CLASS ACTIVITY: DESCRIPTIVE MEASURES - INSTAGRAM

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 difference between measures of central tendency and the need for measures of variation, and they will review sampling techniques.
2. Define and apply the descriptive statistics vocabulary.
3. Gain a basic understanding of Z-scores.
4. Collect data and calculate mean, standard deviation, and Z-scores.

# Teacher Preparation

In this activity, students are asked to collect data of likes/views from the last 10 posts on Instagram of athletes and movie stars and use this data to calculate mean, standard deviation, Z-scores, and compare group averages (athletes vs movie stars). They also recall randomization and non-random sampling. Below is a suggested list of items for preparation.

* **Randomize students into two different groups.** We need a random process to separate people into two 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 half of the names into group 1 (athletes) and the second half into group 2 (movie stars).

* **Data Collection.** We need to record the answers/scores 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 an Instagram account. [Here](https://docs.google.com/spreadsheets/d/1ZxvDUlBe_ZWP8LKsE2VxJP1wxHa9Qw6eJ0dYBB7sswI/edit?usp=sharing) is one possible spreadsheet that can be downloaded and modified.

* **Exercise Introduction.**

1. Bring up the question "Who are, on average, more popular: athletes or movie stars? Can we assess this by their likes on Instagram?”
2. Ask each student, based on their group, to choose an athlete or movie star, and record the number of likes/views on that person’s last 10 instagram posts. Question to students: “What kind of sampling is this?”

# In-Person Classroom

1. Give everyone time (5-7 minutes) to record the number of likes/views on the last 10 posts.
2. Ask students to calculate the mean and standard deviation corresponding to their chosen athlete/movie star on the spreadsheet;
3. Bring up the question: “Do 10K likes of person A and person B tell us the same story? Are they comparable? Would it be nice to bring them to a same/similar scale? What are the ways we can do it?”
4. Introduce Z-scores, and calculate the Z-scores corresponding to each post of each athlete/movie star using the sample mean and standard deviation calculated in step 2. Here, one can mention that we are assuming the last 10 observations are the entire population for the purposes of this step.
5. Notice that if the spreadsheet data entry is not available, steps 2 to 4 can be done in the following class, after the teacher enters the collected paper data into a spreadsheet.

# 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 mean and standard deviation corresponding to their chosen athlete/movie star on the spreadsheet.
3. Create a discussion board on compatibility of means of different athletes, movie stars, followed by Z-score calculations following the questions described for in-person step 3.

# Follow up

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

1. While talking about calculating Normal probabilities, we can refer to the Z-scores from this exercise. This may include a discussion of the appropriateness of the Normal model for a particular Instagram content creator. If the model is deemed reasonable, then Z-score calculations can be used together with the Normal probability calculations.
2. While talking about the two population t-tests, we can compare average likes of athletes vs movie stars. Notice that our data is not a random sample, and based on the class size, we may have a very small sample size. This means that while this exercise is a good way of demonstrating the mechanics of hypothesis testing, the results from the test are not meaningful. However, it will be beneficial in creating a discussion on when to and when not to perform a hypothesis test.
3. While talking about ANOVA.