Random Sampling

When researchers have a question concerning a large population, they obtain a **sample** (a part) of that population. That is because it is typically impossible to study the entire population.

For example, if you want to know how the citizens of France feel about climate change, you cannot just go and ask every person in France about it. You would choose for example 600 French citizens as your sample and ask them your question.
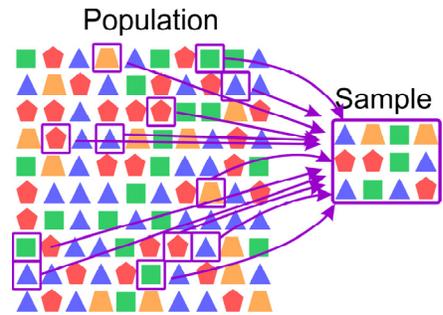
The way a sample is chosen is very important. Some methods of sampling may produce a sample that is *not* representative of the entire population. We call that a **biased sample**.

For example, if you are studying a student population of 630 in a school with close to an equal number of boys and girls, and you happen to choose a sample of 20 boys, then your sample is biased. It doesn't represent the entire population well.
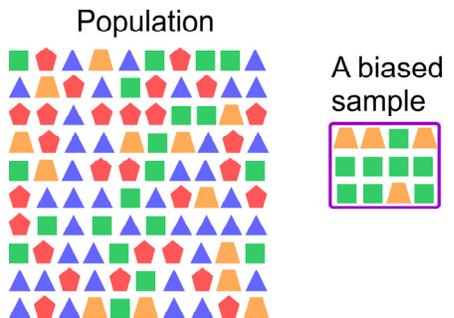
We need to use **unbiased sampling methods** in order to get a sample that truly represents the population being studied. The best way to avoid biased samples is to select a **random sample.**

The main characteristics of a random sample are:

1. **Randomness:** each member of the population has an equal chance of being selected.

   Let's say a researcher is studying the types of cars Americans own. He decides to interview only people he finds at a local mall because that mall is close to where he lives, so it is convenient for him. His sample is biased because not every member of the US population even has a chance to be selected in his sample. Maybe the people at his local mall are predominantly rich people who own several cars per family, so in that respect those people would not be a good representation of the entire population of the US.

   We call this type of sample a **convenience sample** because it is convenient or easy to obtain.

2. **External selection:** respondents must be chosen by the researcher, not self-selected.

   If our researcher mails a questionnaire to various people across the US asking them to fill it out and return it, his sample is a **voluntary response sample**, which is a biased sample. Some people volunteer to return the questionnaire, but others don't. The people themselves decide whether or not to be a part of the sample.

   Why might this be a problem? Some of the people who would choose to take part may have an external reason to do so. They might want to show off how "good" they are in the particular aspect being studied, or they might just like to speak out about their opinions.

Our researcher could get a true random sample by choosing people randomly from a list of people living in the US and calling them. That way, each person has an equal chance of being selected in the sample (it is random), and the people cannot self-select to take part (the researcher chooses who takes part).

An unbiased sampling method is more likely to produce a representative sample.

1. You are studying whether students in a large college prefer to drink coffee black, with milk, with cream, or with sweetener, or whether they prefer not to drink coffee at all.

   a. Which of the six sampling methods listed below produce a voluntary response sample?

   b. Which methods don't give each member of the student population an equal chance to be selected for the sample?

   c. Which method is likely to produce a sample with only coffee drinkers, overlooking those who don't drink coffee?

   d. Which method will be the most likely to give you a representative (unbiased) sample?

   <u>Sampling Methods</u>

   (1) You interview 80 students in a cafe on the campus.

   (2) You interview 80 students who come in at the main door of the campus.

   (3) You interview the first 80 students you happen to meet on a certain day.

   (4) You choose 80 names randomly from a list of all the students. You call them to interview them.

   (5) You send an email to all the students in the college, asking them to fill in a form on a web page you have set up. You hope to get at least 80 responses.

   (6) You choose 80 names randomly from a list of all the students. You send them an email, asking them to fill in a form on a web page you have set up.

2. A recipe website posts a poll on their home page that any visitor to that website can take. In it, they ask if people are looking for a recipe for a dessert, a main dish, a side dish, bread, or salad. During the course of one Sunday, 4,600 people visit the page, and 252 of them fill in the poll. Explain why the poll results will be based on a biased sample.

**Sample worksheet from
https://www.mathmammoth.com**

43

Statistics & Probability (Blue Series)