Survey research in the digital age

Bernhard Clemm von Hohenberg
Department of Computational Social Science
GESIS

Summer Institutes in Computational Social Science
July 28, 2023

Schedule

- ▶ 9.00-9.45 Introduction & total error survey framework
- ▶ 9.45-10.15 Probability and non-probability sampling

- ▶ Coffee break

- ▶ 10.30-11.00 Computer-administered interviewing
- ▶ 11.00-11.30 Linking surveys to big data
- ▶ 11:30-13:00 Intro and begin group exercise

- ▶ Lunch (or Eisbach plunge)

- ▶ 14:00-15:45 Continue group exercise

Credits

These materials build heavily on Matthew Salganik's 2019 SICSS class as well as Chapter 3 of "Bit by Bit: Social Research in the Digital Age".

Background readings

- Salganik, Matthew J. *Bit by bit: Social research in the digital age.* Princeton University Press, 2019. Chapter 3. Link.

- Berinsky, Adam J. "Measuring public opinion with surveys." *Annual Review of Political Science* 20 (2017): 309-329. Link.

- Barabas, Jason and Jennifer Jerit. "Are Nonprobability Surveys Fit for Purpose?" Forthcoming in *Public Opinion Quarterly*. Link.

- Barabas, Jason and Jennifer Jerit. "Are Nonprobability Surveys Fit for Purpose?" Forthcoming in *Public Opinion Quarterly*. Link.

- Goel, Sharad, et al. "Online, Opt-in Surveys: Fast and Cheap, but are they Accurate?". *Working Paper*, 2017. Link.

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

Surveys and big data

Digital age provides tons of data about people, e.g.,

- ▶ Online manifestations of attitudes and behaviors

Surveys and big data

Digital age provides tons of data about people, e.g.,

▶ Online manifestations of attitudes and behaviors
▶ Structure of social networks (literally speaking)

Surveys and big data

Digital age provides tons of data about people, e.g.,

▶ Online manifestations of attitudes and behaviors

▶ Structure of social networks (literally speaking)

▶ Digitized administrative data

But what can I really learn from this?

https://www.ebay.com/sch/i.html?_from=R40&_trksi...	1	2019-04-04 18:21:59	0AcWoYUnHhtzmqfsEfmp
https://www.facebook.com/profile.php?id=1272416...	1	2019-04-05 08:39:53	0AcWoYUnHhtzmqfsEfmp
https://www.swagbucks.com/surveys?ls=7&m=1&s=...	1	2019-04-05 10:32:19	0AcWoYUnHhtzmqfsEfmp
http://www.ilabinsight.com/mrIWeb/mrIWeb.dll?I.Pro...	1	2019-04-05 13:26:47	0AcWoYUnHhtzmqfsEfmp
https://surveys2.sentientdecisionscience.com/LAB/1...	1	2019-04-05 17:16:17	0AcWoYUnHhtzmqfsEfmp
https://www.instagram.com/stories/jameypricephoto/	1	2019-04-05 19:31:34	0AcWoYUnHhtzmqfsEfmp

Limits of big data

We can learn some things from big data—but there are limits:

▶ Inaccuracy, inaccessibility and incompleteness

Limits of big data

We can learn some things from big data—but there are limits:

- ▶ Inaccuracy, inaccessibility and incompleteness
- ▶ Internal states vs. external states

Limits of big data

We can learn some things from big data—but there are limits:

- ▶ Inaccuracy, inaccessibility and incompleteness
- ▶ Internal states vs. external states
    - ▶ Social desirability bias in the online sphere

Limits of big data

We can learn some things from big data—but there are limits:

- ▶ Inaccuracy, inaccessibility and incompleteness
- ▶ Internal states vs. external states
    - ▶ Social desirability bias in the online sphere
    - ▶ Measuring the "subconscious" from digital traces

We will always need to ask—but how we ask will change.

Eras of survey research

	Sampling	Interviews	Data environment
1st era	Area probability	Face-to-face	Stand-alone

Eras of survey research

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

Eras of survey research

	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

TSE framework

Total survey error = representation error + measurement error

TSE framework

Total survey error = representation error + measurement error

Who we ask (representation)

- ▶ sampling error
- ▶ coverage errors
- ▶ non-response error

TSE framework

Total survey error = representation error + measurement error

Who we ask (representation)
- ▶ sampling error
- ▶ coverage errors
- ▶ non-response error

How we ask (measurement)
- ▶ question wording
- ▶ question ordering
- ▶ social desirability bias
- ▶ mode/device effects

TSE framework

Total survey error $=$ representation error $+$ measurement error

Who we ask (representation)

- ▶ sampling error
- ▶ coverage errors
- ▶ non-response error

How we ask (measurement)

- ▶ question wording
- ▶ question ordering
- ▶ social desirability bias
- ▶ mode/device effects

TSE framework

TSE framework
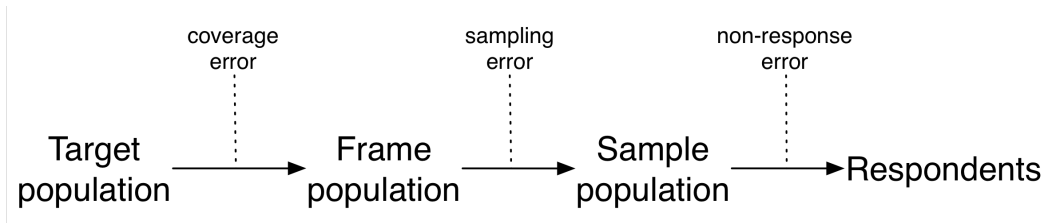
The Literary Digest

NEW YORK   OCTOBER 31, 1936

Topics of the day

LANDON, 1,293,669; ROOSEVELT, 972,897

Final Returns in The Digest's Poll of Ten Million Voters

Well, the great battle of the ballots in the Poll of ten million voters, scattered throughout the forty-eight States of the

lican National Committee purchased THE LITERARY DIGEST?" And all types and varieties, including: "Have the Jews purchased

returned and let the people of the Nation draw their conclusions as to our accuracy. So far, we have been right in every Poll. Will we be right in the current Poll? That, as Mrs. Roosevelt said concerning the President's reelection, is in the 'lap of the gods.'

"We never make any claims before election but we respectfully refer you to the opinion of one of the most quoted citizens

TSE framework

Total survey error = representation error + measurement error

Who we ask (representation)
- ▶ sampling error
- ▶ coverage errors
- ▶ non-response error

How we ask (measurement)
- ▶ question wording
- ▶ question ordering
- ▶ social desirability bias
- ▶ mode/device effects

TSE framework

📄 PDF   ❚❚ Split View   " Cite   🔧 Permissions   ◁ Share ▾

"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"

`https://academic.oup.com/poq/article/82/1/1/4837043`

TSE framework

In sum, total survey error framework...

- ▶ ... helps us organize all the things that can go wrong

TSE framework

In sum, total survey error framework...

- ▶ ... helps us organize all the things that can go wrong
- ▶ ... also helps us think about how digital age can create new opportunities

Questions