

Statistical Reasoning

Sampling Methods

Name: _____

Date: _____

Class: _____

The entire group of individuals that we want information about is called the

population.

A census is an attempt to gather information about every member of population.

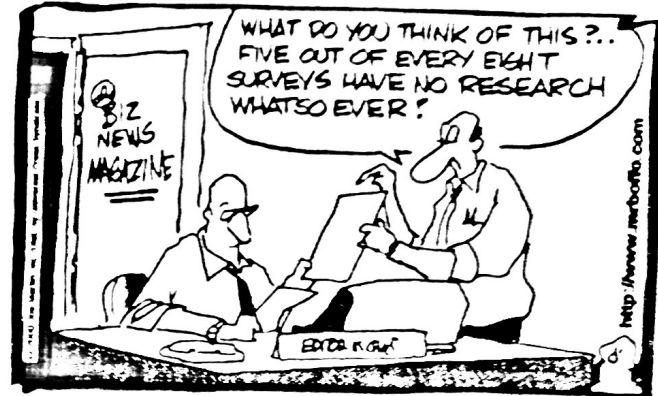
Problems with census—costs; time needed to complete; sometimes testing can destroy item.

A sample is a part of the population that we actually examine in order to gather information.

The design of a sample refers to the method used to choose the sample from the population.

The design of a study is **biased** if it systematically favors certain outcomes.

MISTER BOFFO By Joe Martin



I. Probability sampling:

Obtaining a randomly chosen sample refers to the way in which the sample is chosen, not the specific members of the population that happen to end up in the sample.

A. A sample is a simple random sample ^{SRS} if it is selected so that:

- each member of the population is equally likely to be chosen;
- the members of the sample are chosen independently of one other;
- every set has an equal chance to be the sample actually selected.

"Classic" method—put names in hat and draw until have desired sample size; more commonly, number names and use random number table or other source of random numbers to select sample.

Sources of bias in probability samples:

- Selection Bias (under-coverage) occurs when some groups in the population are left out of the process of choosing the sample (i.e. telephone

survey excludes those without phones). Occasionally random sampling may give a sample that is not representative of the population—but it is still random.

- Measurement Bias occurs when the method of observation tends to produce values that systematically differ from the true value in some way. (i.e. improperly calibrated scale is used to weigh items; poorly worded survey questions)
- Response Bias occurs when individuals chosen for the sample can't be contacted or refuse to cooperate.

Difficulties with probability sampling—may be no easy way to list all members of a population or to contact some people. Can use: voter registration lists—not everybody is registered; list of all addresses—homeless, multiple people at same address; telephone numbers—unlisted phone numbers, people with multiple numbers are overrepresented cell phones

B. Other probability sampling methods for obtaining large samples:

1. Systematic sampling—randomly choose some starting point; then select every k^{th} element in the population. Easier than random sampling; also guarantees sample is taken from throughout the population—be careful how population is ordered. (i.e. take every 10th student at CVHS; start at randomly selected point—still requires list)
 2. Cluster Sampling—divide the population into sections or clusters; randomly select a few of those sections and then choose all members from the selected sections (i.e. randomly select 10 third period classes—survey all students in each class selected—faster way to obtain sample; assumes classes are relatively heterogeneous).
 3. Stratified random sample—subdivide the population into at least two different subpopulations (strata) that share some characteristic; then draw a random sample from each stratum. The purpose is to insure that the sample is more representative of the population than a SRS might be; also might be interested in results from separate strata. (i.e. select SRS from each grade-level at CVHS—9th, 10th, 11th, 12th—in proportion to number of students in each level)
- multi-stage design—uses multiple strata (i.e. stratify by gender as well as grade-level—select proportionately sized SRS from each).

II. Other sampling techniques (non-probability—BAD!): (All hypothesis testing that we do requires probability sampling.)

Convenience sampling—easiest way to obtain a sample is to choose it without any random mechanism (also called haphazard sampling). Simply uses a sample that is readily available—often biased.

self-selected (voluntary) —when people participate in a survey by voluntarily responding. People who care enough to respond usually are not representative of the whole population. Not random. STATISTICAL INFERENCE SHOULD NOT BE DONE ON A VOLUNTARY POLL.

2. judgement sampling —a form of convenience sampling where an "expert" selects a sample s/he considers representative. Not random—no objective way to quantify this kind of sample.
3. Quota sampling —type of convenience sampling using clusters and strata. Interviewers have detailed strata definitions, assigned locations and must find fixed number of subjects for each stratum; still not random.

Random sampling is needed for further calculations, but it is often difficult to do, especially for large populations. (Convenience sampling is easy to do but not very useful—may be representative/ balanced for variables in strata but not for population as a whole.)

The wording of questions is the most important influence on the answers given to a sample survey. Confusing or leading questions can introduce measurement bias.

Even minor changes in wording can change the outcome of a survey—opinion survey (often in interview format) to check clarity of questions.

Never trust the results of a sample survey until you have read the exact questions posed. Also look at the sampling design, the size sample (larger samples generally give more accurate results), the amount of non-response and the date of the survey.