CLASS ACTIVITY: CORRELATION - MOVIES

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. Define and apply correlation vocabulary.
2. Calculate correlation coefficient.
3. Understand that correlation is not causation.

# Teacher Preparation

In this activity, students are asked to rate movies or TV series, 10 being the best and 1 being the worst. Then, they are asked to interview a classmate for their ratings and calculate the correlation coefficient between two ratings. Below is a suggested list of items for preparation.

* **List of movies or TV series to rate.** We need a list of movies/series for students to rank. Some examples are IMDb "Top Rated Movies", IMDb "Best of 2020", or Netflix top 10 list:

1. Create a list of movies/series on a spreadsheet;
2. Add columns for each person's ratings. Here is an example [spreadsheet](https://docs.google.com/spreadsheets/d/1b8xNKVOA2TKIR3JFRZgR2k_nAiKsjJ3hGu1MiQF5tN8/edit?usp=sharing).

* **Someone to compare to (peer).** Students need to be matched with someone else to be able to generate data and calculate a correlation coefficient. Here are some options:

1. One option is to ask students to simply choose someone else in class to ask for their rankings.
2. If the class is online, or the teacher chooses to randomize pairs, the teacher can simply randomly assign people to calculate correlation coefficients together. (Randomization will better emphasize that correlation does not mean causation.)

* **Exercise Introduction.**

1. Ask each student to rate the movies from 1 to 10, 10 being the best.
2. Bring up the questions, "What do you think a correlation of 1 implies between your and someone else's movie ratings? How about a correlation of -1? Would it be easier to choose which movie to watch if your rankings are positively correlated or negatively correlated?"

# In-Person Classroom

1. Give everyone time (4-5 minutes) to record their peer's ranking.
2. Give everyone time (4-5 minutes) to calculate the correlation coefficient between their and their peer's rankings.
3. Ask peers if they ended up with the same correlation coefficient, i.e., if we interchange x and y in the correlation calculation, do we end up with a different number?
4. Bring up the question "Based on the correlation you calculated, how hard would it be for you two to choose a movie to watch?"
5. Bring up the idea of association by asking questions like "How did you decide on your ratings? Did you rate a movie 10 because of your peer's ranking or did you use your own judgement? If you did influence each other, can we go ahead and generalize this to everyone?"
6. Bring up examples of how we can in fact measure the causal effect of one variable on the other.

# Online Classroom

If the class is synchronous, the 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 ratings.
2. Ask students to calculate the correlation coefficient between their ratings and their pre-assigned peer's ratings.
3. Create a discussion board to discuss the meaning of high positive correlation (movie buddies), the meaning of high negative correlation (don't even try being movie buddies), and the interpretation of the correlation in general (correlation does not equal to causation).