VII. WHY SAMPLE?

THIS CHAPTER COVERS:

- The importance of sampling
- Populations, sampling frames, and samples
- Qualities of a good sample
- Sampling size
- Ways to obtain a representative sample based on probability methods
- “Galton’s Problem”

ALL RESEARCHERS SAMPLE.

Consciously or unconsciously, all researchers choose some cases out of a universe of cases to study. It may seem preferable to study all the cases in a population, but sampling has distinct advantages.

- It is more economical to study a sample of cases than it is to study all the cases available. Studying all the cases is often impractical or prohibitively expensive.

- Sampling may yield higher quality data. If you choose to study more cases than you reasonably have time for, you will likely feel rushed and make mistakes. It is better to sample and spend an adequate amount of time on each case.

- Very accurate results can be obtained from samples if they are representative. For example, political polling in the U.S. has predicted voting results accurately using very small samples of the voting population (samples are usually in the hundreds to low thousands).

LIKE THE U.S. POPULATION, THE ETHNOGRAPHIC RECORD IS TOO LARGE TO STUDY IN ITS ENTIRETY.

But what is the right sample size and how do we draw a sample out of thousands of cases?

In the following slides, we briefly discuss important concepts and a variety of probability sampling techniques that allows us to create **representative samples**.

CONCEPTS TO KNOW:

- The universe of cases that one wants to generalize to is called a **population** (e.g. all cultures ethnographically described, all countries, or all adults in a society).

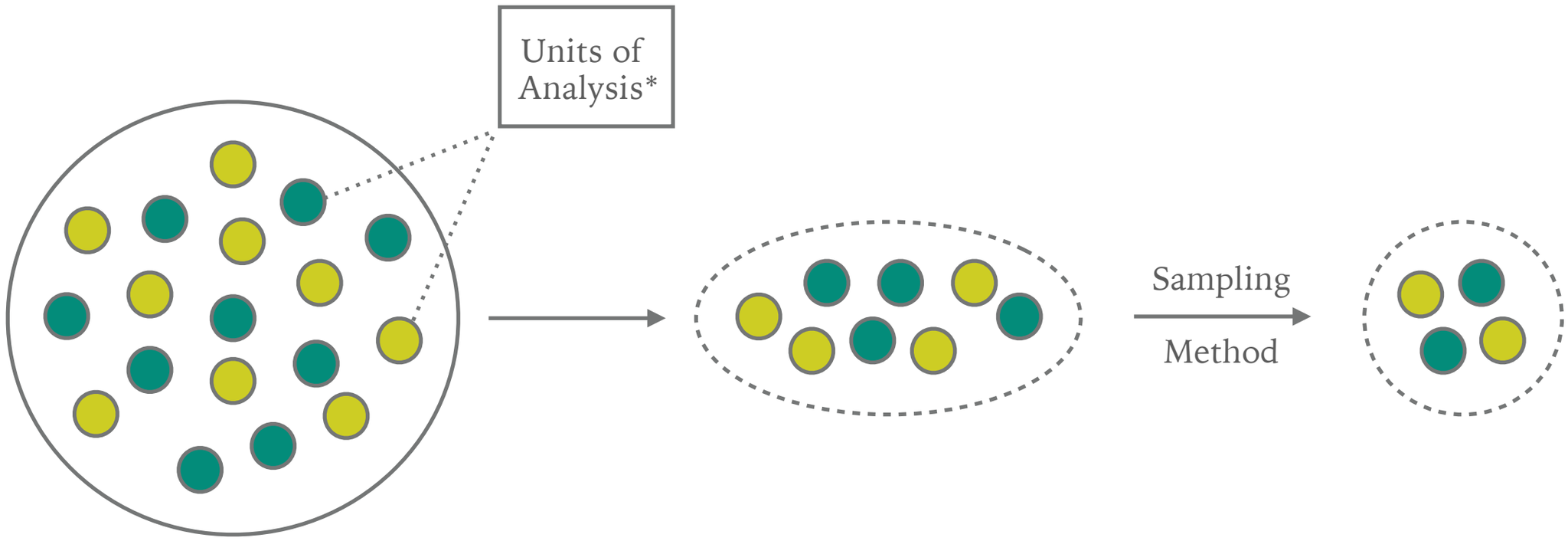
- A population is made up of cases, or more formally, **units of analysis**. The units of analysis are the units of the population that are actually studied (e.g. societies, countries, individuals).

- To sample from a population, we need to know all the units of analysis. At best, we have a list, called a **sampling frame**, that approximates a complete list of the cases to be sampled.

POPULATION OR UNIVERSE

SAMPLING FRAME

SAMPLE



EXAMPLE: ALL THE WORLD'S ETHNOGRAPHICALLY-DESCRIBED CULTURES

LIST OF CULTURES WE SAMPLE FROM

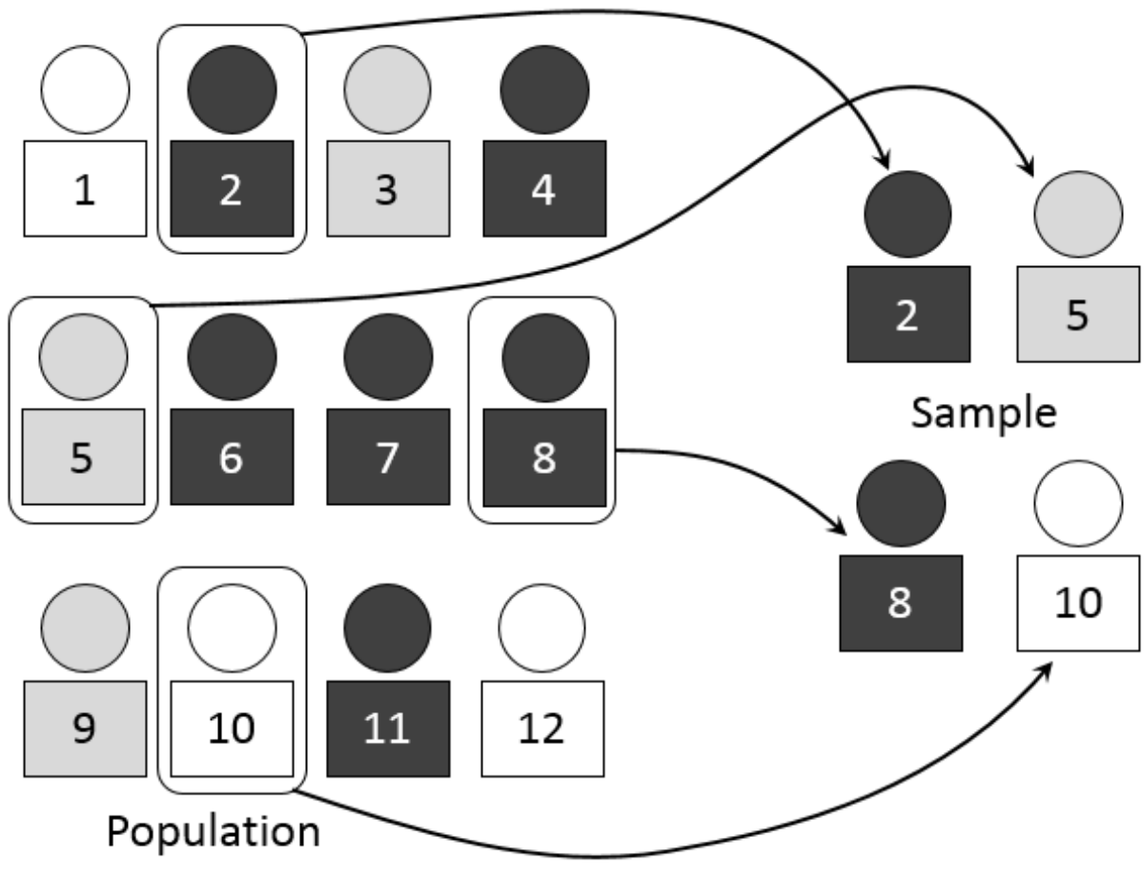
THE SAMPLE

**The diagrams are symbolic. There are almost always many more cases or units of analysis in the population, the sampling frame, and the final sample.*

WHAT MAKES A GOOD SAMPLE?

A good sample should represent a larger set of cases in an unbiased and representative way; this makes it possible to generalize results to the larger set from which the sample was drawn.

Example: A researcher has a population of 12 people and randomly chooses 4 by using a table of random numbers (the first set of random numbers is 2, 5, 8, and 10).



Visualization of simple random population sampling.

HOW DOES A RESEARCHER DETERMINE APPROPRIATE SAMPLE SIZE?

THE SIZE OF YOUR SAMPLE DEPENDS ON WHAT YOUR PURPOSE IS, THE TYPE OF MEASURES YOU HAVE, THE STATISTICAL TEST YOU PLAN TO USE, YOUR CRITERION FOR SIGNIFICANCE, AND HOW ACCURATE YOU HAVE TO BE.

Representativeness is almost always more important than sample size. You can consult with advanced reference works to find formulas that help you determine the sample size you will need. There are different formulas for different types of statistical tests, and you can specify how accurate you want your results to be.

Here are some general guidelines:

- The **more accuracy you require, the larger the sample needs to be**.
  - If you want to claim that a culture trait is universal, a relatively large sample is required.
  - If two presidential candidates are close in the polls, a larger representative sample will be needed to correctly predict the outcome; if the candidates are far apart, smaller samples will be accurate enough.
- If you are looking for **strong associations between culture traits**, relatively small representative cross-cultural samples (e.g., between 20-50 cases) are probably sufficient.
- If you **expect weak associations between culture traits** or you **want to control on many variables**, larger samples are needed.

WHICH TECHNIQUE SHOULD YOU USE?

In order to create a representative sample, we must adopt a sampling procedure that is not based on subjective criteria. Some type of probability sampling is the sampling method of choice for cross-cultural research. Probability sampling means that every case has a known (and non-zero) chance of being chosen.

SAMPLING FOR COMPARISONS USING SECONDARY DATA

What population should you use cross-cultural research methods to generalize to? This can be referred to as your **scale of analysis** and depends upon your research question. Some cross-cultural researchers might want to generalize to all of the world's cultures of the recent past and present. Or, for example, a regional researcher might want to generalize to the cultures of Africa specifically.

The problem is that the "population" of all the world's cultures is not all that clear. Descriptive information about the entirety of the world's cultures does not exist; nor is there a complete list of cultures that we could use as a "sampling frame" that represents all of the world's cultures. The *Outline of World Cultures* (Murdock 1983) was intended to be a complete list (or sampling frame) of ethnographically-documented cultures, but it incomplete and needs updating.

Lacking a complete list, most cross-cultural researchers opt to use an existing published sample that claims to be either to be relatively complete or representative of the population of world's cultures. That published sample becomes their **sampling frame** from which they can draw a smaller final **sample.**

COMMONLY-USED SAMPLES IN CROSS-CULTURAL RESEARCH

There are about 8 published cross-cultural samples [for a more complete discussion, see Ember and Ember (2009: 79-84). Below we discuss a few of the most commonly used samples today in order of their publication:

- **The HRAF Collection of Ethnography** is derived from the Cross-Cultural Survey, a project established at Yale University's Institute of Human Relations in 1935. Societies were chosen from Murdock's *Outline of World Cultures*. The paper/fiche version contains 385 cultures. The online version, **eHRAF World Cultures**, contains data from almost 300 cultures and is growing annually.
  a. One distinctive feature of the HRAF collections is that its ethnographic data is subject-indexed to the paragraph level, facilitating rapid retrieval of information.
  b. The HRAF collection grew somewhat opportunistically; for this reason, it is not claimed that the entire collection is representative of all world cultures. However, the subset of cultures called the Probability Sample Files (PSF) does make such a claim (see the next slide).

COMMONLY-USED CROSS-CULTURAL SAMPLES, CONT'D

- **The Ethnographic Atlas (EA)**, with over **1200 societies**, was intended by Murdock (1962-1980) to be a relatively complete list of the world's described societies, each of which was pinpointed to a time and place focus. A considerable number of coded variables are included. The EA was later truncated to better-described cases and was published as the **Ethnographic Atlas Summary** (Murdock 1967).
- **Probability Sample Files (PSF)**, a subset of the HRAF collection, was designed by Naroll (1967) to be a representative sample of **60 world culture areas**. Ethnographic information had to meet certain data quality control criteria to be included. After identifying which cultures met these criteria, one case was randomly chosen from each area.
- **The Standard Cross-Cultural Sample (SCCS)** of 186 societies (Murdock and White 1969) claimed to have chosen the best-described society for each of **186 culture areas**. There are now **over 2000 coded variables** for this sample. The sampling method was subjective.

SAMPLING TYPES INCLUDE:

- **Simple random sampling:** guarantees that every case in the sampling frame has an equal chance of being chosen, often by assigning random numbers to the cases. If you proceed in a random order and run out of time before studying all the cases, your results are still generalizable to the sample frame. **This is considered the method of choice.** Online generators can be used to obtain a set of random case numbers.
- **Systematic sampling:** every *n*th case is chosen after a random start. All the cases within the sample must be studied to maintain generalizability to the sample frame.
- **Stratified sampling:** the sample is first divided into subgroups or strata and then randomly sampled from each. There are two types of stratified sampling:

    A. **Proportionate stratified sampling:** each subgroup is represented in proportion to its occurrence in the total population.

    B. **Disproportionate stratified sampling:** some subgroups are overrepresented and some are underrepresented. This technique is used when a researcher needs to over-represent a rare type of case in order to have enough cases of that type to study.

GALTON'S PROBLEM

Some researchers have suggested that the validity of cross-cultural findings may be threatened due to cases in the sample being historically related (sharing a recent common origin or being near enough to each other for cultural diffusion to occur).

When Edward Tylor presented the first cross-cultural study in 1889, **Francis Galton** (1889) suggested that many of Tylor's cases were duplicates of one another because they had similar histories, and for this reason many correlations could be falsely inflated.

This idea is now known as **Galton's Problem**. Researchers disagree about the seriousness of Galton's problem as a threat to cross-cultural research. The following slide presents the two sides of the argument.

Galton's Problem is a serious threat	Galton's Problem is not a serious threat
Statistical associations cannot be causal if they can be attributed to common ancestry or diffusion.	The societies in most cross-cultural studies speak mutually unintelligible languages. If language has diverged, other aspects of culture must have diverged as well.
Steps must be taken to ensure that samples contain one culture per identified culture area. (The Standard Cross-Cultural Sample and HRAF Probability Sample both take this approach.)	Random sampling of cases is the best way to prevent sampling bias; culture areas are chosen subjectively.
Tests must be done to test for the effects of diffusion and common ancestry.	There are two requirements of statistical inference: that sample cases be independent and the measures on them be independent. To achieve independence of sample cases, the choice of one case for a sample must not influence the choice of another. To achieve independence of measures, each case's score must be arrived at separately.

WAYS TO ENSURE THAT GALTON'S PROBLEM IS NOT AFFECTING YOUR RESULTS:

Test for the effects of spatial distance or linguistic distance.

Use a small random sample selected from a larger list. This method makes it unlikely that more than a few cultures, at worst, will be selected from the same culture area.

If you randomly sample from a larger sampling frame, **redo your analyses** by randomly omitting more than a single case from the same culture area. If the results are not substantially different, your original results are not affected by historical relatedness.

SUMMARY

- **Sampling from a larger population** is economical in terms of time and expense and often yields higher quality data.
- **Sample representativeness** is almost always more important than sample size
  - ➤ **Larger samples** are needed for greater accuracy, for small differences or weak associations, for many control variables
  - ➤ **Probability sampling** is the method of choice
- Lacking a complete list of the world's societies, most cross-cultural researchers opt to use a published sample that claims to be relatively complete or representative.
- The **most commonly used cross-cultural samples and sampling frames** are:
  - ➤ The **HRAF Collection of Ethnography** (the digital version on eHRAF World Cultures). The subset called the **Probability Sample Files** is based on probability sampling
  - ➤ The **Ethnographic Atlas** with over 1200 societies claims to be relatively complete and can act as a sampling frame
  - ➤ The **Standard Cross-Cultural Sample**
- **Types of probability sampling** include: simple random sampling, systematic sampling, proportionate stratified sampling, and disproportionate stratified sampling.
- The idea that correlations may be distorted by having multiple cultures with similar languages and histories in your sample is known as **"Galton's Problem."** There is disagreement about how serious the problem may be. Various solutions have been proposed for those who are concerned about this issue.

REFERENCES

Ember, Carol R. and Melvin Ember. 2009. *Cross-Cultural Research Methods,* 2nd edition. AltaMira/Roman & Littlefield.

Galton, Francis. 1889. Comment in "Discussion" after Edward B. Tylor (1889).

Murdock, George P. 1962-1980. Ethnographic Atlas. *Ethnology* 1-20.

Murdock, George P. 1983. *Outline of World Cultures,* 6th rev. ed. Human Relations Area Files.

Murdock, George P. and Douglas R. White. 1969. Standard Cross-Cultural Sample. *Ethnology* 8: 329-69.

Naroll, Raoul. 1967. The Proposed HRAF Probability Sample. *Cross-Cultural Research* 2:70-80.

Tylor, Edward B. 1889. On a Method of Investigating the Development of Institutions Applied to the Laws of Marriage and Descent. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 18: 245-72.

CITATION AND TERMS OF USE

Using this course: The material in this course is intended for educational purposes only--either by individuals for personal use or for classroom use by instructors with appropriate attribution. For any commercial use or other uses please contact HRAF. (Email: <hraf@yale.edu>).

This chapter is from: Carol R. Ember. 2016. Introducing Cross-Cultural Research. Human Relations Area Files. <http://hraf@yale.edu/ccc/>