Population and Sample

One goal of every good survey is to learn certain information about a *population* of people. A *population* is every person with a characteristic of interest. For example, the Campus Experiences Survey is interested in the experiences of all current UTEP students. In this case, the *population* includes every current UTEP student. In a presidential election, pollsters are often interested in the opinions of people who might vote in the upcoming election. In this case, the *population* would include all registered voters.

It is often difficult to measure every member of the population of interest. During presidential elections, many organizations are interested in which candidate people are likely to vote for; however, it would be nearly impossible to survey every person who intended to vote in the election. In cases where the entire population cannot be measured, a *sample* of the population is used.

A *sample* is a subset of the population of interest. If the sample *represents* the population, information from the sample can be used to draw conclusions about the population of interest. For example, if we are interested in knowing the average height of UTEP students, using the women's basketball team as a sample of the UTEP population would probably not provide accurate information about the UTEP population as a whole. The women's basketball team is probably not *representative* of the entire UTEP student body in terms of height.

Random Sampling

One way to ensure a representative sample is to use random sampling. In *random sampling*, every member of the population has the same chance of being part of the sample. This means that the tallest person on campus, the shortest person on campus, and a person of exactly the average height on campus all have the same chance of having their height measured.

Sampling Error

Since a sample does not include every member of the population of interest, the sample value may differ from the population value. In other words, even if we achieve a representative sample of UTEP students, the average height of our sample of students is likely to differ from the actual average height of all UTEP students.

The discrepancy between our sample value and the population value is called *sampling error*. Differences in sample and population values are expected by chance alone. That is, we don't expect to draw a sample of UTEP students whose mean height perfectly match the mean height of all UTEP students.

Margin of Error

One way to express sampling error is by using the margin of error. The *margin of error* is a measure of the precision of a sample estimate of the population value. It uses probability to demonstrate the precision of a sample estimate by providing a range of values in which a sample value would be

expected to fall. In general, margin of errors provide a 95% confidence interval. That is, if 100 more surveys of UTEP student height were undertaken, 95 out of 100 times, the mean height would fall into the range given by the margin of error from the first survey.

Let's return to the idea of an election survey for an example. Knowing that a survey found that 60% of people support Candidate X doesn't tell us much without knowing the margin of error. If the margin of error is +/- 15%, we would expect the true population support for Candidate X to be between 45% and 75%. Though at first it would appear that the majority of people support Candidate X, the large margin of error casts doubt on this conclusion. If, however, we know that 60% of people support Candidate X with a margin of error or +/- 3% we would expect the true population support for Candidate X to be between 57% and 63%. We would have much more confidence in the conclusion that the majority of people support Candidate X.

Caveats for interpreting the Margin of Error

There are several cautions for interpreting a margin of error. First, the margin of error does not include *nonsampling error*; it only includes *sampling error* (differences from the population value due to chance). *Nonsampling error* may occur when respondents lie to the surveyor (e.g., say they plan to vote for Candidate X when they actually plan to vote for Candidate Y) or when people who are not members of the population of interest are included in the survey (e.g., including the height of a UTEP faculty member in a survey of UTEP student height).

Another concern is the *nonresponse rate*. This is the number of people who refused to complete either the entire survey or a certain question of interest. It can be difficult to draw strong conclusions from surveys with high nonresponse rates, especially if the people who did not respond differ in opinion from the people who did respond.

References

Scheuren, Fritz (Ed.). (1997). What is a survey? Alexandria, VA: American Statistical Association.

       Available online at http://www.whatisasurvey.info/.