

Umfragen verstehen und beurteilen

Sozial- und Politikforschung für die politische Praxis

Prof. Dr. R. Schnell
Universität Duisburg-Essen
www.methodenzentrum.de

Berlin, 8.10.2019
BVM - Kompaktseminar



- Der allgemeine Teil basiert auf:
Gawiser, S.R./Witt,G.E.: 20 Questions a Journalist
Should Ask about Poll Results. American Association
for Public Opinion Research (AAPOR), 3rd. ed.
- Der spezielle Teil über Internetbefragungen basiert auf:
Schnell, R. (2019): Survey-Interviews. Methoden
standardisierter Befragungen, Wiesbaden, 2. Auflage.

- Sehr viele verschiedene Dinge werden als 'Umfrage' bezeichnet.
- Dazu gehören z.B. Pseudo-Umfragen, wie z.B.
 - Befragungen auf der Straße,
 - in Einkaufszentren,
 - Befragungen auf besuchten Webseiten (Pop-Up-Surveys) usw.
- Solche Pseudo-Umfragen sind prinzipiell nicht verallgemeinerungsfähig.
- Sie taugen auch nicht als Stimmungsbild.
- Daher sollte über diese Art von Umfragen auch nicht so berichtet werden wie über wissenschaftliche Befragungen.

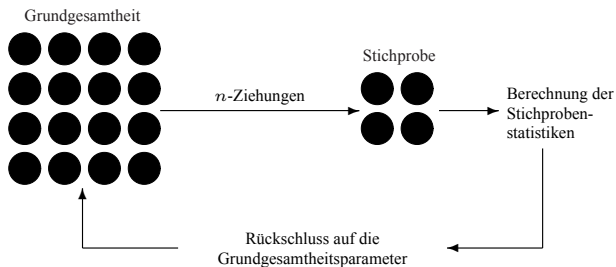
Kriterien für wissenschaftliche Befragungen

- Das Ziel ist die Berechnung statistischer Kennwerte.
- Es wird eine Stichprobe aus einer exakt definierten Menge („Grundgesamtheit“) befragt.
- Die Stichprobe ist eine echte Zufallsstichprobe.
- Es werden besondere Methoden zur Verringerung des Ausmaßes und des Effekts von Ausfällen (Nonresponse) auf die Schätzungen eingesetzt.
- Es werden besondere Methoden zur Verringerung von Fehlern (durch Frageformulierung, Reihenfolge der Fragen und Antworten, Verhalten der Interviewer und Befragten, Dateneingabe und Erhebungsmodus) verwendet.
- Alle technischen Details der Datenerhebung werden öffentlich dokumentiert.

Quelle: American Association for Public Opinion Research (Santos 2001 in Public Opinion Quarterly)

Noch einmal zum Ziel

- Denken Sie immer daran, dass es das Ziel einer Umfrage ist, Rückschlüsse auf die Bevölkerung zu ziehen, nicht auf die Stichprobe.
- Der Rückschluss auf die Grundgesamtheit ist mathematisch **nur** dann angebar, wenn es sich um eine Zufallsstichprobe handelt.¹



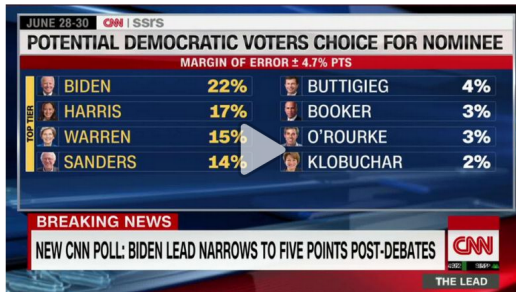
¹Beispiel: Fallgeschwindigkeit im luftleeren Raum. Wenn man die Formel auf der Erde anwendet, ist nicht die Mathematik falsch.

CNN debuts new polling standards as 2020 race heats up



Analysis by [Jennifer Agiesta](#), CNN Polling Director

Updated 1449 GMT (2249 HKT) July 9, 2019



New CNN poll: Biden lead narrows, Harris and Warren surge



Trump lashes out at Mitt Romney on Twitter



Mike Pompeo addresses defying subpoena deadline



Trump tells reporters he doesn't care about Biden



Rep. Trum scan

(CNN) — Today, [CNN](#) is launching new standards for reporting on polling. These standards will



CNN's Transparency Questionnaire

1. What survey firm conducted the poll?
2. How were respondents interviewed – by live interviewers on the phone, IVR, online, self-administered questionnaire or another method?
3. Who paid for the survey and why was it done?
4. How many people were interviewed for this survey?
5. In what language(s) were respondents interviewed?
6. Please provide a copy of the full text and interviewer instructions/programming for all questions included in this survey release.
7. When was your survey conducted?
8. What is the source of your sample for this survey, and by what method were respondents selected? Please be as specific as possible, and if via web panel(s), please include a description of how the panelists were recruited. If your study was conducted online and included respondents chosen via routers, approximately what percentage of respondents were directed to the survey via routers?
9. If any quotas were applied to sampling or interviewing, at what stage were they applied, what variables and targets were used, and what is the source of your estimate of the target quota?
10. What is the universe of people you are trying to survey, and what makes you confident that the sample source represents that universe?
11. If surveys were conducted by telephone, what percentage of interviews were conducted via calls to cellphones? If surveys were conducted online, were respondents allowed to complete the survey via mobile browsers, and approximately what share of your respondents did so?
12. If surveys were conducted by telephone, how many callback attempts did a sampled number receive before being retired?
13. If surveys were not conducted by a live interviewer, what do you do to ensure your respondents are real people and are paying attention to the survey?
14. What is your estimate of this survey's error, how is it calculated, and why is this an appropriate error estimation for your survey? If you are reporting a margin of sampling error, has it been adjusted for design effects?
15. If your survey has been weighted, please list the weighting variables and the source of the weighting parameters. If your survey has not been adjusted for education, please explain why and provide an unweighted frequency for education distribution among your respondents.
16. Is there a minimum unweighted sample size you require before releasing any subset estimates, and if so, what is it?

20 Fragen zu Umfragen

1. Wer hat die Umfrage durchgeführt?

- Welches Erhebungsinstitut hat die Umfrage durchgeführt?
- Dies muss immer die erste Frage sein.
- Wenn man nicht angeben kann, wer die Umfrage durchgeführt hat, kann man die folgenden 19 Fragen auch nicht beantworten.
- Dann kann man die Gültigkeit der Ergebnisse nicht beurteilen.
- Langjährig bekannte Erhebungsinstitute haben einen Ruf zu verlieren.
- Daher werden diese alle Informationen, die man zur Beurteilung einer Studie benötigt, auch liefern.

2. Wer hat die Umfrage bezahlt und warum?

- Umfragen werden immer aus einem bestimmten Grund durchgeführt.
- Das ist meist ein politisches oder wirtschaftliches Motiv (weniger als 5% des Umsatzes der Marktforschung gehen auf öffentliche Auftraggeber zurück).
- Alle diese Gründe sind prinzipiell legitim.
- Allerdings finden sich bei Befragungen für (oder noch öfter durch) Interessengruppen häufiger Erhebungsfehler.
- Dazu gehören vor allem ungeeignete Stichprobenverfahren oder einseitige Fragestellungen.

3. Wie viele Personen wurden befragt?

- Umfrageergebnisse schwanken immer durch Zufall.
- Das Ausmaß dieser Schwankungen lässt sich bei echten Zufallsstichproben exakt berechnen.
- Diese zufälligen Schwankungen werden bei Zufallsstichproben mit steigender Zahl der Befragten kleiner.
- Wenn alles andere unverändert bleibt, führt eine Vervierfachung der Zahl der Befragten zu einer Halbierung der Schwankungsbreite.
- Handelt es sich nicht um eine Zufallsstichprobe bedeutet eine größere Stichprobe keineswegs höhere Genauigkeit.

4. Wie wurden die Befragten ausgewählt?

- Bei wissenschaftlich begründbaren Befragungen erfolgt die Auswahl durch einen Zufallsprozess.
- Dies führt dazu, dass sich die Wahrscheinlichkeit für die Auswahl einer Person exakt berechnen lässt.
- Dies gilt nicht für Stichproben, bei denen
 - der Interviewer die Befragten auswählt oder
 - der Befragte sich selbst auswählt.
- In der wissenschaftlichen Literatur werden diese beiden Formen als willkürliche Stichproben bezeichnet und fast nie verwendet.
- Auf beide Formen werde ich später ausführlich eingehen.

5. Aus welcher Grundgesamtheit wurde ausgewählt?

- Es ist absolut wichtig zu wissen, aus welcher Grundgesamtheit die Befragten ausgewählt wurden.
- Sie müssen wissen, ob aus allen Erwachsenen, nur in einem Bundesland oder nur einer Stadt eine Stichprobe gezogen wurde.
- Wenn man Unternehmen befragt, muss man wissen, ob es sich um eine Stichprobe aus allen Unternehmen, den Mitgliedern eines bestimmten Verbandes oder um eine Stichprobe von sich selbst für eine Befragung gemeldeten Unternehmen handelt.

6. Beruhen die Ergebnisse auf den Antworten aller Befragten?

- Eine einfache Möglichkeiten, die Ergebnisse einer Umfrage falsch darzustellen, besteht darin, nur die Antworten einer Untergruppe zu berichten.
- Ein Beispiel wären nur die Ergebnisse für FDP-Wähler oder Hausbesitzer.
- Solche Teilmengen müssen explizit benannt werden, sonst ist die Veröffentlichung irreführend.

7. Wer hätte interviewt werden sollen?

- Bei keiner Umfrage werden alle interviewt, die ausgewählt wurden.
- Manche Personen sind schwer zu erreichen.
- Andere sind erkrankt oder dauerhaft schwer befragbar.
- Natürlich kann auch eine Sprachbarriere vorliegen.
 - Viele Studien beschränken sich z.B. auf die „deutschsprachige Bevölkerung“.
- Der häufigste Grund für ein nicht stattfindendes Interview (einen Ausfall) ist seit mehr als 30 Jahren die Tatsache, dass Personen eine Befragung ablehnen.

8. Wie geht man mit Verweigerungen um?

- Für die Beurteilung einer Befragung muss man wissen, welche Schritte unternommen wurden, um die Nichtbeantwortung zu minimieren.
- Welche Schritte das sind, hängt von der Ursache für den Ausfall ab.
- Keinesfalls kann man einfach alle Ausfälle hinnehmen oder durch andere Personen ersetzen.
- Je nach Erhebungsinstitut, Auftraggeber, Gegenstand der Befragung und Erhebungsmodus kann der Anteil der Verweigerungen so groß werden, dass die Ergebnisse nicht mehr verallgemeinerungsfähig sind.
- Das hängt nicht allein vom Anteil der Verweigerer ab, sondern vor allem an Unterschieden zwischen Verweigerern und Teilnehmern.

9. Was sind Ausschöpfungsraten?

- Eine Ausschöpfungsrates ist der Anteil der tatsächlich durchgeführten Interviews an der Ausgangsstichprobe.
- Es gibt verschiedene Definitionen der Ausschöpfungsrates, die zu unterschiedlichen Ergebnissen führen.
- In der wissenschaftlichen Literatur gibt es zwei Standarddefinitionen der AAPOR, die in der Regel zu kleineren Zahlen führen als die in Deutschland häufig verwendeten Definitionen. Wenn man es wirklich genau wissen will, fragt man nach der Ausschöpfung nach AAPOR-2 oder AAPOR-3.
- Die genaue Definition erfordert ca. 30 Seiten Text, daher sollten wir die Details hier überspringen.

10. Ist die Ausschöpfungsrage wichtig?

- In den letzten 30 Jahren hat der Anteil der Befragten, die auf Umfragen antworten, abgenommen.
- Vor allem die Zahl derer, die sich weigern, daran teilzunehmen, hat zugenommen.
- Die Ursachen sind vielfältig (Telemarketing, Betrugsversuche, Zeitmangel, Zunahme des Bedürfnisses nach Privatsphäre, Vertrauensverlust in Wissenschaft, Politik etc.).
- Wichtig ist es, dass Verweigerung kein stabiles Persönlichkeitsmerkmal ist.
- Ebenso gibt es keinen notwendigen Zusammenhang zwischen Ausschöpfung und verzerrten Ergebnissen, falls zuvor alle Standards bei der Durchführung der Studie beachtet wurden.

11. Wann wurde die Umfrage durchgeführt?

- Politische oder politisch relevante Ereignisse haben einen dramatischen Einfluss auf Umfrageergebnisse. ,
- Ihre Interpretation einer Umfrage sollte davon abhängen, wann sie relativ zu wichtigen Ereignissen durchgeführt wurde.
- Neue Ereignisse können auch die neuesten Umfrageergebnisse sinnlos werden lassen.
- Umfrageergebnisse, die mehrere Wochen oder Monate alt sind, können durchaus gültig sein, aber das hängt davon ab, ob es zwischenzeitlich relevante Ereignisse gab.

12. Wie wurden die Interviews geführt?

- Es gibt vier Hauptmöglichkeiten („Modi“): persönlich, telefonisch, online oder per Post.
- Die meisten Umfragen werden telefonisch durchgeführt.
- Einige Umfragen werden jedoch auch persönlich (F2F) durchgeführt. Hierbei werden Personen in Ihren Wohnungen durch persönlich anwesende Interviewer durchgeführt.
- Einige Umfragen werden per Post durchgeführt, spezielle Gruppen (z.B. Politiker) werden gelegentlich per Email befragt.
- Zunehmend finden sich online Befragungen (dazu gleich mehr).
- Der Erhebungsmodus sollte immer berichtet werden.

13. Was ist der Standardfehler? (S. 1/2)

- Interviews in einer echten Zufallsstichprobe von 2.000 Erwachsenen können die Meinungen von Millionen von Person korrekt wiedergeben.
- Das bedeutet, dass Interviews mit allen Erwachsenen eines Landes ungefähr die gleichen Ergebnisse liefern würden wie eine Zufallsstichprobe, die auf 2.000 Interviews basiert.
- Was passiert, wenn eine andere Zufallsstichprobe etwas andere Ergebnisse liefert als die erste Umfrage? Keine der Umfragen ist „falsch“. Dieser Bereich möglicher Ergebnisse wird in der Statistik durch den „Standardfehler“ gekennzeichnet.
- Der Standardfehler ist kein „Fehler“ im Sinne eines Irrtums. Er ist ein Maß für die zu erwartende Schwankungsbreite der Ergebnisse aufgrund der Tatsache, dass eine Stichprobe verwendet wurde.

13. Was ist der Standardfehler? (S. 2/2)

- Die Schwankungsbreite hängt von der Streuung des Merkmals, der Stichprobengröße und dem Ausmaß der gewünschten Irrtumswahrscheinlichkeit ab.²
- Die gewünschte Irrtumswahrscheinlichkeit wird meist mit 5% festgelegt.
- So bedeutet z.B. eine „Fehlerquote von 3 Prozentpunkten“: Die Antworten der Umfrage würden bei 95% aller Stichproben innerhalb von plus oder minus 3 Prozentpunkten der Grundgesamtheit liegen.

²Bei Anteilswerten ergibt sich ein entsprechendes Intervall bei einem Anteilswert in der Stichprobe p mit

$$p_{1,2} = p \pm 1.96 \sqrt{\frac{p * (1 - p)}{n}}$$

Bei $p=0.5$ und $n=1600$ ergeben sich entsprechend die Grenzen des Intervalls mit ca. $\pm 2.5\%$.

Beispiel: Konfidenzintervalle

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

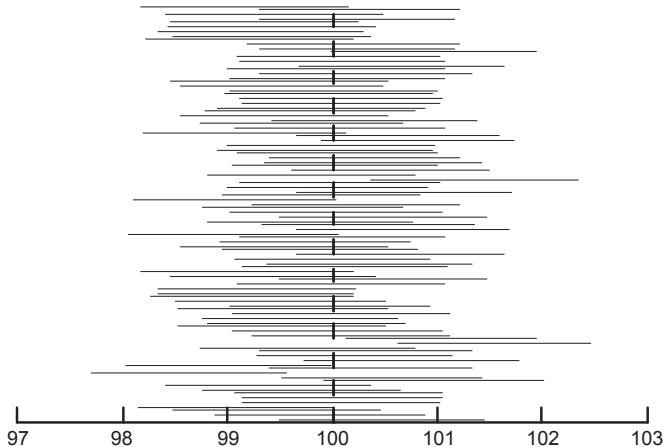
CNN 2019

20 wichtige Fragen

Wie sollte
berichtet werden?

Vertiefungen

Exkursion



14. Was ist mit der „Sonntagsfrage“?

- Schätzungen auf der Basis der Sonntagsfrage sind keine Wahlprognosen.
- Zumindest einige Institute bezeichnen dies mittlerweile auch anders, z.B. als „Projektion“.
- Für eine Wahlvorhersage muss man sowohl die Präferenzen als auch die Wahrscheinlichkeit einer Stimmabgabe korrekt schätzen.
- Dazu kommen Probleme mit Unentschiedenen, Meinungswandel in letzter Minute und möglicherweise falschen Auskünften (z.B. bei Präferenz extremer Parteien).
- Daher sollten die Ungenauigkeiten bei Wahlvorhersagen größer sein, als es die Berücksichtigung des Standardfehlers allein es nahelegt.
- Dies ist sowohl in UK, US als auch in Deutschland empirisch gezeigt worden.

Genauigkeit von Wahlvorhersagen 1957-2013

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

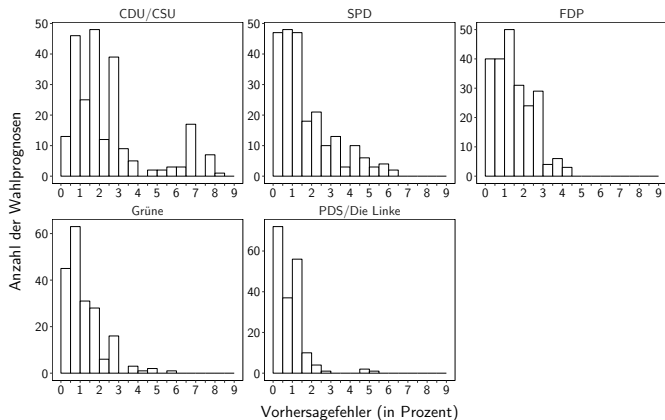
CNN 2019

20 wichtige Fragen

Wie sollte
berichtet werden?

Vertiefungen

Exkursion



Quelle: Schnell/Noack 2014, mda. Datenbasis: 232 Befragungen zu den Bundestagswahlen 1957-2013, maximal ein Monat vor der Wahl.

15. Wie gut sind Hochrechnungen bei Wählerbefragungen?

- Korrekt durchgeführte Wählerbefragungen direkt nach einer Wahl (Exit-Polls) liefern in vielen Ländern ausgezeichnete Ergebnisse.
- Sie sind die einzige Gelegenheit, (nahezu) die tatsächlichen Wähler zu befragen.
- In Ländern, bei denen viele Personen per Brief wählen, wird gelegentlich eine telefonische Befragung mit einer Exit-Umfrage kombiniert. Das ist aber in Deutschland bislang weder üblich noch notwendig.

16. Welche anderen Faktoren können die Ergebnisse der Umfrage verzerren?

- Der Standardfehler ist nur eine mögliche Quelle für Ungenauigkeiten.
- Er ist nicht notwendigerweise die Quelle des größtmöglichen Fehlers.
- Er wird vor allem deshalb verwendet, weil er leicht quantifiziert werden kann.
- Die Frageformulierung und die Reihenfolge der Fragen sind ebenfalls mögliche Fehlerquellen. Eine unzureichende Schulung und Überwachung der Interviewer, Datenverarbeitungsfehler und andere Probleme können ebenfalls zu Fehlern führen.

17. Welche Fragen wurden gestellt?

- Man muss den genauen Wortlaut der Fragen kennen.
- Schon minimale Unterschiede in der Frage können das Antwortverhalten wesentlich ändern.
- Prüfen Sie Ihre Reaktion auf eine Frage:
 - Erscheint die Frage auf den ersten Blick fair und unvoreingenommen?
 - Sind die Antwortmöglichkeiten ausgewogen?
 - Könnten die meisten Leute die Frage beantworten?
- Bei strittigen Themen (z.B. Tempolimit) könnte es sinnvoll sein, den vollständigen Wortlaut der Frage in Ihren Bericht aufzunehmen.
- Bei solchen Themen kann ein Vergleich der Ergebnisse verschiedener Umfragen selbst eine Story werden.

18. In welcher Reihenfolge wurden die Fragen gestellt?

- Manchmal kann die Reihenfolge der Fragen die Ergebnisse beeinflussen.
- Diese Auswirkung kann beabsichtigt sein oder nicht.
- Fragt man in den USA, z.B. nach der wirtschaftlichen Lage und dann nach der Beliebtheit des Präsidenten wird man vermutlich andere Ergebnisse erhalten als bei der umgekehrten Reihenfolge der Fragen.
- Im Beispiel kann die Stärke und Richtung des Effekts von der wirtschaftlichen Lage zum Zeitpunkt der Befragung abhängen.
- Generell ist es wichtig zu prüfen, ob ein Effekt vorhergehender Fragen plausibel ist.

19. Gibt es mehr als eine Umfrage zum Thema?

- Ergebnisse anderer Befragungen sollten verwendet werden, um die Ergebnisse einer vorliegenden Studie zu überprüfen
- Falls sich die Ergebnisse unterscheiden, überprüfen Sie zuerst den Zeitpunkt der Befragung.
- Falls die Umfragen ungefähr zum gleichen Zeitpunkt durchgeführt wurden, fragen Sie jeden Auftraggeber nach einer Erklärung für die Unterschiede.
Widersprüchliche Ergebnisse führen oft zu interessanten Veröffentlichungen.

20. Was sollte noch berichtet werden?

- Wichtig für die Interpretation eines Umfrageergebnisses ist der Kontext.
- Das ist einerseits der Vergleich mit anderen oder früheren Befragungen.
- Das sind andererseits aber auch Ereignisse, die die Umfrageergebnisse beeinflusst haben könnten.
- Gab es z.B. zuvor ein besonderes Ereignis, wie z.B. ein Verbrechen, eine Katastrophe oder ein politisches Ereignis, das zu einem – vielleicht nur kurzzeitigen – veränderten Antwortverhalten geführt haben könnte?
- Eine Umfrage in einen Kontext zu stellen ist in der Regel ein wesentlicher Bestandteil einer Berichterstellung.

Wie sollte berichtet werden?

Berichterstattung: Kodex vs. Praxis

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

CNN 2019

20 wichtige Fragen

Wie sollte
berichtet werden?

Vertiefungen

Exkursion

„Bei der Veröffentlichung von Umfrageergebnissen teilt die Presse die Zahl der Befragten, den Zeitpunkt der Befragung, den Auftraggeber sowie die Fragestellung mit. Zugleich muss mitgeteilt werden, ob die Ergebnisse repräsentativ sind.“

Quelle: <https://www.presserat.de/pressekodex/pressekodex>, Abschnitt 2.1

Information	Prozent
Institut	85
Stichprobe	57
Auftraggeber	56
Fragewortlaut	45
Erhebungszeitraum	40
Grundgesamtheit	30

n=293. Quelle: Raupp, J. (2007): Politische Meinungsforschung, Konstanz, S. 154.

Datenbasis: FAZ, FR, Welt, SZ; drei Monate vor und einen Monat nach den Bundestagswahlen am

22.9.2002 und dem 18.9. 2005.

Was bedeutet „Repräsentativität“?

- In der Mathematik gibt es diesen Begriff nicht.
- Für verallgemeinerbare Stichproben muss man für jedes Element der Grundgesamtheit die Auswahlwahrscheinlichkeit vor der Ziehung kennen.
- Dann sind unklare Begriffe wie 'Repräsentativität' unnötig.
- Der Begriff tauchte fast nur in zwei Kontexten auf:
 - weil man glaubte, einem Publikum den Begriff der bekannten Auswahlwahrscheinlichkeit nicht erklären zu können
 - bei dem Versuch, Quotenstichproben zu rechtfertigen.
- Neuerdings taucht der Begriff wieder auf, um selbstrekrutierte Internet-Stichproben zu rechtfertigen.
- Der wissenschaftliche Name für diese Art von Stichproben ist „non-probability sample“.
- In der deutschen Literatur heißen diese Stichproben „willkürliche Stichproben“.

Checkliste

- ① Zahl der Befragten
- ② Zeitraum der Befragung
- ③ Auftraggeber
- ④ Institut
- ⑤ Fragewortlaut
- ⑥ Fragekontext
- ⑦ Erhebungsmodus
- ⑧ Grundgesamtheit
- ⑨ Ausschöpfung
- ⑩ Falls nicht Einwohner- oder Telefonstichprobe:
 - Quotenstichprobe? (z.B. Allensbach)
 - sich selbst meldende Panel-Teilnehmer? (z.B. Civey, Yougov)

Vertiefungen

Was sind „Push-Polls?“

- Vor allem in den USA haben einige politische Kampagnen sogenannte „Push-Polls“ eingesetzt, um Gerüchte über Gegner zu verbreiten.
- Das sind keine Umfragen, sondern politische Manipulationen.
- Bei einem „Push-Poll“ wird eine große Anzahl von Personen angerufen und um Teilnahme an einer vorgeblichen Umfrage gebeten.
- Die „Fragen“ enthalten verschleierte Anschuldigungen gegen einen Gegner.
- Ein „Push-Poll“ soll dazu führen, dass der Befragte die Anschuldigungen in der Frage hört.
- „Push-Polls“ sind unethisch und werden von allen professionellen Organisationen verurteilt.

Was ist Demoskopie?

- Das Wort wird international nicht verwendet.
- Das Wort wird in Deutschland fast ausschließlich vom Institut für Demoskopie in Allensbach verwendet.
- In den Medienwissenschaften wird der Begriff gelegentlich von Personen verwendet, die selbst kaum Forschung mit Umfragen durchführen.
- Der Begriff spielt in der wissenschaftlichen Diskussion keine Rolle.
- Der einzige Lehrstuhl weltweit für Demoskopie ist in Konstanz – weil Frau Noelle den Ministerpräsidenten überzeugte. Der Lehrstuhl verwendet das Wort nicht zur Beschreibung seiner Tätigkeit.
- Wir nennen das Feld Meinungs-, Wahl-, Politik- oder Sozialforschung, das Feld der Methoden dieser Art von Forschung nennt sich „Survey-Methodology“.

Wie werden Telefonstichproben gezogen?

- Bei Telefonstichproben handelt es sich um Personen, die über ein Telefon erreicht werden können.
- Man kann sich dabei weder auf Mobiltelefone noch auf Festnetznummern allein verlassen.
- In Deutschland basieren professionelle Befragungen auf Nummernblöcken aus beiden Nummernräumen.
- Bei Festnetznummern muss eine Auswahl aus den im Haushalt lebenden Personen erfolgen.
- In der Regel sollte die Anzahl der Nummern, unter denen jemand erreicht werden kann, erfragt und für eine Gewichtung verwendet werden.

Ist ein größere Stichprobe immer besser als eine kleine Stichprobe?

- nein.
- Lassen Sie sich niemals von einer hohen Zahl von Befragten täuschen. Entscheidend vor allem, ob es sich um eine echte Zufallsstichprobe handelt.
- Das Kriterium dafür ist allein, ob vor der Befragung die Wahrscheinlichkeit, dass jemand für die Stichprobe ausgewählt wird, exakt berechenbar ist.
- Falls Sie jemand mit einer großen Anzahl von Befragten beeindrucken will, die nicht aus einer echten Zufallsstichprobe stammen, denken Sie an das berühmte Beispiel des „Literary Digest“ von 1936: Auf der Basis von 2.300.000 Befragten sagte die Zeitschrift vorher, dass Landon Roosevelt bei den Präsidentschaftswahlen von 1936 schlagen würde.

Was sind Gewichte?

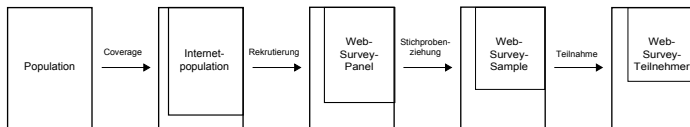
- Mit Gewichten werden Umfrageergebnisse an bekannte Randgrößen angeglichen.
- Es gibt verschiedene Arten und Möglichkeiten der Gewichtung, z.B. Design- oder Nonresponsegewichte.
- Man kann Gewichte mißbräuchlich verwenden, das ist aber bei renommierten Instituten nicht der Fall.
- Ob Gewichte funktionieren oder nicht, bedarf im Einzelfall der Beurteilung von Experten.
- Kritisch ist die Schätzung von Teilnahmewahrscheinlichkeiten. Einige Websurveys basieren auf solchen Gewichten.
- Sehr vorsichtig formuliert: Das ist Gegenstand der Forschung. Manchmal funktionieren solche gewichteten Websurveys, manchmal nicht.

Was ist mit Umfragen im Internet? (S.1/2)

- Man muss trennen zwischen Antwortmodus und Auswahlverfahren.
 - Prinzipiell ist der Antwortmodus bei Internetbefragungen weitgehend unproblematisch.
 - Das Auswahlverfahren ist kritisch, da es keine Liste von Internet-Nutzern gibt, aus denen man auswählen kann.
- Falls die Befragten sich selbst auswählen oder nur aus den Besuchern einer Webseite ausgewählt wurden, sollten die Ergebnisse ignoriert und nicht berichtet werden. Das sind Beispiele für Pseudo-Umfragen.
- Falls man eine Zufallsstichprobe (z.B. aus den Melderegistern) zieht und dann per Post um Teilnahme im Internet bittet, kann das hervorragend funktionieren.
 - Die holländische Censusbehörde macht das so. Nicht-Antwortende werden erneut aufgefordert und schließlich persönlich kontaktiert und befragt.
 - Man kann so auch ein „probability based“ Web-Access-Panel aufbauen.

Was ist mit Umfragen im Internet? (S.2/2)

- Fast alle kommerziellen Webpanels sind **keine** Zufallsstichproben: Das Hauptkennzeichen ist die Tatsache, dass Personen sich selbst freiwillig melden.
- Damit sind es „non-probability samples“: Willkürliche Stichproben.



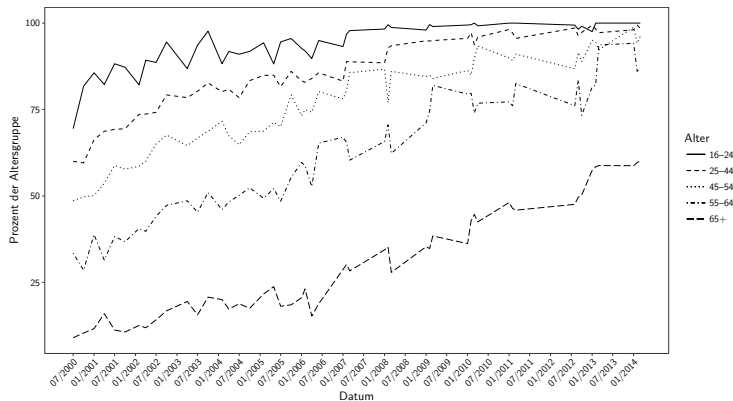
- Auswahlwahrscheinlichkeiten lassen sich nicht berechnen.
- Alle diese Panels sind gewichtet. Die technischen Details, um die Gewichtung nachzuvollziehen, stehen – in Deutschland – nicht zur Verfügung.
- Es sollte immer erwähnt werden, wenn eine Umfrage auf einem Web-Panel beruht.

- Yougov führt in UK Befragungen unter selbstrekrutierten Befragten im Internet durch.
- Yougov gibt selbst für seine Erhebungen eine Fehlermarge von $\pm 3\%$ an.
- Noch am Morgen der Brexit-Abstimmung gab Yougov an, dass „Remain“ gewinnen würde.
- Die Abweichung zwischen dem Ergebnis und der Yougov-Schätzung lag bei 4%.

Gibt es nicht ganz neue Methoden, mit denen das alles einfacher, schneller und billiger geht?

- nein.
- Statistik kann nicht zaubern.
- Auch nicht Artificial Intelligence, Machine Learning, Post-Stratification, Propensity-Score Matching oder sonst ein mathematisches Verfahren.
- Personen mit nicht getesteten Methoden haben immer ein finanzielles oder ein bestimmtes inhaltliches Interesse.

Prozent der Internetnutzer („ever used“)



UK National Omnibus-Survey 2000–2014

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

CNN 2019

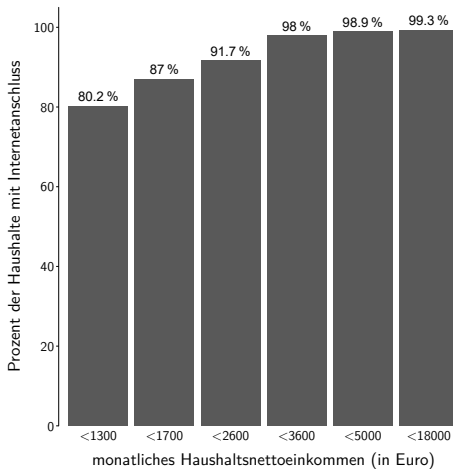
20 wichtige Fragen

Wie sollte
berichtet werden?

Vertiefungen

Exkursion

Haushalte mit Internet nach Einkommen



Datenbasis: Destatis, laufende Wirtschaftsrechnung 2017

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

CNN 2019

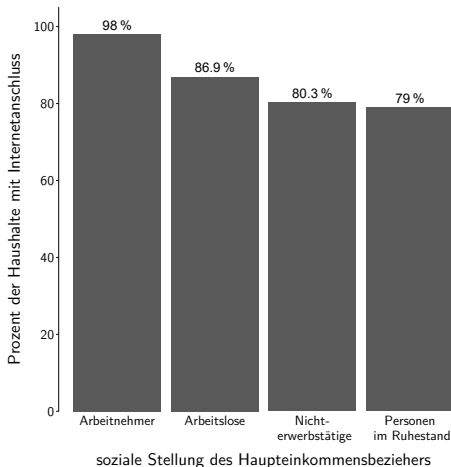
20 wichtige Fragen

Wie sollte
berichtet werden?

Vertiefungen

Exkursion

Haushalte mit Internet nach sozialer Stellung



Datenbasis: Destatis, laufende Wirtschaftsrechnung 2017

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

CNN 2019

20 wichtige Fragen

Wie sollte
berichtet werden?

Vertiefungen

Exkursion

Altersgruppen in Deutschland (65+)

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

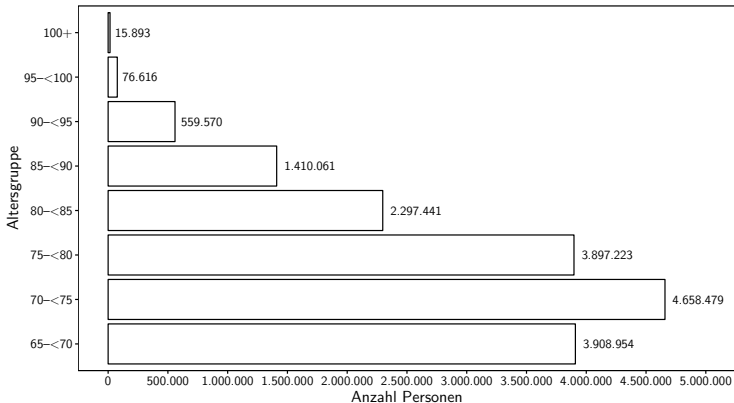
CNN 2019

20 wichtige Fragen

Wie sollte
berichtet werden?

Vertiefungen

Exkursion



Datenbasis: Destatis, Sonderauswertung

- Fast 17 Millionen Personen (mehr als ein Fünftel der Bevölkerung) sind älter als 65.
- Die Wahlbeteiligung liegt bei den über 70-Jährigen bei 75.8% (Destatis: Wista 3/2018).

Internetnutzung 2015 in Deutschland (50+)

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

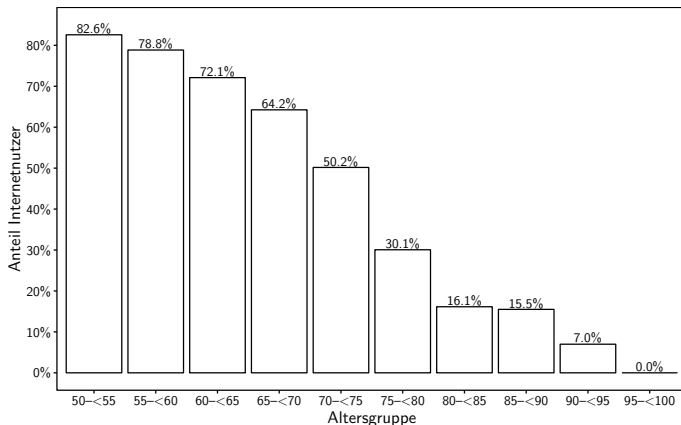
CNN 2019

20 wichtige Fragen

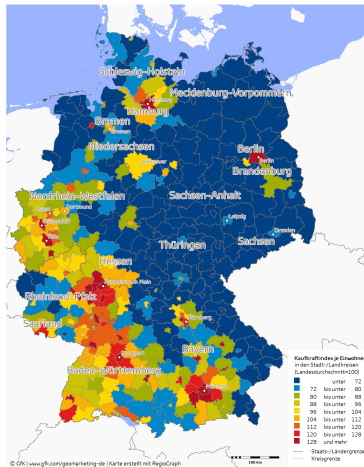
Wie sollte
berichtet werden?

Vertiefungen

Exkursion



Datenbasis: SHARE-6



www.gfk.com/de/insights/news/

bild-des-monats-online-kaufkraft-foodbereich-deutschland-2019

Ausschöpfungsrate und Bias

- Bias ist die Differenz zwischen der Schätzung und dem Grundgesamtheitsparameter.
- Der Bias ist umso höher, je größer der Unterschied zwischen den Respondenten R und den Nichtrespondenten N ist.
- Weiterhin ist der Bias umso höher, je mehr Nonrespondenten es gibt.

$$Bias_{\bar{x}} = (\bar{x}_R - \bar{x}_N) * \frac{n_N}{n_N + n_R} \quad (1)$$

- Ein Bias von Null ergibt sich also entweder
 - wenn es keine Unterschiede zwischen Respondenten und Nonrespondenten gibt,
 - der Anteil der Nonrespondenten gegen Null geht,
 - oder beides der Fall ist.

Nonresponsebias und Nonresponse-Rate

Umfragen
verstehen

Prof. Dr. Schnell

Einleitung

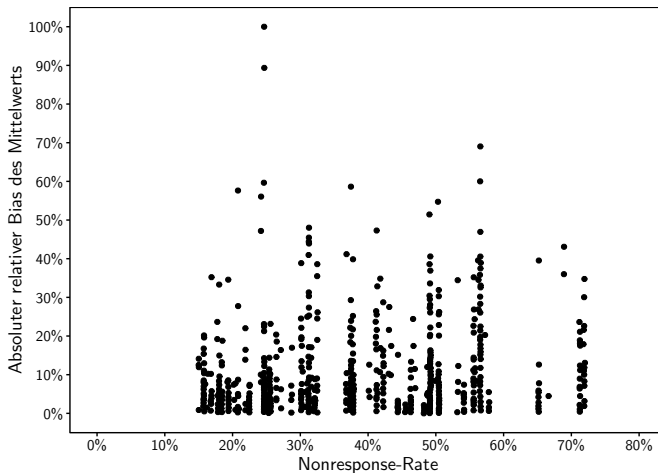
CNN 2019

20 wichtige Fragen

Wie sollte
berichtet werden?

Vertiefungen

Exkursion



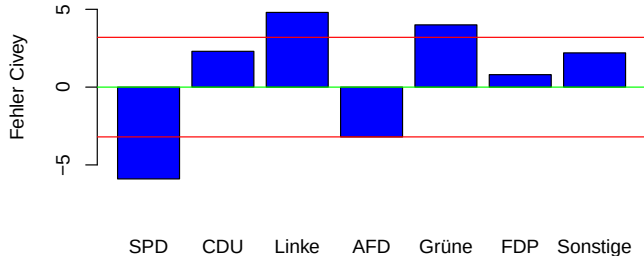
n=355 Schätzungen aus 30 Studien. Datenbasis: Groves/Peytcheva (2008, Public Opinion Quarterly)

- Wir wollen den Mittelwert $\bar{Y}$ schätzen.
- Der Bias eines Websurveys hängt von drei Größen ab:³
 - den Responsepropensities ρ ,
 - der Korrelation zwischen Responsepropensity und der zu schätzenden Variablen $R_{\rho Y}$
 - der Streuung der Responsepropensities S_{ρ} .
- Daher sinkt der Bias, wenn die Beteiligung am Websurvey steigt.
- Der Bias verschwindet, wenn es keinen Zusammenhang mit der Responsepropensity und der interessierenden Variablen gibt.
- Der Bias steigt, je größer die Streuung der Responsepropensities ist.

$$B(\bar{y}) = \frac{R_{\rho Y} S_{\rho} S_Y}{\bar{\rho}},$$

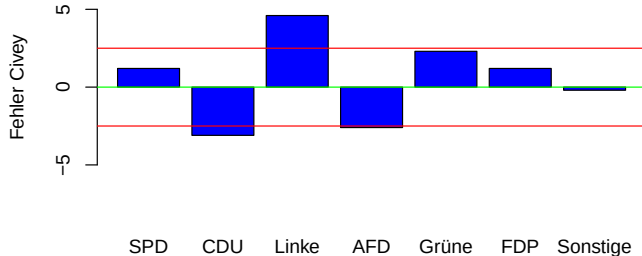
Landtagswahl Brandenburg 2019

- Wahl am 1.9.2019
- Civey am 28.8.2019
- Angegebene Fehlerquote $\pm 3.2\%$



Landtagswahl Sachsen 2019

- Wahl am 1.9.2019
- Civey am 28.8.2019
- Angegebene Fehlerquote $\pm 2.5\%$



Harry's New York Bar (Paris, 5 rue Daunou)

- Erfolgreiche Wahlprognosen sind prinzipiell kein Erfolgskriterium.
- Die amerikanischen Besucher von „Harry's New York Bar“ haben seit 1924 über den zukünftigen Gewinner der amerikanischen Präsidentenwahl abgestimmt.
- Zwischen 1924 und 2000 haben sie 18 von den 19 Wahlen korrekt vorhergesagt (die Ausnahme war 1976).
- Auch eine lange Erfolgsgeschichte bei „Wahlprognosen“ gibt keine Garantie für zukünftige Prognosen:
 - Die Wahlen 2004 und 2016 wurden inkorrekt vorhergesagt.
 - Die Vorhersagen für 2008 und 2012 waren richtig.
- Ein Verfahren, das manchmal funktioniert und manchmal nicht, ohne dass man dafür – am besten vorher – eine Erklärung hat, ist eben kein wissenschaftliches Verfahren.