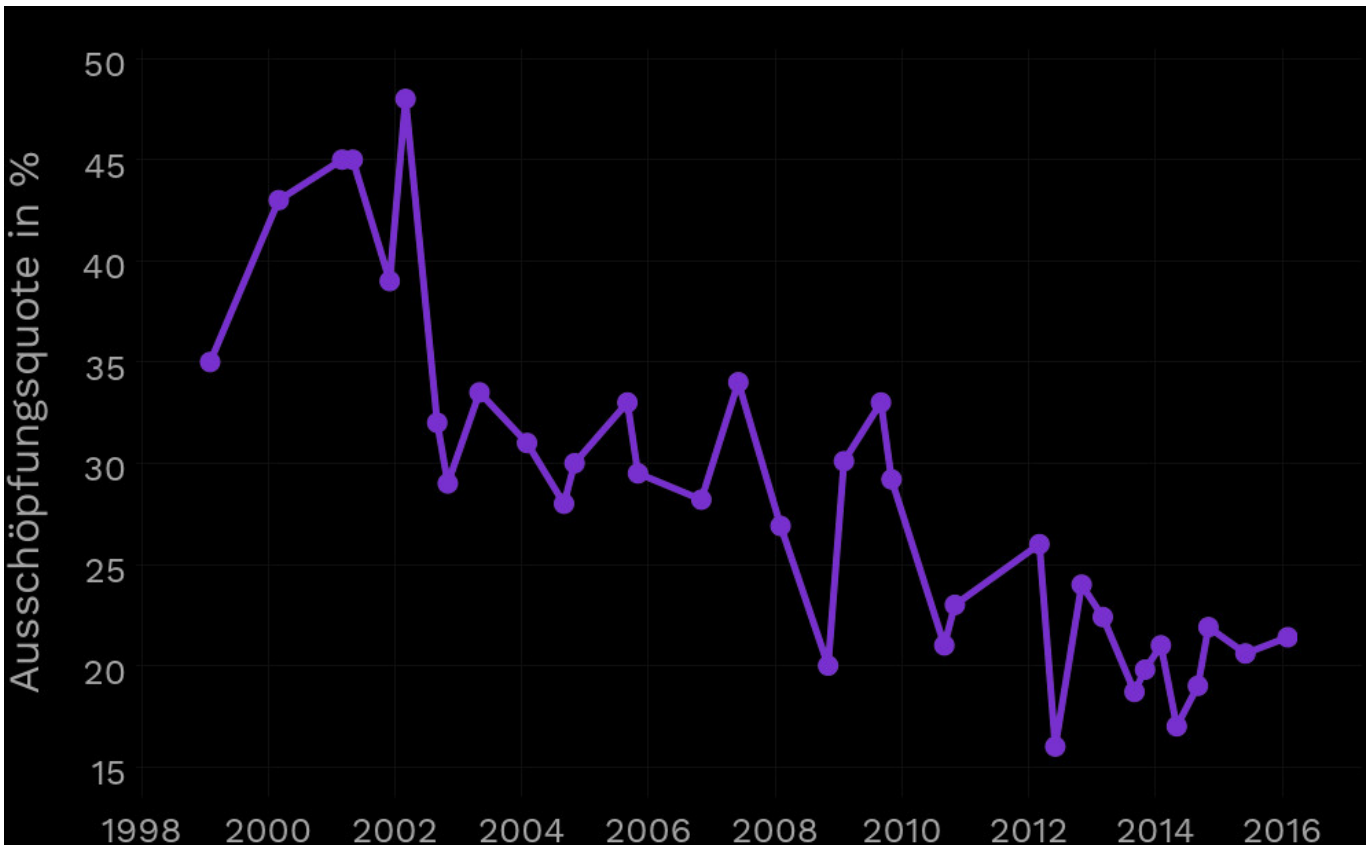
# Was Bevölkerungsumfragen wirklich über die Bevölkerung aussagen

Thomas Milic
16th November 2016




**Die US-Präsidentschaftswahlen kamen anders heraus, als von den meisten Prognostikern angenommen. Es ist nicht das erste Mal, dass bei vielbeachteten Wahlen oder Abstimmungen viele Umfragen ziemlich daneben lagen. Dadurch sinkt das Vertrauen in Umfragen. Wie kann dieses Vertrauen zurückgewonnen werden? Davon handelt mein Beitrag.**

Die Serie zum Jubiläum
50
année politique suisse

Die Frage, was Bevölkerungsumfragen über die Bevölkerung aussagen, hat durch die überraschende und von den meisten Meinungsforschern nicht vorhergesehene Wahl Donald Trumps an Brisanz gewonnen. Sind (politische) Umfragewerte seither tot, wie der konservative Politik-Analyst Mike Murphy in einem Kommentar meinte?

In der Tat ist es nicht das erste Mal, dass das Wähler- und Stimmenpotenzial rechtskonservativer Parteien oder Anliegen unterschätzt wurde – ob in Grossbritannien, Deutschland oder Österreich. Und auch hierzulande rieb man sich nach Abstimmungen über SVP-Initiativen schon verwundert die Augen ob des unerwarteten Resultats.

Vergleicht man etwa die in den Vox-Nachbefragungen ermittelten Resultate bei migrationspolitischen Vorlagen mit dem jeweiligen effektiven Resultat, so entdecken wir teils erhebliche Differenzen, die zudem einseitig zulasten des «rechten» (Protest-)Stimmenpotenzials ausfallen.

**Bemerkung: Angegeben ist das Ausmass (in Prozentpunkten), mit der das migrationsskeptische Lager jeweils unterschätzt wurde.**

**Gründe, wieso manche Stimmbürger in Befragungen nicht mehr gehört werden**

Ein solches Muster ist überzufällig. Woran liegt es aber, dass diese Stimmenden kein Gehör (mehr) in Umfragen finden? Liegt es am Stichprobenrahmen, an der Realisierung der Befragung, an der Gewichtung oder zensurieren sich die «Proteststimmenden» dadurch, dass sie sich Befragungen konsequent verweigern, am Ende selbst?

Zur Beantwortung dieser Frage muss ich ein wenig ausholen. Die Kommunikationsgewohnheiten der Menschen haben sich in den vergangenen zwanzig Jahren rasant gewandelt. Dieser Wandel stellte die Branche der Sozialforschungsinstitute vor nicht unerhebliche Probleme.

**1. Erreichbarkeit**

Die «goldenen» Zeiten, in denen das offizielle Telefonverzeichnis eine beinahe vollständige Liste der Grundgesamtheit der Wohnbevölkerung darstellte, sind vorbei – und dies wohl unwiderruflich. Mobiltelefonie und Internet verdrängen die Festnetztelefonie zusehends und der Wegfall der Eintragspflicht von Telefonnummern riss weitere Lücken.

Der Stichprobenrahmen des Bundesamtes für Statistik (SRPH), der seit 2010 zur Verfügung steht, konnte diese Lücke zwar schliessen. Doch die Nutzung dieses Stichprobenrahmens ist nur Erhebungen vorbehalten, die den strikten Kriterien in Art. 13c, SR 431.012.1 genügen.

Kommerziellen Befragungen – und Vorwahlbefragungen gehören dazu – steht dieser Stichprobenrahmen nicht zur Verfügung. Darüber hinaus wurde damit ein weiteres und viel schwerwiegenderes Problem von Befragungen nicht gelöst: Die hohen Verweigerungsraten.

**2. Verweigerung**

Verweigerung ist im Prinzip nur dann ein Problem, wenn sich die Befragungsteilnehmer von jenen, die nicht teilnehmen wollen, auch systematisch unterscheiden. Tun sie das? Das lässt sich oftmals erst im Nachhinein beurteilen. Aber es leuchtet sofort ein, dass zumindest das *Risiko*

der Verzerrung steigt, je höher die Verweigerungsquote ausfällt. Einer der Doyens der amerikanischen Umfrageforschung, Robert M. Groves[1], sagte dazu:

> «The risk of failures of surveys to reflect the facts increases with falling response rates. The risk is not always realized, but with the very low response rates now common, we should expect more falied predictions based on surveys.»
>
> Robert M. Groves, Umfrageforscher

Wie hoch ist eine gerade noch «akzeptable»  Ausschöpfungsrate? Dazu gibt es nun (fast) so viele Meinungen wie Köpfe. Auffallend ist jedoch, dass der Schwellenwert in den achtziger und neunziger Jahren deutlich höher angesetzt wurde als in jüngerer Zeit. Stellvertretend dafür steht das folgende Zitat von Cliff Zukin, Professor für *Public Policy and Political Science* an der *Rutgers University* und ehemaliger Präsident der *American Association for Public Opinion Research*[2]:

> «When I first started doing telephone surveys in New Jersey in the late 1970s, we considered an 80 percent response rate acceptable, and even then we worried if the 20 percent we missed were different in attitudes and behaviors than the 80 percent we got. Enter answering machines and other technologies. By 1997, Pew's response rate was 36 percent, and the decline has accelerated. By 2014 the response rate had fallen to 8 percent.»
>
> Cliff Zukin, Umfrageforscher

Auch in der Schweiz haben Umfrageinstitute mit sinkenden Antwortraten zu kämpfen. Sie liegen nicht unter zehn Prozent, aber Anlass zur Besorgnis geben auch die Antwortraten bei Schweizer Umfragen. So ist etwa die Ausschöpfungsquote bei den VOX-Nachbefragungen zwischen 2000 und 2016 von rund 40-45 Prozent auf etwa zwanzig Prozent gefallen.

Nun haben Studien aufzeigen können, dass Antwortraten und Datenqualität nicht notwendigerweise in einem linearen Verhältnis zueinander stehen. Man kann trotz vergleichsweise tiefen Ausschöpfungsraten eine hohe Datenqualität erzielen[3].

Aber bei Antwortraten von unter zehn Prozent fällt es schwer zu glauben (und noch viel schwerer, dies Journalisten und Konsumenten von demoskopischen Resultaten weiterzuvermitteln), dass die Verweigerung *nicht* systematisch erfolgt. Im Gegenteil: Angesichts der tiefen Antwortraten sind die Teilnehmenden an Umfragen beinahe schon als Extrempopulation zu betrachten.

**Wie man die Datenqualität verbessern kann**

Lässt sich Remedur schaffen? Eine Möglichkeit, Nonresponse zu verringern, besteht darin, die Kooperationsbereitschaft zu erhöhen – beispielsweise durch Anreize. Solche Anreize sind meist materieller Natur, etwa ein Gutschein von 10 Franken. Solche Inzentivierungsmassnahmen haben durchaus Erfolg und wirken sich nachweislich positiv auf die Datenqualität aus. Allerdings ist damit ein vergleichsweise hoher Aufwand und entsprechende Kosten verbunden. Viele Medienhäuser können oder wollen keine solchen Kosten nicht tragen.

Eine andere, weniger «kostspielige» Methode, ist die nachträgliche Fehlerkorrektur, sprich die Gewichtung von Daten.[4] Gewichtungen können viele der oben genannten Probleme lindern und ermöglichen unter Umständen präzise Schätzungen.[5] Aber auch Gewichtungen sind kein Allheilmittel.

Um eine Gewichtung vornehmen zu können, braucht es zunächst entsprechende Informationen über die Grundgesamtheit. Diese sind längst nicht immer vorhanden. Wenn man etwa eine Stichprobe von voraussichtlich an den Wahlen *Teilnehmenden* nach Alter und Geschlecht gewichten will, so muss man strenggenommen wissen, wie sich der Stimmkörper (voraussichtlich) nach Alter und Geschlecht zusammensetzen wird.

Woher aber soll man das wissen? Modellschätzungen aufgrund von Registerdaten mögen da allenfalls weiterhelfen. Auf jeden Fall aber ist die Gewichtung nicht nur profanes Handwerk, sondern setzt allerlei Annahmen voraus, die – wie jede Hypothese – zutreffend sein können oder auch nicht. Des Weiteren kann eine (unsorgfältige) Gewichtung unter Umständen absurde Effekte erzeugen, wie das Beispiel des inzwischen zu einer gewissen Berühmtheit gelangten 19-Jährigen Teenagers aus Illinois zeigt, der dadurch, dass ihm in einer Panel-Befragung[6] das 30-fache Gewicht eines durchschnittlichen Befragten zugewiesen wurde, am Ende gar die gesamte Schätzung massgeblich beeinflussen konnte.

**Wie begegnet man der Verweigerung?**

All das wird jedoch nicht genügen, um die Interviewverweigerung von Anti-Establishment-Wählenden zu korrigieren. Und genau ihr Stimmenpotential ist es, welches mit hoher Regelmässigkeit unterschätzt wird. Doch warum verweigert sich diese Gruppe von Wählenden den Befragungen konsequent? Die Antwort lautet, so ist zumindest zu vermuten: Sie halten Umfragen für ein (Manipulations-)Instrument des ungeliebten Establishments. Bereits im Vorfeld der Wahlen beschwerten sich deshalb auch viele Trump-Anhänger über die ihrer Ansicht nach gefälschten und zurechtgebogenen Umfrageresultate.

Nach der Wahl waren sie davon *noch überzeugter* als zuvor. Sie sahen ihren Verdacht durch das Ergebnis bestätigt. Auch in Europa hören wir denselben Vorwurf oft und in aller Regel ebenfalls von rechter Seite. Und auch hier kommt es nicht selten dazu, dass sich die «Umfrageverdrossenen» anschliessend durch die Wahlergebnisse in ihrem Verdacht abermals bestätigt sehen. Um diesem Problem beizukommen, gibt es nur ein Rezept: Das Vertrauen in Befragungen muss zurückgewonnen werden. Wie? Erstens: Die (Vor-)Umfragen müssen genauer werden. Zweitens: Es muss mehr Transparenz geschaffen werden. Ein Anfang würde darin bestehen, die Daten zu veröffentlichen.

**Referenzen und Anmerkungen:**


- [1] Zit. in: Bardes, Barbara und Robert W. Oldendick (2016). Public Opinion: Measuring the American Mind. 5th Edition. London: Rowman & Littlefield.
- [2] http://www.nytimes.com/2015/06/21/opinion/sunday/whats-the-matter-with-polling.html?_r=0
- [3] Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer und Roger Tourangeau. 2009. *Survey Methodology*. Zweite Edition. Hoboken, NJ: John Wiley & Sons.
- [4] So gut wie alle Umfragedaten werden heutzutage gewichtet. Im Übrigen: Eine Gewichtung ist eine Korrektur und demnach nur dann nötig, wenn Korrekturbedarf besteht – demnach eine Verzerrung (Nonresponse) vorliegt. Sind die Befragungsdaten unverzerrt, erübrigt sich selbstredend eine Korrektur. Der Umstand, dass fast überall gewichtet wird, deutet darauf hin, dass solche Verzerrungen weit verbreitet sind – weiter als man angesichts der häufigen Selbstdeklaration als «repräsentative Befragung» meinen könnte.
- [5] Wang, W., D. Rothschild, S. Goel und A. Gelman (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*.
- [6] Bei dieser Panelbefragung handelt es sich um den Daybreak Poll von USC Dornsife/LA Times – also ausgerechnet jene Befragung, die als eine der Wenigen am Ende auf der richtigen Seite lag. Sie hatten einen Wahlsieg Trumps vorhergesagt. Sonderlich genau war diese Befragung im Übrigen nicht. Sie lag auf nationaler Ebene klar daneben, denn Hillary Clinton erzielte national die höheren Wähleranteile als Donald Trump. Eines hatte diese Befragung jedoch anderen voraus: Die Daten wurden öffentlich zugänglich gemacht.


**Graphiken:** Salim Brüggemann