

Alternative Selectionmechanisms in Online Samples

Franz Prücklmair*

Ulrich Rendtel†

June 21, 2022

Manuscript under construction

Do not cite

Contents

1	Introduction: Online Samples and their peculiarities	3
2	Arten von Online Samples und Ihre Entwicklung	3
3	Ansatz und Methodik	4
3.1	Theoretische Fundierung der freiwilligen Teilnahme	4
3.1.1	Theoretische Fundierung der freiwilligen Teilnahme	4
3.2	Neue Selektionsmechanismen	4
4	Simulation	4
4.1	Der ESS	4
4.2	Beschreibung des Versuchsaufbaus	4
5	Ergebnisse	4
5.1	SAR-Annahme	4
5.2	Selektivität des Internetzugangs	4
5.3	Vergleich der Performance der Selektionsmechanismen	4
6	Discussion	4

*Universität Bamberg

†FB Wirtschaftswissenschaft, Freie Universität Berlin

Abstract

Online-Surveys gehören aufgrund der relativ günstigen und schnellen Umsetzung zu den weitverbreitetsten Datenerhebungsmethoden. Allerdings weisen die daraus gewonnenen Daten, sogenannte Non-probability Samples einige Probleme auf. Dazu gehören mögliche Verzerrungen durch Selektivität, also des systematischen Ausschlusses von Personengruppen aus der Grundgesamtheit. Auch Erhebungsmethoden, wie z.B. River-Sampling, bei dem Befragte oftmals durch Implementation von Widgets auf Webseiten unkontrolliert rekrutiert werden, führen dazu, dass die für die Gewichtung benötigte inclusion probability des Befragten unbekannt ist. Somit ist die klassische design-basierende Inferenz wie im Probability Sample für Non-probability Samples unmöglich. Der Ansatz der Quasi-randomisation versucht durch Kombination eines Online-Samples mit einem Probability (Reference) Sample den Selektionsmechanismus zu modellieren und mithilfe von Pseudogewichtung die Verzerrung zu korrigieren. Ein häufig modellierter Selektionsmechanismus ist bislang der Internetzugang. Mit fortschreitendem Netzausbau ist jedoch davon auszugehen, dass der Ausschluss von Personen ohne Internetzugang immer mehr vernachlässigbar wird. Es könnte sich daher lohnen, den QR-Ansatz unter der Modellierung alternativer Selektionsmechanismen zu untersuchen. Für Analysen dient die 9. Runde des European Social-Survey (ESS) aus dem Jahr 2018 als Datengrundlage. Dieser übernimmt im QR-Ansatz die Rolle als klassisches Reference Sample und dient gleichzeitig, auf Basis alternativer Selektionsmechanismen wie der „Mindestdauer im Internet“ oder der „Bereitschaft, digitale Inhalte zu verbreiten“, als Quelle für die Simulation der Online-Samples. Die daraus resultierenden pseudo-gewichteten Ergebnisse zur Bundestagswahl 2017 sollen anhand der vom ESS geschätzten und dem tatsächlichen Wahlergebnis verglichen werden. Zudem sollen die Korrekturen mit klassischen Kalibrationsmethoden wie dem Raking gegenübergestellt werden. Erste Ergebnisse zeigen, dass der klassische Internetzugang nicht mehr selektiv ist. Dies spricht für die Verwendung alternativer Selektionsmechanismen, um den QR-Ansatz zu evaluieren. Die alternativen Modelle zeigen zudem, dass sich starke Verzerrungen in den Wahlergebnissen teilweise sehr gut korrigieren lassen. Gleichzeitig können durch den Ansatz auch vergleichbare Grenzen bei der Korrektur von Einstellungen und Meinungen deutlich gemacht werden.

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 is becoming 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)

Translated with www.DeepL.com/Translator (free version)

- Mittlerweile gehört Online Erhebungen zu den etablierten Datenerhebungen
- (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.

2 Arten von Online Samples und Ihre Entwicklung

- 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

3 Ansatz und Methodik

3.1 Theoretische Fundierung der freiwilligen Teilnahme

3.1.1 Theoretische Fundierung der freiwilligen Teilnahme

3.2 Neue Selektionsmechanismen

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

4 Simulation

4.1 Der ESS

4.2 Beschreibung des Versuchsaufbaus

5 Ergebnisse

5.1 SAR-Annahme

5.2 Selektivität des Internetzugangs

5.3 Vergleich der Performance der Selektionsmechanismen

6 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

Statistisches Bundesamt (Destatis), 2022 *Ausstattung privater Haushalte mit Informations- und Kommunikationstechnik - Deutschland* <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum/Ausstattung-Gebrauchsgueter/Tabellen/liste-infotechnik-d.html;jsessionid=C45B0A5AEE64C46398A1live731#fussnote-2-115502> (Accessed 21.6.2022)

Jackob N. , Schoen H. , and Zebrack T. (2009): Vorwort. *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>