

Alternative Selectionmechanisms in Online Samples

Franz Prücklmair*

Ulrich Rendtel†

June 24, 2022

Manuscript under construction

Do not cite

Contents

1	Introduction: Online Samples and their peculiarities	4
2	pros and cons of online samples	5
2.1	pros	5
2.2	cons	5
2.2.1	coverage and sampling error	5
3	Arten von Online Samples und Ihre Entwicklung	5
4	Ansatz und Methodik	6
4.1	Theoretische Fundierung der freiwilligen Teilnahme	6
4.1.1	Theoretische Fundierung der freiwilligen Teilnahme	6
4.2	Neue Selektionsmechanismen	6
5	Simulation	7
5.1	Der ESS	7
5.2	Beschreibung des Versuchsaufbaus	7

*Universität Bamberg

†FB Wirtschaftswissenschaft, Freie Universität Berlin

6	Ergebnisse	7
6.1	SAR-Annahme	7
6.2	Selektivität des Internetzugangs	7
6.3	Vergleich der Performance der Selektionsmechanismen	7
7	Discussion	7

Abstract

Online surveys are among the most widespread data collection methods due to their relatively inexpensive and quick implementation. However, the data obtained from them, so-called non-probability samples, have some problems. These include possible bias due to selectivity, i.e. the systematic exclusion of groups of people from the population. Also survey methods such as river sampling, in which respondents are often recruited uncontrollably by implementing widgets on websites, lead to the fact that the inclusion probability of the respondent required for the weighting is unknown. Thus classical design-based inference as in the Probability Sample is impossible for non-probability samples. The approach of quasi-randomization tries to model the selection mechanism by combining an online sample with a Probability (Reference) Sample and using pseudo-weighting to correct the bias. A frequently modeled selection mechanism to date has been Internet access. However as network expansion progresses, it can be assumed that the exclusion of individuals without Internet access will become increasingly negligible. It may therefore be worthwhile to examine the QR approach under the modeling of alternative selection mechanisms. For analyses, the 9th round of the European Social Survey (ESS) from 2018 serves as the data basis. In the QR approach, this takes on the role as a classic reference sample and at the same time, based on alternative selection mechanisms such as "minimum time spent online" or "willingness to distribute digital content," serves as a source for simulating the online samples. The resulting pseudo-weighted results for the 2017 federal election will be compared using the ESS estimated and actual election results. In addition, the corrections will be compared with classical calibration methods such as raking. First results show that classical internet access is no longer selective. This argues for the use of alternative selection mechanisms to evaluate the QR approach. The alternative models also show that strong biases in the election results can be corrected very well in some cases. At the same time, the approach can also highlight comparative limitations in correcting attitudes and opinions.

Abstract: Keywords: Sampling frame, Access Panel, River Sampling, Non-probability samples

1 Introduction: Online Samples and their peculiarities

There is little doubt that at the beginning of the 21st century, few inventions will have such a strong impact on society as the Internet. This formative influence, which enables the exchange of data between devices through the networking of computers, is primarily due to the continuous increase in the spread of digital devices in private households. In 1998 the percentage of German private households equipped with personal computers was only 38.7 percent, in 2008 it was already 75.4 percent, and in 2018 it will be a full 90.4 percent (Statistisches Bundesamt 2022). A similar development can also be observed in household Internet access, which was just 8.1 percent in 1998, 64.4 percent in 2008, and an incredible 92.7 percent in 2018. In this context, the rapidly increasing number of users of mobile devices, which make location-independent use possible, should be mentioned in particular. It is therefore not surprising that the Internet has become more and more important as a survey mode for a wide range of interests. Market researchers and pollsters use the access of the masses to collect information about products and attitudes, large Internet journals offer real-time opinions on their articles, universities and companies use it to get a picture of the mood of their students and employees, and methodologists investigate the limits and potentials that the variety of technical and methodological features of this medium brings with it (Jackob et al. 2009: S.9). In the scientific context, there are also some peculiarities to consider. For example, although there has been a relative increase in online surveys in top journals, this was initially small in absolute terms. For example, in an analysis of established publications in 2005, less than one Internet-based contribution per journal was measured (Zeprack et al. 2009: p.19). In addition, this number varies greatly across disciplines. In journalism, for example, the proportion of online surveys (approx. 70 percent) was much higher than in sociology and political science (p. 20). According to the authors of the study, this may be due to the different objectives of the different disciplines. While the communication science is mainly investigating small populations (journalists and people in charge), the sociology and political science are mainly interested in the transferability to society. The weaknesses of online surveys in terms of representativeness, which could explain a hesitant behavior in these areas, will be discussed later in the following chapters. Another feature that the authors noticed in the analysis was the decline in methodological approaches. Between 2002 and 2006, they accounted for only 19 percent of the publications surveyed during that period. In general, this can be explained by established methodological research, which has already clarified the most important questions, but also by negligence due to broad acceptance and easy implementation. However, as already described in the first lines, the technology itself, as well

as the spread of the Internet in the last decade was under strong growth. In the case of the former, it is important to check the already established methods for actuality in order to exclude negligence in the case of the latter.

2 pros and cons of online samples

Even though online research is changing over time, the fundamental criteria by which we evaluate survey quality remain unchanged (Couper 2000 p.456).

2.1 pros

2.2 cons

Even if the entire population had access to the Internet, the drawing of samples and the recruitment of test persons would be a potential source of problems (Fass and Schoen 2009 p. 145, Couper 2000 p. 467 , Hauptmanns and Lander 2003)

2.2.1 coverage and sampling error

Coverage error is one of the biggest threats to online sampling in general as long as the sample also aims to provide information about people who may not be able to access it..¹

- (Vorteil) Leicht zu handhaben Zeitgemäß, schnell und bequem
- (Nachteil) Selektion Bias – fehlender Zugang, fehlendes Technisches Verständnis Datenschutz?
- Technologie liegt starken Schwankungen Modernisierung verbreitete Akzeptanz, Generation Digital Natives.

3 Arten von Online Samples und Ihre Entwicklung

Criticisms and glorifications of web surveys, however, should always keep in mind the purpose and context of the online approach. Quality criteria should therefore not be based on a single implementation, nor should all web surveys be treated equally (Couper 2000 p.466). So it is not possible to say that there are simply high quality surveys based on established sampling methods on the one hand and error-prone

¹As already mentioned, there are also studies that relate their statements only to their included group.

fast and cheap implementation on the other hand. So there is wine wide range of qualities in different approaches. Just as the Internet has a wide range of possibilities, so do the different variants. However, there are some main points in which the approaches can be fundamentally divided. It has been shown in the past that these approaches differ greatly in their quality and correction factors (Fass and Schoen 2009).

- Sampling von Email Adressen
- Sampling von Online Panel
- River Sampling Internet via Gadgets

Widgets Mit und ohne Nachverfolgung der Probanden Entscheidend hiebi ist vorallam der Selektionsmechanismus

4 Ansatz und Methodik

4.1 Theoretische Fundierung der freiwilligen Teilnahme

4.1.1 Theoretische Fundierung der freiwilligen Teilnahme

4.2 Neue Selektionsmechanismen

- Internetdauer
- Internetaktivität
- Simulation von Internet mit Interesse
- Simulation auf Basis von Civey

5 Simulation

5.1 Der ESS

5.2 Beschreibung des Versuchsaufbaus

6 Ergebnisse

6.1 SAR-Annahme

6.2 Selektivität des Internetzugangs

6.3 Vergleich der Performance der Selektionsmechanismen

7 Discussion

Netzzugang nicht mehr aussagekräftig Korrektur funktioniert bei etablierten Parteien sehr gut bei kleineren weniger. Messungen von Einstellungen schwierig Gilt für alle Selektionsmechanismen Alternative Selektionsmechanismen Identifizieren

References

- Couper, M.P. (2000): *Web Surveys: A Review of Issues and Approaches*. In: Public Opinion Quarterly, 64, S. 464-494.
- Statistisches Bundesamt (Destatis), (2022): *Ausstattung privater Haushalte mit Informations- und Kommunikationstechnik - Deutschland* <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Ausstattung-Gebrauchsgueter/Tabellen/liste-infotechnik-d.html;jsessionid=C45B0A5AEE64C46398AB4889C9B79006.live731#fussnote-2-115502> (Accessed 21.6.2022)
- Jackob N., Schoen H., and Zebrack T. (2009): Vorwort., In: Jakob N., Schoen H., and Zebrack T. (Hrsg.) *Sozialforschung im Internet, Methodologie und Praxis der Online-Befragung*, 1, 9 - 11 <https://link.springer.com/content/pdf/10.1007/978-3-531-91791-7.pdf>
- Zebrack T., Schoen H., Jakob N., and Schlereth S. (2009): Zehn Jahre Sozialforschung mit dem Internet - eine Analyse zur Nutzung von Online-Befragungen in den Sozialwissenschaften, In: Jakob N., Schoen H., and Zebrack T. (Hrsg.): *Sozialforschung im Internet, Methodologie und Praxis der Online-Befragung*, 1, 15 - 31 <https://link.springer.com/content/pdf/10.1007/978-3-531-91791-7.pdf>
- Faas T., Schoen H. (2009): Fallen Gewichte ins Gewicht? Eine Analyse am Beispiel dreier Umfragen zur Bundestagswahl 2002, In: Jakob N., Schoen H., and Zebrack T. (Hrsg.): *Sozialforschung im Internet, Methodologie und Praxis der Online-Befragung*, 1, 145 - 158 <https://link.springer.com/content/pdf/10.1007/978-3-531-91791-7.pdf>