[Survey research in the digital age], [Probability and non-probability sampling], [Computer-administered interviews], [Combining surveys and big data], [Additions and extensions]

Matthew J. Salganik
Department of Sociology
Princeton University

1) Introduction
2) Observing behavior
3) Asking questions
4) Running experiments
5) Mass collaboration
6) Ethics
7) The future


R. MUTT
1917


R.MUTT
1917

readymades

custommades

https://commons.wikimedia.org/wiki/File:Duchamp_Fountaine.jpg
https://commons.wikimedia.org/wiki/File:%27David%27_by_Michelangelo_JBU0001.JPG

A few notes on my teaching:

- Anti-status quo bias

A few notes on my teaching:

- Anti-status quo bias
- Anti-formality bias (formality is important, but just not right now)

A few notes on my teaching:

- Anti-status quo bias
- Anti-formality bias (formality is important, but just not right now)
- Very brief, more information in Ch. 3 of *Bit by Bit*

Why should I care about surveys?

Why should I care about surveys
in the age of big data?

We will always need to ask

- limitations of big data (fubu vs. nufu-nubu)

We will always need to ask

- limitations of big data (fubu vs. nufu-nubu)
- internal states vs. external states

We will always need to ask

- limitations of big data (fubu vs. nufu-nubu)
- internal states vs. external states
- inaccessibility of big data

We will always need to ask

- limitations of big data (fubu vs. nufu-nubu)
- internal states vs. external states
- inaccessibility of big data

But how we are going to ask is going to change

	Sampling	Interviews
1st era	Area probability	Face-to-face

	Sampling	Interviews
1st era	Area probability	Face-to-face
2nd era	Random digital dial probability	Telephone

	Sampling	Interviews
1st era	Area probability	Face-to-face
2nd era	Random digital dial probability	Telephone
3rd era		

	Sampling	Interviews
1st era	Area probability	Face-to-face
2nd era	Random digital dial probability	Telephone
3rd era	Non-probability	Computer-administered

	Sampling	Interviews	Data environment
1st era	Area probability	Face-to-face	Stand-alone
2nd era	Random digital dial probability	Telephone	Stand-alone
3rd era	Non-probability	Computer-administered	Linked

	Sampling	Interviews	Data environment
1st era	Area probability	Face-to-face	Stand-alone
2nd era	Random digital dial probability	Telephone	Stand-alone
3rd era	Non-probability	Computer-administered	Linked

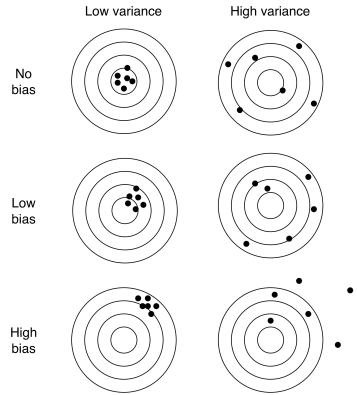
Total survey error framework

Total survey error framework

Insight 1: Errors can come from bias or variance

Total survey error framework

Insight 1: Errors can come from bias or variance

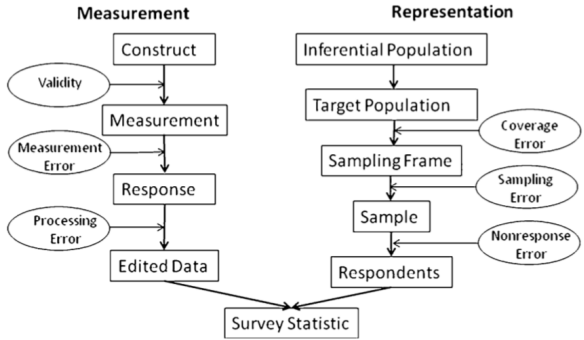


Total survey error framework

Insight 2: Total survey error = measurement error + representation error

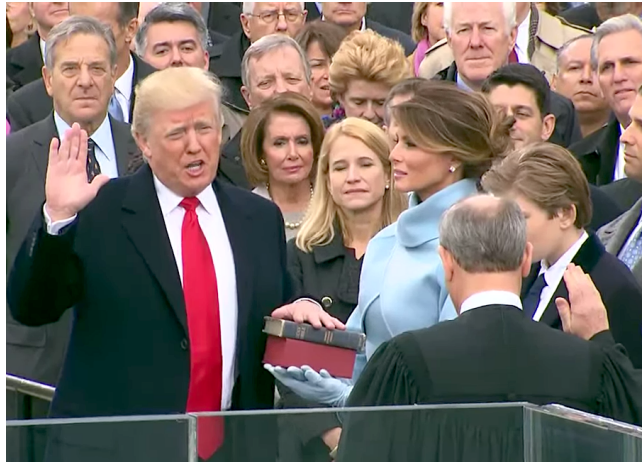
Total survey error framework

Insight 2: Total survey error = measurement error + representation error



Groves and Lyberg 2010, Fig 3

Case study of total survey error framework



https://commons.wikimedia.org/wiki/File:Donald_Trump_taking_his_Oath_of_Office.png

Case study of total survey error framework

An Evaluation of 2016 Election Polls in the U.S.

Ad Hoc Committee on 2016 Election Polling

Courtney Kennedy, Pew Research Center
Mark Blumenthal, SurveyMonkey
Scott Clement, Washington Post
JoshUA d. Clinton, Vanderbilt University
Claire Durand, University of Montreal
Charles Franklin, Marquette University
Kyley McGeeney, Pew Research Center[1]

Lee Miringoff, Marist College
Kristen Olson, University of Nebraska-Lincoln
Doug Rivers, Stanford University, YouGov
Lydia Saad, Gallup
Evans Witt, Princeton Survey Research Associates
Chris Wlezien, University of Texas at Austin

http://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx

Case study of total survey error framework

- National polls were generally correct and accurate by historical standards.

Case study of total survey error framework

- National polls were generally correct and accurate by historical standards.
- State-level polls showed a competitive, uncertain contest . . .

Case study of total survey error framework

- National polls were generally correct and accurate by historical standards.
- State-level polls showed a competitive, uncertain contest . . .
- . . . but clearly under-estimated Trump's support in the Upper Midwest.

Case study of total survey error framework

"There are a number of reasons as to why polls under-estimated support for Trump. The explanations for which we found the most evidence are:"

- "Real change in vote preference during the final week or so of the campaign"

Case study of total survey error framework

"There are a number of reasons as to why polls under-estimated support for Trump. The explanations for which we found the most evidence are:"

- "Real change in vote preference during the final week or so of the campaign"
- "Adjusting for over-representation of college graduates was critical, but many polls did not do it"

Case study of total survey error framework

"There are a number of reasons as to why polls under-estimated support for Trump. The explanations for which we found the most evidence are:"

- "Real change in vote preference during the final week or so of the campaign"
- "Adjusting for over-representation of college graduates was critical, but many polls did not do it"
- "Some Trump voters who participated in pre-election polls did not reveal themselves as Trump voters until after the election, and they outnumbered late-revealing Clinton voters"

Full report: http://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx

Conclusion

Wrapping up:

Conclusion

Wrapping up:

- Total survey error framework 1st key insight: errors can be caused by bias or variance

Conclusion

Wrapping up:

- Total survey error framework 1st key insight: errors can be caused by bias or variance
- Total survey error framework 2nd key insight: errors can be related to representation and measurement

Conclusion

Wrapping up:

- Total survey error framework 1st key insight: errors can be caused by bias or variance
- Total survey error framework 2nd key insight: errors can be related to representation and measurement
- Total survey error framework also helps us think about how digital age can create new opportunities (who to ask and how to ask)

Conclusion

Wrapping up:

- Total survey error framework 1st key insight: errors can be caused by bias or variance
- Total survey error framework 2nd key insight: errors can be related to representation and measurement
- Total survey error framework also helps us think about how digital age can create new opportunities (who to ask and how to ask)
- To learn more: Groves et al (2009)

[Survey research in the digital age], [Probability and non-probability sampling], [Computer-administered interviews], [Combining surveys and big data], [Additions and extensions]

Matthew J. Salganik
Department of Sociology
Princeton University