EDUR 7130
Presentation 3b

1. Sampling

1a. Identifying a Sample

1b. What is the purpose of sampling?
To select a sample (i.e., subgroup of population) to represent a population so we can learn about the population.

The goal of sampling is to make inferences from sample to population, so it is important that the sample be representative of the population.

1c. What is a representative sample?
Sample represents or reflects a population when the sample "looks like" or "resembles" the population in most ways possible.

	Population	Sample 1	Sample 2	Sample 3
	N = 1000	n = 100	n = 40	n = 4
Male	70%	67%	60%	50%
Female	30%	33%	40%	50%

The larger the sample, the closer should be the match between sample and population.

1d. What is meant by inference in sampling?
Sampling is used to learn about populations; what we find in sample we infer to population (we assume what was found in sample applies to population).

It is critical to get a representative sample so inferences are accurate.

1e. When can inferences from the sample to the population be wrong?
Unrepresentative sample = inaccurate inferences

Key is representative nature of sample, not size of the sample (although larger samples tend to be better at representing the population).

Example 1:

National surveys of US voters usually contain between 600 and 900 respondents, and that small sample is used to generalize to 50+ million voters.

Example 2:

- Population = an all-female class
- Researchers know class is all one sex, but do not know which sex

What size sample would be needed to determine sex composition of this class?

Since everyone is female, we need only a sample of size 1 to get a perfectly representative sample for this hypothesized class. Taking a larger sample is pointless since everyone is female; we won't learn anything else about the population regarding sex of population by selecting more.

1f. Why must we work with samples at all -- why not use a population so we don't have to make inferences?
Usually populations are

- too big to study entire population
- too costly
- too impractical
- and maybe population is difficult to define, e.g., what is a USA resident?

Note, however, that not all populations are too big and force one to use a sample. A population can be whatever you define it if you clearly define the characteristics of the population.

1g. How is a sample defined?
A sample occurs when one selects a group from a defined population to study and does not include everyone in the population for the study. A sample is any number from the population; any number less than the population is a sample.

1h. Census vs Sample
A sample is a subgroup from a population.
A census includes everyone in the population.

1i. Sample size = n
The symbol for a sample is n, which denotes the number selected for a sample. For example, n = 10 means 10 participants included in the sample.

2a. Sampling Methods
Two general sampling methods in quantitative research (qualitative sample approaches will be covered later in semester)

- Probability
- Non-probability

How do these differ?

With probability:

- known chance of selection (i.e., we can calculate probability one will be selected)
- use of randomized selection procedure
- known list of population units (i.e., can identify distinctly all population units so population size is known).

With non-probability sampling, one or more of the features noted above will be missing.

Known chance of selection means that with probability sampling we can calculate or determine everyone's odds or chances of being selected for the sample.

Use of randomized selection means that probability-sampling procedures use some form of randomized selection (such as a table of random numbers or blindly selecting names from a hat) to select from a known list of subjects. Known list simply means we know and can identify everyone in the population.

2b. Types of Populations in Sampling
Note the difference between **target** and **accessible** populations.

> **Target Population** – that population to which we wish to generalize findings from a sample.
>
> **Accessible Population** – Population from which we may select a sample. May be the same as the target population, or a restricted subset of the target population. Hope is that accessible population is representative of the target population.

3. Types of Probability Sampling Techniques

3a. Type 1: Simple random sampling (SRS)
Selection of subjects in SRS:

- Selection such that everyone in the population has a
- calculable,
- equal, and
- independent chance of being selected for the sample.
- Technically, every possible sample of size N has equal chance of selection.

SRS can be thought of as the sampling strategy where everyone has their name placed in a box once, then names are blindly selected from that box. Or, everyone has a unique number assigned to them, then a table of random numbers is consulted to select individuals.

While SRS (Simple Random Sampling) requires that all in population have an equal chance of selection, for general random sampling it is NOT necessary that all have an EQUAL chance of selection.

> Example of Non-equal Chance
>
> > If I place one person's name in the box twice, but all others have their name placed in a box once, then that first person has a greater chance of being selected. If I blindly reach in the box and select a name, then we still have random selection, but with slightly unequal probabilities. This would be an example of random sampling, but not SRS (simple random sampling). Although typically, when one thinks of random sampling, one thinks of everyone having an equal chance of selection, so SRS is the type of sampling most associated with “random sampling”.

Process for Selecting a SRS

1. **Purpose of Selection**: Maybe survey people to learn about their demographics or backgrounds. If enough folks in chat, we could ask of our sample the following:
    - “Are you married?”
    - Respond either as yes or no.

2. **Sample Size**: Determine appropriate sized sample for study. This is a complex issue and beyond level of this course; it requires understanding of statistics and hypothesis testing. For chat example, we could select a small sample of 4 or 5.

3. **Assign Unique Identification**: Old fashion method is to assign unique, consecutive numbers starting at 1. This step encompasses uniquely identifying everyone in the population. Newer approach is to create database of population and use computer to randomly select, or maybe use random digit dialing to phone participants. In

chat assign each present a unique number starting at 1 and proceeding until all are provided a number. For example, Beth = 1, Sue = 2, Keesha = 3, Bryan = 4, etc.

4. **Randomly Select Sample**: Consult a table of random numbers (such as in your textbook), or use computer generated random numbers to pick students for the sample. Use Google to find table and illustrate selection of small sample. Here is one example:

https://mathbitsnotebook.com/Algebra2/Statistics/STrandomtable.html

Independent Selection

What does it mean for one to have an "independent" chance of selection?

Independence means that a given person's chance for selection does not depend upon anything else except that the person is a member of the accessible population.

If selection of people is random and independent, then I won't be able to predict who is likely to be picked next.

Example of Dependent Sampling; or
Non-example of Independent Sampling

- Conducting survey of USA adults asking about options of political issues.
- Use random digital dialing to randomly select participants.
- Pollster is lazy and decides to expedite selection by interviewing individuals and their spouses.
- This means that once one person in a household selected, the pollster automatically attempts to interview that person's spouse.
- This is **dependent sampling** because one person's chance of selection depends upon someone else being selected.

How is this dependent sampling?

I can predict who will be selected next once I know who was selected just prior. If you tell me that you selected Mr. Jones, then I can predict you will likely select Mrs. Jones for interviewing. The dependence in selection means that I can predict who may come next in the selection process. With independence of selection, there are no strings tied to selection, and no way to predict who will be selected next if truly random.

3b. Type 2: Stratified Sampling

What is the difference between stratified and random sampling?

With stratified sampling, *one organizes the population into groups or strata, then randomly select people from* ***each and every group/strata***.

Example

Stratified

- Divide class by sex (females and males),
- sex is the **stratification variable** for this sampling approach
- determine number to select from each sex (e.g., 4 females, 3 males),
- and select randomly these two groups (i.e., first randomly select 4 females from the female group, then randomly select 3 males from the male group).

SRS

Just select randomly from the entire group (i.e., the population) and do NOT make selections from unique groups of males and females as we did with stratified.

Why use stratified -- why not use only random sampling?

Stratified ensures selection from all targeted strata; each and every subgroup forming the strata will be included in the final sample.

Question – if we randomly selected 4 people in this class, is there a chance that no males would be selected?

Yes. Unlike SRS, with stratified we ensure that all groups in the strata will be selected and represented in the sample.

What is proportional and non-proportional stratified sampling?

Proportional stratified – One selects from strata in the same proportions in the population.

Non-proportional stratified – One selects from strata in proportions that do not necessarily match the population proportions.

Example

- Population has 60% male and 40% female.
- **Proportional stratified** sample, one would select a sample in which 60% are male and 40% are female, so our sample proportions match those found in the population.
- **Non-proportional stratified** sample, one may decide to select a sample with 45% males and 55% females, so our sample proportions do not match those found in population.
- Non-proportional stratified usually employed to oversample extreme or small groups (e.g., 95% female and 5% male in population, so over sample males to obtain larger sample for better estimates, maybe take sample of 50% female and 50% male, or maybe 70% female and 30% male).
- Non-proportional stratified sampling allows one to increase the group sizes in the sample to allow for study of the groups individually; this allows us to understand the subgroups better. (Most federal studies using non-proportional stratified samples, but use *design effects* to help adjust for this in generalizing to population.)

Two questions:

(a) which is best used for making inferences to the overall population, and
(b) which is best for making inferences to the individual groups, proportional or non-proportional?

(a) Proportional is best for overall representation, and
(b) non-proportional best for representation of unique groups.

- Proportional sampling best represents overall because the percentages of each group selected matches the population,
- with non-proportional the percentages selected for each group does not match the population so it is no good for inferences to the overall population.
- Example
    - suppose there are 15 females in this course and 5 males and we asked whether they are married;

- further suppose that all females are married and none of the males are married;
- in the population of this course we know that 75% of the class is married (15/20 = .75, all the females) and that 25% are not married (5/20 = .25, all the males);
- With **proportional stratified sampling**, we may select, randomly, 3 females and 1 male. This proportional sampling would still give us 3 married and 1 unmarried, thus 75% married and 25% not married in our sample. Note that these proportional-sampling figures agree perfectly with population figures provided above.
- With a **non-proportional stratified sample**, the results would differ from the population figures.
- Suppose with our non-proportional stratified sample we selected 2 females and 2 males randomly. In this non-proportional sample 50% of the sample is female and 50% is male (recall that in our hypothetical class example the population is 75% female). Now the sample figures for marriage status is 50% married (the 2 females) and 50% unmarried (the 2 males).
- These sample figures disagree with the known population figures of 75% married and 25% unmarried.

Process of Selecting a Stratified Sample

1. **Purpose of Selection**: Survey people to learn something about them. If enough folks in chat, we could ask of our sample the following:
   - "Have you watched at least 15 minutes of a fishing program in the past year?"
   - Respond either as yes or no.
   - Among the purposes for this sample is the belief that differences will exist in response patterns among the subgroups belonging to the stratification variable. For example, will there be a sex difference to this question? --> Are there differences in rates of watching fishing programs between males and females? Is one sex more likely to watch fishing than the other? If yes, then stratification would be important as a sampling approach.

2. **Identify Stratification Variable** – sex of student.

3. **Proportional or Non-proportional**, which to choose? Depends upon purpose of study, in this case to learn about subgroups so use non-proportional to get acceptable sample sizes per group.

4. **Sample Size**: Determine appropriate sized sample for study. Again, complex decision, but for our purposes we select small numbers, maybe 3 or 4 from each stratum.

5. **Assign Unique Identification**: Make two lists, one for females and one for males. Within each stratum separately assign unique numbers, start with on group, then move to next group starting numbering system at 1 with each group.

6. **Randomly Select from within Each Strata**: Randomly select from females and then from males.

3c. Type 3: Cluster Sampling

What is the difference between cluster and stratified sampling?

With stratified, you randomly select INDIVIDUALS from **each and every group**, but with cluster you randomly select GROUPS and **not all groups are selected**.

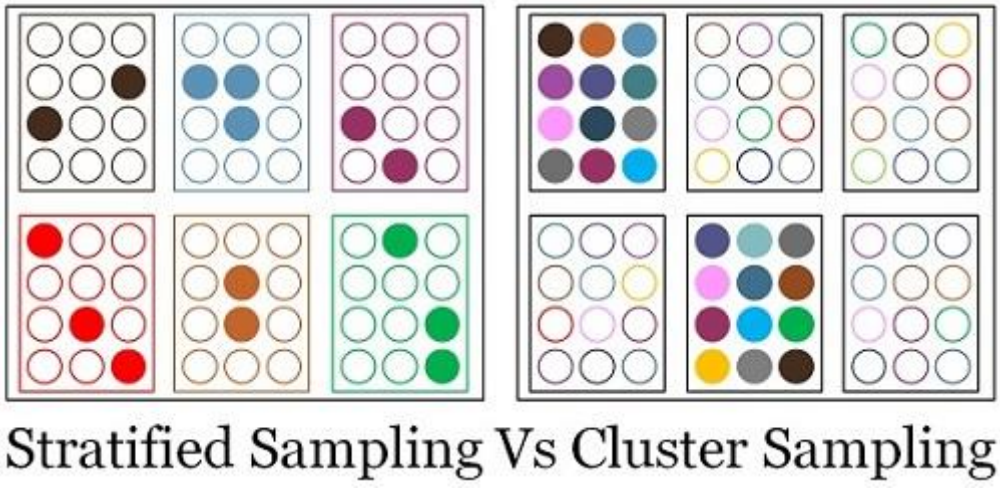
Stratified

- Individuals are randomly selected, not groups
- Individuals are selected from every group formed by the stratification variable (e.g., random selection of males from the male group, and females from the female group)

Cluster

- Groups are randomly selected, not individual units within groups
- Not all groups will be selected (if all groups were selected with cluster, you may have a census, not a sample except in two-stage sampling)

Illustrated: Stratified vs Cluster



Source: "Difference Between Stratified and Cluster Sampling" Keydifferences.com, 23 Jan 2018. http://keydifferences.com/difference-between-stratified-and-cluster-sampling.html

Which type of sampling is this, stratified or cluster?

Example 1
I randomly select 25 counties in Georgia and interview the town mayors in those counties selected. Note that there are 159 counties from which the 25 were randomly selected.

> It is cluster because we first select counties randomly, and each county will represent a cluster of town mayors with several mayors found in each county. Note that Georgia has 159 counties, so if we randomly select 25 counties, then we have performed cluster sampling (random selection of groups first, in this case groups of mayors).

Example 2
If I randomly selected mayors, from every county in Georgia, that would be which type of sampling: simple random, stratified, or cluster?

> This is stratified because mayors were selected from each and every county. Selecting from each and every county is unlikely to occur with random sampling; not all counties would be represented probably. It is not cluster because mayors, not counties, were randomly selected, so no random selection of clustering units occurred.

In the above example, what is the **stratification variable** used (i.e. what is the grouping variable)? Note that stratification variable is that which creates the groups.

> County represents the stratification variable. We first break population (all mayors) into groups according to which county they live, then we randomly select a few mayors from each county. So remember that with

stratified sampling, people are randomly selected, but with the special condition that people are selected from each and every group identified.

Example 3
Which is this? I want to select teachers to interview. Since teachers are located within schools, I will randomly select, let's say, 10 schools located throughout the state.

Since I randomly selected the schools, and not the teachers, I have performed cluster sampling.

Recap

- **Random** is the random selection of individuals without regard to which groups they may belong;
- **stratified** is the random selection of individuals with particular attention paid to group membership with goal to represent each group;
- **cluster** is random selection not of individuals but of a few groups from among several groups.

3d. Type 4: Systematic Sampling
What is systematic sampling?

It is making a list and picking people by skipping a certain number each time, such as picking every 5th person -- 5, 10, 15, 20, etc.

4. Types of Non-probability Sampling Techniques

4a. What is **convenience** sampling?

Convenience is selection of subjects without regard to a systematic, randomized procedure of selecting people and without regard to making a list of all people in the population.

Example

I may be interested in a study to learn whether using one textbook produces better achievement than another in educational research. I may use one text in this class and another text in another section this semester or next semester. Note that I did not randomly select students in any way; this illustrates convenience sampling. Another example would be surveying all faculty members about their opinions regarding some policy in the local district.

Here is a practice test question:

Carfax publishers will offer a new journal entitled "Teaching in Higher Education" next year. They ask a professor from Leeds University to serve as editor. This professor needs support. For example, he will need co-editors and manuscript reviewers. He seeks reviewers by placing advertisements in several education related newspapers (e.g., The Chronicle, Education Week), related journals, and in related electronic discussion lists.

What type of selection procedure is he using - - - random, stratified, cluster, systematic, or convenience?

Yes, convenience

Why not random, stratified, or cluster?

None of the characteristics of probability sampling included in example.

Recall characteristics of probability sampling:

1. known chance of selection → is this present in current example?
   Can he calculate probability that each potential reviewer will be selected?

2. use of randomized selection procedure → is this present in current example?
   Was a systematic randomized selection process used? List of random numbers assigned to potential reviewers?

3. known list of population units (i.e., can identify distinctly all population units so N is known) → is this present in current example?

Random, stratified, and cluster are eliminated as options because the editor does NOT randomly select people in any way -- he asks for volunteers -- that is convenience. Note also that no aspect of probability sampling took place: no randomized selection occurred and a population was not clearly identified and listed.

4b. What is **judgment sampling** (sometimes called purposive sampling)?

This is when a researcher uses his or her judgment to select participants the researcher thinks are representative of the population he or she wishes to study.

Example

If I wanted to study the teaching strategies of strong instructors, I might select a handful of master teachers rather than select, randomly, many teachers.

Another type of purposive selection procedure is called **maximum variation sampling**. What do you think this would entail?

Maximum variation is selecting folks who are at extremes on variables important to the research purpose.

Example

If I wanted to study factors that distinguish the better teachers from the weaker teachers, I may select a number of teachers who have been identified as strong and a number who have been identified as weak. From these two groups, I may then study characteristics that differ between them in an attempt to learn how and why these two groups differ. For example, I may enter these teachers' classrooms and observe their teaching behaviors for several weeks in an attempt to note important differences between the two groups.

4c. What is **quota** sampling?

With quota sampling, one selects certain number of participants based upon predefined characteristics.

Quota sampling may look like a probability sampling technique, but it is not because there is no randomized selection from a defined population.

Example

Once I was involved in quota sampling of voting behavior. I was told to interview voters as they exited a polling station. I was to interview 10 Black males, 10 Black females, 5 White males, and 5 White females. I asked whoever came next out of the station, and once I met my quota for a particular group, I stopped asking those individuals.

5. Sample Size

Why are larger samples better?

The larger the sample, the better the chance of good representation (assuming selection is based upon some type of randomized selection scheme, if there is built-in bias in the way individuals are selected, then larger samples do not help).

	Population	Sample 1	Sample 2	Sample 3
	N = 1000	n = 100	n = 40	n = 4
Male	70%	67%	60%	50%
Female	30%	33%	40%	50%

The larger the sample, typically the closer should be the match between sample and population if randomize selection occured.

6. Error in Sampling

We expect for a sample to be somewhat unrepresentative of the population; we expect samples to be different from the population to some degree; we expect there to be some error in estimating population characteristics from samples.

In sampling, what is meant by "sampling error" and what is "bias"?

Sampling error is any random difference between sample and population characteristics. Sampling error is expected and will occur every time a sample is taken. Researchers know that no sample will exactly represent the population perfectly in every way. There is nothing wrong with sampling error---it is a natural part of research.

Example,

If the mean age of a population is 43 and the mean age of our sample is 41, that difference of 2 years, if due solely to random chance---random fluctuations in who we sample---is called "sampling error."

A **bias** occurs when samples systematically differ from a population in a specific way.

Example,

I was interested in learning of the typical household pre-tax income of people in Savannah, what would be the likely results if I interviewed people as they left a local brokerage house during a typical workday?

Typically people with higher than average incomes use brokerage services. Often folks in what we typically label blue-collar jobs won't see much action in brokerages. If I sample people as they leave brokerages, will my sample income be systematically higher than the true population income for people in Savannah?

And if other researchers replicated my study---sampling people from brokerages--- their results will also be systematically higher than they should be.

As this example illustrates, a study with this design would systematically over-represent the household income of Savannah residents. As a result, this study design introduced what is known as “bias” --- systematic differences between sample and population characteristics.