Introduction to group activity

Matthew J. Salganik
Department of Sociology
Princeton University

Summer Institute in Computational Social Science
June 21, 2018

The Summer Institute in Computational Social Science is supported by grants from the Russell Sage Foundation and the Alfred P. Sloan Foundation.

Online, Opt-in Surveys: Fast and Cheap, but are they Accurate?

Sharad Goel
Stanford University
scgoel@stanford.edu

Adam Obeng
Columbia University
adam.obeng@columbia.edu

David Rothschild
Microsoft Research
davidmr@microsoft.com

https://5harad.com/papers/dirtysurveys.pdf

ABSTRACT

It is increasingly common for government and industry organizations to conduct online, opt-in surveys, in part because they are typically fast, inexpensive, and convenient. Online polls, however, attract a non-representative set of respondents, and so it is unclear whether results from such surveys generalize to the broader population. These non-representative surveys stand in contrast to probability-based sampling methods, such as random-digit dialing (RDD) of phones, which are a staple of traditional survey research. Here we investigate the accuracy of non-representative data by administering an online, fully opt-in poll of social and political attitudes. Our survey consisted of 49 multiple-choice attitudinal questions drawn from the probability-based, in-person 2012 General Social Survey (GSS) and select RDD phone surveys by the Pew Research Center. To correct for the inherent biases of non-representative data, we statistically adjust estimates via model-based poststratification, a classic statistical tool but one that is only infrequently used for bias correction. Our online survey took less than one-twentieth the time and money of traditional RDD polling, and less than one-hundredth the time and money of GSS polling. After statistical correction, we find the median absolute difference between the non-probability-based online survey and the probability-based GSS and Pew studies is 7 percentage points. This difference is considerably larger than if the surveys were all perfect simple random samples drawn from the same population; the gap, however, is comparable to that between the GSS and Pew estimates themselves. Our results suggest that with proper statistical adjustment, online, non-representative surveys are a valuable tool for practitioners in varied domains.

https://5harad.com/papers/dirtysurveys.pdf

Activity:

- Design a questionaire using questions already asked on high quality surveys

Activity:

- Design a questionaire using questions already asked on high quality surveys
- Recruit participants from Amazon Mechanical Turk and have them complete your questionnaire

Activity:

- Design a questionaire using questions already asked on high quality surveys
- Recruit participants from Amazon Mechanical Turk and have them complete your questionnaire
- Compare results from your survey to the results from the high-quality survey

Activity:

- Design a questionaire using questions already asked on high quality surveys
- Recruit participants from Amazon Mechanical Turk and have them complete your questionnaire
- Compare results from your survey to the results from the high-quality survey
- Try different approaches to weighting and see how the change the estimates

Activity:

- Design a questionaire using questions already asked on high quality surveys
- Recruit participants from Amazon Mechanical Turk and have them complete your questionnaire
- Compare results from your survey to the results from the high-quality survey
- Try different approaches to weighting and see how the change the estimates
- De-identify and open-source data

This activity will give you practice:

- Collecting survey data

This activity will give you practice:

- Collecting survey data
- Analyzing survey data (data wrangling and post-stratification)

This activity will give you practice:

- Collecting survey data
- Analyzing survey data (data wrangling and post-stratification)
- Working with Amazon Mechanical Turk

This activity will give you practice:

- Collecting survey data
- Analyzing survey data (data wrangling and post-stratification)
- Working with Amazon Mechanical Turk
- Archiving data for other researchers

Remember: This is a learning activity so try whatever you want.

Our recommended work flow:

- Create survey on Google Forms
- Deploy to MTurk
- Take a break

Allison Morgan
@alliecmorgan

Following

Just wrapped up the first week of #SICSS2017! On Thursday, we got 50+ online survey responses, all while frolicking in a fountain.

3:24 PM - 24 Jun 2017

Our recommended work flow:

- Create survey on Google Forms
- Deploy to MTurk
- Take a break
- Validate and pay workers
- Analyze the much larger sample that we have collected for you
- De-identify and open-source the data that you collected