# MA 207 – Spring 2018
### Lab #1: Descriptive Statistics, Histograms, and Statistical Thinking

- **Your work should be compiled in a Word document and turned in via Moodle.**
- Please write all of your answers in complete sentences using appropriate grammar, punctuation, and terminology.
- You should work in groups of two. You must both work on all problems.
- You will probably not be able to complete it within class time.
- Both students should submit the Word document, with both names at the top.

Refer to the *Instructions for Lab Write-ups* document on Moodle for information on how your lab should be formatted.

In an earlier class we gathered data from the M&M cups experiment. Each participant was given a cup to scoop some M&Ms out of a bowl. The data we gathered has been combined with data from the other sections of MA-207 and is compiled in the Excel spreadsheet *Lab1Data.csv* that is available on the Moodle page. Download the spreadsheet. In this lab we use Microsoft Excel and TinkerPlots to study this data set.

## Part 1: Getting Started.

For each of the following questions, write complete sentences explaining your answer:

a. What is the population of this study?

b. What was the sample in the study?

c. What was the primary research question we were trying to answer?

d. Why was randomization used in this study?

e. Before truly examining the data, do you anticipate there will be a difference in the number of M&Ms selected depending on the size of the assigned cup? If so, how big of a difference do you anticipate?

## Part 2: Explore the data.

Now we are going to examine the data, and do some basic computations

a. Sort the data by cup size so the people who were assigned large cups are listed first. *You may find this video helpful:* `https://youtu.be/bZvBaWQ1hnE`

b. Use Excel to compute the five-number summary for the number of M&Ms chosen by people with large cups. Also compute the five-number summary for the number of M&Ms chosen by people with small cups.

c. Use Excel to compute the mean and standard deviation of the large cup group and the small cup group.

d. Organize your answers to parts a. and b. in a nicely formatted table.

e. Now that we have some descriptive statistics for our experiment, do you believe that we have answered the primary research question? Thoroughly defend your answer using the descriptive statistics that you have calculated.

## Part 3: Creating Visualizations.

Now we are going to make a visual display of the data. One of the most useful visual displays used in Statistics is called a *histogram*. Begin by watching this video to learn how to use Excel and TinkerPlots together to conveniently make a histogram. If you would like your histogram to have rectangular bars, you can shift from *circle icon* to *fuse rectangle*. To get a copy of your histogram into your lab write-up just click on the plot in TinkerPlots, click Edit, select *Copy as Picture*, and paste it into your Word document.

a. Create histograms showing the distribution of the number of M&Ms chosen for the large and small cup groups.

b. Write a few sentences discussing the features you see in each of the plots. When we discuss plots in statistics, we can use the acronym C.U.S.S.: discuss the Center of the data, the Unusual features of the data, the Shape of the data, and the Spread of the data.

c. Compare and contrast the two plots you generated in part a. It is possible that there are differences between the genders in the *big cups* experiment. Create histograms to show the distributions of M&M counts for male students and for female students. Compare and contrast the data.

---

**Part 4: Giving estimates.**

We would like to estimate the number of M&Ms that will be chosen by a Carroll student who was assigned a large cup. Instead of giving a single number as an estimate (e.g. "exactly 50 M&Ms will be chosen") statisticians frequently give an interval or range of numbers as an estimate (e.g. "between 30 and 50 M&Ms will be chosen"). The larger the interval, the more likely it is that the statistician is correct, but the less useful the prediction. (The prediction "a student will choose between 0 and 1 million M&Ms" is absolutely true, but completely useless.)

a. Give an interval estimate for the number of M&Ms we expect a Carroll student who is not from Montana to choose if they were assigned a large cup and if they were assigned a small cup. Be sure to explicitly state how you are using descriptive statistics to create your range estimate. (Hint: You will want to do another sort and you will definitely want to find some descriptive statistics.) Include an appropriate TinkerPlots graph to support your reasoning.

b. Suppose we are interested in the variable of whether the student has eaten breakfast on the day of the M&Ms experiment. For students who ate breakfast, give an interval estimate for the number of M&Ms we expect the student to take (regardless of the size of their cup.) Similarly, provide an interval estimate for students who did not eat breakfast. Include an appropriate TinkerPlots graph to support your reasoning.

---

**Problem 5: Inference.**

Suppose that a student in your class claims to have collected 200 M&Ms with a small cup. Is your class data consistent with this claim or would your class data cause you to reject this claim? Explain your thinking using the evidence that we have gathered in this lab.