CLASS ACTIVITY: SAMPLING DISTRIBUTION - M&M

Burcu Eke Rubini  
Available under a [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) license

***To use this activity, start by making your own copy of this document by going to the menu above and selecting File > Download and choosing an option.***

# Learning Outcomes

Students will:

1. Understand the sample mean is a statistic.
2. Observe sample to sample variability.
3. Gain a basic understanding of a sampling distribution.
4. Investigate how the mean and standard deviation of the sample mean change as sample size increases.

# Teacher Preparation

In this activity, students are asked to collect data on the number of candies in a small sized candy bag (M&M fun size packs, smarties mini rolls, etc), form samples, calculate the sample mean, and the sample standard deviation. Students are also prompted to recall randomization and random sampling. Below is a suggested list of items for preparation.

* **Randomize students into samples.** We need a random process to sample students into different groups. If this step is shown to the students, face-to-face or online, it will be a review of randomization. One way this could be done is on Google Sheets:

1. Create a list of names of the students on a Google Sheets spreadsheet;
2. Select the list of names, and then right-click, and choose “Randomize Range”;
3. Now, assign the first two of the names into group 1 and the next two into group 2 and so on;
4. Next, select the list of groups on the Google Sheets spreadsheet, and then right-click, and choose “Randomize Range” again;
5. Then, assign the first two of the groups into sample 1 and the next two into sample 2 and so on;
6. Follow this procedure until the whole class is one big sample.

* **M&M Data.** We need to supply candy bags. Here are some options:

1. One option would be to bring a big bag of fun size M&M’s or any other candy for the students to count the number of candies of a specific color, say orange.
2. If the class is online, or the teacher chooses not to distribute candy bags in class, the initial data can be simulated. For the simulation, if the teacher is following the labs associated with OpenIntro Statistics, one can refer to the [Probability lab](https://www.openintro.org/book/statlabs/?statlab=probability). If not, a random number generator in any spreadsheet can be used. For reference, [here](https://docs.google.com/spreadsheets/d/1tDKWqqkjaM-hFN6N-ejVA4N_3sL6cGem-jBaR5WvTS0/edit?usp=sharing) is an example.

* **Data Collection.** We need to record the answers of the students so that we can refer back to this data set. Here are some options:

1. If it is available, teachers can create an editable spreadsheet like Google Sheets, Office 365, etc. This way, students can enter their data. If this is not an option, they can fill out a paper form.
2. Complete one row on the spreadsheet or show a filled out paper form as an example. Each row of the sheet will correspond to a student. [Here](https://docs.google.com/spreadsheets/d/1DKutALshCmBlfd_8m_CDdASYRJtJ_dqoudjQDrzSkTI/edit?usp=sharing) is one possible spreadsheet that can be copied/downloaded and modified.

* **Exercise Introduction.**

1. Bring up the question "Do you think all fun size M&M bags (or any other candy) have the same number of orange M&M’s?”
2. Ask each student to open their bag of M&M’s and count the number of orange M&M’s and record on the excel sheet.

# In-Person Classroom

1. Bring up the question “Do you think all M&M bags (or insert your type of candy) have the same number of orange M&Ms?”
2. Give everyone time (4-5 minutes) to record the number of orange candies in their bag. We will pretend that this is the population data.
3. Calculate the population average and population standard deviation.
4. Plot a histogram of this population data.
5. Randomly assign students into groups of two (samples of size 2) and ask the students to calculate the sample mean and sample standard deviation for each color. At this point, we will have class size (n) divided by 2 samples of size 2. Bring up the question “Do you think we have all possible samples of size 2?” (It may be a good idea to bring up the number of samples of size 2 - n choose 2.)
6. Calculate the sample average and sample standard deviation for all the samples of size 2.
7. Plot a histogram of the sample averages corresponding to samples of size 2.
8. Randomly assign the previous samples of size 2 to each other to form samples of size 4 and ask the students to calculate the sample mean and sample standard deviation for each color, again.
9. Calculate the sample average and sample standard deviation for all the samples of size 4.
10. Plot a histogram of the sample averages corresponding to new samples of size 4.
11. Repeat this process to samples of size 8, 16, 32, until you have the entire class in one group.
12. Bring up the questions “What differences do you observe for the sample means in larger numbers vs from smaller samples? How about the differences with respect to standard deviation and histograms?”
13. Bring up the question “Is there a pattern how the standard deviation of sample mean changes as sample size increases? Would it be nice to know exactly how much the change will be?”
14. Bring up the Central Limit Theorem so that the class can look at how the standard deviation of sample mean is related to the standard deviation of our entire population, i.e., it changes with

# Online Classroom

If the class is synchronous, activity can be carried out as part of an in-person class. If the class is asynchronous, we can:

1. Ask the students to fill out an online editable spreadsheet with their collected data.
2. Ask students to calculate means and standard deviations corresponding to samples of size 2, 4, 8, and so on.
3. Create a discussion board on similarities and differences of samples at a fixed sample size and similarities and differences of histograms for different sample sizes.

# Follow up

This exercise and its data could be reused under the following scenarios:

1. While talking about calculating Normal probabilities using the sampling distribution of .
2. While talking about one population confidence interval, we can calculate a confidence interval for the mean number of orange M&M’s per bag.