A sample

- A finite part of a statistical population whose properties are studied to gain information about the whole
  - A set of respondents selected from a larger population for the purpose of a survey or experiment.

Sampling

- The act, process, or technique of selecting a suitable sample, or a representative part of a population for the purpose of determining parameters or characteristics of the whole population.

Types of samples

- Probability sampling
  - Simple random sample
  - Systematic random sample
  - Stratified sample
  - Cluster sample
  - Multistage sampling
- Nonprobability sampling
  - Convenience sampling
  - Purposive sampling

The probability sample

- Any method of sampling that utilizes some form of random selection.
- Select among participants randomly based on some probability.
- Requires setting up some process or procedure that assures that the different units in your population have equal probabilities of being chosen.

Simple random sample

- Obtained by choosing elementary units in search a way that each unit in the population has an equal chance of being selected.
- Advantages

- Disadvantages

Stratified sample

- A population can be divided into different groups (or strata) based on some characteristic or variable
- A stratified sample is obtained by selecting a separate simple random sample from each population stratum.
- Why you might prefer stratified sampling over simple random sampling:
  - It assures that you will be able to represent not only the overall population, but also key subgroups of the population, especially small minority groups.
  - If you want to be able to talk about subgroups, this may be the only way to effectively assure you'll be able to.

Example

- We're interested in satisfaction with our clinic
- Assume the population of clients can be divided into three groups: Caucasian (C), African-American (AA) and Hispanic-American (HA).
- Both AA and HA are relatively small minorities of the clientele (10% and 5% respectively).
- With a simple random sample of n=100, we would expect we would only get 10 AA & 5 HA. And, by chance, we could get fewer.
- If we stratify, we can do better, and we know we'll have enough people in each subgroup to make meaningful inferences.

Systematic random sample

- Obtained by selecting one unit on a random basis and choosing additional elementary units at evenly spaced intervals until the desired number of units is obtained.
- Example:
  - There are 100 students in your class.
  - You want a sample of 20.
  - You have a class listing in alphabetical order.
  - Divide 100 by 20, you will get 5.
  - Randomly select any number between 1 and five. That is your starting number.
  - From there select every 5th name until you reach the last one.

Systematic random sample, cont.

- Disadvantages:

- Advantages:

Example: Library study

- How frequently do books at the library circulate?
- You have a card catalog where the entries are arranged in the order in which they're on the shelf
- To do a random sampling, you could estimate the total number of books and generate random numbers to draw the sample
  - But how would we find book #74,329 easily if that is the number we selected? Count items in catalog until we come to 74,329?
- Stratifying will not solve this problem.
  - We could stratified based on drawers in card catalog, but we'd still be stuck counting within that drawer.

Library study, cont.

- Systematic random sample:
  - If there are 100,000 books, and you want a sample of 1000, you'd want to select books at 100-book intervals (100,000 / 1000).
  - Select a random integer between 1 and 100 for your starting point, and then go every 100 books from then on.
  - Since it is the same number each time, you could figure out how many inches in card catalog it took to do 100 books, and then you'd know how far along in your list to jump to - so you wouldn't have to count each time.

Cluster sample

- Obtained by selecting clusters from the population on the basis of simple random sampling.
- Identify groups within the population, and then randomly sample groups.
- Example: if you want to look at schoolchildren in MD, rather than sample randomly among all kids in the state, you might consider each school to be a cluster and randomly select certain schools – you then test all the kids within that school.
- Though very economical cluster sampling is susceptible to sampling bias.

Multistage sampling

- Combine different methods of probability sampling.
- Example: Using cluster sampling to select certain schools, and then doing random samples within that school.

Probability sampling in hearing & speech

- In our field, probability sampling is often used for clinical field tests
- But other than that, probability sampling is rarely used
  - It is almost impossible to get a complete listing of all members of the population in order to select among them
  - The cost of doing so would be prohibitive (both in terms of time and money)

Nonprobability sampling

- Does not involve random selection.
- Can be divided into two broad types: accidental/convenience or purposive.

Probability vs. nonprobability

Probabilistic samples

- We know the odds or probability that we have represented the population well.
- We are able to estimate confidence intervals for the statistic.
- Generally more accurate and rigorous, but not always feasible

Nonprobability samples

- We may or may not represent the population well
- It will often be hard for us to know how well we've done so.

Convenience sample

- The more convenient elementary units are chosen from a population for observation.
- Examples:
  - Traditional "man on the street" interviews by TV programs
  - Use of college students in research
  - Using clients who attend our clinic as our sample.
  - Using local volunteers.
- The problem with all of these types of samples is that we have no evidence that they are representative of the populations we're interested in generalizing to -- and in many cases we would suspect that they are not.
- This is the kind of sampling we do most often.

Purposive sampling

- We have one or more specific predefined groups we seeking.
- Can be very useful for situations where you need to reach a targeted sample quickly and where sampling for proportionality is not the primary concern.
- There are many different subcategories of purposive sampling. In all of these methods we know what we want -- we are sampling with a purpose.

Modal instance sample

- Sampling the most frequent case, or the "typical" case.
- Example: interviewing a "typical" voter.
- Problem: How do we know what the "typical" or "modal" case is?
  - We could say that the modal voter is a person who is of average age, educational level, and income in the population.
  - But are averages the best? (consider the skewed distribution of income, for instance).
  - How do we know that those three variables -- age, education, income -- are the most relevant for classifying the typical voter? What if religion or ethnicity is important?

Expert sampling

- Assembling of a sample of persons with known or demonstrable experience and expertise in some area - a "panel of experts."

Quota sampling

- Select people nonrandomly according to some fixed quota.
- Proportional quota sampling:
  - Represent the major characteristics of the population by sampling a proportional amount of each.
- Nonproportional quota sampling:
  - Specify the minimum number of sampled units you want in each category.
  - Typically used to assure that smaller groups are adequately represented in your sample.

Heterogeneity sampling

- Sampling so as to include all opinions or views, but not representing these views proportionately.
- Purposefully picking a wide range of variation on dimensions of interest.
  - AKA sampling for diversity, maximum variation sampling.
- Documents unique or diverse variations that have emerged in adapting to different conditions.
- Identifies important common patterns that cut across variations.
- Example: In interviewing MD students, you may want to get students of different nationalities, backgrounds, cultures, work experience…

Snowball sampling

- Begin by identifying someone who meets the criteria for inclusion in your study.
- Then ask them to recommend others who they may know who also meet the criteria.
- Does not lead to representative samples.
- May be the best method available for reaching populations that are inaccessible or hard to find.
- Example: Studying the homeless
  - You will not find lists of homeless people.
  - However, if you identify one or two, they may know the other homeless people in their vicinity and how you can find them.

Extreme and deviant case sampling

- Learning from highly unusual manifestations of the phenomenon of interest
- Example: studying outstanding successes, notable failures, top of the class, dropouts, exotic events, crises, CI stars

Criterion sampling

- You set a criteria and pick all cases that meet that criteria
- Example: all women six feet tall, all white cars, all farmers that have planted onions.
- Most often used in quality assurance.

Problems with nonprobabilistic sampling

- Even if the sample size is large enough, the data may not generalize if it isn't representative of the whole.
  - Ex: If you asked shoppers in an organic food store what they thought of pesticide use, their replies would not be representative of the rest of the country.

Text from Jim Norton, http://www.info-pollution.com/evidence.htm

Problems with nonprobabilistic sampling

- Here are some examples from news stories:
- A survey of college students found that music was the most important thing in their lives.
  - Problem: It was soon discovered that it was a survey of music majors.
- Another survey found that Hawaii was the favorite vacation destination for Americans.
  - Problem: the survey used data from travel agencies, and not all vacationers use travel agencies. According to most surveys, Florida is the favorite vacation destination for Americans.

Text from Jim Norton, http://www.info-pollution.com/evidence.htm