HW #1.1: Exploring Oreo Data (Baseline)

Available: 8/25/23; Due: 9/1/23

Last Updated: 08/25/2023

The intent/goal for this assignment is to act as a baseline for you to demonstrate what you know how to do at the start of the course. This assignment will *not* impact your grade in the course. However, at the end of the semester, you will have an assignment where you must reflect upon your growth over the course of the semester. This assignment (HW #1.1) provides you with a clear point from which to see your growth.

There are three assignment links in Canvas for you to be aware of:

- HW #1.1–Oreo Comparison Value Check
  - This is where you'll get access to your assigned data set and where you'll be able to enter various values you're asked to calculate.
- HW #1.1–Oreo Comparison Write Up
  - This is the grading rubric that goes with the report portion of HW #1.1. There is actually nothing that you need to do in this area.
- HW #1.1–Oreo Comparison Submission Portal
  - This is where you'll actually upload your written report (DOC, DOCX, or PDF).

Please be aware that there are multiple data sets–you must use the one that is assigned to you in HW #1.1–Oreo Comparison Value Check

Background

Food product marketing often invokes strong feelings from consumers. For example, Subway was sued by consumers over their "foot-long" sub sandwiches not being 1 foot/12 inches long. In a similar vein, we might ask whether the marketing around one version of a product being "double" another version is accurate. Thus was born an investigation.

In 2017, I worked with an Introductory Statistics class to test out whether Double Stuf Oreos have twice the amount of stuf as Regular Oreos. ("Stuf" is the official term for the créme filling.) We designed a stochastic process (see end of document for more details) that allowed us to 1) select a grocery store in Tempe, AZ, via a lottery, 2) conduct a stratified random sampling of Regular and Double Stuf Oreo packages (9 packages each), and 3) weigh the amount of créme filling (in grams) for all cookies in each package. We separated the packages into 9 separate pairings (one for each team) of Double and Regular for data collection and analysis.

In this assignment, you will explore the same question as these students did (i.e., Do Double Stuf Oreos have double the créme filling mass as Regular Oreos?). Each student will be assigned one of three selected data sets. I will use another data set as a demo. **See the Getting Started Guides on the R/RStudio Resources and Example Code page in Canvas; these were built with this assignment in mind.**

Caution: don't assume that the data is error free.

Task Summary

For this assignment, you'll need to complete the following tasks. [MOM] indicates that there is something you need to do in the MyOpenMath; [Report] indicates that there is something you need to in your final report;

[MOM/Report] indicates that you need to do something in both the MyOpenMath and Report portions of the assignment.

1. Get and Read in the Appropriate Data using the software of your choice [MOM]
2. Explore the Data and Write a Data Narrative
    a. Clean your data
    b. Make appropriate data visualizations [Report]
    c. Find and interpret the values of various statistics [MOM/Report]
    d. Write your data narrative [Report]
3. Develop and Carry Out a Plan to Answer the SRQ
    a. State the Hypotheses/Models you will test [MOM/Report]
    b. Explain how you will test and *why* this test is appropriate [Report]
    c. Explain the results of the test [MOM/Report]
    d. State your decision [MOM/Report]
    e. Discuss any issues [Report]
4. Code Appendix–won't be graded, only examined [Report]

What follows are some notes (and hints) for approaching each of the tasks.

Task 1-Read in the Data.

Your first task is to get the data. Again, you'll need to check out the first question of the MOM component to see which data set you've been assigned. Be sure that you click the appropriate choice in the question to ensure that the rest of the questions are graded appropriately.

Once you have the data (you've downloaded the data or you've copied the URL to the file), you'll need to open it up in your chosen software.

Once you have the data saved, you'll need to read the data into your chosen software program. If you are using something other than R, I suggest that you copy/paste the appropriate URL into your browser and download the data file. Then open that in your selected program. Several software packages will import the data for you from the URL (e.g., R, JMP).

Hint 1–Use the URL to Read in Data.

If you are using **R**, then you can read the data with the following:

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

Note: you'll need to replace the example URL with yours.

Task 2-Explore your data and write a data narrative.

Explore your assigned data using your selected software package. Your answer here should come in form of a data story, complete with data visualizations, interpretations of various statistics, and plenty of discussion of both.

Hint 2

Be sure that you have checked your data for any cleaning that should be done. Look at data visualizations. Are there any cookies that seem to not make sense with the rest of their type? Anything looked switched? Be sure you describe any such issues and the remedy you implemented in your narrative.

Hint 3

When making data visualizations, be sure that you use an appropriate title and label the axes. Some useful commands to add to a `ggplot` call:

- `ggtitle("your title here")`
- `xlab("label for horizontal axis")`
- `ylab("label for vertical axis")`
- `theme_*()` (invokes a pre-built style; several options, plus you can build your own)

Hint 4

Control the size and position of your plots by adding options to code chunk header.

That is, `{r chunkName, option1=, option2=, etc.}`

- `fig.align='left/center/right'`
- `fig.height=7` and `fig.width=7` (in inches)

Hint 5

You can reference plots in your narrative in a couple of easy steps.

R Markdown

1. Build your code chunk to make a plot and give the chunk a name (no spaces in the chunk name).
2. In the chunk options, you'll need to define a caption: `fig.caption="caption text goes here"` (If you do this, you can omit the graph title.)
3. In your text type `\ref{fig:chunkName}` to call the reference to your plot.

For example, if I want to talk about the scatterplot (Figure 1), I can have the system keep track of the numbering for me.

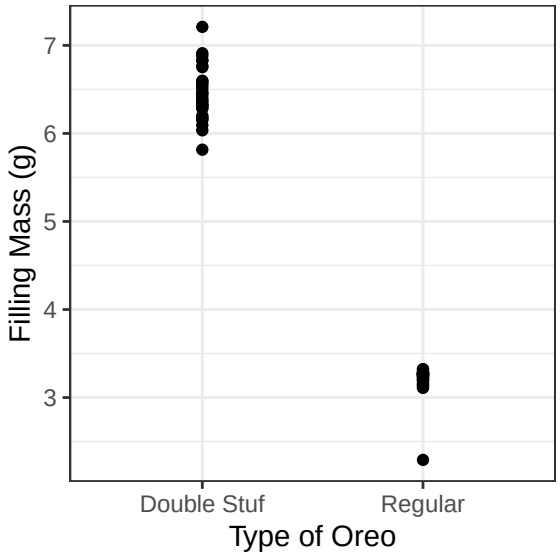

Figure 1: Scaterplot of Filling Mass (g) by Type of Oreo

Microsoft Word-Current Version

1. Insert your graphic into Word.
2. Click/Select your graphic, and then from the Insert Menu select Caption...

3. Fill in the appropriate information and press OK.
4. In a sentence where you want to add a reference, select Cross-reference... from the Insert Menu
5. Make the appropriate selections in the drop down menu and selection field. Press Insert when finished.

*Note: Word will update the numbering when you call up the Print menu (even if you don't print). You'll need to do this if you change the order of your figures, add/remove some, etc.

Requested statistics

You must include values of the following statistics in your narrative (and in MOM). You may include additional statistics as you wish. However, **you must provide an interpretation of one value from each statistic**. Please include the appropriate units of measurement where applicable. You may place them in a table BUT you must also incorporate them into your data narrative.

statistic	Regular Oreos	Double Stuf
Sample Max		
Third Quartile		
Sample Median		
First Quartile		
Sample Min		
Sample Arithmetic Mean		
Sample Arithmetic Variance		
Sample Skewness		

Task 3-Develop and Carry Out a Plan to Answer the SRQ

In this task, you'll need to explain how you're going to approach answering the SRQ using the tools of hypothesis testing. Then, you'll need to conduct a test and report all appropriate values. Finally, you'll need to discuss these results.

You will need to report any values of a test statistic (as well as degrees of freedom, if applicable), your p-value, and a **97%** confidence interval. You must interpret both the p-value and the confidence interval in your report.

Hint 6

Challenge yourself here. Think through what would be good methods to answer the question.

Hint 7

You know what they say about assumptions... CHECK THEM! In your narrative, discuss the assumptions, how you checked each one, and what you learned by checking them.

Hint 8

Type set your mathematical statements. In Microsoft Word use the Equation Editor. In R Markdown, use LaTeX expressions.

Task 4-Code Appendix

This will help us see if there are any issues with the code you ran (i.e., why your values might be incorrect). Your choice of software package will dictate what you put here.

GUI Bases Package–SPSS, JMP, Minitab, etc.

Please include a statement of which package you used, including version number (check out the "About" menu option for information). You may also copy your log/journal to the appendix.

Write a brief description of what parts of the package you used, but do not give a click-by-click accounting.

Command Line Package–R, SAS, etc.

Include the actual code you used, being sure to comment and organize your code.

Hint 9

Use the Homework Template RMD file (Also on GitHub.) to automatically create a code appendix for you.

Further Details on the Stochastic Process

The class designed the following stochastic process to generate the date for this study:

1) We complied a list of grocery stores located in Tempe, AZ through a Google search.
2) After giving each listed store a unique identifier, we placed those identifiers into a vector in an R session.
3) Using the `sample` function of [base] R, we selected one element from this vector.
4) Upon looking up the selected identifier, we traced back to the store and then made plans to go to the store.
5) Prior to going to the store, we used R to generated two lists (one for each type of Oreo) of sampled integers for package selection.
6) Upon getting to the store, we went to the main shelving area for Oreo cookies and assigned the packages an integer sequentially from the front-left-bottom package to the back-right-upper package. Assignment moved up through a stack, then back through the shelf, before moving right across the shelf. This was done separately for each kind of Oreo.
7) After packages where numbered, we used the max number as the limit and looked at the first pre-generated list of sampled integers. Integers higher than the max number were skipped, and we continued through the list until 9 packages were selected.
8) Cookies were purchased.
9) Each team arbitrarily selected one package of each kind of cookie to collect data from.
10) To get the mass of créme filling, the teams weighed each cookie as a whole (in grams), and then carefully split the cookies apart. The créme filling was removed (with some scrapping of the cookies) and the wafers of the cookies were then re-weighed. The difference in the two weights was then attributed to the créme filling.
11) Step 10 was repeated with each cookie in the package and for both types of Oreos.