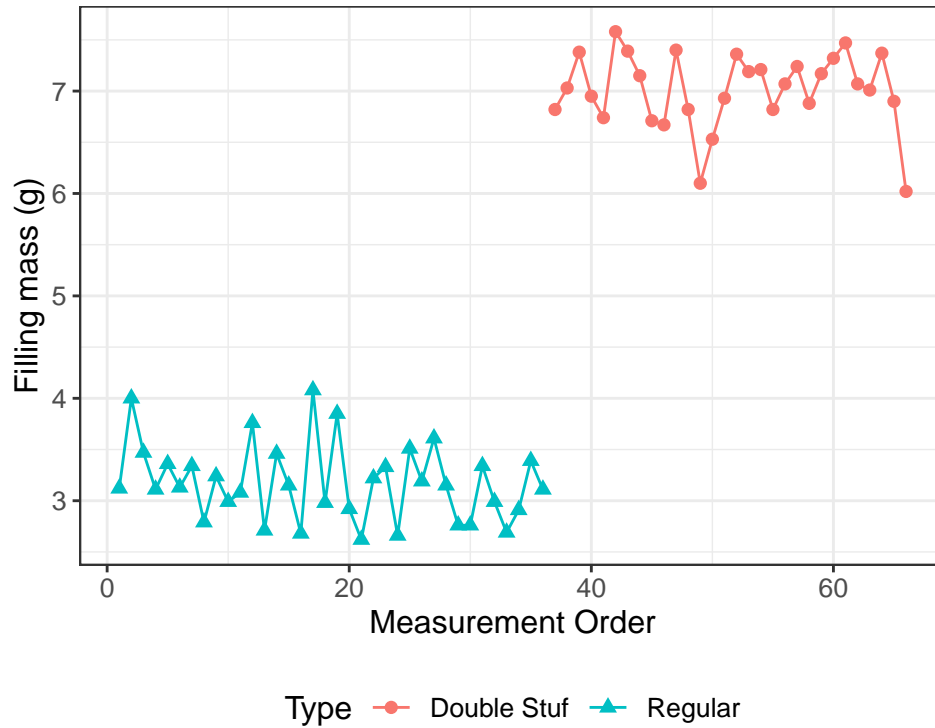[6]https://youtu.be/b__hDOHmF9nE?t=196

Figure 6: Index Plot for Log Filling Mass

larger than 2.

Bootstrapping

Given the potential issues with the t test, I decided to run a quick check via a bootstrap simulation.

Using a stratified resampling (by Type), I looked at 10,000 values of the ratio of values of the *SAM* (Double Stuf with respect to Regular). The 97% *percentile* confidence interval is (2.1, 2.311). This tells us that the middle 97% of the boostrapped sampling distribution for the ratio of *SAM* values is between 2.1 and 2.311. The Bias Corrected and Accelerated (BCa) version of this interval is (2.097, 2.308). Taken together along with the parametric test, I feel much more confident about my decision to reject the null hypothesis and decide to act as though Double Stuf Oreos have more than twice the filling of Regular Oreos.

Final Thoughts

After exploring the data and doing inference, I believe that Oreos are not engaged in false advertisement with their Double Stuf Oreos. While I can always find a pair of Oreos (one Regular, one Double Stuf) that shortchanges the consumer, I can also find pairs of Oreos where the consumer will have a Double Stuf oreo with more than twice the créme filling of a regular oreo. However, there is a preponderance of evidence here to convince me to give the benefit of the doubt to Nabisco. Future work could take this study as a starting point and expand out. For example, we could take a random sampling of Oreo packages from across the country to limit any geographical influences. Additionally, we might consider the age of the Oreos when we take measurements. Oreos that are older might have altered masses. The number of experiments abounds when we think about Oreos and their compositions.

Code Appendix

```r
# Setting Document Options
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center"
)

# Add additional packages by name to the following list
packages <- c("tidyverse", "knitr", "kableExtra", "psych", "car")
lapply(
  X = packages,
  FUN = library,
  character.only = TRUE
)

oreoData <- read.table(
  file = "https://raw.github.com/neilhatfield/STAT461/master/dataFiles/oreo3.dat",
  header = TRUE,
  sep = ","
)

# Initial Box plot
ggplot(
  data = oreoData,
  mapping = aes(x = Filling.Mass, y = Type, fill = Type)
) +
  geom_boxplot() +
  theme_bw() +
  xlab("Filling mass (g)") +
  ylab(NULL) +
  theme(
    legend.position = "none",
    text = element_text(size = 12)
  )

# Fix the data entry errors
oreoData[13, "Type"] <- "Regular"
oreoData[39, "Type"] <- "Double Stuf"

# Recreate the box plots
ggplot(
  data = oreoData,
  mapping = aes(x = Filling.Mass, y = Type, fill = Type)
) +
  geom_boxplot() +
  theme_bw() +
  xlab("Filling mass (g)") +
  ylab(NULL) +
  theme(
    legend.position = "none",
```

```r
    text = element_text(size = 12)
  )

# Load shadowgram function
source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/shadowgram.R")

# Create a shadowgram for the filling masses
shadowgram(
  dataVec = oreoData$Filling.Mass,
  label = "Filling mass (g)",
  layers = 60,
  aStep = 4,
  color = "blue"
)
# Create side-by-side histograms
ggplot(
  data = oreoData,
  mapping = aes(x = Filling.Mass)
) +
  geom_histogram(
    binwidth = function(x){
      return(
        ifelse(
          test = IQR(x) == 0,
          yes = 0.1,
          no = 2 * IQR(x) / (length(x)^(1/3))
        )
      )
    },
    col = "black",
    fill = "blue",
    closed = "left",
    boundary = 0
  ) +
  facet_wrap(
    facets = ~Type,
    scales = "free"
  ) +
  theme_bw() +
  xlab("Filling mass (g)") +
  ylab("Frequency") +
  theme(
    text = element_text(size = 12)
  )

# Get descriptive statistics by type of Oreo
groupStats <- psych::describeBy(
  x = oreoData$Filling.Mass,
  group = oreoData$Type,
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
```

```r
    IQR = TRUE,
    mat = TRUE,
    digits = 4
)

# Set row names as type of cookie; select useful columns
groupStats <- groupStats %>%
  tibble::remove_rownames() %>%
  tibble::column_to_rownames(
    var = "group1"
  ) %>%
  dplyr::select(
    n, min, Q0.25, median, Q0.75, max, mad, mean, sd, skew, kurtosis
  )

# Generate a professional looking table
groupStats %>%
  knitr::kable(
    caption = "Summary Statistics for Oreo Filling Masses",
    digits = 3,
    format.args = list(big.mark = ","),
    align = rep('c', 11),
    col.names = c("n", "Min", "Q1", "Median", "Q3", "Max", "MAD", "SAM", "SASD",
                  "Sample Skew", "Sample Ex. Kurtosis"),
    booktabs = TRUE
  )  %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("scale_down", "HOLD_position")
  )

# Transform the filling masses by applying the log
oreoData <- oreoData %>%
  dplyr::mutate(
    logMass = log(Filling.Mass)
  )

car::qqPlot(
  formula = ~ logMass,
  data = oreoData,
  subset = Type == "Regular",
  distribution = "norm",
  envelope = 0.90,
  id = FALSE,
  pch = 19,
  ylab = "Empirical quantiles"
)

car::qqPlot(
  formula = ~ logMass,
  data = oreoData,
  subset = Type == "Double Stuf",
  distribution = "norm",
```

```r
  envelope = 0.90,
  id = FALSE,
  pch = 19,
  ylab = "Empirical quantiles"
)

ggplot(
  data = oreoData,
  mapping = aes(
    x = seq_along(Filling.Mass),
    y = Filling.Mass,
    color = Type,
    shape = Type
  )
) +
  geom_point(
    size = 2
  ) +
  geom_line() +
  theme_bw() +
  xlab("Measurement Order") +
  ylab("Filling mass (g)") +
  theme(
    legend.position = "bottom",
    text = element_text(size = 12)
  )

# Perform t test on transformred data
tOut <- t.test(
  formula = logMass ~ Type,
  data = oreoData,
  mu = log(2),
  alternative = "two.sided",
  var.equal = FALSE,
  paired = FALSE,
  conf.level = 0.97,
  na.action = "na.omit"
)

# Set type as Factor since I hadn't already done so
oreoData <- oreoData %>%
  dplyr::mutate(
    dplyr::across(where(is_character), as_factor)
  )

# Define my estimator
ratioSAM <- function(data, index){
  means <- aggregate(
    Filling.Mass ~ Type,
    data = data,
    subset = index,
    FUN = mean,
    na.action = "na.omit"
```

```
  )
  return(means$Filling.Mass[2] / means$Filling.Mass[1])
}
# Set seed for reproducibility
set.seed(461)

# Run the bootstrap
bootOut <- boot::boot(
  data = oreoData,
  statistic = ratioSAM,
  strata = oreoData$Type,
  R = 10000
)

# Construct the confidence intervals
bootCI <- boot::boot.ci(bootOut, conf = 0.97, type = c("perc", "bca"))
```