Recitation 3: Introduction to Research Methods for Politics

Dept. of Politics, NYU

POL-850

Spring 2020

Agenda

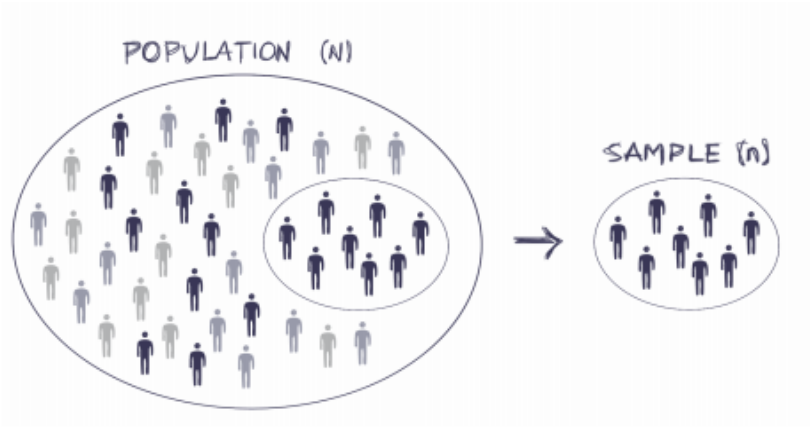
1. Review: Survey Research
2. Population Characteristics Measuring Brexit Support
3. Functions (table, prop.table)
4. Handling Missing Data
5. Two-Way Frequency and Propotion Tables

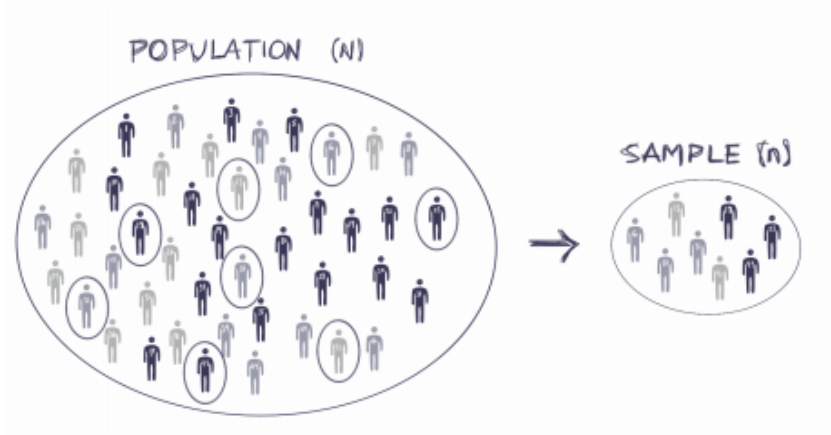
Survey Research

Survey Research

1. Goal of social science: estimate population characteristics using surveys
2. Necessary because collecting data from every person can be costly or infeasible
3. Solution? Collect data from a small subset of observations (sample) to understand target population
4. What does it mean for a sample to be representative? How do social scientistics attempt to collect a representative sample?

Representative Sample





What are the consequences for inference if a survey is unrepresentative?

Challenges to Random Sampling

1. Obtaining the correct sampling frame
2. Unit non-response
3. Item non-response
4. Misreporting
5. How might these challenges bias survey results?

Measuring Brexit Support

British Election Survey (BES) Variable Description

variable	description
vote	vote intention in the EU referendum: "leave", "stay", "don't know", or "won't vote"
leave	leave voters: 1=intends to vote "leave" or 0= intends to vote "stay"; (NA=doesn't know or won't vote)
education	highest educational qualification: 1=no qualifications, 2=general certificate of secondary education (GCSE), 3=general certificate of education advanced level (GCE A level), 4=undergraduate degree, or 5=postgraduate degree; (NA=no answer)
age	age (in years)

Load and Understand the Data

1. Load the data
2. Display the first 5 observations
3. Exercise #1: Identify each variable's *type*
4. Discuss results from part (3) with a partner

Functions (table, prop.table)

Introduction: table

Syntax: table(dataframe$variable)

1. A frequency table for a variable shows (a) the **value** a variable takes and (b) the **number of times** each value appears

```
> freq_table <- table(bes$leave) # object with frequency table
> freq_table

    0     1
14352 13692
>
```

2. How do we interpret these values?
3. Exercise #2: Create a frequency table using the variable **vote**

Introduction: prop.table

syntax: prop.table(table(dataframe$variable))

1. Shows the proportion of observations each value of a variable takes using two different methods

```
41  ## option a: create frequency table first
42  ## then create proprotion table
43  freq_table <- table(bes$vote)
44  prop.table(freq_table)
45
46  ## option b: do it all at once
47  prop.table(table(bes$vote)) # creates table of proportions
```

2. Exercise #3: Using either method of your choosing, create a proportion table using the **education** variable

Handling Missing Data

Handling Missing Data

- In survey data, we often encounter missingness in the data. In R, missing values are represented using **NA**
- Some functions automatically remove missing values while others do not

```
57  # table() including NAs
58  table(bes$education, exclude=NULL)
59
60  # mean() without removing NAs
61  mean(bes$leave)
62
63  # mean() removing NAs
64  mean(bes$leave, na.rm=TRUE)
65
66  # removes observations with NAs if
67  ##at least ONE column contains an NA value
68  bes1 <- na.omit(bes)
```

Two-Way Tables

Two-Way Frequency Tables

syntax: table(dataframe$variable1, dataframe$variable2)

1. Show the # of observations that take each combination of values of two variables

```
> ## Two-way frequency tables
> table(bes1$leave, bes1$education)

     1    2    3    4    5
0  498 1763 3014 6081 1898
1 1356 3388 2685 3783  631
```

2. Exercise #4: With a partner, discuss how to interpret the output of the two-way frequency table

Two-Way Proportion Tables

- syntax: prop.table(table(dataframe$variable1, dataframe$variable2))
- optional arguments: margin = 1 or margin = 2

1. Shows the proportion of observations that take each combination of values of the two variables of interest

```
78
79  ## Two-way tables of proportions
80  prop.table(table(bes1$leave, bes1$education))
81
82  ## Advanced: margins
83  ## Proportion of different education levels
84  ## within Brexit supports and non-supporters
85
86  prop.table(table(bes1$leave, bes1$education), margin=1)
87  ## Proportion of Brexit supporters and non-supporters
88  ## within different education levels
89  prop.table(table(bes1$leave, bes1$education), margin=2)
90
```

2. Margins = 1(2) $\rightarrow$ Variable #1 (2) is reference group