

First Day of Class: Observational Units and Variables

Allan Rossman – Cal Poly, San Luis Obispo (www.causeweb.org/sbi/?p=295)

I think the most important goals for the first day of class are to give students a sense for what statistics is and to excite them about studying statistics.

So, you might find it surprising that I spend much class time on the first day with what might appear to be a boring task: identifying observational units and variables.

Why do I do this? For one thing, I think that identifying observational units and variables is a critical step in conducting statistical investigations. This step is key for determining how to analyze the resulting data: what graphs to produce, what statistics to calculate, what inference procedure to perform. Moreover, identifying observational units and variables can be surprisingly difficult for students, who need considerable practice to become comfortable with this. Students tend to make some common mistakes with this that I like to address head-on right away. For instance, three common mistakes involve confusing a variable with a research question, with a statistic that summarizes the distribution of the variable, and with an individual category of the variable. Another misconception to be confronted early is the tendency to think that observational units are always people.

But more importantly, this exercise allows me to introduce students to diverse and fascinating kinds of research questions that can be addressed with a statistical study. I present students with a series of research questions, asking them to identify the relevant observational units and variable(s) in a statistical study to address those questions. These research questions all come from studies that students will examine later in the course, either through examples or assignments.

Let me illustrate with by showing three examples that I ask students to work through on the first day of class. They discuss these with nearby students, and then we have a class discussion to compare answers.

Example 1: Variables about You Consider the students in this class as the observational units in a statistical study. For each of the following variables, indicate whether the variable is *categorical* or *numerical*.

1. How many *Harry Potter* books have you read?
2. Have you ever been outside the United States?
3. What would you guess for the instructor's age?
4. How many units of classes are you taking this quarter?
5. On what day of the week were you born?
6. Do you consider yourself an "early bird" or a "night owl"?

Example 2: More about You Continue to consider students in this class as the observational units in a statistical study. Explain why the questions in #1 and #2 are not variables.

1. What is the average number of *Harry Potter* books read by a student in this class?
2. What percentage of students in this class have been outside the United States?
3. What would the *observational units* have to be in order for these questions in #1 and #2 to be legitimate variables? (*Hint*: The observational units would *not* be people.)

Explain why the questions in #1 and #2 below are not variables.

1. Have female Cal Poly students read more *Harry Potter* books, on average, than males?
2. Do the percentages of Cal Poly students who have been outside the U.S. differ among the colleges?

Example 3: Preview of Coming Attractions We will study examples that address the following statistical questions in this course. For each one, identify the observational units and variable(s). Also classify each variable as numerical or categorical.

1. How much did an average American consumer spend on Christmas presents in 2013?
2. Are Reese's Pieces candies equally likely to be orange, brown, and yellow?
3. Do college students who pull all-nighters tend to have lower grade point averages than those who do not pull all-nighters?
4. Do cars driven by students on your campus tend to be newer than cars driven by faculty on your campus?
5. Is the residence situation of a college student (on-campus, off-campus with parents, off-campus without parents) related to how much alcohol the student consumes?
6. Can you predict how far a cat can jump based on factors such as its length and weight and takeoff velocity?
7. Are students' scores on essay assignments affected by whether they get to choose their own deadlines, have specific deadlines imposed, or have no deadlines at all?
8. In what percentage of baseball games does the winning team score more runs in one inning than the losing team scores in the entire game?
9. Is the age at which a child first speaks related to his/her cognitive aptitude?
10. Do people tend to scoop larger helpings of ice cream for themselves if they are given larger spoons and larger bowls, as opposed to smaller spoons and smaller bowls?
11. Does wearing socks over footwear decrease the risk of falling while walking on an icy incline?
12. Among heterosexual couples, who is more likely to say "I love you" first: the man or the woman?
13. Statistical evidence was used in the murder trial of Kristen Gilbert, a nurse who was accused of killing patients. More than one thousand 8-hour shifts were analyzed. Was the proportion of shifts with a death significantly higher for the shifts that Gilbert worked?
14. Are teenagers in the United Kingdom more likely to have read a *Harry Potter* book than teenagers in the United States?
15. Are kissing couples more likely to lean their heads to the right than to the left?

After I distribute and discuss the course syllabus, discussion the questions above takes most of a 50-minute first class session. Then, to emphasize that students are responsible for learning something in every class meeting, including the very first one, I typically end the class with a short quiz that students can take in groups, along the lines of the following.

Quiz 1: Suppose that I examine information for all purchases that I have made on amazon.com this year, so those purchases are the observational units in a statistical study. For each of the following, indicate whether it is a categorical variable, a numerical variable, or not a variable at all.

1. How much did I spend on the purchase?
2. Was the purchase shipped to me or to someone else?
3. Do I tend to spend more on purchases sent to me than on purchases sent to others?
Now consider the observational units in a statistical study to be all visitors to Yosemite National Park in the past month.
4. Identify a *categorical* variable that could be recorded on these observational units.
5. Identify a *numerical* variable that could be recorded on these observational units.

I believe that this opening day of class helps students to develop an important skill that will serve them well throughout the course, while at the same time giving them a sense for some interesting questions that can be addressed in a statistical study. I hope that this makes students look forward to the rest of the course.