Random Sampling

Benjamin Graham
Office Hours: M 11:30-12:30, W 10:30-12:30
SSB 447

What is wrong with this question?

How would you rate the quality and difficulty of this course?
A. Supremely awesome
B. Very awesome
C. Awesome
D. Terrible
Housekeeping

- Jake Bowers talks next week
  - Noon Wednesday, 3:30pm Thursday
  - Must RSVP for Weds (you get lunch)
  - Substantial extra credit
- Today:
  - Finishing up some stuff on gathering data (questions)
  - Random Sampling
  - Next Week: Nonrandom Sampling
How many kids?

- Fertility rate in the US. Could be interesting as an independent or a dependent variable.

- How many children did your mother have?
Landon in a Landslide!

• 1936 Literary Digest Poll: nationwide sample
  • Ralph Landon: 1,296,669 (57%); Franklin Roosevelt 972,897
  • Broken down by state it predicted 370 electoral college votes for Landon

• What went wrong with that poll?
The Power of Random Sampling

- We often can’t measure the entire population we want to know about
  - Instead we measure a sample
  - A census vs. a poll
- We can make valid inferences about the population, based on the sample, if and only if:
  - The sample is a miniature version of your population
- The best way to do this is by random Sampling (AKA probability Sampling)
  - Every unit in the population has the same probability of being chosen
Some terminology

- **Sampling frame**: A list of all the units in the population
  - Sampling frame for the analysis is the only population we can make valid inferences about

- **Units**: Sometimes we sample at multiple levels, we have to keep our units straight
  - Households vs. individuals
  - Schools vs. classrooms vs. students

- **Variance**: A statistical way to talk about the diversity of the population
Which sample of heights has a higher variance?

- Group 1: 5’10” 6’ 5’11” 6’1” 6’2”
- Group 2: 6’10” 7‘ 4’11” 6’1” 5‘

- A: Group 1 has a higher variance
- B: Group 2 has a higher variance

- Which group has the higher mean?
The Weak Law of Large Numbers

• **The Weak Law of Large Numbers:** If you take a random sample of observations from a population....

• The mean of the sample approaches the mean of the population as the size of the sample approaches infinity (or approaches the size of the population).

• That’s great, but we have finite samples...
  • So how big does our sample have to be?
Central Limit Theorem

• **The Central Limit Theorem:** The mean of a sample of independent draws from the same population with expected value \( \mu \) and finite variance \( \sigma^2 \) will be normally distributed with an expected value of \( \mu \) and a variance of \( \sigma^2/n \).

• What that means in practice: If we take a random sample from a population and measure something about that sample:
  • The *expected value* of the mean in that sample is the mean of the population.
  • If we calculate the variance in the sample, so we know how close to the population mean our sample mean is going to be.
Random Sampling Error

• The error caused by observing a random sample, instead of the whole population.
  • The bigger the sample, the smaller the error.
  • The smaller the variance, the smaller the error.

![Diagram showing normal distribution with different variances](image.png)
How Do We Measure Unemployment?

• # looking for work/(# of employed people + # looking for work)
  • What is the appropriate population we’re Sampling from in this measure?

• How do we know how many people are looking for work?
The Power of Random Assignment

• Random Sampling is tied directly to random assignment
  • In an experiment, we randomly assign individuals to the treatment and the control group.

• Here, we have two random samples from the population
  • Across all variables the mean of the sample matches the population
  • Also matches the mean of the other sample

• This means that (with a few caveats) we can attribute any difference between the two groups to the treatment.
Providing Good Answer Choices

• Must be **exhaustive** and **mutually exclusive**

  • Order should be logical
  • Phrasing should be neutral

A. Very Happy
B. Happy
C. Very Unhappy
D. Unhappy
E. Neither Happy nor Unhappy
Providing Good Answer Choices (2)

- Must be **exhaustive** and **mutually exclusive**
  - Order should be logical
  - Phrasing should be neutral

A. Very Happy
B. Happy
E. Neither Happy nor Unhappy
D. Unhappy
C. Very Unhappy
Providing Good Answer Choices (3)

• Likert scales should be balanced
  A. Very Happy
  B. Happy
  E. Neither Happy nor Unhappy
  D. Unhappy
  C. Very Unhappy

  A. Extremely Happy
  B. Very Happy
  E. Happy
  D. Not Happy

• To give the neutral option or not

• Should you make “I don’t know” an option?
Framing Effects: It’s more than just one question

- Framing effects are all the other cues that researchers provide respondents (intentionally or not) that affect the way they answer questions.
  - Remember: The respondent is always trying to find the “right” answer.

- For in-person interviews:
  - Who is the enumerator? What are they wearing? What mood are they in? Where is the interview taking place?

- In written surveys: What questions has the interviewee already been asked?
  - How have they been “primed?”