

Fall 2022

PSO700 Research Method in Political Science

[Recitation 9]

Survey Research/Sampling and Big Data

October 28

Jungmin Han

The Outline of Today's Recitation

- Research Exercise Outline
- A Brief Review
 - Survey Research and Sampling
 - Benefits and Limitations of using Big Data
- Discussion
 - “Why It’s So Difficult to Poll Americans’ Feelings About Abortion”, Washington Post
 - “Eight (No, Nine!) Problems with Big Data”, New York Times
 - “The Battle to Prevent Another January 6 Features a New Weapon: The Algorithm”, Washington Post

Research Paper Outline (1)

Requirements

- 7 pages (double spaced, 12 pt font), excluding tables and references
- Due Sunday, December 11
- Test two hypotheses using R
- Interpret the test results
- Discuss strengths and weaknesses of your research design and strategies for improving the design in future research

Research Paper Outline (2)

- Hypotheses
 - Why do you expect IV #1 to cause DV? Discuss and provide a citation to *one* work in political science (journal article/book) that relates to your idea
 - Why do you expect IV #2 to cause DV?
- How are your variables measured? What weaknesses do you see in these measures in terms of validity and reliability?
- What is the appropriate statistical procedure for testing each hypothesis? (T-test of group means? Chi-Square for crosstabulation? Regression for continuous variables?)
- What do the R results tell you about each hypothesis? Can you reject the Null Hypothesis for each? How strong are the relationships that you have found?
- What do you conclude substantively about all of these relationships? What do they tell you about the way the political world works?
- How do you think all of these variables are related in a causal sense?
- What should future research do to improve these tests and improve on what you have done?

Research Paper Outline (3)

- *How to install R Studio?*

1. Download R which is a pre-requisite for R Studio:

<https://www.r-project.org>

2. Download R Studio:

<https://www.rstudio.com/products/rstudio/download/>

- DO NOT PAY ANYTHING: Just choose FREE source

What is Survey Research?

- Survey research is a method
 - For **collecting data**
 - By **asking questions**
 - To **a sample of individuals**
 - In a structured fashion and coding the response
- We use surveys as one of the “**passive observation designs**”
 - To obtain information on individual attitudes, perceptions, and behaviors
 - To generalize the information and the relationships we find in our sample to an overall population of interest

Some Terminologies in Survey Research

- **Population**

- The people (or other units of analysis) we want to generalize to
- E.g. U.S. potential voters; all democratic countries

- **Sample**

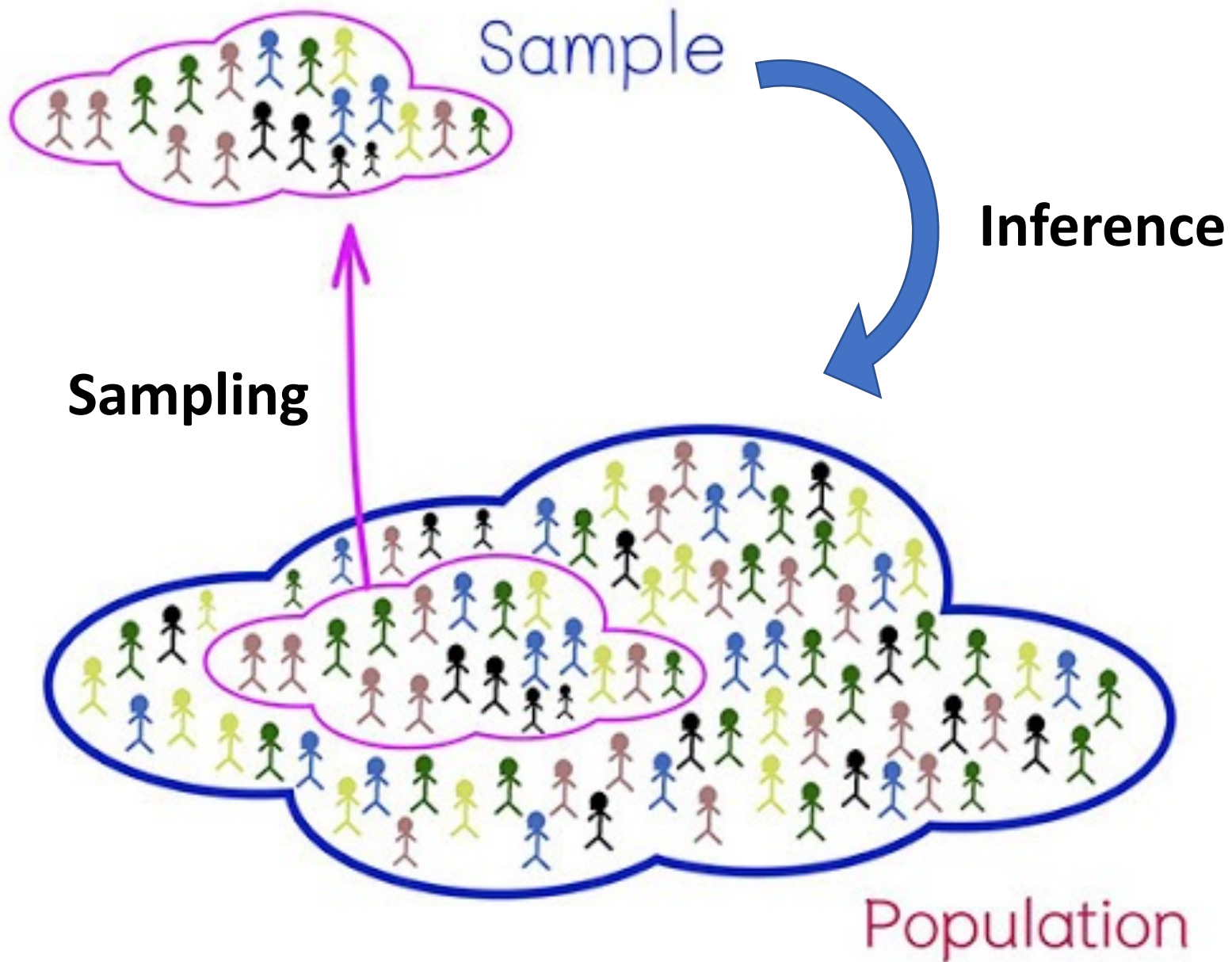
- The people (or other units of analysis) in our study
- E.g. people who completed the survey; countries included in the study

- **Probability sample**

- Each person/unit in the target population has a *known* chance of being selected for inclusion in the sample

- **Non-probability sample**

- The chance for each person/unit in the target population to be selected for inclusion in the sample is *unknown*



Random Sampling vs. Random Assignment

	Random Sampling	Random Assignment
What is “random” about the process?	Given a population of interests, how the sample is selected to be included in the study	Given a sample we have selected, how the treatments are assigned to the sample
What does the process help to achieve?	Sample is representative of the population	Treatment and control group are similar (in terms of confounders) to each other prior to the treatment
What advantage do we get from having this in our research design?	Increase external validity (conclusion can be generalized to the population)	Increase internal validity (can make credible causal inference)

General Strengths of the Survey Methods

- With proper sampling methods and good questionnaire design, surveys can yield:
 - High **descriptive** accuracy of attitudes, perceptions, and behaviors
 - High **predictive** accuracy for many future behaviors
 - Accurate assessments of the **covariation** between independent variables and dependent variables
- A good deal of **external validity**

Sources of Error in Surveys

- There are many potential sources of error in the survey process, many biases that may distort the findings
- Some relate to mechanics of survey research
 - Sample procedures and sampling error
 - Mode of data collection (telephone, internet, in-person)
 - Questionnaire construction
- Some are inherent to the method of survey research itself
 - Reliance on verbal reports
 - Passive observational data

Big Data: Strengths and Limitations

- Strengths of using Big Data in Political Science
 - Non-intrusive measurement of behavior and public opinion
 - Provision of information that are usually not approachable and observable.
 - Social networks and online base that provide an easy way to map network structure and density
- Limitations of Big Data in Political Science
 - Non-representativeness
 - Interpretation
 - Measurement Validity
 - Causal Inference
 - Data Quality

Article 1: “Why It’s So Difficult to Poll Americans’ Feelings About Abortion”, Washington Post

- According to the article, what makes it difficult to sort through Americans’ attitudes on abortion?
- Can you think of any way that helps us overcome the limitations of surveying people’s attitudes on abortion?

Article 2: “The Battle to Prevent Another January 6 Features a New Weapon: The Algorithm”, Washington Post

- According to the article, how can Big Data and Machine Learning contribute to the studies of social conflicts and violence?
- Still, there seem to be some limitations in using Big Data to predict social conflicts and violence. What are they?

Article3: “Eight (No, Nine!) Problems with Big Data”, New York Times

- Which do you think is most harmful to “scientific research” on political phenomenon among the suggested problems with Big Data in this article? Why?