Lecture Agenda - Research Design

The components of research design

Bias and Research Validity

- External and Internal Validity

Qualitative and Quantitative Divides

Types of Qualitative Sampling

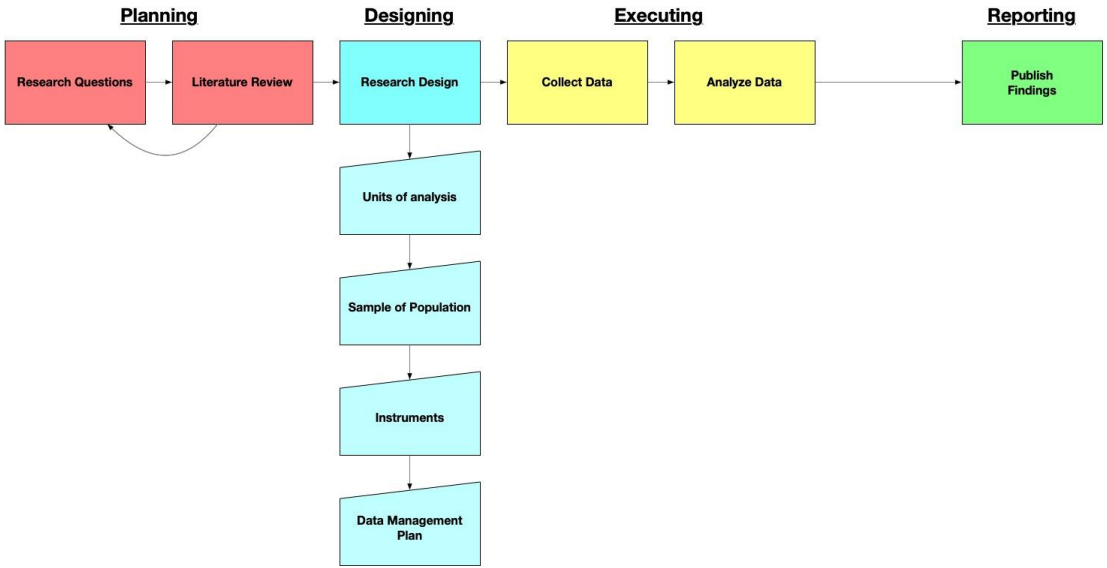
- Unit of Analysis and Unit of Observation
- Random, snowball, purposeful, chaining,

Types of Quantitative Sampling

- Simple, stratified, systematic, cluster, multi-stage

Research Design

Components of Research Design: Questions, justification (literature review), **plans for bias reduction and validity**, choice of methods, choice of analysis, and plans for managing data.



Bias and Research Validity

All research is inherently biased.

By being transparent in our intent and method we reduce bias.

We also reduce bias through the concept of validity

- **Internal validity** is about the design and rigor of our method
  - Follow community norms, best practices, and ethical standards
- **External validity** is about the representativeness of our findings.
  - Check with participants; statistical tests that are replicated

Sampling and bias

You may be saying… this all sounds very positivistic Dr. Weber.

Maybe. It depends on our goals.

> Qualitative analysis may not aim for generalization.
>
> Quantitative analysis may be exploratory and suggestive.

The point of sampling and bias reduction is not to have an **unimpeachable** (!) research finding, but to adhere to some form of rigor in how we design a research project so we can communicate clearly about what we know (**justified beliefs**).

Qualitative and Quantitative Divides

The ‘logic’ of representativeness is all about framing our intent of subject.

- Qualitative sampling is purposeful sampling. Participants, cases, etc are chosen for what will be most valuable to RQs.
- Quantitative sampling is based on probability of representation.

The two could not be more different in their logic.

- Qualitative research assumes bias, and designs sample to effectively answer a research question based on research intent. (small n)
- Quantitative research attempts to remove bias, and assumes that a sample is not representative unless it clearly follows a logic that is probable. (proportional n)

Qualitative sampling

Central to the logic of qualitative sampling is the idea of Unit of Analysis and Unit of Observation.

- Unit of analysis: What will be studied as representative of a phenomena
  - Cliques in High School cafeterias (tables)
- Unit of observation: What will be observed (collected, interrogated, interviewed, etc)
  - Individuals that sit at particular tables in the cafeteria as representing a clique

A purposeful sample could be designed how…?

- Observe a cafeteria for three days. Watch cliques emerge by who sits where.
- Interview students that sit at different tables
- Develop a rich description of seat choice, temporal variation, self-identity, etc.

Qualitative sampling

So we select samples *purposefully* or *randomly* on their representativeness.

Some examples:

- Deviance (e.g. crises); Intensity; Critical case; Theory-based; Snowball or chaining

Quantitative sampling

Just because we assume a bias exists doesn't mean we can't be rigorous and transparent in our research design.

Quantitative sampling is based the assumption that we can meaningfully **infer** things about a populaton by gathering and analyzing data about a **represetative subset**.

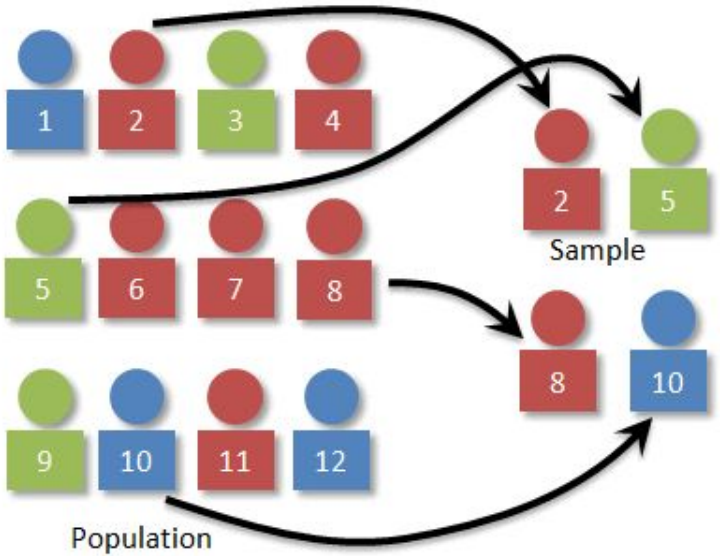
Some examples:

- Polling, American Community Survey, Sentiment Analysis, GDP, etc.

We are going to cover - random, systematic and cluster sampling techniques for

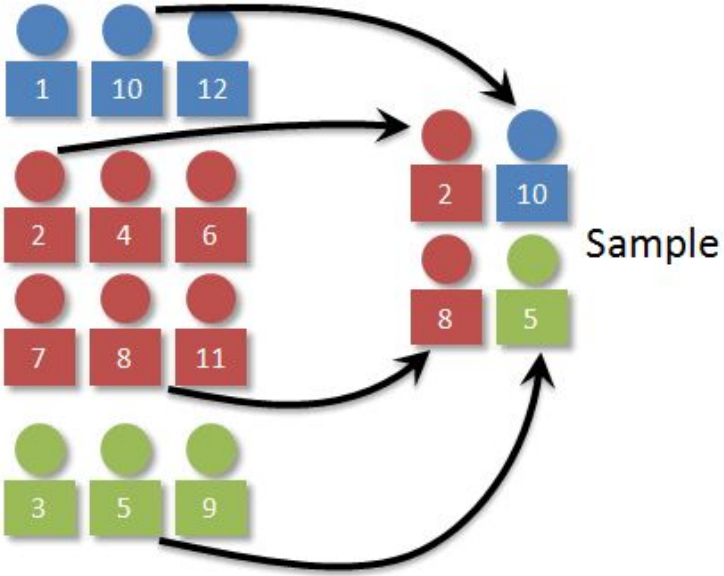
Simple random sampling

- The key idea is that each member of a population is **equally likely** to be selected for a sample
- We could randomly draw numbers from a hat, or use a computer generated selection
  - But, something has to intervene and introduce randomness
- In advance we have to know what our sample size should be.

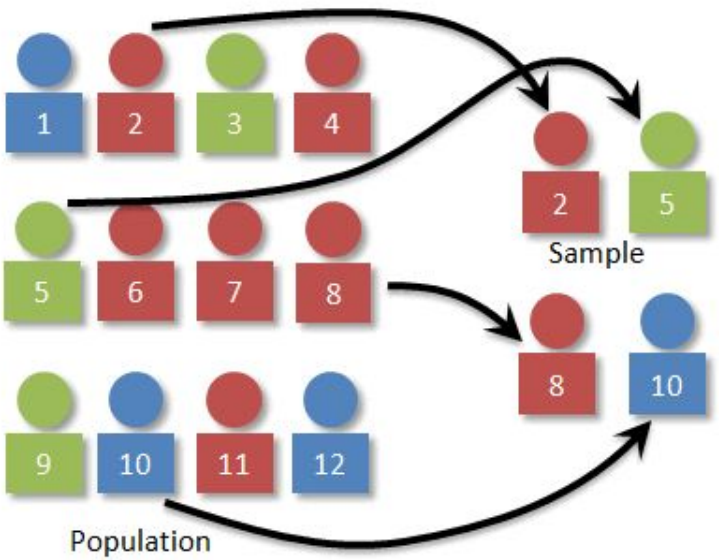


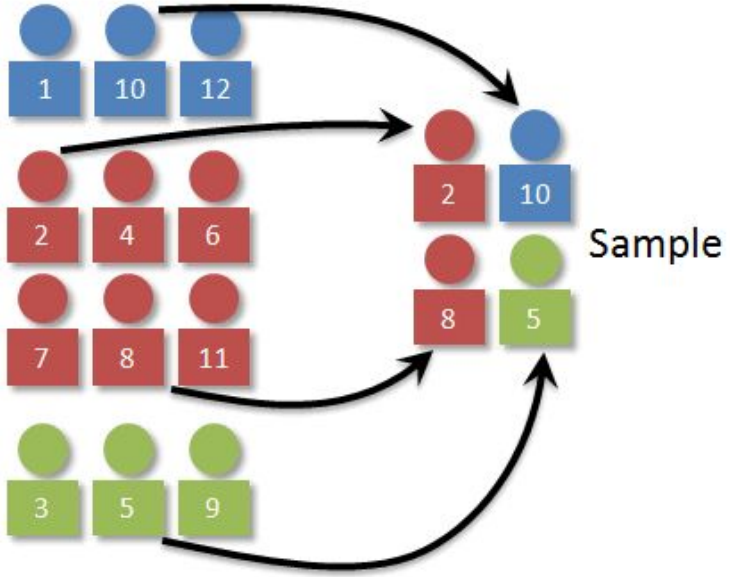
Stratified sampling

- We may identify, in advance, a construct or variable that is meaningful to our study
- We want to use a method of sampling that guarantees we have equal representation of different constructs
- We stratify our population (non-overlapping categories)
- We want a sample that is **proportional** to each ‘stratum’
- We select from each stratum proportionally



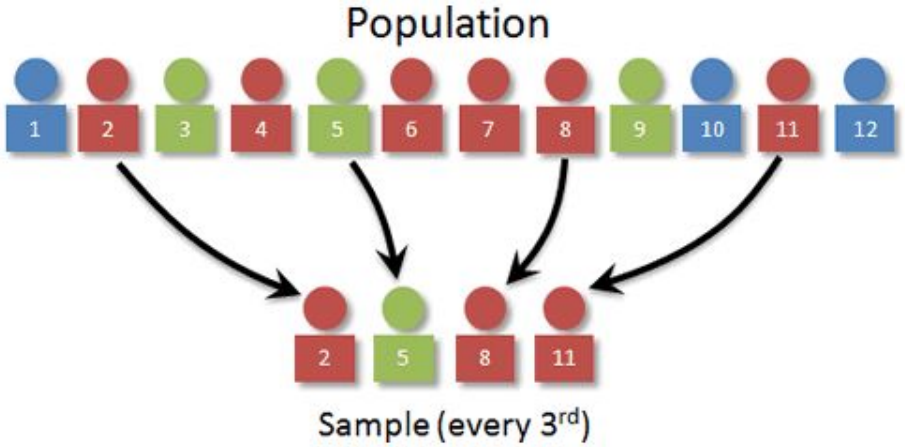
Random + Stratified sampling





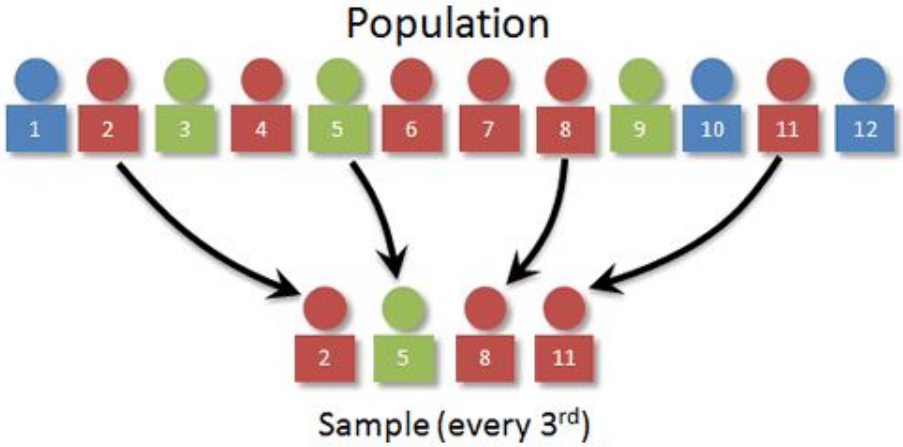
Systematic Sampling

- We may not know, in advance, much about our population of study.
- We could choose to sample simply with a number in mind (random)
- Or, we could introduce a logic of how to select the sample.
- The easiest way to do this is selecting the k-th number in a sample



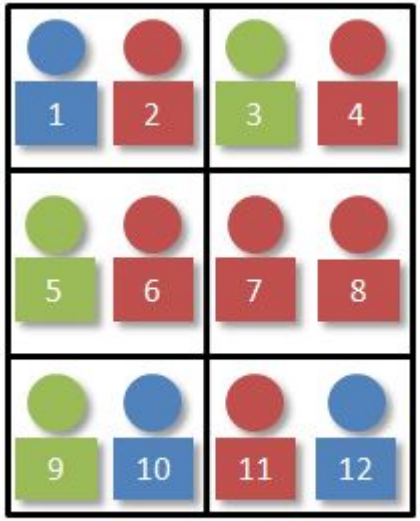
Systematic Sampling

- Systematic sampling can be used by both qualitative and quantitative researchers
- Cliques in a cafeteria - we may choose to interview the 4th student at every table. (Note there are no blues in our sample)



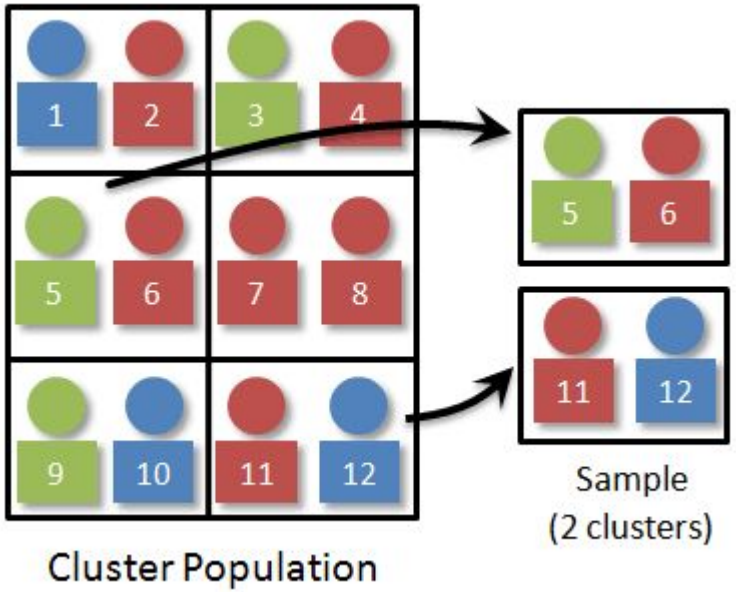
Cluster sampling

- Our population may have already been sorted or divided (clustered) in some meaningful way
  - Cafeteria tables
- We want to effectively sample for representation from existing clusters using the logic that the total number of clusters represents a population.



Cluster sampling

- To do this, we can choose clusters to sample.
- Doing so allows us to retain the internal cohesion of a cluster, and generalize from clusters instead of individuals
- This is valuable in census studies (block level sampling), manufacturing (packaged goods), and for evaluation of services (classes)



Random, stratified, and cluster sampling

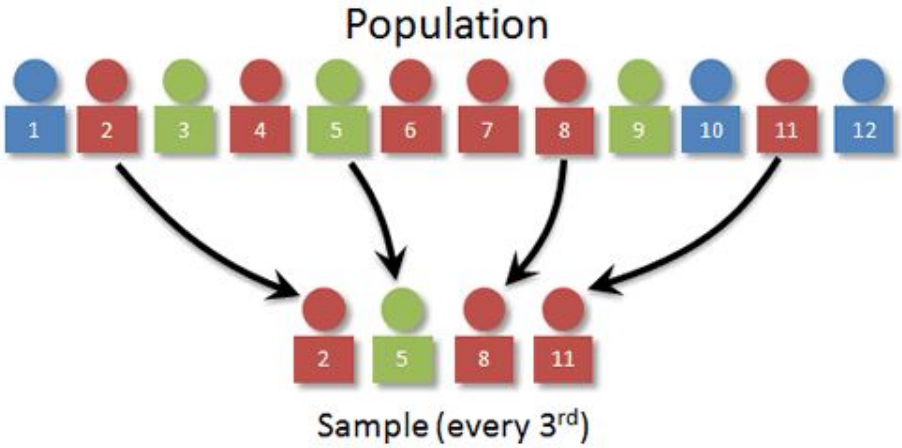
The differences in these methods of sampling

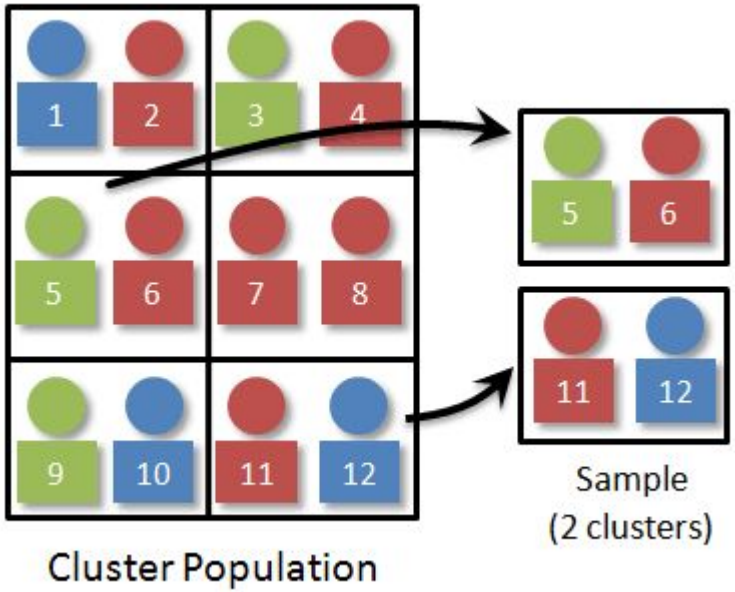
- How a population is arranged
- How we select individuals from the population

The similarities

- Attempting to control the **likelihood** that a sample is representative.
- Has a logic that is internal, but well established. The debates around validity of sampling is usually in execution and interpretation.
- Are very amenable to combination

Multi-stage sampling





Summing up - Be Weary of the Sample

- Bias is inherent in research
- When research is not transparent, or attempts to generalize from non-representative samples we encounter confusion and sometimes fraud
- **Pay attention to sampling in your everyday news consumption and research reading**.
- Sampling is easy to manipulate and shade.
    - Polls, statistics, and other forms of quantitative reporting often describe a sample without *justification* for why the sample is representative
    - This gets harder with more data and more visualizations