

# Running Online Surveys with Nonprobability Samples

Thomas J. Leeper

Department of Political Science and Government  
Aarhus University

10 December 2015  
EUI Quantitative Methods Working Group

# Overview

- What is this course about?
- Who are you?

# Learning Goals

By the end of today you should be able to:

- Describe logic of design-based and model-based representativeness
- Evaluate the quality of a convenience sample
- Evaluate trade-offs between client-side and server-side technologies for behavioral research
- Design simple web forms using several tools
- Apply all of this to your own research

- 1 "The Gold Standard"
- 2 Web Questionnaires
- 3 Recruitment in Practice
- 4 Challenges and Opportunities

# Introductions

- Who are you?
- What field are you from?

# About Me

- Assistant Professor at London School of Economics since September
- Postdoc at Aarhus University 2012–2015
- PhD in Political Science from Northwestern University (2015)
- Interested in:
  - Political psychology
  - Survey–experimental methods
  - Reproducible computational social science

**How many of you have...**

# How many of you have...

- taken a course on survey methods?



# How many of you have...

- taken a course on survey methods?
- taken a course on experimental design?

# How many of you have...

- taken a course on survey methods?
- taken a course on experimental design?
- written HTML markup before?

# How many of you have...

- taken a course on survey methods?
- taken a course on experimental design?
- written HTML markup before?
- run a study on MTurk or Crowdfunder, with a vendor like YouGov, or another online platform?

1 "The Gold Standard"

2 Web Questionnaires

3 Recruitment in Practice

4 Challenges and Opportunities

# "The Gold Standard"

*a population-based experiment uses survey sampling methods to produce a collection of experimental subjects that is representative of the target population of interest for a particular theory . . . the population represented by the sample should be representative of the population ot which the researcher intends to extend his or her findings. In population-based experiments, experimental subjects are randomly assigned to conditions by the researcher*

p2. from Mutz, Diana. 2011. *Popuation-Based Survey Experiments*. Princeton University Press.

# Causal Inference in Experiments I

- Causal inference is a comparison of two *potential outcomes*

# Causal Inference in Experiments I

- Causal inference is a comparison of two *potential outcomes*
- A potential outcome is the value of the outcome ( $Y$ ) for a given unit ( $i$ ) after receiving a particular version of the treatment ( $X$ )

# Causal Inference in Experiments

## I

- Causal inference is a comparison of two *potential outcomes*
- A potential outcome is the value of the outcome ( $Y$ ) for a given unit ( $i$ ) after receiving a particular version of the treatment ( $X$ )
- Each unit has multiple *potential* outcomes, but we only observe one of them



# Causal Inference in Experiments I

- Causal inference is a comparison of two *potential outcomes*
- A potential outcome is the value of the outcome ( $Y$ ) for a given unit ( $i$ ) after receiving a particular version of the treatment ( $X$ )
- Each unit has multiple *potential* outcomes, but we only observe one of them
- A *causal effect* is the difference between these (e.g.,  $Y_{X=1} - Y_{X=0}$ ), all else constant

# Causal Inference in Experiments II

- We cannot see individual-level causal effects

# Causal Inference in Experiments II

- We cannot see individual-level causal effects
- We can see *average causal effects*
  - Ex.: Average difference in participation between those with and without university degrees

# Causal Inference in Experiments II

- We cannot see individual-level causal effects
- We can see *average causal effects*
  - Ex.: Average difference in participation between those with and without university degrees
- We want to know:  $TE_i = Y_{1i} - Y_{0i}$

# Causal Inference in Experiments III

- We want to know:  $TE_i = Y_{1i} - Y_{0i}$

# Causal Inference in Experiments III

- We want to know:  $TE_i = Y_{1i} - Y_{0i}$
- We can average:  
 $ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$

# Causal Inference in Experiments

## III

- We want to know:  $TE_i = Y_{1i} - Y_{0i}$
- We can average:  
 $ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$
- But we still only see one potential outcome for each unit:

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0]$$

# Causal Inference in Experiments III

- We want to know:  $TE_i = Y_{1i} - Y_{0i}$

- We can average:

$$ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$$

- But we still only see one potential outcome for each unit:

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0]$$

- Is this what we want to know?



# Causal Inference in Experiments IV

- What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \quad (1)$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \quad (2)$$

# Causal Inference in Experiments IV

- What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \quad (1)$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \quad (2)$$

- Are the following statements true?
  - $E[Y_{1i}] = E[Y_{1i}|X = 1]$
  - $E[Y_{0i}] = E[Y_{0i}|X = 0]$

# Causal Inference in Experiments IV

- What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \quad (1)$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \quad (2)$$

- Are the following statements true?

- $E[Y_{1i}] = E[Y_{1i}|X = 1]$

- $E[Y_{0i}] = E[Y_{0i}|X = 0]$

- Not in general!

# Causal Inference in Experiments

## V

- Only true when both of the following hold:

$$E[Y_{1i}] = E[Y_{1i}|X = 1] = E[Y_{1i}|X = 0] \quad (3)$$

$$E[Y_{0i}] = E[Y_{0i}|X = 1] = E[Y_{0i}|X = 0] \quad (4)$$

- In that case, potential outcomes are *independent* of treatment assignment

- If true, then:

$$\begin{aligned} ATE_{naive} &= E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \quad (5) \\ &= E[Y_{1i}] - E[Y_{0i}] \\ &= ATE \end{aligned}$$

# Causal Inference in Experiments VI

- This holds in experiments because of randomization
  - Units differ only in what side of coin was up
  - Experiments randomly reveal potential outcomes

# Causal Inference in Experiments VI

- This holds in experiments because of randomization
  - Units differ only in what side of coin was up
  - Experiments randomly reveal potential outcomes
- Matching/regression/etc. attempts to eliminate those confounds, such that:

$$E[Y_{1i}|Z] = E[Y_{1i}|X = 1, Z] = E[Y_{1i}|X = 0, Z]$$

$$E[Y_{0i}|Z] = E[Y_{0i}|X = 1, Z] = E[Y_{0i}|X = 0, Z]$$



# Surveys Start with an Inference Population

- We want to speak to a population
- But what population is it?



# Surveys Start with an Inference Population

- We want to speak to a population
- But what population is it?
  - A national population?

# Surveys Start with an Inference Population

- We want to speak to a population
- But what population is it?
  - A national population?
  - Adults in Western, industrialized democracies?

# Surveys Start with an Inference Population

- We want to speak to a population
- But what population is it?
  - A national population?
  - Adults in Western, industrialized democracies?
  - All human beings?

# Surveys Start with an Inference Population

- We want to speak to a population
- But what population is it?
  - A national population?
  - Adults in Western, industrialized democracies?
  - All human beings?
- This is rarely specified, but is important when we think about whether a sample is appropriate

# Think about your own work

Consider the following:

- 1 What is your research about?
- 2 What population do you intend to generalize to?

Discuss with the person next to you.

# A Hypothetical Census

- Advantages

- Disadvantages

# A Hypothetical Census

- Advantages
  - Perfectly representative
  - Sample statistics are population parameters
  
- Disadvantages

# A Hypothetical Census

- Advantages

- Perfectly representative
- Sample statistics are population parameters

- Disadvantages

- Costs
- Feasibility
- Need



# So, instead we sample!

# Sampling Frames

- Enumeration (listing) of all units eligible for sample selection
- Random sample from that list

# Sampling Frames

- Enumeration (listing) of all units eligible for sample selection
- Random sample from that list
- Building a sampling frame
  - Combine existing lists
  - Canvass/enumerate from scratch

# Sampling Frames

- Enumeration (listing) of all units eligible for sample selection
- Random sample from that list
- Building a sampling frame
  - Combine existing lists
  - Canvass/enumerate from scratch
- Concern about coverage: Does frame match population?

# Sample Estimates from an SRS

- Each unit in frame has equal probability of selection
- Sample statistics are unweighted
- Variances are easy to calculate
- Easy to calculate sample size need for a particular variance

# Sample mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (5)$$

where  $y_i$  = value for a unit, and  
 $n$  = sample size

$$SE_{\bar{y}} = \sqrt{(1 - f) \frac{s^2}{n}} \quad (6)$$

where  $f$  = proportion of population sampled,  
 $s^2$  = sample element variance, and  
 $n$  = sample size

# Estimating sample size

If all we cared about was a single proportion:

$$\text{Var}(p) = (1 - f) \frac{p(1 - p)}{n - 1} \quad (7)$$

Given a large population:

$$\text{Var}(p) = \frac{p(1 - p)}{n - 1} \quad (8)$$

Need to solve the above for  $n$ .

(9)

# Estimating sample size

If all we cared about was a single proportion:

$$\text{Var}(p) = (1 - f) \frac{p(1 - p)}{n - 1} \quad (7)$$

Given a large population:

$$\text{Var}(p) = \frac{p(1 - p)}{n - 1} \quad (8)$$

Need to solve the above for  $n$ .

$$n = \frac{p(1 - p)}{v(p)} = \frac{p(1 - p)}{SE^2} \quad (9)$$



# Estimating sample size

Determining sample size requires:

- A possible value of  $p$
- A desired precision (standard error)

If support for each coalition is evenly matched ( $p = 0.5$ ):

$$n = \frac{0.5(1 - 0.5)}{SE^2} = \frac{0.25}{SE^2} \quad (10)$$

# Estimating sample size

What precision (margin of error) do we want?

- +/- 2 percentage points:  $SE = 0.01$

$$n = \frac{0.25}{0.01^2} = \frac{0.25}{0.0001} = 2500 \quad (11)$$

# Estimating sample size

What precision (margin of error) do we want?

- +/- 2 percentage points:  $SE = 0.01$

$$n = \frac{0.25}{0.01^2} = \frac{0.25}{0.0001} = 2500 \quad (11)$$

- +/- 5 percentage points:  $SE = 0.025$

$$n = \frac{0.25}{0.000625} = 400 \quad (12)$$

# Estimating sample size

What precision (margin of error) do we want?

- +/- 2 percentage points:  $SE = 0.01$

$$n = \frac{0.25}{0.01^2} = \frac{0.25}{0.0001} = 2500 \quad (11)$$

- +/- 5 percentage points:  $SE = 0.025$

$$n = \frac{0.25}{0.000625} = 400 \quad (12)$$

- +/- 0.5 percentage points:  $SE = 0.0025$

$$n = \frac{0.25}{0.00000625} = 40,000 \quad (13)$$

# Sampling Considerations. . .

# Sampling Considerations. . .

- More complex designs possible, all based on each unit having a *known, non-zero* probability of being sampled
  - Stratified sampling can produce lower variances

# Sampling Considerations...

- More complex designs possible, all based on each unit having a *known, non-zero* probability of being sampled
  - Stratified sampling can produce lower variances
- Random sampling ensures that samples are, *in expectation*, representative of the population *in all respects*
  - Demographics
  - Psychological traits
  - Covariances
  - Potential outcomes

# Sampling Considerations. . .

- More complex designs possible, all based on each unit having a *known, non-zero* probability of being sampled
  - Stratified sampling can produce lower variances
- Random sampling ensures that samples are, *in expectation*, representative of the population *in all respects*
  - Demographics
  - Psychological traits
  - Covariances
  - Potential outcomes
- Only works in a world of perfect coverage and zero response bias
  - Says nothing about response rate!



# Combining Probability Sampling and Experimental Design

- Sample is representative of population in every respect (in expectation)
- Sample Average Treatment Effect (SATE) is the average of the sample's individual-level treatment effects
  - Unbiased estimate of PATE
  - Not necessarily any unit's individual treatment effect
  - Blocking might reduce variance
- Says nothing about effect heterogeneity
  - Design is optimized for estimating SATE

# Representativeness

What does it mean for a sample to be representative?

# Representativeness

What does it mean for a sample to be representative?

- Census

# Representativeness

What does it mean for a sample to be representative?

- Census
- SRS (or more complex) design

# Representativeness

What does it mean for a sample to be representative?

- Census
- SRS (or more complex) design
- Quota sampling (common prior to the 1940s)

# Representativeness

What does it mean for a sample to be representative?

- Census
- SRS (or more complex) design
- Quota sampling (common prior to the 1940s)
- Demographics match population

# Representativeness

What does it mean for a sample to be representative?

- Census
- SRS (or more complex) design
- Quota sampling (common prior to the 1940s)
- Demographics match population
- Others?

# Representativeness

What does it mean for a sample to be representative?

- Census
- SRS (or more complex) design
- Quota sampling (common prior to the 1940s)
- Demographics match population
- Others?

Which of these matter? Discuss with the person next to you.



# My View

100% design-based inference does not exist

# My View

100% design-based inference does not exist

- All survey designs involve reweighting adjustments

# My View

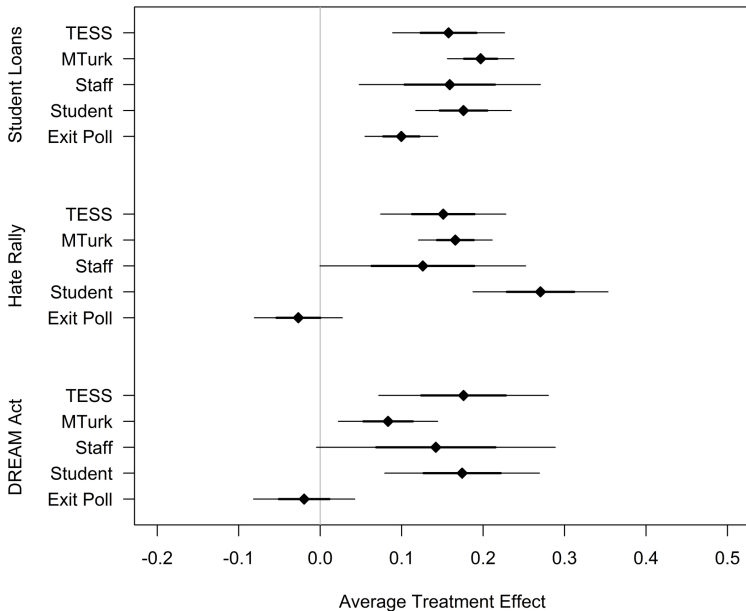
100% design-based inference does not exist

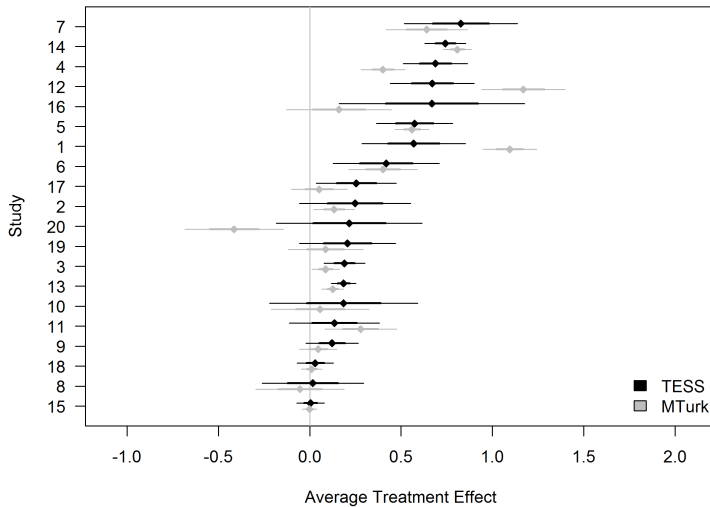
- All survey designs involve reweighting adjustments
- Representativeness is a more complex issue than demographic comparisons

# My Own Research

	GfK	Poll	Student	Staff	MTurk	Ads	ANES
<b>Dem. (%)</b>	51.3	86.1	75.7	66.4	62.1	72.1	46.2
<b>Rep. (%)</b>	46.0	7.7	17.8	16.4	20.3	14.7	39.3
<b>Lib. (%)</b>	27.8	75.4	68.5	62.7	60.4	66.2	23.8
<b>Con. (%)</b>	35.3	9.4	14.7	19.8	19.1	17.7	36.1
<b>Fem. (%)</b>	51.1	60.8	56.4	50.8	41.7	65.3	51.9
<b>White (%)</b>	77.9	67.6	62.9	60.2	76.0	53.8	80.4
<b>Age</b>	49.4	40-49	18-24	25-34	25-34	25-34	50-54
<b>Interest</b>	2.8	3.5	3.2	2.8	2.7	3.0	3.0
<b>N</b>	593	741	299	128	1024	80	—

Mullinix et al. In press. "The Generalizability of Survey Experiments."  
*Journal of Experimental Political Science.*





# SUTO Framework

- Cronbach (1986) talks about generalizability in terms of UTO
- Shadish, Cook, and Campbell (2001) speak similarly of:
  - **S**ettings
  - **U**nits
  - **T**reatments
  - **O**utcomes
- External validity depends on all of these things

## Population

- Setting
- Units
- Treatments
- Outcomes

## Your Study

- Setting
- Units
- Treatments
- Outcomes



## Population

- Setting
- Units
- Treatments
- Outcomes

## Your Study

- Setting
- Units
- Treatments
- Outcomes

In your study, how do these correspond?

## Population

- Setting
- Units
- Treatments
- Outcomes

## Your Study

- Setting
- Units
- Treatments
- Outcomes

In your study, how do these correspond?  
how do these differ?

## Population

- Setting
- Units
- Treatments
- Outcomes

## Your Study

- Setting
- Units
- Treatments
- Outcomes

In your study, how do these correspond?  
how do these differ?  
do these differences matter?

# Common Differences

- Most common thing to focus on is demographic representativeness
  - Sears (1986): "students aren't real people"
  - Western, educated, industrialized, rich, democratic (WEIRD) psychology participants

# Common Differences

- Most common thing to focus on is demographic representativeness
  - Sears (1986): "students aren't real people"
  - Western, educated, industrialized, rich, democratic (WEIRD) psychology participants
- But do those characteristics actually matter?

# Common Differences

- Most common thing to focus on is demographic representativeness
  - Sears (1986): "students aren't real people"
  - Western, educated, industrialized, rich, democratic (WEIRD) psychology participants
- But do those characteristics actually matter?
- Shadish, Cook, and Campbell tell us to think about:
  - Surface similarities
  - Ruling out irrelevancies
  - Making discriminations
  - Interpolation/extrapolation

# Focus on effect heterogeneity

- Think about and make an evidence-based argument for why you think there are (or are not) heterogeneous effects

# Focus on effect heterogeneity

- Think about and make an evidence-based argument for why you think there are (or are not) heterogeneous effects
- If you think there is heterogeneity, then we probably do not care about the SATE anyway
  - Focus on CATE instead
  - $E[Y_{1i}|X = 1, Z = z] - E[Y_{0i}|X = 0, Z = z]$



# Focus on effect heterogeneity

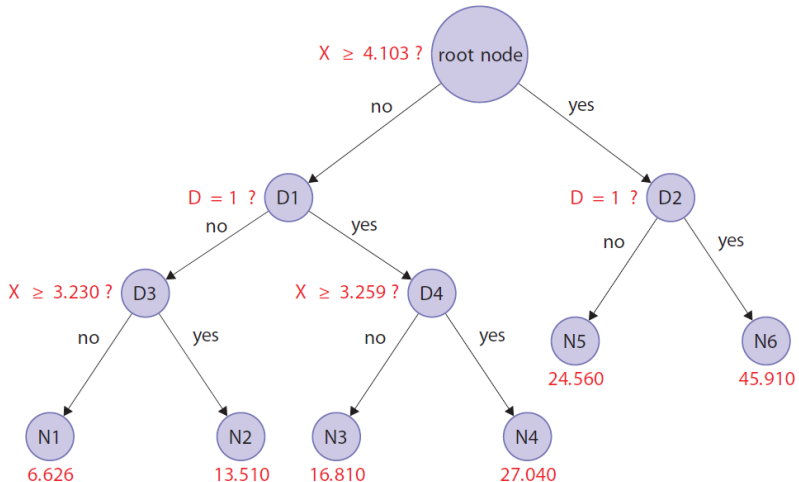
- Think about and make an evidence-based argument for why you think there are (or are not) heterogeneous effects
- If you think there is heterogeneity, then we probably do not care about the SATE anyway
  - Focus on CATE instead
  - $E[Y_{1i}|X = 1, Z = z] - E[Y_{0i}|X = 0, Z = z]$
- Two basic analytic strategies
  - Regression with large number of interactions
  - Bayesian Additive Regression Trees

# Focus on effect heterogeneity

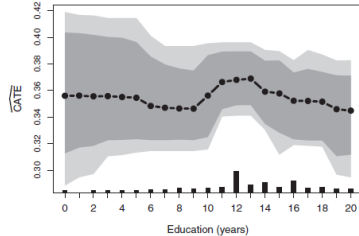
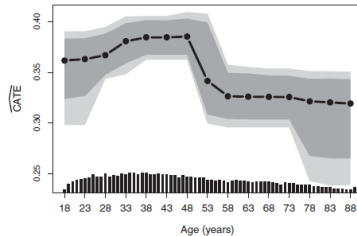
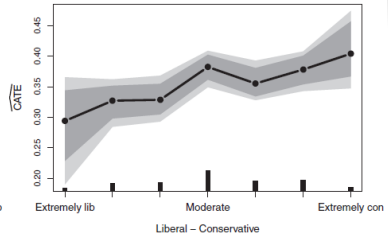
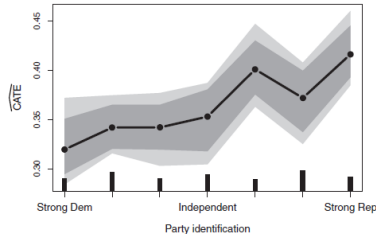
- Think about and make an evidence-based argument for why you think there are (or are not) heterogeneous effects
- If you think there is heterogeneity, then we probably do not care about the SATE anyway
  - Focus on CATE instead
  - $E[Y_{1i}|X = 1, Z = z] - E[Y_{0i}|X = 0, Z = z]$
- Two basic analytic strategies
  - Regression with large number of interactions
  - Bayesian Additive Regression Trees
- But remember: you have to convince reviewers!

# BART

- Estimate *conditional average treatment effects*
- BART is essentially an ensemble machine learning method
- Iteratively split a sample into more and more homogeneous groups until some threshold is reached using binary (cutpoint) decisions
- Repeat this a bunch of times, aggregating across results



Green and Kern. 2012. "Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees." *Public Opinion Quarterly*.



Green and Kern. 2012. "Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees." *Public Opinion Quarterly*.

# Stratification/Blocking

As soon as we care about heterogeneous effects, it makes sense to stratify and block on factors that might moderate the treatment effect.

# Stratification/Blocking

As soon as we care about heterogeneous effects, it makes sense to stratify and block on factors that might moderate the treatment effect.

As soon as we identify all sources of heterogeneity, it doesn't matter what sample we use because effects are *by definition* homogeneous within such strata.

# Stratification/Blocking

As soon as we care about heterogeneous effects, it makes sense to stratify and block on factors that might moderate the treatment effect.

As soon as we identify all sources of heterogeneity, it doesn't matter what sample we use because effects are *by definition* homogeneous within such strata.

But, we never know when we've reached that point!



## Aside: Induced Value Theory

- Incentivized (economic) experiments rely on induced value theory
- This is a way to reduce heterogeneity
  - Incentives reduce variation across individuals
  - Sample characteristics should matter less (than in other types of research)
- Actually merits empirical testing, though

If we acknowledge and start thinking about effect heterogeneity, does this mean we can use any convenient group of participants as if they were probability samples?

No. Of course not.

# Not All Convenience Samples Are Alike

- Different types:

# Not All Convenience Samples Are Alike

- Different types:
  - Passive/opt-in/“river sampling”

# Not All Convenience Samples Are Alike

- Different types:
  - Passive/opt-in/"river sampling"
  - Sample of convenience (not a sample per se)
    - Email list
    - Snowball sample
    - Respondent-driven Sampling
    - Students
    - Crowdsourcing

# Not All Convenience Samples Are Alike

- Different types:
  - Passive/opt-in/"river sampling"
  - Sample of convenience (not a sample per se)
    - Email list
    - Snowball sample
    - Respondent-driven Sampling
    - Students
    - Crowdsourcing
- Differ in numerous ways
  - Cost
  - "Experience"
  - Attentiveness
  - Demographics

# Costs per participant

From one of my studies:

Sample	Cost	n	Cost/participant
National	\$13200	593	\$22.26
Exit Poll	\$3000	741	\$4.05
Students	\$0	299	\$0
Staff	\$1280	128	\$10.00
MTurk	\$550	1024	\$0.54
Ads	\$636	80	\$7.95

# Participant Experience

- A lot of growing concern about experience
- Larger literature on "panel conditioning"
  - Inconclusive evidence



# Participant Experience

- A lot of growing concern about experience
- Larger literature on "panel conditioning"
  - Inconclusive evidence

# Participant Experience

- A lot of growing concern about experience
- Larger literature on "panel conditioning"
  - Inconclusive evidence
- Some numbers:
  - MTurk workers are doing 100+ studies per month

# Participant Experience

- A lot of growing concern about experience
- Larger literature on "panel conditioning"
  - Inconclusive evidence
- Some numbers:
  - MTurk workers are doing 100+ studies per month
  - Numbers are the same for YouGov panelists

# My Advice

- Only work with enumerated populations
  - Each unit is uniquely identifiable

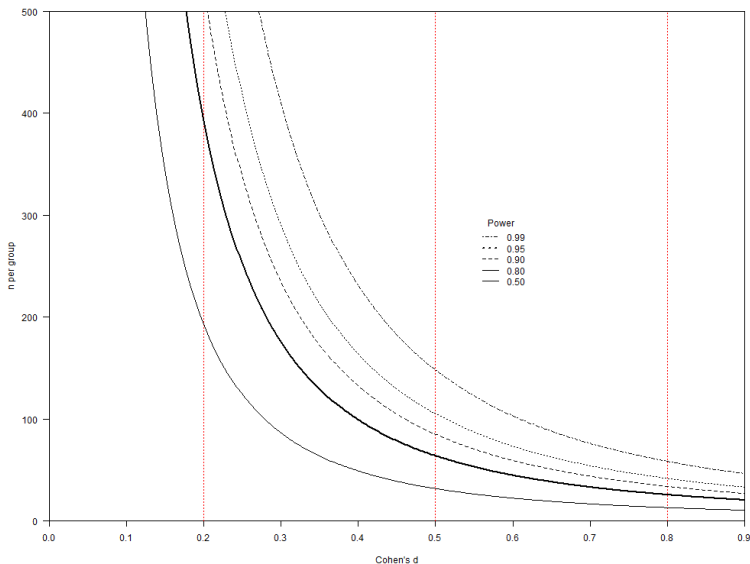
# My Advice

- Only work with enumerated populations
  - Each unit is uniquely identifiable
- Without this, you risk many things:
  - Ambiguous eligibility
  - Retakes, treatment crossover
  - No way to evaluate response rates/bias

# My Advice

- Only work with enumerated populations
  - Each unit is uniquely identifiable
- Without this, you risk many things:
  - Ambiguous eligibility
  - Retakes, treatment crossover
  - No way to evaluate response rates/bias
- Know something about your sample
  - How does it differ from your target of inference?
  - What theories or evidence would suggest those differences should matter?
  - What can you do to adjust or control for those *consequential* differences?

# Don't forget statistical power...







1 "The Gold Standard"

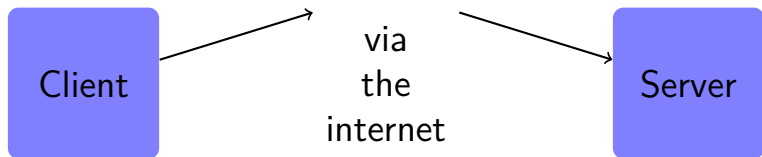
2 Web Questionnaires

3 Recruitment in Practice

4 Challenges and Opportunities

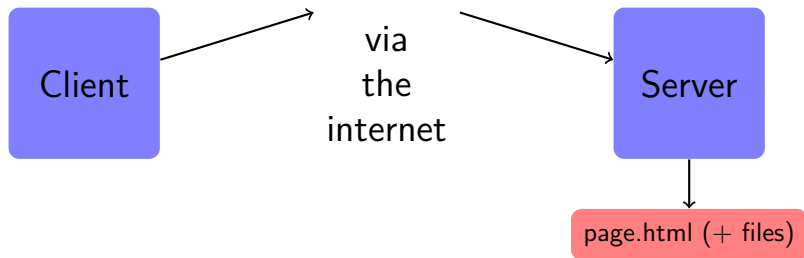
When you navigate to a URL...

# When you navigate to a URL...



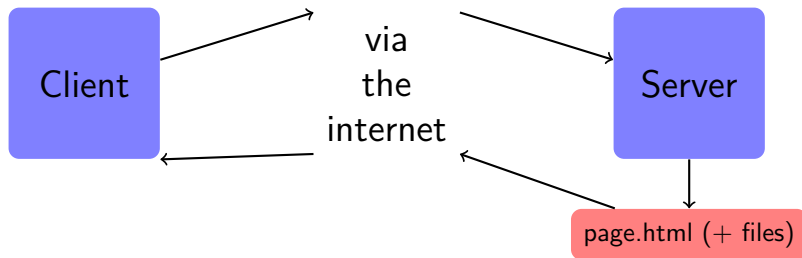
- 1 your browser sends an HTTP request to a server

## When you navigate to a URL...



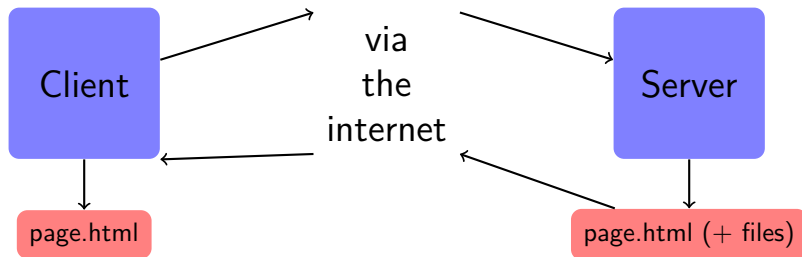
- 1 your browser sends an HTTP request to a server
- 2 the server processes the request and executes server-side code

## When you navigate to a URL...



- 1 your browser sends an HTTP request to a server
- 2 the server processes the request and executes server-side code
- 3 the server replies with the contents of the page

## When you navigate to a URL...



- 1 your browser sends an HTTP request to a server
- 2 the server processes the request and executes server-side code
- 3 the server replies with the contents of the page
- 4 your browser executes client-side

# Questionnaires are client side

A web page consists of four things:

- HTML describing content
- Cascading style sheets (CSS) to style that content
- Images or other multimedia content
- Javascript code that makes a page dynamic

```
<html>
<head>
  <title>Survey</title>
</head>
<body>
  <form action="http://httpbin.org/post" method="POST">
    <p>
      <label for="q1">Name:
        <input type="text" id="q1" name="q1" />
      </label>
    </p>
    <p><input type="submit"></p>
  </form>
</body>
</html>
```



# Getting a grip on HTML

- Every element should have an opening `<tag>` and closing `</tag>`
- Necessary tags:
  - `<html></html>`
  - `<head></head>`
  - `<body></body>`

# Getting a grip on HTML

- Every element should have an opening `<tag>` and closing `</tag>`
- Necessary tags:
  - `<html></html>`
  - `<head></head>`
  - `<body></body>`
- Use an intelligent text editor (*not Notepad*)

# Getting a grip on HTML

- Every element should have an opening `<tag>` and closing `</tag>`
- Necessary tags:
  - `<html></html>`
  - `<head></head>`
  - `<body></body>`
- Use an intelligent text editor (*not Notepad*)
- Use a validator:  
<https://validator.w3.org/nu/>

# Getting a grip on HTML

- Every element should have an opening `<tag>` and closing `</tag>`
- Necessary tags:
  - `<html></html>`
  - `<head></head>`
  - `<body></body>`
- Use an intelligent text editor (*not Notepad*)
- Use a validator:  
<https://validator.w3.org/nu/>
- Remember that browsers differ

# Test Yourself!

Make a simple complete HTML document that passes displays a paragraph of text and passes the validator

`https://validator.w3.org/nu/`

# Getting a grip on HTML

- Common elements

- `<div></div>`
- `<span></span>`
- `<p></p>`
- `<h1>`, `<h2>`, ...
- `<br />`

- Tag attributes describe each element:

- `id`: unique identifier for each element
- `class`: grouping identifier for elements (useful for CSS)
- `style`: in-line CSS styling

# Getting a grip on HTML

- Common form elements

- `<form></form>`
- `<input />`
- `<label></label>`
- `<button />`

# Getting a grip on HTML

- Common form elements
  - `<form></form>`
  - `<input />`
  - `<label></label>`
  - `<button />`
- Attributes specific to form elements
  - type: the kind of input
    - "radio"
    - "checkbox"
    - "text"
  - name: the "variable" being recorded
  - value: the default variable value



# Test Yourself!

Make a simple HTML form that displays a question and a free response answer and passes the validator

`https://validator.w3.org/nu/`

# Test Yourself!

Make a simple HTML form that displays a question and a multiple choice answer and passes the validator

`https://validator.w3.org/nu/`

## Some other elements:

- Bullet list: `<ul></ul>`
- Enumerated list: `<ol></ol>`
  - List element: `<li></li>`
- Tables:
  - Table: `<table></table>`
  - Header: `<th></th>`
  - Row: `<tr></tr>`
  - Cell: `<td></td>`

# Getting a grip on HTML

HTML files can also contain other content

- Style sheets (CSS) in `<style></style>` elements in head
- Javascript in `<script></script>` elements in head and/or body
- Images (``)
- HTML5 features (e.g., `<canvas>`, `<svg>`)

# CSS is Elegant

- HTML originally (until 1996) had to be styled manually



## White House Virtual Library

### You can search the White House web site for:

- [All White House web features combined](#): Press releases, Radio Addresses, photographs and Web Pages.
  - [White House documents](#): Publicly-released documents since the start of the Clinton Administration.
  - [The contents of this web site](#): Just the pages of this service.
  - [Radio Addresses of the President](#): Search and listen to the President's Saturday Radio Addresses.
  - [Executive Orders](#): Official actions, procedural changes, and organizational changes.
  - [White House photographs](#): Search a public collection of photographs.
- You can also search the GovBot database [all government sites](#)

### You can also browse some historic national documents:

- [The Declaration of Independence](#)
- [The Constitution of the United States of America](#)

If you wish to receive White House publications on a daily basis, you can [subscribe to the publications mailing list](#)

Presidential addresses and the White House press releases may be found in the [Briefing Room](#).



# CSS is Elegant

- HTML originally (until 1996) had to be styled manually

# CSS is Elegant

- HTML originally (until 1996) had to be styled manually
- CSS allows you to style a document separately from its content
  - Class- and element-specific styling
  - Change styling without changing the HTML



# CSS is Elegant

- HTML originally (until 1996) had to be styled manually
- CSS allows you to style a document separately from its content
  - Class- and element-specific styling
  - Change styling without changing the HTML
- Can be included inline, in `<head>`, or in separate document

# Test Yourself!

Make a simple HTML form uses CSS to style the answer options to your past survey and passes the validator

<https://validator.w3.org/nu/>

```
<html>
<head>
  <title>Redirect</title>
</head>
<body>
  <script>
    var u = new Array ();
    u[0] = "http://www.google.com";
    u[1] = "http://www.bing.com";
    u[2] = "http://www.yahoo.com";
    var i = Math.floor(u.length*Math.random());
    document.write("Redirecting to " + u[i]);
    window.location.replace(u[i]);
  </script>
</body>
</html>
```

```
<html>
<head>
  <title>Redirect</title>
</head>
<body>
  <p>Please read the following:</p>
  <script>
    var u = new Array ();
    u[0] = "Treatment 1";
    u[1] = "Treatment 2";
    u[2] = "Treatment 3";
    var i = Math.floor(u.length*Math.random());
    document.write("<p><b>" + u[i] + "</b></p>");
  </script>
</body>
</html>
```

# Test Yourself!

Make a simple HTML form that displays a randomly selected piece of text and passes the validator

`https://validator.w3.org/nu/`

```
<html>
<head>
  <title>Survey</title>
</head>
<body>
  <div id="player"></div>
  <script>
    var tag = document.createElement('script');
    tag.src = "https://www.youtube.com/iframe_api";
    var firstScriptTag = document.getElementsByTagName('script')[0];
    firstScriptTag.parentNode.insertBefore(tag, firstScriptTag);

    var player;
    function onYouTubeIframeAPIReady() {
      player = new YT.Player('player', {
        height: '390',
        width: '640',
        videoId: 'OBmhjf0rKe8',
        playerVars: {
          'controls': '0',
          'showinfo': '0',
          'rel': '0'
        },
        events: {
          'onReady': onPlayerReady
        }
      });
    }

    function onPlayerReady(event) {
      event.target.playVideo();
    }
  </script>
</body>
</html>
```

If client-side is so cool. . .

. . . why do we care about server-side  
technology?

## Client-Side

## Server-Side



## Client-Side

- HTML (markup)
- CSS (styling)
- Javascript (scripting)

## Server-Side

## Client-Side

- HTML (markup)
- CSS (styling)
- Javascript (scripting)

## Server-Side

- Python, PHP, ...
- Cookies
- Databases

## Client-Side

- HTML (markup)
- CSS (styling)
- Javascript (scripting)

## Server-Side

- Python, PHP, ...
- Cookies
- Databases

We need a server to record a participant's behavior

# Web Questionnaires

- Google Spreadsheet Forms:  
<https://www.google.co.uk/forms/about/>
- Survey Monkey:  
<https://www.surveymonkey.com/home/>
- Qualtrics:  
<https://www.qualtrics.com/login/>

# Web Questionnaires

- Google Spreadsheet Forms:  
<https://www.google.co.uk/forms/about/>
- Features
  - Free!
  - Somewhat complex branching
  - No randomization

# Google Consumer Surveys

- `http://www.google.com/insights/consumersurveys/home`
- Features
  - Cheap and fast
  - Very limited functionality
  - One-off questionnaires
  - Great for pilot testing

# Survey Monkey

- `https://www.surveymonkey.com/home/`
- Features
  - Free account
  - Limited surveys and respondents
  - No randomization in free account
  - Nice respondent management tools ("collectors")

# Test Yourself!

Create a simple survey and create a panel including yourself and maybe me (thosjleeper@gmail.com) as recipients. Try sending the survey.



# Qualtrics

- <https://www.qualtrics.com/login/>
- Features
  - Free account w/ limited surveys and respondents
  - Much more expensive than SurveyMonkey
  - Powerful randomization functionality
  - Useful “embedded data” controls
  - Optimized for mobile

# Test Yourself!

Create two kinds of randomization:

- 1 Using a random embedded data field
- 2 Using block randomization

Preview the survey to make sure it works.

# Connecting Surveys to MTurk

```
<html>
<head></head>
<body>
  <script>
    function turkGetParam( name ) {
      var regexS = "[\?&]" + name + "=( [^&#]*)";
      var regex = new RegExp( regexS );
      var tmpURL = window.location;
      var results = regex.exec( tmpURL );
      if( results == null ) {
        return "";
      } else {
        return results[1];
      }
    }
    var assign = turkGetParam('assignmentId');
    var worker = turkGetParam('workerId');
    var surveylink = new String("http://httpbin.org/get?" + "assignmentId=" + assign + "&workerId=" + worker);
    if(assign=="ASSIGNMENT_ID_NOT_AVAILABLE") {
      /* DO NOTHING */
    }
    else {
      document.write("<p>Visit <a href='" + surveylink + "' target='_blank'>this link</a></p>");
    }
  </script>
  <form action="http://httpbin.org/post">
    <p><label for="code">Code: <input type="text" id="code" name="code" /></label></p>
    <p><input type="submit"></p>
  </form>
</body>
</html>
```

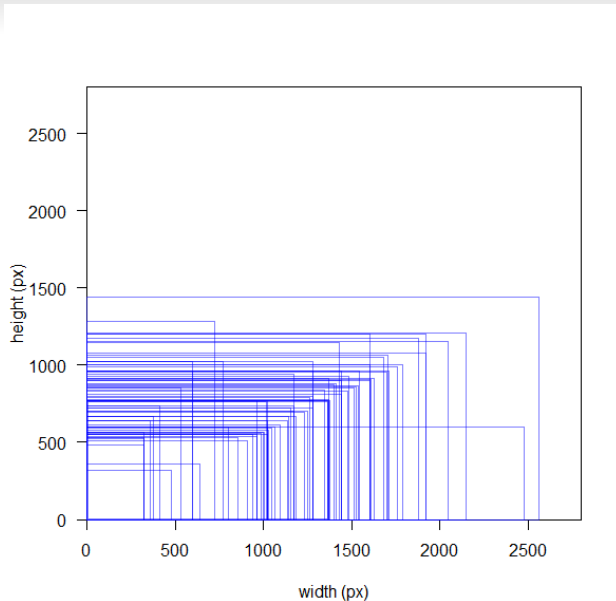
# Test Yourself!

Setup an embedded data field in Qualtrics and then use a webform or simple hyperlink to redirect someone to your survey using that embedded data field.

Try out the form (or link) and see how it is registered in your Qualtrics data.

# Design Considerations

- Qualtrics highlights challenge of modern devices



# Design Considerations

- Qualtrics highlights challenge of modern devices

# Design Considerations

- Qualtrics highlights challenge of modern devices
- But it's not just device size
  - Different browsers
  - Readability
  - Images/video



# Design Considerations

- Qualtrics highlights challenge of modern devices
- But it's not just device size
  - Different browsers
  - Readability
  - Images/video
- Web represents a general loss of control

# Design Considerations

- Qualtrics highlights challenge of modern devices
- But it's not just device size
  - Different browsers
  - Readability
  - Images/video
- Web represents a general loss of control
- So, key is know your sample before you use it

# Two flavors of pretesting

## 1 Technical pretesting

- Make sure your instrument works
- Across browsers/platforms/devices

## 2 Substantive pretesting

- Does your instrument make sense
- Does it make sense for your participants

1 "The Gold Standard"

2 Web Questionnaires

3 Recruitment in Practice

4 Challenges and Opportunities

# For Multi-Person Games

- Simultaneous participation can be challenging
- Best workflow is lab-like:
  - 1 Recruit participants
  - 2 Schedule them for time slots
  - 3 Monitor to ensure participants show up
  - 4 Pay a show-up fee
- So, think of the following as relevant the first two steps in that process
- We'll talk more about running multi-person sessions later

# Professional Panels

- Big players: SSI, YouGov, GfK, TNS/Gallup
- Online panels of respondents
- Respondents participate for incentives
- Study costs are negotiated
  - Sample size
  - Study length (number of survey items)
  - Targetting
  - Timing

# Considerations

- Recruitment
  - Sampling
  - Opt-in
  - A mix of each
- Incentives
- Frequency of participation
- "Profile" variables
- Quotas, post-stratification, weighting
- Respondent "quality"

# Knowledge Networks versus YouGov

- Big debate in early 2000s about online panels
  - KN used ABS to build a representative panel
  - YouGov created an opt-in panel; used "sample matching"



# Knowledge Networks versus YouGov

- Big debate in early 2000s about online panels
  - KN used ABS to build a representative panel
  - YouGov created an opt-in panel; used "sample matching"
- YouGov's process:
  - Randomly sample from a list
  - Match each sampled individual to someone in their opt-in panel
  - Survey the matched individuals

# Knowledge Networks versus YouGov

- Big debate in early 2000s about online panels
  - KN used ABS to build a representative panel
  - YouGov created an opt-in panel; used "sample matching"
- YouGov's process:
  - Randomly sample from a list
  - Match each sampled individual to someone in their opt-in panel
  - Survey the matched individuals
- Evidence inconclusive but many think KN approach is better

# Opt-in Crowdsourcing Sites

- Not exactly a panel (fully opt-in)
- Incentivized participation

# Opt-in Crowdsourcing Sites

- Not exactly a panel (fully opt-in)
- Incentivized participation
- Prominent examples
  - MTurk
  - Crowdfunder
  - Microworkers
  - Prolific Academic

# Other Recruitment Methods

- Online advertising
- Webforums
- Email lists (students, staff, etc.)

# Randomization

- Two flavors of randomness
- Pseudo-random
  - Not actually random
  - Reproducible
  - Implemented everywhere
  - Excel: =RANDBETWEEN(1,3)
  - R: `sample(1:3, 100, TRUE)`
- Truly random
  - Not reproducible
  - <http://www.random.org/>



# Open Science Considerations

- Regardless of how you run studies, try to make them *reproducible*
- What does this mean?
- Why do we care?



# Reproducibility

# Reproducibility

- Everything you do for your study should be publicly shared after publication
  - Dataverse
  - Open Science Framework
  - figshare

# Reproducibility

- Everything you do for your study should be publicly shared after publication
  - Dataverse
  - Open Science Framework
  - figshare
- This helps others build on your work

# Reproducibility

- Everything you do for your study should be publicly shared after publication
  - Dataverse
  - Open Science Framework
  - figshare
- This helps others build on your work
- Makes it easier for you to build on your work

# Reproducibility

- Everything you do for your study should be publicly shared after publication
  - Dataverse
  - Open Science Framework
  - figshare
- This helps others build on your work
- Makes it easier for you to build on your work
- Makes you a more careful researcher

# Some Examples

- Leeper, Thomas J. 2014. "The Informational Basis for Mass Polarization." *Public Opinion Quarterly* 78(1): 27–46.
  - On Dataverse:  
<http://hdl.handle.net/1902.1/21964>
- Mullinix, Kevin J., Leeper, Thomas J., Druckman, James N., and Freese, Jeremy. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science*: In press.
  - On Dataverse:  
<http://dx.doi.org/10.7910/DVN/MUJHGR>

# What should be shared?

- Recruitment protocol and materials
- Complete questionnaire (plain text)
- Web forms/markup
- Data (raw, but anonymized)
- Codebook
- Data Preparation Code
- Analysis Code
- Manuscript pre-print
- Preanalysis plan (if applicable)
- README

# Some Reproducibility Tips



# Some Reproducibility Tips

- 1 Be selfish

# Some Reproducibility Tips

- 1 Be selfish: Be reproducible for yourself first; benefits for science are a positive externality

# Some Reproducibility Tips

- 1 Be selfish: Be reproducible for yourself first; benefits for science are a positive externality
- 2 Start early

# Some Reproducibility Tips

- 1 Be selfish: Be reproducible for yourself first; benefits for science are a positive externality
- 2 Start early: Develop a reproducible workflow from day 1

# Some Reproducibility Tips

- 1 Be selfish: Be reproducible for yourself first; benefits for science are a positive externality
- 2 Start early: Develop a reproducible workflow from day 1
- 3 Save everything

# Some Reproducibility Tips

- 1 Be selfish: Be reproducible for yourself first; benefits for science are a positive externality
- 2 Start early: Develop a reproducible workflow from day 1
- 3 Save everything: Archive frequently so you never lose your work



# Developing Custom Apps

- Apps provide much more data than HTML forms
  - Geolocation
  - Accelerometer (and other sensors)
  - User account data (with permissions)



# Developing Custom Apps

- Apps provide much more data than HTML forms
  - Geolocation
  - Accelerometer (and other sensors)
  - User account data (with permissions)
- MIT AppInventor:  
<http://appinventor.mit.edu/>
- Trinity edX course (link)



1 "The Gold Standard"

2 Web Questionnaires

3 Recruitment in Practice

4 Challenges and Opportunities



# Reweighting

- It may be possible to *reweight* convenience sample data to match a population
- Any method for this is “model-based” (rather than “design-based”)
- Not widely used or evaluated (yet)
- All techniques build on the idea of stratification

# Overview of Stratification

- 1 Define population
- 2 Construct a sampling frame
- 3 Identify variables we already know about units in the sampling frame
- 4 Stratify sampling frame based on these characteristics
- 5 Collect an SRS (of some size) within each stratum
- 6 Aggregate our results

# Estimates from a stratified sample

- Within-strata estimates are calculated just like an SRS
- Within-strata variances are calculated just like an SRS
- Sample-level estimates are weighted averages of stratum-specific estimates
- Sample-level variances are weighted averages of stratum-specific variances

# Post-Stratification

- Used to correct for nonresponse, coverage errors, and sampling errors



# Post-Stratification

- Used to correct for nonresponse, coverage errors, and sampling errors
- Reweight sample data to match population distributions
  - Divide sample and population into strata
  - Weight units in each stratum so that the weighted sample stratum contains the same proportion of units as the population stratum does

# Post-Stratification

- Used to correct for nonresponse, coverage errors, and sampling errors
- Reweight sample data to match population distributions
  - Divide sample and population into strata
  - Weight units in each stratum so that the weighted sample stratum contains the same proportion of units as the population stratum does
- There are numerous other related techniques

# Post-Stratification: Example

- Imagine our sample ends up skewed on immigration status and gender relative to the population

Group	Pop.	Sample	Rep.	Weight
Native-born, Female	.45	.5		
Native-born, Male	.45	.4		
Immigrant, Female	.05	.07		
Immigrant, Male	.05	.03		

- PS weight is just  $w_{ps} = N_I/n_I$

# Post-Stratification: Example

- Imagine our sample ends up skewed on immigration status and gender relative to the population

Group	Pop.	Sample	Rep.	Weight
Native-born, Female	.45	.5	Over	
Native-born, Male	.45	.4	Under	
Immigrant, Female	.05	.07	Over	
Immigrant, Male	.05	.03	Under	

- PS weight is just  $w_{ps} = N_I/n_I$

# Post-Stratification: Example

- Imagine our sample ends up skewed on immigration status and gender relative to the population

Group	Pop.	Sample	Rep.	Weight
Native-born, Female	.45	.5	Over	0.900
Native-born, Male	.45	.4	Under	
Immigrant, Female	.05	.07	Over	
Immigrant, Male	.05	.03	Under	

- PS weight is just  $w_{ps} = N_I/n_I$

# Post-Stratification: Example

- Imagine our sample ends up skewed on immigration status and gender relative to the population

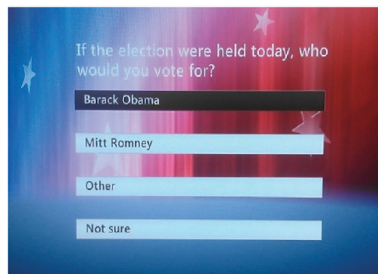
Group	Pop.	Sample	Rep.	Weight
Native-born, Female	.45	.5	Over	0.900
Native-born, Male	.45	.4	Under	1.125
Immigrant, Female	.05	.07	Over	0.714
Immigrant, Male	.05	.03	Under	1.667

- PS weight is just  $w_{ps} = N_I/n_I$

# Post-Stratification

- This is the basis for inference in non-probability samples
  - *Demographic* representativeness
- Online panels will reweight sample based on age, sex, education, etc.
- Purely design-based surveys are increasingly rare

# The Xbox Study



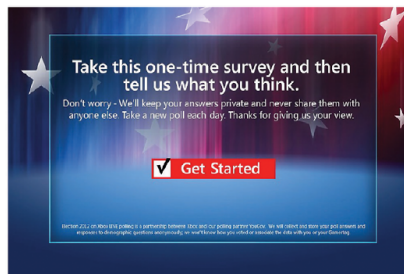
If the election were held today, who would you vote for?

☒ Barack Obama

☐ Mitt Romney

☐ Other

☐ Not sure



**Take this one-time survey and then tell us what you think.**

Don't worry - We'll keep your answers private and never share them with anyone else. Take a new poll each day. Thanks for giving us your view.

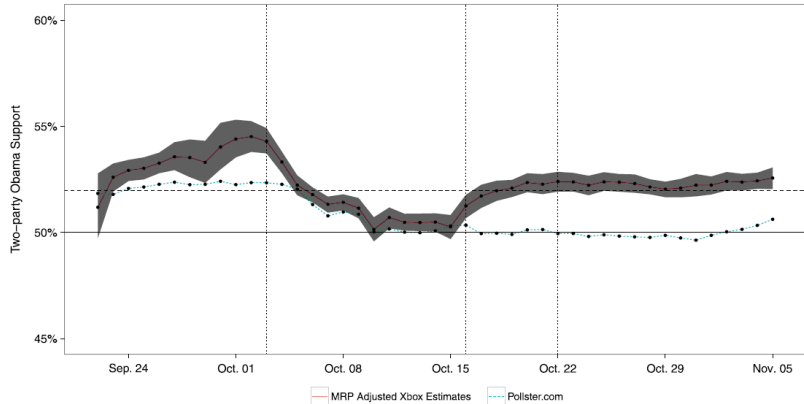
☒ **Get Started**

Xbox Live on Xbox 360 polling is a partnership between Xbox and our polling partner YouGov. We will collect and store your poll answers and responses to demographic questions anonymously, so we'll know how you voted or answered the data with you or your Xboxing.

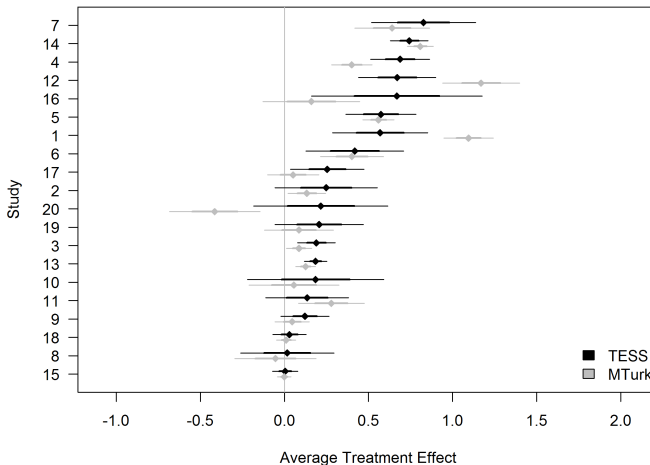
Wang et al. 2015. "Forecasting elections with non-representative polls."  
*International Journal of Forecasting*.



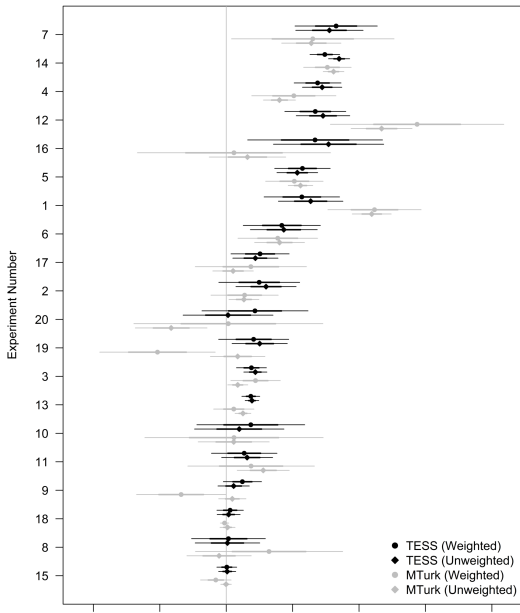
# The Xbox Study



Wang et al. 2015. "Forecasting elections with non-representative polls."  
*International Journal of Forecasting*.



Mullinix et al. In press. "The Generalizability of Survey Experiments."  
*Journal of Experimental Political Science*.



# Propensity Score Approach

- 1 Define a target population to which sample inference is intended to generalize
- 2 Estimate a propensity score model
  - Pool experimental samples and target population units
  - Predict membership of all target and sample units in the experimental sample
- 3 Using fitted logits, divide the population and sample into strata
  - Number of strata is commonly 5 (Cochran, 1968)
- 4 Estimate stratum-specific ATE
- 5 Calculate weighted average of stratum-level estimates

# Propensity Score Approach

Target population average treatment effect:

$$\sum_{v=1}^5 p(v) T(v) \quad (14)$$

where  $p(v)$  is the proportion of the target population in a given stratum,  $v$ , and  $T(v)$  is the estimated effect from stratum  $v$  of the experimental sample

# Propensity Score Approach

Effect variance:

$$\sum_{v=1}^5 p(v)^2 V(v), \quad (15)$$

where  $V(v)$  is the variance of the estimated experimental sample effect for stratum  $v$

# Propensity Score Subclassification Estimator

Stratum	Weights		Estimates			
	Nat'l	Sample	Loan	DREAM (Pro)	DREAM (Con)	Rally (All)
1	0.20	0.83	0.94 (0.08)	0.06 (0.11)	-0.22 (0.12)	0.74 (0.10)
2	0.20	0.11	0.99 (0.26)	0.22 (0.37)	-0.28 (0.36)	0.77 (0.29)
3	0.20	0.04	1.28 (0.43)	-0.61 (0.58)	-1.76 (0.54)	1.00 (0.45)
4	0.20	0.01	1.99 (0.73)	0.29 (1.12)	0.56 (0.89)	1.44 (0.79)
5	0.20	0.00				
Sample	-	-	1.04 (0.30)	-0.01 (0.44)	-0.34 (0.38)	0.79 (0.33)
Nat'l	-	-	1.14 (0.18)	0.02 (0.22)	-0.94 (0.23)	0.94 (0.19)

**So does reweighting solve everything forever?**



## So does reweighting solve everything forever?

- Need well-defined target population
  - and detailed covariate data
  - and large stratum sizes

## So does reweighting solve everything forever?

- Need well-defined target population
  - and detailed covariate data
  - and large stratum sizes
- Purely model-based, so only as good as the model
  - What unobservables might be hiding bias?
  - What reweighting might worsen bias?

## So does reweighting solve everything forever?

- Need well-defined target population
  - and detailed covariate data
  - and large stratum sizes
- Purely model-based, so only as good as the model
  - What unobservables might be hiding bias?
  - What reweighting might worsen bias?
- Non-coverage is a potentially huge problem

## So does reweighting solve everything forever?

- Need well-defined target population
  - and detailed covariate data
  - and large stratum sizes
- Purely model-based, so only as good as the model
  - What unobservables might be hiding bias?
  - What reweighting might worsen bias?
- Non-coverage is a potentially huge problem
- Not well-tested on experimental data

# Mode Effects and Comparisons

- Behavioral research is historically lab-based

# Mode Effects and Comparisons

- Behavioral research is historically lab-based
- Online mode is different in many ways aside from *mode*
  - Self-paced
  - Anonymous
  - Private
  - Computer-based
  - General loss of experimental control

# Mode Effects and Comparisons

- Behavioral research is historically lab-based
- Online mode is different in many ways aside from *mode*
  - Self-paced
  - Anonymous
  - Private
  - Computer-based
  - General loss of experimental control
- Two big consequences
  - Attrition
  - Lower attention

# Attrition

- We care about two issues:
  - Who leaves a study early
  - When they leave a study



# Attrition

- We care about two issues:
  - Who leaves a study early
  - When they leave a study
- We care about representativeness (not just demographically)

# Attrition

- We care about two issues:
  - Who leaves a study early
  - When they leave a study
- We care about representativeness (not just demographically)
- Analyze when participants leave study to identify difficult, confusing, or problematic study elements
  - Ideally, do pilot tests

# Custom Panels

- Creating your own panel is great
  - Carefully sample on specific characteristics
  - Organize repeated interviewing or interaction
- Lots of additional issues
  - Attrition
  - Compensation
  - Panel Conditioning
- See Callegaro et al. 2014. *Online Panel Research: A Data Quality Perspective*. Wiley.

# Attention Checking

- Online mode invites satisficing
- Attention checking can help, but is imperfect

# Apparent Satisficing

- Filter out respondents based on response behavior
- Some common measures:
  - "Straightlining"
  - Non-differentiation
  - Acquiescence
  - Nonresponse
  - DK responding
  - Speeding
- Difficult to detect
- Difficult to distinguish from "real" responses

# Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding

# Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding
- Mousetracking or eyetracking
  - Mousetracking is unobtrusive
  - Eyetracking requires participants opt-in
  - Voice input or interactive (video CASI)

# Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding
- Mousetracking or eyetracking
  - Mousetracking is unobtrusive
  - Eyetracking requires participants opt-in
  - Voice input or interactive (video CASI)
- Record focus/blur browser events



# Direct Measures

- How closely have you been paying attention to what the questions on this survey actually mean?

# Direct Measures

- How closely have you been paying attention to what the questions on this survey actually mean?
  
- While taking this survey, did you engage in any of the following behaviors? Please check all that apply.
  - Use your mobile phone
  - Browse the internet
  - ...

# Substantive Manipulation Check

- Two common approaches:
  - Information recall or understanding
  - Measure level of manipulated treatment variable
- Risky to remove cases based on this because it is a form of conditioning on post-treatment variables
- May be useful to consider either a mediator of effects

# Instructional Manipulation Check

We would like to know if you are reading the questions on this survey. If you are reading carefully, please ignore this question, do not select any answer below, and click "next" to proceed with the survey.

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

# Instructional Manipulation Check

Do you agree or disagree with the decision to send British forces to fight ISIL in Syria? We would like to know if you are reading the questions on this survey. If you are reading carefully, please ignore this question, do not select any answer below, and click "next" to proceed with the survey.

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

# Attention Checking

In summary. . .

- Attention checking can be useful
- Lots of options
- No obvious best metric
- Can be analytically consequential



# To Sum Up. . .

- Nationally representative samples are a hypothetical gold standard for behavioral research
- We can get a lot of leverage from non-representative samples
- Online context also enables innovative designs
- Wide array of tools available to implement experiments and recruit participants



# Thanks!

I will be around for questions.

But don't hesitate to be in touch later on:

- Website: <http://www.thomasleeper.com>
- Email: [thosjleeper@gmail.com](mailto:thosjleeper@gmail.com)
- Twitter: @thosjleeper
- GitHub: @leeper



# Experimental Methods

- Druckman et al. 2011. *Cambridge Handbook of Experimental Political Science*. Cambridge.
- Gerber and Green. 2011. *Field Experiments*. W.W. Norton.
- Mutz. 2011. *Population-Based Survey Experiments*. Princeton.
- Shadish et al. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

# Survey Methods

- Groves et al. 2008. *Survey Methodology*. 2nd Edition. Wiley.
- Lohr. 2010. *Sampling: Design and Analysis*. 2nd Edition. Cengage.

# Online Surveys

- Callegaro et al. 2015. *Web Survey Methodology*. Sage.
- Callegaro et al. 2015. *Online Panel Research: A Data Quality Perspective*. Wiley.