

Midterm Roadmap, Surveys, Prediction

Communication Research Methods

Jennifer Pan

Assistant Professor
Department of Communication
Stanford University

February 3, 2016

Announcements

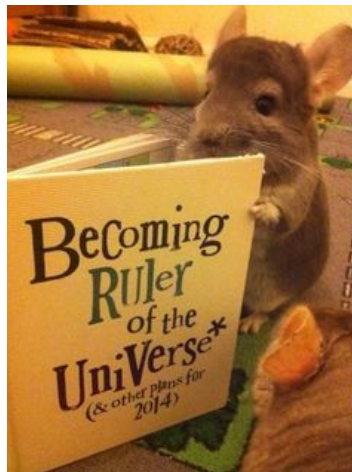
- ▶ Midterm Feb. 5: no books, no laptops, multiple choice
- ▶ Midterm review Feb 4, Feb 5 section (no section Monday Feb 8)

Outline

Our course goals

By the end of the course, you should be able to:

1. explain the basic concepts of research
2. better understand, interpret, and criticize scientific research
3. use basic features of the *R* programming language



Our course goals and course thus far

-
1. explain the basic concepts of research • **Week 1:** research, concepts, measurement
2. better understand, interpret, and criticize scientific research • **Week 2:** introduction to R
3. use basic features of the R programming language • **Week 3:** causality
- Week 4:** Getting and summarizing data
- > Summarizing data w/ descriptive statistics, e.g., visualization, central tendency, spread
 - > Plots in R
 - > Random sampling to get representative data and avoid bias

Our course goals and course thus far

-
1. explain the basic concepts of research
2. better understand, interpret, and criticize scientific research
3. use basic features of the R programming language
- Week 1:** research, concepts, measurement
- ▶ Research is about expanding knowledge; scientific research involves inference, public procedures, estimates of uncertainty
 - ▶ Theories are explanations with properties we like, and concepts are the phenomenon we theorize about
 - ▶ To test theories, we need to measure concepts; measured concepts = data; need to be careful of random error (reliability) and systematic error (validity) when measuring
- Week 2:** introduction to R
- Week 3:** causality
- Week 4:** Getting and summarizing data
- > Summarizing data w/ descriptive statistics, e.g., visualization, central tendency, spread
 - > Plots in R
 - > Random sampling to get representative data and avoid bias

Our course goals and course thus far

-
1. explain the basic concepts of research **Week 1: research, concepts, measurement**
2. better understand, interpret, and criticize scientific research **Week 2: introduction to R**
3. use basic features of the R programming language
- Week 3: causality**
- Week 4: Getting and summarizing data**
- > Summarizing data w/ descriptive statistics, e.g., visualization, central tendency, spread
 - > Plots in R
 - > Random sampling to get representative data and avoid bias

Our course goals and course thus far

-
1. explain the basic concepts of research
 - Week 1: research, concepts, measurement
 - Week 2: introduction to R
 - ▶ R as calculator
 - ▶ R basics: objects, vectors, functions, loading data
 - Week 3: causality
 - Week 4: Getting and summarizing data
 - > Summarizing data w/ descriptive statistics, e.g., visualization, central tendency, spread
 - > Plots in R
 - > Random sampling to get representative data and avoid bias
 2. better understand, interpret, and criticize scientific research
 - Week 2: introduction to R
 - Week 3: causality
 - Week 4: Getting and summarizing data
 3. use basic features of the R programming language
 - Week 3: causality
 - Week 4: Getting and summarizing data

Our course goals and course thus far

-
1. explain the basic concepts of research **Week 1: research, concepts, measurement**
2. better understand, interpret, and criticize scientific research **Week 2: introduction to R**
3. use basic features of the R programming language **Week 3: causality**
- Week 4: Getting and summarizing data**
- > Summarizing data w/ descriptive statistics, e.g., visualization, central tendency, spread
 - > Plots in R
 - > Random sampling to get representative data and avoid bias

Our course goals and course thus far

-
1. explain the basic concepts of research
 - Week 1: research, concepts, measurement
 2. better understand, interpret, and criticize scientific research
 - Week 2: introduction to R
 3. use basic features of the R programming language
 - Week 3: causality
 - ▶ Causal effect = actual – counterfactual
 - ▶ Finding comparable cases (treatments and controls) to statistically test for causal inference with experiments,
 - Week 4: better and summarizing data
 - > observational studies/ descriptive statistics,
 - ▶ Subsetting data in R w/ experiment data
 - > Plots in R
 - > Random sampling to get representative data and avoid bias

Our course goals and course thus far

-
1. explain the basic concepts of research
2. better understand, interpret, and criticize scientific research
3. use basic features of the R programming language
- Week 1: research, concepts, measurement
- Week 2: introduction to R
- Week 3: causality
- Week 4: Getting and summarizing data
- > Summarizing data w/ descriptive statistics, e.g., visualization, central tendency, spread
 - > Plots in R
 - > Random sampling to get representative data and avoid bias

Our course goals and course thus far

-
1. explain the basic concepts of research • **Week 1:** research, concepts, measurement
2. better understand, interpret, and criticize scientific research • **Week 2:** introduction to R
3. use basic features of the R programming language • **Week 3:** causality
- **Week 4:** Getting and summarizing data
- ▶ Summarize data w/ descriptive statistics, e.g., visualization, central tendency, spread
 - ▶ Random sampling to get representative data and avoid bias
 - ▶ Estimate pop. parameter w/uncertainty

Surveys and Sources of bias

- ▶ Sampling bias and random sampling error apply when doing surveys
- ▶ Other sources of bias in surveys:
 - ▶ **Unit non-response**: failure to reach selected units (in Afghanistan survey, 2754 out of 3097 sampled respondents agreed to participate → 11% refusal rate)
 - ▶ **Item non-response**: respondents refuse to answer certain survey questions (in Afghanistan survey, the income variable had a non-response rate of approximately 5%)
 - ▶ **Mis-reporting**



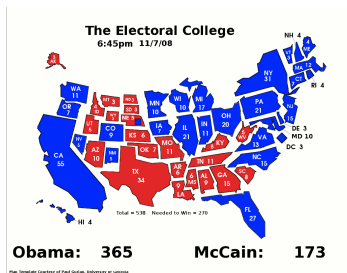
Surveys and Sources of bias: Misreporting

Different types (reasons) for mis-reporting

- ▶ Social desirability bias: respondents choose an answer that is seen as socially desirable regardless of what they really think
 - ▶ Asking question about support for foreign forces in Afghanistan sensitive
 - ▶ Will you vote in the next election?
 - ▶ How often do you lie to people close to you?
 - ▶ How many sexual partners have you had?
 - ▶ Would you be upset if a black family moved next door?
- ▶ Observer effects: people being surveyed are affected (act differently) by being observed

Predict Outcome of the 2008 Presidential Election

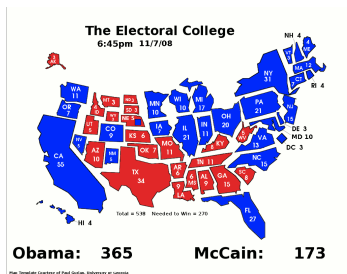
Figure : Map of electoral college votes for the 2008 election



- ▶ Candidate with absolute majority of electoral votes elected as president
- ▶ Total of 538 electoral votes, allocated to states (435 + 100)
- ▶ Electors vote for candidate who won plurality of votes in state they represent: “winner-take-all” system for each state
- ▶ Winning presidential candidate must obtain at least 270 electoral votes
- ▶ To predict presidential election, we must accurately predict winner of each state

Predict Outcome of the 2008 Presidential Election

Figure : Map of electoral college votes for the 2008 election



- ▶ Election results by state: `pres08.csv`
- ▶ Many polls in each state leading up to election: `polls08.csv`
- ▶ Use polls conducted within each state on the day closest to election day
- ▶ Prediction error = actual outcome - predicted outcome; Obama's vote margin
- ▶ Calculate in R: for loop, `as.Date()`, `unique()`, `subset()`, `mean()`
- ▶ Visualize errors: `hist()`, `plot()`