

Introduction to Statistics

Suppose that we decide that we want to find out which brand of soda pop is most popular in the United States today. To do this, we pick an afternoon to take a poll. We stop students as they enter the student union and ask them to indicate whether they prefer Coke, Pepsi, Mountain Dew, or “Other”. Although many people won’t take the time to stop and answer our survey question, we keep at it until we get 50 total responses. We find that of the 50 people we asked, 25 say they prefer Mountain Dew, 10 say they prefer Pepsi, 8 say they prefer Coke, and 7 say that they prefer “Other”.

Here is some vocabulary that we will need:

1. **Statistics** is an area of mathematics which is used to gather, organize, analyze, interpret, and make predictions from information called **data**.
2. The universal set of objects we are interested in getting information about is called the **population**, while the subset we actually collect data from is called the **sample**.
3. Ideally, a sample should be chosen randomly so that every member of the population has an equal chance to be in the sample. That way, the sample will reflect the entire population we are interested in finding out about rather than just some subset of the population.
4. If our sample is *not* random, it will almost certainly be **biased** in some way.