Commonly Used Survey Terms

Bias:

A systematic deviation from a population parameter caused by something other than random error. For example, an interviewer can ask a question such as "who are you supporting in the upcoming election?" as "You are supporting candidate X in the upcoming election aren't you?" Asking the question in that way will systematically bias the responses towards candidate X.

Cluster:

Distinguishable geographical divisions such as counties, cities, census tracts, blocks, etc. within a population that contain sample elements and are used for sampling. Unlike strata, which are also used in sampling, **not all** clusters are chosen to be in the sample.

Confidence Interval:

An interval around a statistic in which the true population parameter would fall "X" percent of the time (for example 90% confidence level or 95% confidence level) if the statistic and interval were constructed the exact same way on many different samples. The statistic plus or minus the standard error is a 68% Confidence Interval. The statistic plus or minus 1.64 times the standard error is a 90% Confidence Interval. The statistic plus or minus 1.96 times the standard error is a 95% confidence interval.

Elements:

Are the basic entity that a researcher would like to study (for example, people, states, organizations, dogs, cars, etc.).

List Assisted Random Digit Dialing:

A modification of the RDD in which numbers are randomly generated within an interval (say 100 numbers) if there is a listed telephone number within the block of 100 numbers. This is a much more efficient way of generating actual telephone numbers within the population. The intervals are often called "banks" and referred to as single banks if there is one listed number within it, zero banks if there are no numbers within it, and 2 banks if there are 2 listed numbers within it.

Parameter:

An actual characteristic (a mean, a median, a rate, a number of count of elements – for example, the number of people uninsured - etc.) of a population. It is estimated using a statistic that is calculated from a sample of population elements.

Population:

The entire group of elements that a researcher is interested in learning about.

Precision:

Is the amount of error associated with a statistic. One way to gauge precision is to divide the standard error by the statistic (also called a Relative Standard Error).

Primary Sampling Unit:

The first division among the distinguishable clusters. For example, if a researcher was doing a state survey and used counties as clusters and randomly selected ten counties for inclusion in the survey (and left the rest of the counties out of the sample), the county would be the primary sampling unit.

Random Digit Dialing (RDD):

A common method used for generating a sampling frame for a telephone survey in which possible phone numbers are randomly generated by using a computer program (the computer is limited to actual area codes, and three digit prefixes).

Random Error:

A non-systematic deviation of a statistic from a population parameter. Sampling error is one type of random error that is controlled by using (inferential) statistics.

Sample:

A subset of elements from a population drawn into a survey.

Sample Size:

The number of elements from the population drawn into the survey.

Sampling Frame:

A list of elements from a population that is used for sampling. The sampling frame almost never covers the entire population of interest. For example, if a researcher were interested in studying people in an organization (the population), an organizational roster (the sampling frame) may be used to sample people within the organization. However the list may not include new members of the organization, and it may not include people who have left.

Sampling Variability:

Variation between a statistic and the population parameter that is likely to occur if a researcher performed the exact same measurement method multiple times on the exact same population. The larger the sample size the smaller the sampling variability.

Standard Deviation:

Is a uniform measure of the "average" amount of **deviation of a population element from a population parameter**. It is always in the same metric as the population parameter's metric (for example, if the parameter is a percentage, then the standard deviation is a percent. If the parameter is a number of people, then the standard deviation is the number of people). It is equal to the square root of the variance.

Standard Error:

A uniform measure (a "standard") of the "average" amount of deviation of a **statistic from the population parameter**. It is used to evaluate the amount of sampling variability ("error") associated with a **statistic** (e.g., mean, rate, coefficient, etc.). The metric is always the same as the population parameter's metric (in other words, if the statistic is a percentage, then the standard error is a percent. If the statistic is a number of people, then the standard error is the number of people).

Statistic:

A characteristic (for example, a mean, a median, a rate, a count of elements with a characteristic – for example, the number of people uninsured - etc.) of a sample, used to infer the population characteristic.

Strata:

Distinguishable groups of elements within the population that a researcher may be specifically interested in (for example, strata could be counties, racial groups, or people without health insurance). Researchers sample within **all** the population strata of interest but often oversample some relative to others in order to get less variability ("error") in a statistic calculated using data from a particular stratum.

Variance:

Is a uniform measure of the squared "average" amount of **deviation of a population element from a population parameter**. Its metric is always the squared metric of the population parameter's (for example, if the parameter's metric is a percentage, then the variance's metric is in a squared percentage. If the parameter is the number of people, then the variance is the squared number of people). All the elements used to construct a statistic are subtracted from the parameter, squared, added up, and divided by the number of elements. It is the square of the standard deviation.