Week 7 Lecture - Sampling Methods

Undergraduate Research Methods in Psychology

Quinton Quagliano, M.S., C.S.P

Department of Psychology

Table of Contents

1 Week Overview

1.1 Learning Objectives

- Explain why external validity is essential for most frequency claims.
- Describe which sampling techniques allow generalizing from a sample to a population of interest, and which ones do not.
- Be able to describe the difference between probabilistic and non probabilistic sampling techniques and the resulting representative-ness of the sample
- Understand the risk and implications of biased samples and be able to apply that knowledge to a reading of a real article

1.2 Week Overview

- We can never study _____ in a **population of interest**
  - Example: ***All*** *veterans* or ***all*** *people with GAD*
- Instead, we can only study a smaller _____ of those groups, and then try to extrapolate/generalize our conclusions to the population we care about.
  - Example: Just 40 veterans or 10 people diagnosed with GAD
- That subset that we study on, from the population, is our _____.

Discuss: Think of some other examples of populations and their respective samples

2 Generalizability

2.1 Overview

- A critical notion to consider in how "good" research is in how we sample really relates to the _____ population and variety of settings, i.e., external validity
- This consideration is present in pretty much all social research, and in all 3 _____ types (frequency, association, causal)

2.2 Relationship Between Populations and Samples

- To "sample" every individual in a population is called a _____.
  - But such a procedure is not really possible in most research, due to time, money, resources, and ethical _____
  - Example: The US census is (theoretically) a count of _____ people in the United States
  - Instead, we may _____ that an adequately representative sample's behavior is carried over to the population. Thus, a study of a sample's behavior is meant to really be a study of a population's behavior!
- A population of _____ is one that we define when we make a hypothesis. Rarely, do we make a hypothesis or theory applicable to everyone. Instead, we may primarily aim to make a theory _____ around individuals with some specified characteristic(s).
  - Example: In my study, I want to _____ study older adults with suspected dementia (that is my **population of interest**)
  - Example: For that _____ study, I gather 100 older adults from the broader population (that is my **sample**)

Discuss: Think of some of the articles you've read this semester - what were their populations of interest?

- A sample, even if accurately taken from the population of interest, is not inherently _____:
  - **Biased/unrepresentative samples** are those that are somehow taken in a way that they do not properly represent the _____
  - Unbiased/_____ samples are those that are a proper representation of the population.
  - Realistically, no sample is _____, and we must be discerning in what flaws a sample may have

2.3 Probability in Sampling Methods

- **Probability/random sampling** is all about using a randomization method to _____ members of a population for the sample who then agree or

disagree to be in the study
  - Good for _____ validity / generalizability
  - Side note: from a mathematical perspective, the type of random sampling we use is technically _____ due to how computers and chance work, but for the purpose of the class, we will call this random.

- **Non-probability/non-random sampling** is the opposite, in which the method is not _____ and therefore, may be biased towards certain individuals
  - Bad for external validity / _____
- We'll talk about specific types and examples for these in the following sections
- In real studies, we may very well use _____ of the following methods to obtain the final sample to run the study on. There can be a lot a grey area on whether a study used a proper sampling method!

2.4 Representative/Probability-based Sampling Types

2.4.1 Simple Random

- The "gold standard" method, this is if/when we have an equal, known chance of selecting every _____ within a population
- Example: I am interested in a population of GV students, so I put every ID number in a list and then have a computer _____ one at random.
  - Each _____ has a $1/n$ chance to be selected, where n is the number of people in the population
- Fantastic _____, but very difficult and rare

2.4.2 Systematic

- Return to our population of GV students: select two random numbers, say a and b.
  - a will be our a^{th} person in the list of IDs.
  - b will be the number of people we increment by
  - We start with the a^{th} person, and then sample each person b away from that person
- Say we have $a = 4$ and $b = 8$, we would _____ the 4th student, then the 12th, then the 20th, etc on the list

Discuss: Try following the procedure above with a different a and b value (e.g. a = 3, b = 2). What people would you sample?

- Still results in a pretty _____ sample comparable to simple random!

2.4.3 Cluster

- This method comes in when we have naturally existing _____ within our population of interest
  - Example: high school students organized into high schools
- We randomly sample from the clusters (e.g., the high school) and then sample _____ students within a cluster
- Still pretty good!

Discuss: What are some other 'clusters' you can think of that naturally occur?

2.4.4 Multistage

- Simply, cluster random sampling _____ by simple random within the selected clusters. A combination of our descriptions of the prior types.
- Example: I randomly select 10 high schools in the state, and then sample randomly 50 students from a list of students at each school.
- Also, good!

2.4.5 Stratified

- This is a method that is done when we are trying to ensure that our sample is representative of proportions of natural _____ in the population.
- Strata must be some _____ grouping variable (e.g., different ethnicities, different genders, different education levels, etc.) with known levels within the population.
  - E.g., class level (First-year, sophomore, junior, senior) at GVSU
- After we have our stratifying variable, we then randomly sample within each _____ of the variable.
- *However,* our goal is to get a sample with _____ proportions of the strata to the population
  - E.g., If GVSU has 30% First-year, 20% Sophomore, 25% Junior, and 25% senior, I want my sample to have that same breakdown!

2.4.6 Oversampling

- This is a modified stratified sampling, where we _____ represent one or more levels of the strata more in sample, than is found in the population
- This is used when we may want to ensure that we still capture the experience of a relatively _____ group of individuals.
- Example: When gathering my sample I know there will be relatively few indigenous/-Native American people. So, I purposefully ensure that at least 10% of the people in my sample are in that group, even though they only represent 5% of the population.

📢 Discuss: What are some other naturally small demographic groups you may encounter in research?

- A related concept for a similar goal is _____ where an unbalanced sample (across some strata) has occurred and we use a statistical technique to give more “value” or weight to less-represented groups.

2.4.7 Randomness in Sampling and Assignment

- Random *sampling* is related to external _____ and how we initially draw our sample from the broader population of people
  - Example: Randomly selecting patients from a hospital for a retroactive study analysis
- Random *assignment* is when we already have our sample and are sorting people into different levels of a manipulated/_____ variable, much more related to experiments and internal validity.
  - Example: after getting my sample, deciding randomly who will get a new trial drug and who will get a placebo during the study

2.5 Biased Sampling Methods

- Biased samples result from poor _____ in the sampling process, which results in the members of the sample being especially "unusual" or those that might behave _____ than other members of the same population
  - Effectively, certain _____ types might mean we are actually gathering individuals of a different population than intended!
- There are _____ issues in doing thorough random sampling techniques
- In some cases, it *may* be _____ to use a method which is *not* random to sample, with some caveats and drawbacks.

Discuss: Take a guess at some 'red flags' of a biased sample. What things would you suspect are telling of bad sampling?

2.5.1 Convenience Sampling

- This occurs when we sample only those people who are _____ to sample
  - Example: Psych 101-mandated research, college students, Amazon MTurk

  - The main problem is that those who are readily _____ may represent individuals who differ from the population of interest - they may be different in motivation, beliefs, status, etc.
- This may also occur _____ in samples that are just hard to track
  - Example: people across a large geographic area
- Be mindful that the method by which we collect data (internet, telephone, etc.) can be un-inclusive and _____ the sample as well.

2.5.2 Self-sampling/selection

- This happens when we rely upon a sample that volunteers or selects itself
  - Example: I put up a poster in the hall with a QR code that says, “take my survey!”
  - Compare this to me sending a randomized email to students of the psychology department
- This is **not** a difference of voluntary vs. involuntary - we still ask for informed consent of everyone regardless of sample type!

2.5.3 Purposeful Sampling

- This is when sampling is done by some method that _____ limits the randomness of who from the population can be sampled, which may be intentional
- Example: Only recruit political science students for a survey by putting posters in the department
  - What about people who don’t go into the department or online students.

2.5.4 Snowball Sampling

- When you have current participants _____ other friends / acquaintances for a study.
- This is especially useful when doing research on especially small or _____-to-contact individuals
- Example: Trying to study autistic adults - I may ask my participant to go share my recruitment link with other adults with that diagnosis

2.5.5 Quota Sampling

- Very similar to stratified _____ that we choose individuals non-randomly for each level of our stratifying variable.

- Example: I send out a survey for people to self-sample if they are interested, and I wait until I get at least 15 people of every class level.

Discuss: What are the benefits you see from using these biased methods?

3 Checking External Validity

3.1 Overview

- In my opinion, external validity should _____ be on our mind, because it affects just how much our research means outside the narrow context of the study (external validity)
- *But,* _____ external validity does not necessarily mean that a study is doomed or of no value
- Like any strength or weakness in research, we assess external validity in light of the claim being made. A weaker claim requires _____ robust evidence.
  - Very specific claims with a very _____ population of interest don't need especially high external validity
  - Very broad claims need more evidence of good external validity

Discuss: Do you reckon that most studies do a good job with sampling? Why or why not?

3.2 Frequency Needs External Validity

- Especially when we make a frequency claim about a population, we _____ need robust sampling techniques.
- That is because almost the entire _____ of a frequency claim is on accurate description of the population - and without good external validity, we have nothing.
- Frequency claims are also “more important” when they _____ better across more people. Limits on external validity mean that the claim will not widely apply across too many people.

3.3 When is External Validity Less of a Focus

- As a general rule of thumb, we should *try* to have high external validity wherever possible. Good sampling will always help the meaning and _____ of research.
- *But,* like all validities, we must _____ maximizing generalizability with the practical limits on our resources and time
- In my opinion, the *most* important factor is that researchers transparently report on the _____ of their sampling and are realistic in what this means for the impact of their research.

3.4 Larger Sample Does Not Always Equal Better

- There is a point of _____ returns on sample size to help external validity of a study. Sampling more and more people with a _____ method will still result in a biased sample.
  - Example: after around 100 people, you’re not necessarily going to get more representative just by having more people
- Instead, we must focus on both sampling many people, **and** doing this _____ well (i.e., with the probabilistic sampling methods described above)
- Larger samples *will* result in higher statistical _____ (i.e., higher chance of significant findings) - but a significant result does not mean that a study is valid and generalizable!

4 Key Points

4.1 Key Points

- Samples are subsets of the population we are attempting to study. Our conclusions about a sample are meant to represent trends present in the population.
- Sampling methodology is the most important factor in determining external validity and generalizability in a study, aided to a lesser extent by sample size
- Sampling can be done in a randomized manner, resulting in a unbiased sample likely to represent the population. However, it can also be done in an non-randomized way, leading to bias.
- External validity is always important, but tends to be more stressed for frequency claims