



Data Acquisition

- With the research plan in place, including a good definition of the research problem, we move along to collect our data
- Q: What data collection experience do we have in this room? What collection methods have we used?

Samples and Sampling

- A key part of data acquisition common to both human and physical geography is our thinking on samples and sampling

Samples and Sampling

- Sampling: how we choose which subjects to include in our study
- Idea covers which system(s) we use to select study subjects
- Which system we choose depends on
 1. The characteristics of what we are studying
 2. The requirements of our analytical methods

Samples and Sampling

- Q: Why is sampling an issue anyways? Why should geographers care?
 - Improper sample can skew results
 - Study results that are not reflective of reality
 - Example: Interviewing only “wealthy white men” for an opinion poll on society in general

Samples and Sampling

- Q: Why is sampling an issue anyways? Why should geographers care?
 - From a geographic perspective: over-represent some places and under-represent others

Sampling Methods

- Two major types of sampling methods
 - 1. Random
 - 2. Systematic
- Can take a spatial or aspatial approach to either (we will discuss implementation in a few minutes)

Sampling Methods

- Random Sampling
 - Maintains two important statistical principles
 - Equality principle: probabilities of inclusion are equal
 - Independence principle: one observation does not influence another

Sampling Methods

- Random Sampling
 - Important point: not “replacing” violates both of these principles
 - Example: draw names of possible survey subjects out of a draw drum
 - Can either replace names (statistically correct) or not replace names (not correct, but makes intuitive sense)
 - Q: Why is not replacing “intuitively” sensible?

Sampling Methods

- Random Sampling
 - Example of correct random sampling: 400 villages, want to select 100 randomly
 - 1. Number the villages 1-400
 - 2. Use a random number generator to select the villages required (possible to include a village twice, but not likely)

Sampling Methods

- Systematic Sampling
 - Violates equality rule
 - Example: choose every 4th village on a list of villages
 - Q: why isn't this OK?
 - Using this rule means certain villages have a 100% chance of inclusion while others have a 0% chance

Sampling Methods

- Systematic Sampling
 - Systematic sampling is obviously easier to implement than random selection
 - But, it introduces bias – strictly speaking, you cannot infer from sample to the population (although many researchers do anyways)

Sampling Methods

- **Stratified Sampling**
 - A “third way”
 - Applies when you know that there are differences within the population; identifiable groups

Sampling Methods

- **Stratified Sampling**
 - **Approach**
 - 1. Group the population into “strata”
 - 2. Sample from each stratum (either randomly or systematically)

Sampling Implementation

- **Simple Random**
 - 1. Set up coordinates
 - 2. Generate random number pairs
- **Problems**
 - Resolution level of grid
 - Origin/angle of grid
 - Clumping (points falling in groups)
 - Poor coverage

Sampling Implementation

- **Simple Systematic**
 - Sample at every grid intersection
- **Good Features**
 - Equal coverage across study area
 - Complete coverage across study area

Sampling Implementation

- **Simple Systematic**
 - Sample at every grid intersection
- **Bad Features**
 - Resolution level of grid
 - Origin/angle
 - Spatial periodicity
 - Can't infer

Sampling Implementation

- **Stratified Random Sampling**
 - 1. Create quadrats (strata)
 - 2. Random sampling inside each quadrat
- **Advantages**
 - Maintains a degree of randomness
 - Also ensure good coverage of entire study area

Sampling Implementation

- **Cluster Sampling**
 - 1. Set up quadrats/zones
 - 2. Select zones randomly
 - 3. Sample within these zones

Sampling Implementation

- **Cluster Sampling**
 - **Problems**
 - Can get poor coverage
 - Map pattern problem (the grid itself is systematic)
 - **Good Features**
 - Random (mostly)
 - Less travel for study (can be important)

Sampling Implementation

- **Cluster Sampling**
 - **Overall**
 - Choose the approach that fits your situation
 - Common sense comes before theoretical correctness