

Studienunterlagen

François Höpflinger (www.hoepflinger.com)

Standardisierte Erhebungen - methodische Hinweise zu Umfragen

Inhalt

Einführende Literatur zu Forschungsmethoden

Zu den gesellschaftlichen Voraussetzungen für allgemeine Bevölkerungsumfragen

Formen von Umfragen

Wissenschaftliche Umfragen - integriert in Forschungsplan und -design

Zu den Frageformen

Befragung: Wichtige Regeln der Fragenbogen-Konstruktion

Stichprobenauswahl und Samplingverfahren

Ausschöpfungs- und Verweigerungsraten: Berechnung und Konsequenzen
von Ausfällen bei Repräsentativumfragen

Interviewer und befragte Personen – Interviews als soziale Situation und Interaktion

Integrative Gesamtschau: Vor- und Nachteile einer standardisierten Erhebung

Einführende Literatur zu Forschungsmethoden

- Atteslander, Peter 2008: Methoden der empirischen Sozialforschung. Berlin, New York: de Gruyter.
- Bortz, Jürgen, Nicola Döring 2006: Forschungsmethoden und Evaluation. Berlin u.a.: Springer.
- Brosius, Hans-Bernd, Friederike Koschel 2008: Methoden der empirischen Kommunikationsforschung – Eine Einführung. Opladen: Westdeutscher Verlag.
- Burton, Dawn (Hrsg.) 2000: Research Training for Social Scientists. London u.a.: Sage.
- Diekmann, Andreas 2010: Empirische Sozialforschung – Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt.
- Kallus, K. Wolfgang 2010: Erstellung von Fragebogen, Wien: Facultas.
- Kirchhoff, Sabine, Sonja Kuhnt, Peter Lipp, Siegfried Schlawin 2010: Der Fragebogen – Datenbasis, Konstruktion, Auswertung. Wiesbaden: Verlag für Sozialwissenschaften.
- Krämer, Walter 2011: So lügt man mit Statistik. Frankfurt/New York: Campus. Siehe auch: Huff, Darrell 1993 (1954): How to Lie with Statistics. New York, London: Norton.
- Kromrey, Helmut, Eckart Struck 2009: Empirische Sozialforschung. Buch und CD-ROM: Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung. Stuttgart: utb.
- May, Tim 1997: Social Research – Issues, Methods and Process. Buckingham, Philadelphia: Open University Press.
- Schnell, Rainer, Paul B. Hill, Elke Esser 2008: Methoden der empirischen Sozialforschung. München, Wien: Oldenbourg.
- Zelazny, Gene 2005: Wie aus Zahlen Bilder werden: Der Weg zur visuellen Kommunikation. Wiesbaden: Gabler.

Zu den gesellschaftlichen Voraussetzungen für allgemeine Bevölkerungsumfragen

Standardisierte Bevölkerungsumfragen setzen gesellschaftlich viel voraus, und sie sind faktisch nur in offenen, demokratischen und modernen Gesellschaften durchführbar (und auch dort nur innerhalb bestimmter Grenzen). Soziale Voraussetzungen, damit grössere und namentlich standardisierte Befragungen überhaupt sinnvoll durchgeführt werden kann, sind:

- a) eine gewisse Offenheit gegenüber Fremden, nur so wird einem Interviewer überhaupt Auskunft gegeben. Starkes Misstrauen oder Angst vor unbekanntem Personen behindert Umfragen. Ebenso muss eine Bereitschaft da sein, über sich Auskunft zu geben. Massenweise Antwortverweigerung - z.B. weil privates Tabu oder weil man Fremden nicht traut - verunmöglicht Umfragen.
- b) eine politisch angstfreie Atmosphäre, in der sich Personen getrauen, ihre Meinung frei und ungehindert zu äussern. Umfragen lassen sich in diktatorischen und autoritären Gesellschaften kaum sinnvoll durchführen (bzw. die Antworten werden dem entsprechen, was die Regierung wünscht). Repräsentativumfragen sind häufig nur in demokratischen Gesellschaften sinnvoll durchzuführen.
- c) eine individualistische Gesellschaft, in der jeder einzelne Bürger bzw. jede einzelne Bürgerin für sich verantwortlich ist und damit für sich antworten kann (in feudalen Gesellschaften oder Gesellschaften mit starken Sippenstrukturen sind Repräsentativumfragen sinnlos. (Die Antwort der Befragten wird sein: Fragen sie doch den Grundherrn, den Stammesoberhaupt, den Familienvorstand). Zudem sollte jeder Befragte eine eigene Meinung haben (was auch in Demokratien kaum der Fall ist). Verbreitete Meinungslosigkeit verhindert Meinungsumfragen.
- d) eine Tendenz, die Wahrheit zu sagen und nicht systematisch zu lügen oder zu übertreiben. Ebenso muss die Bereitschaft da sein, auf Fragen direkt zu antworten (und nicht jede Frage durch eine Gegenfrage oder eine Geschichte zu beantworten). Umfragen erfordern eine recht hohe soziale Disziplin, auch in dem Sinne, dass die Befragten über sich Bescheid wissen (z.B. wissen viele Befragte in wenig industrialisierten Ländern nicht, wie alt sie sind).

Gesamthaft sind repräsentative Bevölkerungsumfragen am ehesten in demokratisch-bürokratisch organisierten Gesellschaften möglich. In diesen Gesellschaften hat es genügend Bürger/innen, die ihre eigene, individuelle Meinung haben und diese auch zeigen. In diesen Gesellschaften sind die Leute aber auch sozialisiert, auf Fragen relativ präzise und zuverlässig zu antworten.

Zur Entstehung der (politischen) Umfrageforschung:

Keller, Felix (2001) Archäologie der Meinungsforschung - Mathematik und die Erzählbarkeit des Politischen, Konstanz: UVK Verlagsgesellschaft.

Formen von Umfragen

Es gibt verschiedene Formen von Umfragen. Die wichtigsten Umfrageformen sind:

- A) Mündliche Interviews (face-to-face-interviews): Befragung durch einen Interviewer. Je nach Grad der Standardisierung und Vorstrukturierung des Fragebogens werden bei mündlichen Interviews unterschieden zwischen:
- a1) Offene, qualitative Interviews: Interviews, die einen offenen Charakter haben, z.B. anhand eines kurzen Gesprächsleitfadens wird offen diskutiert oder erzählt. Eine Form ist das "narrative" (erzählende) Interview (z.B. zum Studium von Lebensgeschichten).
 - a2) Vorstrukturierte Experteninterviews: Befragung von ausgewählten Spezialisten und Experten (z.B. über ihr Fach oder ihre Organisation) gemäss vorbereitetem Frageraster.
 - a3) Standardisierte Interviews mit ausgearbeitetem und voll strukturierten Fragebogen, wobei die Fragen durch Interviewer in vorgegebener Form und Reihenfolge vorgelesen werden.
- B) Schriftliche Befragung: Versand eines schriftlichen standardisierten Fragebogens, der von den Befragten selbst ausgefüllt wird (z.B. zu Hause, im Betrieb, während einer Unterrichtsstunde). Häufig postalischer Versand mit (portofreiem) Rückantwortcouvert. Vorteil: kostengünstiger als mündliche Befragung, Nachteil: Befragungssituation bleibt unkontrollierbar.
- C) Telefonisches Interview: Befragung via Telefon, sachgemäss mit Fragen, die akustisch einfach zu verstehen sind. Wird heute zunehmend benützt, und zwar in der Form von 'computer-aided telephone interviews (CATI), wodurch Erhebung und Auswertung gleichzeitig geschehen. Vorteil: kostengünstig, computerunterstützt und zentral organisierbar, flexible Handhabung der Fragen, Nachteil: akustische Limitierung der Fragen, Sampling-Verfahren nicht einfach (zu dual-frame-Verfahren bei Telefonbefragungen (Festnetz und Mobilfunk, vgl. Häder, Häder 2009).
- D) Internet-Umfragen (elektronische Befragung): Elektronische Befragungen haben den Vorteil von Kostengünstigkeit und Raschheit. Sie sind zudem flexibel zu handhaben, und je nach Befragungsgruppe lassen sich ohne grossen Aufwand etwa unterschiedliche Frageblöcke stellen. Vom Format entsprechen Internet-Umfragen schriftlichen Befragungen (und es gelten deshalb dieselben Regeln klarer schriftlicher Darstellung). Angesichts der Tatsache, dass weiterhin nur ein ausgewählter Teil der Bevölkerung dem Internet angeschlossen sind, ergeben sich Probleme ihrer Repräsentativität bei allgemeinen Bevölkerungsumfragen. Zudem besteht kein fixes 'Adressbuch' an Internet-Adressen, die ein Zufallsampling erleichtern.

Befragungsformen A, C und D erlauben es, Reihenfolge von Fragen bzw. Items zu wechseln (um Positionseffekte zu minimieren. Bei Befragungsform B (schriftlich) sind Reihenfolgen der Fragen bzw. Items fix. Befragungsformen C und D erlauben auch sogenannte 'splitting sampling'-Verfahren, bei dem einem Teil der Befragten eine Frageform, einem anderen Teil der Befragten eine andere Frageform bzw. -formulierung vorgelegt wird. Dies erlaubt etwa die Auswirkungen verschiedener Frageformen und Frageformulierungen auf das Antwortverhalten zu überprüfen, vgl. dazu: Petersen, Thomas (2002) Das Feldexperiment in der Umfrageforschung, Frankfurt: Campus.

Ein neuer Trend in der Forschung besteht darin, dass vor allem grössere Erhebungen (Mikrozensus, internationale Studien) zentral von einer Forschergruppe bzw. einem statistischen Amt organisiert und durchgeführt werden, wogegen die Auswertung der Befragungsdaten dezentralisiert an verschiedene Forschergruppen delegiert wird. Beispiele sind etwa die Daten des schweizerischen Haushalts-Panel (SHP) oder auch Mikrozensus-Daten des Bundesamts für Statistik (Arbeitskräfteerhebung SAKE, Gesundheitsbefragung SGB u.a.).

Eine zentrale Dienstleistungsstelle für sozialwissenschaftliche Forschung in der Schweiz ist FORS - Schweizer Kompetenzzentrum für die Sozialwissenschaften in Lausanne (www.fors.unil.ch).

Quantitative Interviews der Mitgliedsinstitute des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM) nach Befragungsart:

	1990	2000	2005	2007
Persönliche Interviews	65%	34%	24%	26%
Telefon-Interviews	22%	41%	45%	41%
Schriftliche Interviews	13%	22%	9%	6%
Online-Interviews	-	3%	22%	27%

Quelle: Häder, Michael; Häder, Sabine (Hrsg.) (2009) Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung, Wiesbaden: VS Verlag für Sozialwissenschaften.

Wissenschaftliche Umfragen - integriert in Forschungsplan und -design

Bei wissenschaftlichen Forschungsgesuchen wird schon bei der Projekteingabe ein klarer und detaillierter Forschungsplan verlangt (inkl. Kostenaufschlüsselung und genauem Zeitplan). Dasselbe gilt für manche Auftragsforschung.

Ein klares Forschungsdesign zu Beginn einer Forschungsarbeit erspart später viel Unsicherheit, Unklarheiten oder massive Zeitüberschreitungen.

Ein Forschungsplan, der eine grössere Befragung vorsieht, schliesst nachfolgende Schritte ein:

1. Festlegung und Präzisierung der Forschungsfragen z.B. Untersuchung der Lebensbedingungen von allein erziehenden Müttern, Studie sozialer Mobilität bei Arbeitern, Determinanten von Drogenkonsum usw.

1.1. Festlegung der Stichprobe bzw. der zu befragenden Gruppe (z.B. Jugendliche ab 15 bis 20, Mütter zwischen 20 bis 45 oder repräsentativer Querschnitt aller Schweizer Bürger). Es lohnt sich, diese Festlegung auch an pragmatischen Gesichtspunkten (Adressenbeschaffung, Kostenaufwand) zu messen (z.B. bedeutet eine gesamtschweizerische Befragung, dass der Fragebogen ins Französische und Italienische übersetzt werden muss; eine heterogene Gruppe erfordert eine höhere Zahl an Interviews; Personen im Gefängnis oder Spital sind oft nicht erreichbar usw.).

2. Auflösung der Forschungsfragen in ein Modell operationalisierbarer Variablen (= Variablenmodell). D.h.: Die Frage "Was wollen wir wissen?" muss umgesetzt werden in: "Wie können wir das erfassen, was wir wissen wollen?" Dazu gehört, dass die Grenzen der Befragungstechnik erkannt werden. Erfragen lässt sich grundsätzlich nur, was die Befragten auch effektiv wissen oder was sie zugeben wollen. .

Eine detaillierte Operationalisierungsliste dient zur Disziplinierung der Arbeit der Fragebogenkonstruktion. Die einzelnen Fragen erhalten somit einen klaren theoretischen Stellenwert.

2.1. Aufarbeitung der bestehenden Forschungsliteratur zum gewählten Thema (soweit nicht schon vorher daran gearbeitet wurde). Das "Kolumbus-Syndrom" (=Glaube, man sei die Ersten, die dieses Thema erforschen) sollte vermieden werden. Heute sind in dieser Phase internationale Kontakte unerlässlich. Hilfreich ist oft auch die Berücksichtigung bisheriger Befragungen (und Fragebogen). Dies erlaubt einen Vergleich z.B. mit früheren Perioden. Auch eine Sekundäranalyse einer schon vorliegenden Befragung kann sinnvoll sein (und für Sekundäranalysen stehen heute aufgearbeitete Datenbanken zur Verfügung).

Teilweise zeigt sich, dass eine eigene (kostspielige) Befragung gar nicht notwendig ist, da die Forschungsfrage international schon beantwortet ist.

2.2. Exploration des Forschungsgegenstandes, um das eigene Primärwissen zu erweitern. Dazu können Gespräche bzw. Interviews mit Experten, aber auch Gruppendiskussionen dienen (z.B. wer das Thema "beruflicher Wiedereinstieg von Frauen" untersucht, tut gut daran, Kontakt mit entsprechenden Frauengruppen zu suchen). Auch Beobachtungen oder die Analyse von Texten kann explorativ wertvoll sein. Vor allem wird ein zu abstraktes Vorgehen vermieden. Befragte sind meistens keine Studenten oder Intellektuelle, und eine Befragung muss sich auf dem Niveau der Alltagssprache und des Alltagsverständnis der Befragten bewegen.

Je weniger Forschungserfahrungen zum Thema, desto mehr Zeit muss für Phase 2 eingesetzt werden.

3. Revision des Variablenmodells (und vielleicht auch einiger Forschungsfragen) aufgrund der in Phase 2 gewonnenen Erkenntnisse. Ein ständiges Überarbeiten und Überdenken bisheriger Theorien oder Vorstellungen gehört zu jedem Forschungsprozess (Prinzip: learning by doing, trial-and-error). Oft muss das Forschungsthema eingegrenzt bzw. stärker fokussiert

werden. Teilweise muss auch die Stichprobe präzisiert werden (z.B. anstatt alle Frauen im gebärfähigen Alter 15-49 Jahren nur Frauen zwischen 20 bis 35 J., anstatt Süchtige allgemein spezifisch Alkoholiker, anstatt Schüler allgemein zwei Gruppen bilden (Gewerbeschüler/Mittelschüler) usw.).

4. Vorbereitung der Befragung: Die Vorbereitungsphase einer Befragung umfasst mindestens zwei verschiedene Schritte, die teilweise parallel laufen. Dazu gehören namentlich:

4.1. Fragebogenkonstruktion: Frageformulierung und Aufbau des Fragebogens. Durch eine vorherige Operationalisierungsliste (Phase 2) wird dies erleichtert. Allerdings muss beachtet werden, dass viele gute Fragen noch keinen guten Fragebogen garantieren. Auch dem Aufbau (Ablauf und Rhythmus der Befragung, keine Eintönigkeit usw.) muss genügend Aufmerksamkeit gewidmet werden. Je nach Form der Befragung (mündlich, schriftlich, telefonisch) ist Ablauf und Frageform anders zu gestalten.

Oft können Fragen können auch aus früheren Befragungen übernommen werden (gerade was sozio-demografische Standardfragen betrifft). Wichtig ist: Fragebogenkonstruktion ist nie Einzelarbeit, sondern sinnvollerweise immer Teamarbeit (Auf Frageformulierung und Fragebogenkonstruktion wird später noch eingehender eingegangen).

4.2. Festlegung der Stichprobenstrategie: Wie viele Personen sollen befragt werden? Wie sollen die Leute ausgewählt werden, die befragt werden sollen? Wichtig: Nur eine Zufallsauswahl garantiert Repräsentativität. Die genaue Samplingstrategie muss festgelegt werden (einstufiges/mehrstufiges Sampling, Repräsentativumfrage, Quotenumfrage). Auch die Adressenbeschaffung muss vorbereitet werden. Die Adressen müssen zwar oft möglichst aktuell sein, aber die Vorbereitung braucht oft viel Zeit (z.B. Anfragen an Gemeinden, ob sie Adressen liefern). Aus Datenschutz-Gründen ist bei der Beschaffung von Adressen aus Einwohnerkarteien oft eine offizielle Genehmigung notwendig.

5. Pretest: In dieser Phase wird der erste Fragebogenentwurf getestet und wenn nötig revidiert. Z.B. werden 10, 30 oder auch mehr Interviews durchgeführt, um zu sehen, wie der Fragebogen funktioniert. Dabei ist es sinnvoll, wenn alle am Forschungsprojekt beteiligten Forscher selbst auch Interviews unternehmen oder die Interviews via Tonbänder aufgezeichnet werden. Durch den Pretest wird geprüft, ob und welche Fragen auf Schwierigkeiten stossen (z.B. Verständnisprobleme, Missverständnisse, falsche Filterfragen usw.) und wie ein Interview abläuft (Dauert es zu lange, ist es zu eintönig, gibt es Brüche). Durch einen guten Pretest lässt sich auch feststellen, ob die verwendeten Fragebatterien befriedigend eindimensional sind (d.h. ob das gemessen wird, was man sich erhofft).

Je neuer und unerforschter das Forschungsthema ist, desto wichtiger ist ein Vortest, und in allen Fällen gibt der Pretest Anlass zu wichtigen Revisionen und Verbesserungen des Fragebogens (wozu auch eine Straffung und Kürzung gehört).

6. Organisation und Durchführung der Befragung: Dazu gehört selbstverständlich ein graphisch sauberes Lay-Out - bei dem z.B. Fragen und Interviewer-Anweisungen optisch klar abgegrenzt sind - und der Druck der Fragebögen. Daneben müssen bei einer mündlichen Befragung die Interviewer/innen organisiert sein (sofern man die Durchführung der Befragung nicht einem Umfrage-Institut überlässt). Auch müssen die Adressen an die jeweiligen Interviewer/innen verteilt werden und Anweisungen zur Durchführung der Befragung weitergeben ist. Sinnvoll ist auch eine klare Interviewer-Schulung am spezifischen Fragebogen (empfohlen wird ein zweitägiger Kurs).

Eine grössere mündliche Befragung dauert im Allgemeinen zwei bis drei Monate (teilweise auch länger). Bei schriftlichen Befragungen muss nach etwa drei Wochen meist auch eine zweite Aufforderung verschickt werden. Telefonische Interviews können hingegen innerhalb von ein bis zwei Wochen organisiert werden.

Wichtig ist auch, dass die Befragung terminlich günstig vorgeplant wird (Frühjahr und Herbst sind besser als Sommer, wenn manche Leute ferienhalber verreisen).

Zur Feldarbeit gehört auch eine Nachkontrolle der eingesandten Fragebogen (auf richtiges Kodieren) sowie eine stichprobenweise Überprüfung (z.B. Nachfragen ob Befragung tatsächlich stattfand).

Falls sich Probleme mit zu grossen Ausfallsquoten und Verweigerungen ergeben, muss unter Umständen gezielt "nachgehakt" werden (z.B. Kurzbefragung von Verweigern).

7. Aufbereitung der gewonnenen Daten: Sobald die Fragebögen eintreffen, werden die Daten aufbereitet (d.h. in den Computer eingelesen). Ein gut vorbereiteter Fragebogen sollte so aufgebaut sein, dass das Einlesen der Fragecodes keine Probleme aufwirft. Oft ist es sinnvoll, zuerst die geschlossenen Fragen einzulesen, und die offenen Fragen erst später zu kodieren (anhand eines Codeplans). Elektronische Formen einer Befragung - z.B. computerunterstützte Telefonerhebung CATI, Interview mit Laptop - erleichtern die Datenaufbereitung.

8. Auswertung der Daten: Als erstes wird eine einfache Randauszählung erstellt (auch um Codierungsfehler zu finden und zu korrigieren. Empfehlenswert sind auch eine Kontrolle der Repräsentativität (etwa durch Vergleich mit statistischen Daten).. Danach werden häufig Fragebatterien ausgewertet und Skalen gebildet bzw. getestet, mit denen weiter gearbeitet wird. Die Ermittlung von Zusammenhängen zwischen interessierenden Variablen, etwa auf der Basis des theoretischen Modells, geschieht mit Hilfe der verschiedenen heute verfügbaren multivariaten Auswertungsverfahren (multiple Regression und Pfadmodelle, Varianzanalyse, Clusteranalyse, Loglineare Modelle usw.). Je nach Thema ist es sinnvoll, die Fragebogendaten durch weitere Daten (statistische Daten, Dokumente usw.) zu ergänzen.

9. Forschungsbericht: Der wissenschaftliche Bericht über die wichtigsten Ergebnisse bildet den Abschluss der Forschungsarbeit. Je nach Auftraggeber werden mehr die wissenschaftlichen oder mehr die politisch-planerischen Konsequenzen ausgearbeitet. Häufig werden die Forschung in Rahmen von Zeitschriftenartikeln oder Buchpublikation veröffentlicht (wobei primär 'peer-reviewed-Beiträge karrierefördernd sind). Bei angewandten Forschungsprojekten - etwa im Rahmen Nationaler Forschungsprogramme - wird auch eine praxisorientierte Umsetzung zentraler Forschungsergebnisse angestrebt.

Zu den Frageformen

Zum Messniveau von Fragen:

Schon bei der Vorbereitung des Fragebogens sollte man sich klar sein, über das gewünschte bzw. mögliche Messniveau zumindest der wichtigsten Fragen. Denn vom Messniveau hängt Art und Weise der späteren Auswertung ab. Häufig ist allerdings das Messniveau von vornherein gegeben:

1. Nominale Fragen: nominales bzw. kategorielles Messniveau (z.B. Geschlecht: männlich/weiblich, Zivilstand: ledig, verheiratet, geschieden, verwitwet). Nominale Fragen können allerdings - umgewandelt in 'dummy'-Variablen auch für multivariate Auswertungen eingesetzt werden.

2. Ordinale Fragen: Fragen, bei denen die Antworten rangiert werden können (z.B. immer, oft, selten, nie; sehr einverstanden, einverstanden, nicht einverstanden, gar nicht einverstanden usw.)

Die meisten Einstellungs- und Bewertungsfragen innerhalb eines Fragebogens haben ordinalen Charakter. Damit fallen manche Auswertungsverfahren, die auf (normalverteilten) quantitativen Daten beruhen, weg. Es gibt allerdings diverse methodische "Kniffe", um auch ordinale Daten zu Intervallvariablen umzuwandeln.

3. Quantitative Fragen: Fragen, die auf einer Intervallskala oder Ratioskala messen. Dazu gehören etwa Fragen nach dem Alter (in Jahren), nach der wöchentlichen Arbeitszeit oder nach dem Einkommen u.a. Teilweise lassen sich auch Einstellungen als Intervallskala konzipieren (z.B. Thermometer-Fragen).

Offene Fragen versus geschlossene Fragen

Strukturierte Interviews können offene oder geschlossene Fragen einschliessen. Offenheit bzw. Geschlossenheit einer Frage bezeichnet den Spielraum, der dem Befragten gelassen wird. Bei offenen Fragen unterliegt der Antwortende keinen Beschränkungen. Inhalt, Form, Spezifität und Ausführlichkeit der Antwort liegen in seinem Ermessen. Bei offenen Fragen ist der Spielraum für Interviewereinflüsse grösser. Bei geschlossenen Fragen ist der Spielraum der Antwortmöglichkeiten durch vorgegebene Antwortkategorien mehr oder minder stark eingeschränkt. Beispiel: Flüchtlingspolitik: Offene Frage: "Was halten Sie von der jetzigen Flüchtlingspolitik?" (Befragte antworten frei und Interviewer notiert sich Antwort). Geschlossene Frage: "Was halten Sie von der jetzigen Flüchtlingspolitik? Finden Sie diese zu engherzig, gerade richtig oder zu grosszügig? Antwortkategorien: zu engherzig, gerade richtig, zu grosszügig, keine Antwort.

Für die Benützung geschlossener Fragen sprechen technische Vorzüge: schnelle und genaue Antwortprotokollierung und -Kodierung. Geschlossene Fragen erbringen eine grössere Einheitlichkeit und Eindeutigkeit der Antworten, und sie erleichtern die Vergleichbarkeit. Geschlossene Fragen sind besser und einfacher auszuwerten. Der Nachteil liegt darin, dass der Reichtum an Antworten verkürzt wird. Zudem wird nicht klar, ob der Befragte die Frage auch richtig verstanden hat.

Bei offenen Fragen werden die spezifischen Ansichten und Deutungen der Befragten (und vor allem auch ihre Wortwahl und ihr Bezugsrahmen) genauer erfasst. Die technischen Nachteile sind jedoch erheblich (höherer Protokoll- und Auswertungsaufwand, funktioniert nur bei kompetenten Interviewern). Offene Fragen stellen zudem höhere Anforderungen an das Ausdrucksvermögen des Befragten. Befragte mit geringer Schulbildung und aus unteren sozialen Schichten sind häufiger überfordert (damit ergibt sich bei offenen Fragen meist ein "Mittelschicht-Bias").

Empfohlen wird daher bei mündlichen wie schriftlichen Interviews primär geschlossene Fragen zu stellen. Generell sollte die Strategie verfolgt werden, für die Zieldimension, die man messen will, quantitative Fragen zu formulieren. Offene Fragen sollten höchstens ergänzend gestellt werden, z.B. wenn es um noch nicht genügend vorab geklärte Sachverhalte geht. Of-

fene Fragen können aber auch im Anschluss an geschlossene Fragen gestellt werden, um nachzuhaken oder die Gründe für ein Ja/Nein etwas genauer zu erfassen (Beispiel: Hat Ihr Vater Sie bei der Wahl Ihres Studienfaches beeinflusst oder nicht? Antwortkategorien: ja, nein. Wenn ja: In welcher Weise hat er sie beeinflusst?).

Offene Fragen stehen allerdings bei narrativen (=erzählenden) Interviews, aber auch bei Expertenbefragungen im Zentrum. Auch in Pretests sind offene Fragen nützlich, um für später die Antwortkategorien zu erfassen.

Frageformen

Es bestehen verschiedene Arten von Fragen, wobei wir vereinfacht von folgenden, sich teilweise überlappenden Frageformen ausgehen können:

1. Faktfragen: Bei Faktfragen erkunden wir objektive Tatbestände, die den Befragten, seine Familie, seinen Betrieb, seine Umwelt usw. betreffen. Wichtig ist allerdings, dass man nur Faktfragen stellen kann, die Befragte im allgemeinen auch wissen. So sind manche Befragte bei Detailfragen zu wirtschaftlichen und sozialen Lebensbereichen (Haushaltsausgaben, erhaltene oder gegebene soziale Unterstützung sowie zu Verhalten wie Konsum- oder Freizeitverhalten) überfordert. Entsprechend sind die entsprechenden Angaben häufig ungenau und ungenau.

1.1. Sozio-demografische Daten: In allen Meinungs- und Bevölkerungsumfragen werden einige wichtige sozio-demografische Hintergrunddaten erfasst. Häufig werden diese Fragen erst am Schluss des Fragebogens gestellt. Sie sind meist unproblematisch, mit Ausnahme des Einkommens (hohe Verweigerungsquote speziell in der Schweiz).

Wichtige sozio-demografische Hintergrundvariablen sind etwa:

Alter: am besten nach genauem Geburtsdatum fragen (Monat, Jahr)

Geschlecht: männlich/weiblich

Zivilstand: klassisch: ledig, verheiratet, geschieden, verwitwet.

heute wird vermehrt zwischen "ledig/allein lebend, ledig/mit Partner(in) wohnend" unterschieden. Um die Lebenssituation zu erfassen, kann zusätzlich zum Zivilstand die Frage nach der Haushaltsform bzw. Partnerschaft gestellt werden ("Leben Sie mit einem Partner/einer Partnerin zusammen?")

Haushaltsgrösse: "Wie viele Personen leben im Haushalt, ausser Ihnen?", eventuell: "Wer lebt mit Ihnen im gleichen Haushalt?" (notieren der Personen (Ehemann, Freund, Tante, Kinder usw.).

Nationalität: zunehmend wichtig bei multi-kulturellen Gesellschaften. Allerdings sagt die formelle Nationalität oft wenig aus. Eventuell ergänzen durch Frage nach Geburtsland, Dauer des Aufenthalts in der Schweiz (Seit wann leben Sie in der Schweiz?)

Erwerbstätigkeit: "Sind Sie momentan berufstätig oder nicht?" (nein, arbeitslos, nein, in Ausbildung, nein, bin Hausfrau/ Hausmann, nein, pensioniert, Ja), Wenn ja: "Arbeiten Sie ganztags, halbtags oder stundenweise". Oder nach Wochenarbeitszeit fragen.

Problem: Definition, was als Berufs- bzw. Erwerbstätigkeit gilt, wird meist den Befragten überlassen. (ev. präzisieren: Berufstätig= bezahlte Arbeiten).

Ausbildung: oft erfasst durch "zuletzt besuchte Schule" oder durch Schul- bzw. Bildungsabschluss. Problem: Schulkategorien variieren je nach Land bzw. je nach Kanton. Wichtig: Berufslehre miteinfassen.

Beruf: oft offene Frage: "In welchem Beruf sind Sie momentan tätig?" (bei allgemeinen Antworten nachhaken). Bei momentan nicht berufstätigen Befragten nach zuletzt ausgeübtem Beruf fragen.

Eventuell Nachfragen nach Stellung im Beruf ("Nehmen Sie in Ihrem Beruf eine leitende Stellung ein?"/ "Arbeiten Sie selbständig erwerbend oder sind Sie angestellt?")

Einkommen: z.B. "Wie hoch ist Ihr persönliches Monatseinkommen?" bzw. "In Ihrem Beruf: Wie viel Franken verdienen Sie im Monat?" Problem: Befragte können meist nicht zwischen

Brutto- und Netto-Einkommen unterscheiden. Je nach Studie müssen unterschiedliche Einkommensstypen erfasst werden (Erwerbseinkommen, Haushaltseinkommen, Vermögenseinkommen usw.). Zahl der Antwortverweigerungen hier relativ hoch (ev. Liste mit Einkommenskategorien vorgeben). Teilweise wird deshalb die subjektive finanzielle Lage erfragt (Wie beurteilen Sie Ihre gegenwärtige finanzielle Situation: gut, mittelmässig, schlecht?)

Wohnform: erfasst wird häufig: Zahl an Zimmern, Eigentum/Miete, Einfamilienhaus/ Mehrfamilienhaus. Eventuell auch Eigentumsstatus (Mieter, Eigentümer).

Wohnort (sofern nicht von Adresse sichtbar), eventuell auch: Wohnort vor 5 Jahren.

Generell zu empfehlen: Soweit möglich: Übernahme analoger Fragen und Kategorien wie bei der Volkszählung bzw. den Mikrozensus des Bundesamtes für Statistik. Erlaubt eine Kontrolle von Verzerrungen bei der Befragung aufgrund von Ausfällen oder Verweigerungen und ermöglicht Vergleiche zwischen der befragten Population mit gesamtschweizerischen Daten.

1.2. Retrospektive Fragen: Beziehen sich auf früheres Verhalten und frühere Situationen (z.B. frühere Berufsstellungen, Wohnort während Kindheit, Beruf des Vaters usw.). Soweit es markante biographische Daten (Heirat, Berufswechsel usw.) betrifft, sind Erinnerungsfragen wenig problematisch. Aber je länger etwas zurückliegt, desto eher wird vergessen oder falsch datiert (z.B. Erkrankungen und Spitalaufenthalte, die längere Zeit zurückliegen, bleiben unerwähnt). Auch können sich bei retrospektiven Fragen Probleme mit dem Zeithorizont von Befragten ergeben (problem of telescoping). Sinnlos sind allgemeine retrospektive Fragen, wie: "Wann haben Sie das erste Mal von X gehört?", "Wie oft haben Sie letztes Jahr ferngesehen?", "Wie viele Stunden pro Woche haben Sie letztes Semester für Ihr Studium aufgewendet?" usw.

Es ist nahe liegend anzunehmen, dass die Gültigkeit der Antworten zu retrospektiven Fragen sowohl vom erfragten Zeithorizont als auch von der Art der Ereignisse bestimmt wird. Allerdings zeigen entsprechende Studien, dass die Gültigkeit der Antworten nicht einfach linear mit dem Zeithorizont reduziert wird, sondern dass der erinnerte Zeithorizont eng mit der Art von Ereignissen bzw. Aktivitäten verknüpft sind. Je nach Art der retrospektiver Frage können auch weit zurückliegende Ereignisse zuverlässig erinnert und datiert werden, wogegen bei anderen Fragen schon nach kurzer Zeit (z.B. zwei oder drei Wochen) Erinnerungsprobleme auftauchen. Ein sehr bedeutsamer Faktor ist namentlich die Zentralität (saliency) eines Ereignisses, wobei die Zentralität eines Ereignisses mit mindestens drei Faktoren assoziiert ist:

- die emotionale Intensität eines Ereignisses: Emotional intensive Ereignisse und Erlebnisse - etwa aus der Kindheit oder der Jugendzeit - werden weniger vergessen als emotional weniger intensive Ereignisse oder Handlungen. Andererseits werden emotional intensive Ereignisse und Erlebnisse unter Umständen - sofern es sich um negativ erlebte Ereignisse (Unfall, Gewaltakte, Arbeitslosigkeit usw.) handelt - eher verdrängt. Alltägliche oder relativ häufig vorkommende Ereignisse oder Handlungen - wie Urlaubsreisen, Konsumverhalten, Medienkonsum usw. - sind hingegen schon nach recht kurzer Zeit nicht mehr differenziert abfragbar.

- das Ausmass, in dem ein Ereignis einen zentralen biographischen Wendepunkt oder lebenszyklische Weichenstellung darstellt. Solche Ereignisse (wie Schuleintritt, Heirat, Geburt von Kindern, Berufseintritt, Pensionierung usw.), werden im allgemeinen relativ zuverlässig erinnert und datiert (wenn auch die damaligen Umstände oder Motive in der Rückerinnerung beschönigt und idealisiert werden). Auch wichtige lebenskritische Ereignisse (Scheidung, Tod eines Partners) werden im allgemeinen zuverlässig erinnert (auch wenn dafür teilweise die Antwortbereitschaft reduziert wird).

- der finanzielle und soziale Impakt einer Handlung. So werden grosse Konsumeinkäufe besser erinnert als kleinere Einkäufe. Ebenso werden Handlungen mit weittragenden Wirkungen - z.B. besonders ereignisreiche Urlaubsreisen, weit reichende Stellenwechsel usw. - eher erinnert.

Je alltäglicher und routinemaßiger ein Ereignis, ein Erlebnis oder eine Handlung ist, desto kürzer ist allerdings im allgemeinen der Zeithorizont, der retrospektiv zuverlässig erfasst werden kann.

Die Beantwortung von Erinnerungsfragen lässt sich teilweise durch gezielte Frage-Strategien" verbessern (z.B. Verwendung von biographischen Markierungen und Verknüpfung mit anderen Ereignissen). Bei Fragen, die biographische Aspekte und Fakten detailliert zu erfassen suchen - etwa im Rahmen von Lebenslaufanalysen - muss das Interview so gestaltet werden, dass

- den Befragten und dabei auch älteren Personen genügend Zeit zur Rückerinnerung gelassen wird. Das Tempo des Interviews - oft ganz allgemein zu schnell - muss bei solchen Fragen gedrosselt werden.

- den Befragten explizit die Möglichkeit eröffnet und gelassen wird, im Verlaufe des Interviews auf frühere Erinnerungsfragen zurückzukommen, sei es um Daten zu korrigieren oder Ergänzungen anzubringen.

- assoziative Querverbindungen (cross-cutting) zwischen Erinnerungsfragen vorgesehen werden, um via assoziatives Verfahren die Qualität der Rückerinnerung zu vergrößern, Vgl. etwa: Auriat, Nadia (1991) 'Who Forgets?' An Analysis of Memory Effects in a Retrospective Survey of Migrations History, *European Journal of Population*, 7,4: 311-342; Freedman, David, Thornton, Arland; Camburn, Donald, Alwin; Duane; De-Marco, L.Young (1988) The Life History Calendar: A technique for collecting retrospective data, in: C.C.Clogg (ed.) *Sociological Methodology*, Pp. 37-68, San Francisco: Jossey-Bass. Robinson, John A. (1986) Temporal Reference Systems and Autobiographical Memory, in: David C. Rubin (ed.) *Autobiographical Memory*, Pp. 159-188, Cambridge: University Press; Thompson, Charles P.; Skowronski, John J.; Lee, John D. (1991) Telescoping in Dating Naturally Occurring Events, *Memory and Cognition*, 16: 461-468.

1.2. Heikle Fragen: Verschiedene Fragen stoßen auf Antworthemmungen, z.B. weil sie sehr private Aspekte oder Verbotenes ansprechen (z.B. Fragen zur Sexualität, Vergewaltigungen, Abtreibung, Gewalttätigkeit, Drogenkonsum, Diebstahl, Schwarzarbeit). Entsprechende Studien zeigen, dass auf heikle Fragen entweder nicht oder beschönigend geantwortet wird (Problem des "underreporting"). Eine Strategie liegt in der Verharmlosung der Frage ("Die meisten machen heute irgendwie Schwarzarbeit. Leisten Sie selbst auch Schwarzarbeit?"). Eine andere Strategie liegt darin, dass die Antwort in einem verschlossenen Couvert übergeben wird (so dass der Interviewer die Antwort nicht kennt). Eine weitere Strategie ist die sogenannte "randomized response technique": Nach Zufallsprinzip muss der Befragte entweder eine harmlose Frage ("Sind Sie im Sommer geboren?") oder die heikle Frage ("Haben Sie schon jemals Heroin konsumiert?") beantworten, ohne dass der Interviewer weiß, ob die harmlose oder die heikle Frage gestellt wurde. Zu Frageformen bei heiklen Themen vgl. James A. Fox, Paul E. Tracy, *Randomized Response. A Method für Sensitive Surveys*, Beverly Hills, Sage 1986.

1.3. Filterfragen: Handelt sich um eine Faktfrage, die das weitere Vorgehen des Interviewers bestimmt. Zum Beispiel werden einige Fragen nur erwerbstätigen Befragten gestellt. Mittels der Filterfrage werden die nicht-erwerbstätigen Befragten ausgesondert. Filterfragen zu Beginn des Fragebogens können auch sinnvoll sein, um zu überprüfen, ob die betreffende Person überhaupt zur Untersuchungsgruppe gehört, etwa bei telefonischen Suchinterviews bzw. bei Interviews spezieller sozialer Gruppen (Alleinlebende). Filterfragen werden namentlich bei telefonischen Umfragen häufig eingesetzt, wogegen sie bei schriftlichen Fragebogen auf enge Grenzen stoßen.

1.4. Handlungsfragen: Bei Handlungsfragen - als einer speziellen Untergruppe von Faktfragen - wird nach dem (aktuellen, früheren oder zukünftigen Verhalten des Befragten nachge-

sucht: "Wohin fahren Sie im letzten Urlaub?", "Wie oft hören Sie DRS1?", usw. Wichtig bei Handlungsfragen ist ein klar definierter Zeithorizont: "letzter Urlaub" kann dieses Jahr, letztes Jahr oder auch vor 10 Jahren sein. Besser wäre: "Fahren Sie 2004 in den Urlaub? Wenn ja, wohin?", "Und 2003: Fahren Sie dann in den Urlaub, und wenn ja wohin? Auch Fragen nach der Häufigkeit eines Tun müssen sorgfältig formuliert werden. "Wie oft hören Sie DRS1?" kann je unterschiedliches bedeuten: jeden Tag stundenlang, jeden Tag einmal, öfters als DRS2 etc. Die Frage etwa nach der Häufigkeit der Mithilfe im Haushalt führte bei Ehemännern zu signifikant anderen Antworten als bei den gleichzeitig befragten Ehefrauen.

Bei vergangenem Handeln stellen sich die üblichen Probleme retrospektiver Fragen (Erinnerungslücken, Zeithorizont wird verwechselt usw.). Bei Fragen nach zukünftigem Verhalten kommt Verhaltensunsicherheit dazu. Wunsch und Wirklichkeit fallen oft auseinander. So zeigen Längsschnittstudien, dass erwartete Kinderzahl und tatsächliche Kinderzahl häufig divergieren.

Auch Fragen zu "hypothetischem Handeln" sind mit Vorsicht zu interpretieren (Hypothetische Fragen geben hypothetische Antworten). z.B. "Was würden Sie tun, wenn Sie 1000 Franken mehr verdienen?" "Würden Sie das Produkt X kaufen, wenn es erhältlich wäre?" usw. Im allgemeinen bringen hypothetische Fragen wenig.

Das gleiche gilt für Fragen nach den Motivationen bzw. Gründen für ein bestimmtes Verhalten (Warum?-Fragen). Gründe/Motivationen können selten durch direkte Fragen erfasst werden. Zudem können nur die interpretierten, rationalisierten und bewussten Gründe via Fragebogen erfasst werden.

2. Wissensfragen: Bei Wissensfragen wird ermittelt, ob Befragte bestimmte nachprüfbare Fakten weiss (z.B. um Informationsgrad der Bevölkerung zu prüfen). Durch entsprechende Batterien von Wissensfragen kann etwa ein Indikator politischer Information konstruiert werden. Wissensfragen können auch dazu dienen, zu erfahren, ob und in welcher Form staatliche Kampagnen (z.B. zu AIDS, zur Altersversorgung) überhaupt rezipiert worden sind. Bei Wissensfragen ist zu achten, dass das Interview nicht zu stark "Prüfungscharakter" aufweist. Um das Gefühl von Versagen zu vermeiden, werden teilweise auch Wissensfragen gestellt, die fast alle Befragten beantworten können. Oft stellt sich heraus, dass der Informationsstand von Befragten geringer ist als die Forscher erwarteten.

3. Einschätzungsfragen: Hier wird nach der Einschätzung einer relativ ungewissen Sachlage gefragt, z.B. "Welche Partei wird Ihrer Ansicht nach die nächste Wahl gewinnen?", "Welches Droge ist Ihrer Meinung nach gefährlicher: Heroin oder Haschisch?" usw. Einschätzungsfragen sind meistens analytisch von eher geringem Wert.

4. Einstellungsfragen: (=Meinungsfragen) Es geht um die Einstellung bzw. Meinung/ Ansicht zu bestimmten Sachverhalten (z.B. für oder gegen die Abschaffung der Armee, für oder gegen die Todesstrafe). Formelle Definition: Einstellung ist die affektuelle Beziehung des Befragten zu einem materiellen, personellen, sozialen, idealen Umweltobjekt (z.B. zu einem Gegenstand, einer Person, einer Gruppe, einer Ideologie). Einstellungsfragen stehen häufig im Zentrum der politischen oder sozialwissenschaftlichen Meinungsforschung. Erfragt wird etwa die Einstellung (oder Identifikation) mit politischen Parteien, Verbänden, aktuellen Sachfragen (Umweltschutzmassnahmen, Gleichberechtigung usw.). Die Annahme, dass Leute überhaupt eine Meinung bzw. Einstellung haben, ist allerdings teilweise fraglich. Zudem sind bei Einstellungs- und Bewertungsfragen die genaue Frageformulierung und Interviewsituation besonders bedeutsam. "Ja-Sage-Tendenz" ist hier besonders wirksam, und bei manchen politischen Meinungsumfragen werden Fragen so gestellt, dass das erwünschte Antwortverhalten resultiert (1. Frage: "Sie sind für oder gegen eine Einschränkung der persönlichen Freiheit? 2. anschliessende Frage: "Welches ist Ihre Meinung zur Einschränkung der Tempo-Freiheit auf Autobahnen?"

Bei wissenschaftlichen Umfragen ist es meist sinnvoll, Einstellungen durch mehrere Fragen - bzw. durch ganze Fragebatterien - zu erfassen. Die gleiche Frage kann wiederholt werden (z.B. zu Beginn des Fragebogens, und am Schluss), um Inkonsistenzen zu eruieren.

4.1 Bewertungsfragen: Bei diesen Einstellungsfragen geht es um die Bewertung - z.B. im Sinne etwa von "gut/schlecht" - etwa von Parteien, Konsumgüter usw. Zum Beispiel: "Wie beurteilen Sie die Politik der folgenden Parteien? Geben Sie jeder eine Note von 6 (sehr gut) bis 1 (miserabel)", "Wie zufrieden bzw. unzufrieden sind Sie mit Ihrem Vorgesetzten? usw. Zu beachten ist, dass die Bewertung einer Angelegenheit oder Gruppe von verschiedenen Dimensionen abhängt (Identifikation, Interesse, Information, aktuelle Stimmung). Die Interpretation von Bewertungsfragen ist deshalb nicht einfach. So kann Zufriedenheit unterschiedliche Bedeutung haben (aktive Zustimmung oder resignative Anpassung). Befragte haben oft die Tendenz, zu positiver Bewertung (weil negative Bewertung sozial mehr Intensität voraussetzt).

Ganz allgemein ist zu beachten, dass der Begriff der Einstellung (engl.: attitudes) mehrere Dimensionen umfasst. "Einstellungen" umfassen beispielsweise die Dimension "positive versus negativ Einschätzung/Beurteilung" und die Dimension der Intensität der Einstellung (emotional-affektiver Aspekt). Beispiel EU-Beitritt der Schweiz: einige Befragte sind für oder gegen einen EU-Beitritt, ohne emotional engagiert zu sein (z.B. weil das Thema ihnen persönlich gleichgültig oder fremd ist). Andere Befragte sind intensiv und engagiert für oder gegen einen EU-Beitritt. Sinnvoll ist daher oft ein Nachfragen: 1.Frage: "Wie stellen Sie sich zum Beitritt der Schweiz zur EU. Sind Sie dafür oder dagegen?" 2. Frage: "Wie wichtig bzw. unwichtig ist Ihnen die ganze Diskussion um einen EU-Beitritt? (oder spezifischer: "Würden Sie an einer Kundgebung für/gegen einen EU-Beitritt teilnehmen?")

Attitudes: Five different dimensions of attitudes

We will focus on just five of many dimensions of strength suggested, because this five possess three great virtues: "they are the easiest to measure in surveys, they are the easiest to comprehend conceptually, and they are the most extensively validated as measures of the fixedness and consequentiality of attitudes."

1) Extremity: "Attitude extremity is the degree of favorableness or unfavorableness of an individual's evaluation of a given object. The more extreme an individual's attitude is, the further it is from neutrality. Therefore, attitude extremity has typically been operationalized as the deviation of an individual's attitude rating from the midpoint of a pro-contra dimension.

2) Intensity: "Some attitudes involve strong, affective responses to objects, whereas other attitudes involve little or no emotional reaction. Attitude intensity is defined as the strength of an individual's feelings about an attitude object. Typically, it has been measured by asking people how strong or intense their feelings are toward a particular object.

3) Certainty: "Attitude certainty refers to the degree to which an individual is certain that his or her attitude toward an object is correct. It has usually been measured by asking respondents how sure they are of their opinion, how easily their opinion could be changed, or how confident they are that their opinion is correct. Other studies have asked respondents how difficult they found it to report an opinion on an issue.

4) Importance: "Attitude importance has been defined as the degree to which an individual considers an attitude to be personally important to him or her. Attitude importance has generally been measured by asking people how personally important their attitude or the attitude object is, how concerned they are about it, or how much they care about it.

5) Knowledge: "Some attitudes are accompanied in memory by relatively little information about the attitude object, whereas other attitudes are linked to large stores of beliefs about the object. Amount of attitude-relevant knowledge has usually been measured by asking respondents to list everything they know about an attitude object. This dimension has also been assessed simply by asking individuals to report how knowledgeable they feel they are about an issue.

Relations between these dimensions: A number of studies have documented low to moderate positive correlations between intensity and extremity, extremity and certainty, importance and extremity, importance and intensity etc. Thus, it seems that these various dimensions are generally empirically quite separable from one another, so they should be viewed as distinct constructs. This conclusion is further reinforced by a small body of evidence suggesting that these various dimensions of attitude strength each have unique causes.

Quelle: John A. Krosnick, Robert P. Abelson (1992) The Case for Measuring Attitude Strength in Surveys, in: Judith M. Tanur (ed.) Questions about Questions. Inquiries into the Cognitive Bases of Surveys Russell Sage Foundation, New York: 177-203.

Befragung: Wichtige Regeln der Fragebogen-Konstruktion

Die bisherigen Erfahrungen mit Umfragen, aber auch gezielte Studien zum Effekt verschiedener Frageformulierungen haben zu einer ganzen Reihe von Empfehlungen und Erfahrungsregeln bei der Formulierung von Fragen bzw. Fragebogen geführt. Nachfolgend sind die wichtigsten Regeln bei der Konstruktion eines standardisierten schriftlichen, mündlichen, telefonischen oder elektronischen Fragebogens aufgelistet. Allerdings sind nicht alle Regeln unumstritten, namentlich was Aspekte der Operationalisierung und Antwortvorgaben betreffen. Die nachfolgend aufgeführten Regeln sind entsprechend situationsgerecht der jeweiligen Forschungsfrage und Erhebungsmethode anzupassen. Die nachfolgenden Regeln beziehen sich - wie erwähnt - explizit auf strukturierte Interviews bzw. standardisierte Befragungen. Bei qualitativen Befragungen (Expertengesprächen, narrative Interviews usw.) gelten andere Regeln.

1 Regeln der Operationalisierung von Zieldimensionen bei standardisierten Befragungen

1. Es sollte möglichst die Strategie verfolgt werden, für die Zieldimension (= den Forscher/die Forscherin interessierende Sachverhalte und Einstellungen) quantitative Fragen zu formulieren. Die Möglichkeiten, aber auch die Grenzen der Quantifizierbarkeit sind sorgfältig auszuloten.

2. Wenn nicht gewichtige Gründe dagegen sprechen, dann sollte eine Zieldimension so fein wie möglich gemessen werden, um den Messfehler klein zu halten.

Beispiel: Alter erfassen durch Erfragen des genauen Geburtsdatums. D.h. nicht "Wie alt sind Sie?", sondern: "Wann sind Sie geboren? (Jahr/Monat).

3. Die Feinheit, mit der wir messen wollen, muss allerdings mit dem Differenzierungsvermögen von Befragten korrespondieren. Speziell bei Fragen nach Häufigkeiten ausgewählter Tätigkeiten oder bei komplexen Sachverhalten ergeben sich oftmals enge Grenzen im Differenzierungsgrad. Die Frage "Wie viele Minuten pro Woche lesen Sie?" kann nicht gestellt werden. Eben so wenig wissen die meisten Leute über ihre Wohnungsgröße in Quadratmeter Bescheid.

4. Um die Chance zu erhöhen, dass man auf Fragen Antworten mit befriedigender Zuverlässigkeit und Gültigkeit erhält, sollte man für eine Zieldimension verschiedene Fragen formulieren. Gilt insbesondere bei Einstellungen, die im Zentrum der Untersuchung stehen. Dafür werden häufig ganze 'Frage-Batterien' sowie Skalen mit unterschiedlichen Items verwendet.

5. Je länger die Frageformulierung und je stärker der Reizwert der verwendeten Worte, umso grösser ist die Fremdbestimmtheit einer Frage.

Fremdbestimmtheit = Mass, in dem eine Frage (unerwünschten) Fremddimensionen mitefasst. Tatsächlich werden viele sozialen Sachverhalte (und ihre Erfassung) durch mehrere Dimensionen beeinflusst. Beispiel: 'Arbeitszufriedenheit' kann nicht nur von Arbeitssituation, sondern auch von Lebenssituation mitgeprägt werden. Antworten zur Arbeitszufriedenheit werden durch Zeitpunkt der Befragung beeinflusst (z.B. Antworten kurz vor oder kurz nach dem Mittagessen können unterschiedlich sind).

6. ForscherInnen sollen zu ermitteln versuchen, ob die Zieldimension bei Befragten verbalisiert und vorformuliert vorliegt. Ist dies nicht der Fall, muss die Zieldimension über verbalisierte Ersatzdimensionen gemessen werden.

Beispiel: Unsicherheit als Zieldimension liegt nicht vorformuliert vor. Sie muss indirekt, über Indikatoren wie "Zukunftsangst, Gefühl von Traditionsverlust usw. erfasst werden. Dasselbe gilt für viele sozial-psychologische Gefühlszustände (Ängstlichkeit, depressive Stimmungslage usw.).

7. Wenn man nicht gezwungen ist, indirekt zu fragen - weil die Zieldimension nicht verbalisiert ist oder weil Befragte sonst eine sozial erwünschte Antwort gibt - sollte man sich bemühen, so direkt wie möglich zu fragen. Dadurch wird die Fremdbestimmtheit (= Beeinflussung der Antwort durch "fremde" Faktoren) verringert.

8. Wenn hingegen Faktoren der sozialen Wünschbarkeit oder fehlender Vorverbalisierung bei direkter Frage zu bedeutsamen Verzerrungen führen, kann es sinnvoll sein, indirekte Fragen zu stellen, unter Umständen so, dass den befragten Personen die eigentliche Zieldimension unbekannt bleibt. (Beispiele: sozio-psychische Skalen zur Erfassung depressiver Stimmung, neurotizierender Persönlichkeitsmerkmale, Skalen zur Erfassung von Fremdenfeindlichkeit, Suchtgefährdung).

9. Da oft nicht klar ist, ob soziale Wünschbarkeit und/oder fehlende Verbalisierung zur indirekten Frageformulierung zwingen oder nicht, werden teilweise, um alle Eventualitäten zu berücksichtigen, zu wichtigen Zieldimensionen (namentlich Einstellungsdimensionen) direkte und - im Fragebogen an anderer Stelle - indirekte Fragen gestellt.

10. Abstrakte Zieldimensionen sollten vor Beginn der Fragebogen-Konstruktion in mehrere, spezifischere Zieldimensionen umgesetzt werden. Eine direkte, stufenlose Operationalisierung abstrakter soziologischer Konzepte in Fragen ist meist sinnlos. Operationalisierung ist ein mehrstufiger Prozess (und dazu ist das Instrument eines Indikatorenkatalogs sehr wertvoll).

11. Gemäss William Foddy (1993) ist bei der Konstruktion von Fragen vom 'TAP' Paradigma auszugehen:

- Topic: The topic should be properly defined so that each respondent clearly understands what is being talked about.
- Applicability: The applicability of the question to each respondent should be established: respondents should not be asked to give information that they do not have.
- Perspective: The perspective that respondents should adopt, when answering the question, should be specified so that each respondent gives the same kind of answer.

12. Sudman und Bradburn (1985) betonen, dass bei der Fragebogenkonstruktion die folgenden drei Arbeitsregeln eingehalten werden sollten:

- (a) Do not formulate specific questions until you have thought through the research question.
- (b) Write the research question down and keep it handy while formulating specific questions.
- (c) Keep asking, 'Why do I want to know this?' ('It would be interesting to know' is not an acceptable answer).

2. Regeln zur Frageformulierung

1. Suggestive Formulierungen sind unzulässig. Suggestive Wirkungen werden bei der Frageformulierung dadurch herbeigeführt, dass

- a) dem Befragten die Antwort "in den Mund gelegt" wird,
- b) Stereotype oder emotional geladene Begriffe verwendet werden,
- c) in die Frageformulierung Argumente aufgenommen werden.

Beispiele suggestiver Fragen: "Soll die erfolgreiche Umweltschutzpolitik der Regierung weitergeführt werden oder nicht?", "Eine grosse Mehrheit aller Bürger ist dafür, dass..., Welches ist Ihre Meinung zu...usw.; Beispiele von politischen Reizwörtern mit suggestiver Wirkung sind etwa: Freiheit, Solidarität, Ruhe und Ordnung u.a.).

2. Fragen sollten möglichst neutral formuliert werden, und Fragen sollten ausgewogene Antwortalternativen umfassen. Sofern Antwortvorgaben in die Frage eingebaut sind, sollten die Antwortalternativen gleichermaßen positive und negative Vorgaben enthalten.

Beispiel für unausgewogene Frage: "Sind Sie mit den Verhältnissen in der Schweiz im allgemeinen zufrieden oder wünschen Sie sich vieles grundlegend anders?" ("vieles grundlegend anders" ist als Kategorie extremer als Kategorie "im Allgemeinen". Oder Beispiel aus dem Tages-Anzeiger-Magazin 30/95: 'Wie oft haben Sie das Gefühl, die Politik von National- und Ständerat versage in wichtigen Dingen?')

Unausgewogen ist im Grunde auch eine Frage wie: Wie oft sehen Sie Fernsehen? u.ä.

3. Eine Frage sollte nicht zu allgemein formuliert sein (schon gar nicht im Stile eines Allgemeinplatzes). Sie muss noch so spezifisch und konkret formuliert sein, dass sie in den Befragten eine eindeutige Zieldimension anspricht. Geschieht dies nicht, dann wird eine Ja-Sage-Tendenz erweckt.

Konkrete Fragen sind meist besser als allgemeine Fragen. Dies gilt sowohl für Faktfragen als auch für Einstellungen.

Beispiel einer allgemeinen Frage, die zu nichts sagenden Antworten führt: "Wie ist Ihr Verhältnis zur Demokratie? Sie sind dafür oder dagegen?" Oder: Wie zufrieden bzw. unzufrieden sind Sie mit ihrem Leben? (Lebenszufriedenheit umfasst viele Aspekte, besser: Zufriedenheit mit Gesundheit, mit Arbeit, Partnerschaft, Finanzieller Lage, Wohnsituation gezielt erfassen).

4. Oberstes Gebot der Frageformulierung ist das Prinzip der Einfachheit. Fragen sollten möglichst einfach und verständlich sein. Dies bedeutet:

a) Fragen sollen möglichst kurz sein, das heisst so wenig Worte wie möglich umfassen.

b) Fragen sollen für alle Befragten gut verständlich sein. Befragte dürfen nicht überfordert werden. Zu vermeiden sind: Fremdwörter, Fachausdrücke, komplizierte Sätze und Wendungen (wie z.B. doppelte Verneinungen) usw. Befragte sind meist keine "Intellektuelle". Gerade abstrakte Oberbegriffe werden von Befragten fast ausnahmslos unterschiedlich aufgefasst.

Beispiel von Überforderung: "Wie viele Prozente Ihres Haushaltseinkommens geben Sie durchschnittlich aus für...?", "Wie wird sich Ihrer Ansicht nach das neue Gesetz gegen unlauteren Wettbewerb auswirken?"

Beachten: Befragte werden selten zugeben, dass sie eine Frage nicht verstanden haben, sondern sie werden einfach so antworten, wie sie die Frage verstanden haben (z.B. im Sinne sozialer Erwünschtheit, Ja-Sage-Tendenz usw.).

c) Fragen sollten umgangssprachlich formuliert werden. Verständlichkeit ist wichtiger als grammatikalische Richtigkeit. In der Schweiz kann es sinnvoll sein, Fragen an die Dialektsprache anzunähern bzw. Helvetismen zu benutzen.

5. Fragen sollten eindeutig sein. Mehrdeutige Begriffe und unklare Formulierungen sind zu vermeiden, und vor allem soll jede Frage nur einen Sachverhalt/Einstellungsaspekt erfassen (vermeiden von "double-barreled"-Fragen).

Beispiel: Sind Sie für oder gegen genetische Forschung und Forschung an menschlichen Stammzellen? Hier werden zwei Fragen: genetische Forschung, Forschung an Stammzellen „gemischt“. Damit wird die allgemeine Frage (Genetische Forschung) durch eine spezifische Frage (Stammzellenforschung) überlagert.

Ungewohnte und/oder mehrdeutige Begriffe sind ebenfalls zu vermeiden. Beispiel: Sind Sie dafür oder dagegen, dass der Konsum von Suchtmitteln verboten ist? Problem: Begriff

"Suchtmittel" ist a) zu allgemein, und b) wenig gebräuchlich). Besser ist ein gezieltes Fragen danach, welche konkrete Suchtmittel (Tabak, Heroin, Haschisch) verboten sein sollen oder nicht.

Allerdings gilt, dass viele Alltagsbegriffe mehrdeutig und unscharf definiert sind. Die Frage "Wie oft bzw. selten sind Sie allein?" kann sich beziehen a) auf das Alleinsein als Fakt, oder b) auf Alleinsein als Gefühl von Einsamkeit, Verlassenheit. Auch die Frage nach der Zahl an Personen im Haushalt kann unterschiedlich gesehen werden (z.B. werden einige Leute, nur diejenigen Personen nennen, die gerade jetzt hier leben, andere Befragte werden auch ein zeitweise auswärts wohnendes Kind einbeziehen). Der Begriff "Einkommen" kann sich beziehen auf Erwerbseinkommen, Haushaltseinkommen, Brutto- bzw. Nettoeinkommen usw. Auch eine Spezifikation der Frage wird solche Unklarheiten/ Definitions-Unschärfen nicht lösen.

6. Die Beantwortung von Fragen zu Einstellungen und Bewertungen wird durch die genaue Wortwahl stark beeinflusst, und dies gilt vor allem für Fragen, welche für die Befragten nicht wichtig sind oder wo keine klaren und festen Meinungen bestehen. (Für eine empirische Untersuchung der Auswirkungen unterschiedlicher Frageformen, vgl. etwa Belson, 1981, Petersen 2002).

3. Antwortvorgaben

1. Wird bei einer Testbefragung festgestellt, dass bei Verwendung des Antwortschemas "stimme stark"... "lehne stark ab" viele Befragte in die linke oder rechte äussere Kategorie fallen, dann sollte man versuchen, noch eine weitere äussere Kategorie zu verwenden (wie "stimme extrem zu" bzw. "lehne extrem ab").

Generell ist bei sozialwissenschaftlichen Studien darauf zu achten, dass die Antwortvarianz (bei Einstellungen oder Skalen) hoch ist. Fragen, bei denen fast alle Antworten in eine Antwortkategorie fallen, sind nicht weiter auswertbar.

2. Wenn von Befragten nicht eine unmittelbar quantitative Antwort zu erwarten ist, dann sollte man quantitative Fragen "geschlossen" (= mit vorgegebenen Antwortkategorien) formulieren. Vor allem bei schriftlichen Fragebogen sollte aus Vergleichsgründen primär mit vorgegebenen Antwortkategorien gearbeitet werden. Damit wird die Perspektive, in der eine Antwort erwartet wird, klar gemacht.

3. Eine interne Antwortvorgabe ist nur zulässig, wenn sie aus zwei oder drei leicht merkbaren Alternativen besteht. Sind mehr Alternativen vorhanden, oder bestehen die Alternativen aus relativ komplexen Formulierungen, dann müssen sie extern vorgegeben werden. Beispiel: Interne Vorgabe: "Sind Sie dafür oder dagegen, dass die Schweiz der Europäischen Union beitrifft? Externe Vorgabe: "Wie ist Ihre Haltung zu einem Beitritt der Schweiz zur Europäischen Union?" a) bin stark dafür, b)bin eher dafür, c) bin weder dafür noch dagegen, d) bin eher dagegen, e) bin stark dagegen.

Bei Telefoninterviews sind primär interne Antwortvorgaben zu verwenden.

4. Ein elementarer Fehler liegt vor, wenn Fragen nach Häufigkeit, Dauer oder Grösse und ähnlichem gestellt werden, ohne dass die Einheiten angegeben werden, in denen die Antworten erfolgen soll.

Beispiel: "Wie häufig bzw. selten gehen Sie ins Kino?" Ohne Zeitraum ist diese Frage sinnlos. Bedingt richtig: "In den letzten zwei Wochen, wie oft gingen Sie ins Kino?" (=inhaltlich unausgewogen). Am besten: "In den letzten zwei Wochen, wie viele Mal gingen Sie ins Kino? (nie, einmal, zweimal, ...). Falsch: Seit wann studieren Sie? Besser: In welchem Jahr began-

nen Sie mit Ihrem Studium? (plus Zusatzfrage: Haben Sie Ihr Studium unterbrochen oder nicht? Wenn ja, nachfragen: Wann und wie lange?).

5. Es ist unzulässig, eine (oder mehrere) Antwortalternativen nur implizit intern vorzugeben. Die explizit genannten Alternativen werden favorisiert und die nicht genannten Alternativen werden vernachlässigt.

Beispiel: Falsch ist "Glauben Sie, dass Sie durch Ihr Studium auf Ihren Beruf gut vorbereitet sind? Richtig ist: "Glauben Sie, dass Sie durch Ihr Studium auf ihren späteren Beruf gut, weniger gut, oder schlecht vorbereitet sind?"

6. Die Zahl der (externen) vorgegebenen Antwortalternativen muss für den Befragten gut überschaubar sein. Werden die Antwortmöglichkeiten vom Interviewer vorgelesen, muss sich der Befragte, wenn er die letzte Antwortmöglichkeit hört, noch an die erste Möglichkeit erinnern. Sonst wird die zuletzt gehörte Antwortvorgabe bevorzugt. Bei grösserer Zahl von Antwortvorgaben oder komplexen Antwortvorgaben empfehlen sich schriftliche Listen oder Karten. Bei telefonischen Umfragen sind Fragen mit mehreren Antwortalternativen in mehrere Fragen aufzulösen.

7. Je länger und je komplizierter die Antwortvorgaben, desto stärker ist der Einfluss ihrer Reihenfolge. Dies gilt z.B. auch bei Listen von Antwortvorgaben (vgl. Petersen 2002).

Rangreihen-Effekte können durch Verwendung von Kartenspielen oder durch das "Drehen" von Listen vermieden werden. Bei computerunterstützten Befragungen kann die Reihenfolge etwa von Items durch Zufallsparameter variiert werden.

8. Die vorgegebenen Antwortmöglichkeiten sollten erschöpfend sein (d.h. sie müssen alle relevanten Möglichkeiten enthalten). Ist eine erschöpfende Aufzählung der Antwortalternativen nicht möglich, sollte eine Kategorie "sonstige", "anderes" vorgesehen werden. Dies gilt vor allem bei Faktfragen.

9. Werden mehrere Antwortmöglichkeiten angeboten, dann sollte die Zahl der Antwortmöglichkeiten "rechts" und "links" von der Mittelposition gleich sein, weil von einer Überzahl von Kategorien auf einer Seite suggestive Wirkungen ausgehen. Der Befragte bevorzugt die Seite, die mit mehr Antwortmöglichkeiten vertreten ist.

Beispiel: Falsch ist: sehr gut, gut, weder noch, schlecht. Richtig: sehr gut, gut, weder noch, schlecht, sehr schlecht.

10. Zu beachten ist, dass positive Antwortalternativen häufig negativen Antwortalternativen vorgezogen wird (Tendenz zu "Ja-Antworten"). Speziell bei Fragebatterien müssen deshalb negative und positive Aussagen "gemischt" werden.

11. Auch keine Antwort ist eine Antwort, und Meinungslosigkeit ist eine sozial wichtige Dimension. Zu den Antwortvorgaben gehört oft auch die Vorgabe "keine Meinung", bzw. "unsicher", "ist mir egal", oder "weder noch".

4. Regeln zur Dramaturgie einer mündlichen oder telefonischen Befragung

1. Grundvoraussetzung eines guten Fragebogens (für mündliche oder telefonische Interviews) ist, dass er den Eindruck einer echten Gesprächssituation erzeugt, die der Befragte interessant findet. Bei schriftlichem Fragebogen ist hingegen vor allem ein gutes und klares Layout zentral.

2. Beim Aufbau des Fragebogens für mündliche und telefonische Interviews muss Bedacht auf die spätere Gesprächssituation genommen werden. Die Unterhaltung soll sich für Interviewer und Befragten möglichst mühelos und frei von Peinlichkeiten vollziehen. Bei schriftlichen Befragungen sollte ein Wirrwarr bzw. eine zu hohe Vielfalt an Antwortkategorien vermieden werden (z.B. Beschränkung auf einige wenige Formen von Antwortkategorien, die bei verschiedenen Fragen angewendet werden können).

3. Die vielleicht wichtigste Regel demoskopischer Dramaturgie bestimmt, dass möglichst bald ein Vertrauensklima zwischen Befragten und Interviewer entstehen muss. Die Praxis zeigt, dass der Erfolg einer Befragung wesentlich von der Tauglichkeit der Einleitungsfragen abhängt. Bei telefonischen Befragungen ist der Erstkontakt zentral (was speziell geschulte Interviewerinnen voraussetzt).

Wichtig ist, dass die Einleitungsfragen Probleme ansprechen, die den Befragten unmittelbar ansprechen und interessieren. Sie sollten leicht zu beantworten sein, um die Auskunftsperson von der Vorstellung einer Prüfungssituation zu entlasten.

4. Bei umfangreichen mündlichen Befragungen (die z.B. mehr als dreissig Minuten dauern) sollten schwierige und komplexe Passagen des Fragebogens immer wieder durch einfache oder gar spielerische Fragen unterbrochen werden. "Spielfragen" stellen Erholungspausen dar und zudem neutralisieren sie - als so genannte "Pufferfragen" - die nachfolgende Befragungssituation.

5. Ganz allgemein erfordert die Dramaturgie eines guten mündlichen Interviews einen ständigen Wechsel von Spannung und Entspannung, von "schweren" und "leichten" Fragen, geschlossenen und offenen Fragen usw., nicht zuletzt aber auch einen genügend grossen Wechsel von Themen.

6. Der allgemeinen Erfahrung nach wirken bei mündlichen Interviews vor allem viele geschlossene Fragen, unmittelbar nacheinander gestellt, ermüdend. Daher ist es zweckmässig, solche Serien durch offene Fragen zu unterbrechen. Bei schriftlichen Fragebogen wirkt vor allem ein zu langer (und dicker) Fragebogen 'abschreckend'.

7. Der Eindruck der Vielfalt und Kurzweil, der einen guten Fragebogen auszeichnet, kann durch einen Wechsel von Fragetechniken - inklusive Kartenspiele, Bildblätter - verstärkt werden. Bei telefonischen Interviews kann das Interesse hingegen durch Tempowechsel und Abwechseln von Ja-Nein-Fragen mit anderen Frageformen verstärkt werden. Bei schriftlichen Fragebogen ist wiederum auf eine gewisse Einheitlichkeit der Antwortvorgaben und des Layout zu achten (so dass sich Befragte an spezifische Formen des 'Ankreuzen' gewöhnen können). Zu viel Wechsel (z.B. zu viele Schriftarten) wirkt hier störend.

8. Ein Fragebogen sollte nicht zu lang werden (wegen Ermüdungseffekt). Befragungen, die sich über die Dauer einer Stunde hinziehen, überfordern oft die Konzentrationsfähigkeit und Geduld des Befragten, aber auch des Interviewers. Telefonische Befragungen sollten im Allgemeinen noch kürzer sein, und bei schriftlichen Befragungen führen zu lange bzw. dicke Fragebogen zur Reduktion der Antwortbereitschaft.

9. Einstellungsfragen, die sich gegenseitig stark beeinflussen können, sollten in mündlichen Befragungen durch "Pufferfragen" getrennt werden.

10. "Filter" sollen nur sparsam verwendet werden. Zu viele Filter bzw. Gabelungen sprengen nicht nur die Einheit der Befragungssituation, sondern sie behindern auch die spätere Auswertung.

11. Generell ist zu beachten, dass standardisierte Fragebögen so konstruiert werden müssen, dass sie von verschiedenen Bevölkerungsteilen verstanden und beantwortet werden können. Faktisch bedeutet dies, dass viele Fragen nicht spezifisch auf jeweilige Sondersituationen zugeschnitten sein können.

Fragebatterien

Viele Sachverhalte, Einstellungen oder Motive sind nicht direkt und unmittelbar erfassbar. Oft ist auch unklar, ob eine spezifische Frage die angestrebte Zieldimension erfasst und trifft. Deshalb müssen zum gleichen Thema meist mehrere Fragen gestellt werden. Um die Chancen zu erhöhen, dass man tatsächlich das erfasst, was man wissen will, sollte man für eine Zieldimension möglichst viele Fragen formulieren. Dies gilt vor allem um Einstellungen oder Motive, die für das gewählte Forschungsprojekt zentral sind. Zum einen vermindert mehrfaches Fragen das Risiko, dass man die gewünschte Zieldimension voll verfehlt. Zum zweiten kann getestet werden, inwiefern die Zieldimension tatsächlich ein-dimensional ist. Zum dritten können durch mehrfaches Fragen Motive und Einstellungen erfasst werden, die sich einer direkten Erfragung entziehen (wie z.B. Fremdenfeindlichkeit).

Beispiel: Erfasst werden soll die schulische Zufriedenheit von Schülern/innen. Um dies zu messen, wurden mehrere Fragen gestellt:

- 1) Gehst Du gerne in die Schule? (sehr gerne, etwas gerne, nicht gerne, gar nicht gerne)
- 2) Wenn Du morgen aufwachst, möchtest Du da manchmal ganz gerne krank sein, um nicht zur Schule gehen zu müssen? (das möchte ich jeden Tag, das möchte ich öfters, das möchte ich selten, das möchte ich nie)
- 3) Möchtest Du Deine Hausaufgaben zu Hause machen oder wäre es Dir lieber, wenn Du sie am Nachmittag in der Schule machen könntest? (ich möchte sie lieber zu Hause machen, es ist mir gleich lieb, ich möchte sie lieber in der Schule machen).
- 4) Zähle bitte genau nach, wie oft Du seit den letzten Herbstferien wegen Krankheit nicht in die Schule gehen konntest (ich war ... mal krank).

Bei diesem Beispiel ist offensichtlich, dass Frage 4 nicht allein mit schulischer Zufriedenheit zusammenhängt (auch zufriedene Schüler können krank werden). Auch die Beantwortung von Frage 3 ist in jedem Fall ein Hinweis auf schulische Zufriedenheit. Tatsächlich trafen - wie die Auswertungen zeigten - nur die ersten beiden Fragen die Zieldimension "Schulische Zufriedenheit".

Da Forscher jedoch im Voraus oft nur schlecht abschätzen können, wie gut ihre Fragen sind, sollte man zu wichtigen Aspekten eben mehrere Fragen formulieren (und diese in einem Pre-test testen).

Zur Erfassung komplexer Einstellungen oder Motive werden häufig eigentliche Fragebatterien verwendet. Von Fragebatterien wird dann gesprochen, wenn zu ein und demselben Sachverhalt eine grössere Zahl von Fragen oder Behauptungen (statements) in einem kompakten Block gestellt werden. Für alle Fragen-Items sind die Antwortvorgaben gleich.

In vielen Befragungsstudien werden Fragebatterien vor allem dazu verwendet, um eindimensionale Skalen bzw. Indikatoren zu erstellen (Auf das Thema der Skalen/ Skalierung wird später noch gezielter eingegangen). In der Auswertung interessiert nicht die einzelne Frage bzw. das einzelne Frage-Item, sondern die Summe aller Antworten. D.h. die einzelnen Antworten zu den "statements" werden zu einer Skala addiert (und in folgenden wird nur mit der Skala gearbeitet).

Beispiel: Erfasst werden sollte die Einstellung junger Frauen zu Mutterschaft. Dazu wurde den befragten Frauen verschiedene Aussagen über Kinder-haben vorgelegt:

- Eine Frau kann mit ihrem Leben ganz zufrieden sein, wenn sie eine gute Mutter ist
- Es ist wichtig, Kinder zu haben, damit die Familientradition weitergeführt wird
- Kinder zu haben ist für mich das wichtigste Ziel der Ehe

-Allein schon das Gefühl, dass Kinder einen brauchen, macht es lohnend, Kinder zu haben.

-Es ist wichtig, Kinder zu haben, damit man im Alter nicht allein ist.

Antwortvorgaben jeweils: sehr einverstanden, einverstanden, unentschieden, weniger einverstanden, gar nicht einverstanden.

Es zeigte sich, dass die Items untereinander stark interkorrelierten (Cronbach's Alpha: .72). Bei der Auswertung wurden die Antworten zu den einzelnen Aussagen (Items) deshalb zu einer Skala "Wert von Kindern" addiert. Inhaltlich ist anzumerken, dass diese Skala mehr traditionelle Mutterschaftsaspekte anspricht.

Werden zum gleichen Themenkreis mehrere Fragen nacheinander gestellt, ergibt sich das Problem von "Positionseffekten" (das heisst: die Reihenfolge der Fragen beeinflusst das Antwortverhalten). Positionseffekte werden gerade auch bei Fragebatterien und Listenfragen (Fragen mit langen Antwortlisten) bedeutsam. Positionseffekte können durch (zufälligen) Wechsel der Rangfolge von Items vermieden werden. Z.B. Aussagen werden auf Karten gedruckt, und vom Interviewer vor jedem Interview neu gemischt.

Bei Fragebatterien kommt das Problem dazu, dass Befragte dazu neigen, konsistent und "vernünftig" zu antworten. Dies kann die Korrelationen zwischen Items künstlich erhöhen. Eine Möglichkeit um dies zu verhindern, ist eine Streuung der entsprechenden Items im Fragebogen (Nachteil: Ablauf wird dadurch gestört, das Interview wird aufwändiger). Zudem ist Antwortkonsistenz nicht immer unerwünscht.

Wichtige Zusatzregelung:

Während Fragebogenfragen grundsätzlich neutral und ausgewogen sein müssen, können Fragebogen-Items durchaus einseitig formuliert werden. Items sind immer Teil eines grösseren Ganzen (eindimensionale Skala). Bei Skalen ist allerdings darauf zu achten, dass positiv und negativ formulierte Items abwechseln. Vor allem bei einer Skala mit vielen Items ist es problematisch, nur positive Items einzubeziehen.

Vignetten-Technik:

Eine Form (halb-standardisierter bis standardisierter) Erfassung ganzer Problem- oder Lösungsformen: An einem kurzen Beispiel wird ein Thema, ein Problem angesprochen, das zu beurteilen ist.

Beispiel aus einer Untersuchung zu Familiensituationen:

Vignette Computerspiele: „Normalerweise erledigen die Kinder die Hausaufgaben, bevor sie am Computer spielen oder fernsehen. Kevin (11-jährig) hat für den morgigen Schultag noch eine Menge Aufgaben zu machen. Er zieht es jedoch vor, zuerst am Computer zu spielen.“

Wie reagieren Sie als Eltern (als Grosseltern, Nachbarn etc).

Vignette Kirche: „Für die Familie des 13-jährigen Thomas ist der sonntägliche Gottesdienst sehr wichtig. Thomas möchte aber nicht mehr mit seiner Familie in die Kirche gehen“

Wie würden Sie (als Eltern, Grosseltern, Nachbarn, Lehrkräfte...) reagieren:

- a) Wenn Thomas nicht mehr in die Kirche gehen will, darf er zu Hause bleiben.
- b) Thomas muss weiterhin mit der Familie in die Kirche kommen
- c) Thomas fragen, warum er nicht mehr in die Kirche kommen will und mit ihm zusammen eine Lösung finden.

In welchem Masse sollen frühere Befragungen/Fragen übernommen werden?

Um Vergleiche der eigenen Befragung mit anderen Erhebungen zu erleichtern, kann es sinnvoll sein, Frageblöcke, Skalen usw. von früheren Erhebungen zum Thema zu übernehmen. Damit lassen sich Zeit- oder Kontextvergleiche durchführen. Bei der Verwendung von Fragebögen/Fragen aus anderen Sprachen ist eine sorgfältige Übersetzung notwendig.

Sinnvoll ist es häufig auch, wichtige sozio-demographische Fragen direkt aus vorhandenen Fragebögen von offiziellen Mikrozensus (Schweizerische Arbeitskräfteerhebung, Schweizeri-

sche Gesundheitsbefragung u.a.) zu übernehmen. Dies erleichtert eine externe Validierung der eigenen Erhebung.

‚Cut and Paste‘ ist somit bei der Konstruktion eines Fragebogens durchaus erlaubt bzw. erwünscht (mit entsprechender Quellenangaben).

Ausgewählte Literatur zu Frageformulierung und Fragebogenkonstruktion

- Converse, Jean M.; Presser, Stanley (1986) *Survey Questions. Handcrafting the Standardized Questionnaire*, Beverly Hills: Sage.
- Fielding, Nigel (ed.) (2003) *Interviewing, Sage Benchmarks in Social Research Methods*, London/Thousand Oaks: Sage Publ.: Vol. I-IV.
- Foddy, William (1993) *Constructing Questions for Interviews and Questionnaires. Theory and Practice in Social Research*, Cambridge: University Press.
- Gordon, Raymond L. (1975) *Interviewing: Strategy, techniques, and tactics*, Homewood: Dorsey Press.
- Holm, Kurt (Hrsg.) (1982) *Die Befragung I*, München: Franke.
- Holm, Kurt (1975) *Theorie der Frage*, Kölner Zeitschrift für Soziologie und Sozialpsychologie, S.91-114.
- Hron, Aemilian (1982) *Interview*, in: Huber, Günter L.; Mandl, Heinz (Hrsg.), *Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung*, Pp. 119-140, Weinheim: Beltz.
- Kaase, Max (Hrsg.) (1999) *Qualitätskriterien der Umfrageforschung - Quality Criteria for Survey Research*, Deutsche Forschungsgemeinschaft, Bonn: Akademie-Verlag.
- Kallus, K. Wolfgang (2010) *Erstellung von Fragebogen*, Wien: Facultas.
- Kirchhoff, Sabine, Sonja Kuhnt, Peter Lipp, Siegfried Schlawin 2010: *Der Fragebogen – Datenbasis, Konstruktion, Auswertung*. Wiesbaden: Verlag für Sozialwissenschaften.
- Kreutz, Henrik; Titscher, Stefan (1975) *Die Konstruktion von Fragebögen*, in: Van Koolwijk, Jürgen; Wieken-Mayser, Maria (Hrsg.), *Techniken der empirischen Sozialforschung*, 4. Band, Pp. 24-75. München: Oldenburg.
- Petersen, Thomas (2002) *Das Feldexperiment in der Umfrageforschung*, Frankfurt: Campus.
- Porst, Rolf (2009) *Fragebogen. Ein Arbeitsbuch*, Wiesbaden: VS Verlag
- Schuman, Howard; Presser, Stanley (1981) *Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context*, New York: Academic Press.
- Sudman, Seymour; Bradburn, Norman M. (1985) *Asking Questions. A Practical Guide to Questionnaire Design*, San Francisco: Jossey-Bass.
- Sudman, Seymour; Bradburn, Norman M.; Schwartz, Norbert (1996) *Thinking about answers: The application of cognitive process to survey methodology*, San Francisco: Bass.
- Tanur, Judith M. (ed.) (1992) *Questions about Questions. Inquiries into the Cognitive Bases of Surveys*, Russell Sage Foundation, New York.

„Änderungen im Stichprobendesign können in den Daten deutlichere Änderungen als der soziale Wandel bewirken. Damit ergibt sich das Problem, dass ohne Berücksichtigung des Stichprobendesigns methodische Artefakte als sozialer Wandel interpretiert werden können.“ Hoffmeyer-Zlotnik 2006: 34.

Stichprobenauswahl und Samplingverfahren

1) Problemstellung und Begriffe

Entscheidende Aspekte bei jeder grösseren Umfrage sind Wahl von Stichprobe und Samplingverfahren. Eine optimale Stichprobenauswahl ist umso wichtiger, je eher ein Anspruch auf Repräsentativität gestellt wird.

Zu den Begriffen:

Stichprobe (oder Sample): = befragte Personen aus einer grösseren Menge von Personen. Eine Stichprobe umfasst immer nur einen ausgewählten Teil der interessierenden Bevölkerung. Beispielsweise werden 1000 Personen aus der Wohnbevölkerung einer Region befragt, um Auskunft über die Lebenslage dieser Personen zu erhalten.

Eine Alternative zur Stichprobenerhebung besteht in einer Totalerhebung (etwa im Rahmen von Volkszählungen). Es ist offensichtlich, dass eine Stichproben-Befragung einen wesentlich geringeren Zeit- und Kostenaufwand erfordert als eine Totalerhebung. Aufgrund statistischer Gesetzmässigkeiten (Gesetz der grossen Zahl) können bei zufällig ausgewählten Stichproben trotz einer im Vergleich zur Gesamtbevölkerung relativ kleinen Zahl von Befragungen dennoch gültige Aussagen über die Gesamtbevölkerung gemacht werden.

Samplingverfahren (oder kurz: Sampling): = Verfahren, mit dem Befragte (bzw. Personen der Stichprobe) aus der Gesamtmenge an möglichen Personen ausgewählt werden. Es bestehen - wie nachfolgend gezeigt wird - unterschiedliche Samplingverfahren, wobei nur ein Teil der denkbaren Auswahlverfahren zu einer repräsentativen Erhebung führt.

Repräsentativität: Eine Befragung kann dann und nur dann als repräsentativ angesehen werden, wenn die Stichprobe in ihrer sozialen Verteilung der Grundgesamtheit entspricht. Repräsentativität bedeutet, dass die aus einer Teilbefragung von Personen gewonnenen Informationen für die Grundgesamtheit aller Personen gültig sind. Die Verallgemeinerung von Forschungsergebnissen einer Stichprobenbefragung auf die Grundgesamtheit hat allerdings nur dann Gültigkeit, wenn die Stichprobe gemäss der statistischen Theorie gebildet wurde. Die wichtigste Regel ist, dass die zu befragenden Personen im statistischen Sinne 'zufällig' ausgewählt werden müssen. Und eine reine Zufallsauswahl bedeutet prinzipiell nichts anderes, als dass jede Person aus der Untersuchungspopulation statistisch die genau gleiche Wahrscheinlichkeit aufweist, befragt zu werden.

Repräsentative Umfragen setzen somit in Zufalls-Sampling (engl. random sampling) voraus. Nur eine Zufallsauswahl lässt eine spätere Verallgemeinerung der Forschungsergebnisse und das Testen von allgemeinen Hypothesen zu. Und im Grunde sind Signifikanz-Tests nur bei repräsentativ erhobenen Daten sinnvoll (sowohl bei grob verzerrten, nicht-repräsentativen Stichproben als auch bei Vollerhebungen sind Signifikanz-Tests sinnlos).

Bei der Vorbereitung einer (repräsentativen) Befragung stehen primär drei zentrale Schritte zu entscheiden:

- a) Genaue Definition der Untersuchungspopulation: Wer genau soll befragt werden?
- b) Wahl der Stichprobengrösse: Wie viele Personen sollen befragt werden?
- c) Auswahl der geeigneten Samplingstrategie: Welches Auswahlverfahren ist geeignet bzw. möglich?

Bei allen drei Schritten sind neben theoretisch-statistischen Aspekten immer auch praktisch-pragmatische Aspekte zu berücksichtigen. Die statistisch beste Samplingstrategie nützt wenig, wenn zu wenig Mittel, Zeit oder Grundlagen (z.B. aktuelle Adressen) vorliegen, und auch die Wahl der Stichprobengrösse wird oft von vorhandenen finanziellen Rahmenbedingungen mitbestimmt. Heute ist bei einigen Befragungen vorgängig auch das Einverständnis einer ethischen Kommission notwendig, etwa bei Patientenerhebungen. Der Vorbereitung des Samplings ist bei grösseren Forschungsprojekten auf jeden Fall genügend Zeit einzuräumen.

2) Zur Wahl der Untersuchungspopulation

Der erste Schritt ist sachgemäss eine klare und operationalisierbare Definition der Untersuchungspopulation, d.h. die Bestimmung der für die jeweilige Umfrage interessierenden Gruppen von Personen, Haushalte oder Organisationen. Die Definition der Untersuchungspopulation ist im Wesentlichen von den Forschungszielen und den Fragestellungen abhängig (und sie sollte frühzeitig entschieden werden) (vgl. Forschungsdesign).

Die Untersuchungspopulation muss sowohl sozio-demografisch als auch geographisch klar abgrenzbar sein, wobei diesbezüglich auch forschungspraktische Faktoren mitspielen. So bedeutet eine repräsentative gesamtschweizerische Befragung, dass etwa der Fragebogen in mehreren Sprachen übersetzt werden muss (was sich auch finanziell auswirkt). Eine Umfrage bei städtischen Bevölkerungen muss klar definieren, welche Gemeinden zu den städtischen oder stadtnahen Gebieten gezählt werden. Bei einer Mitgliedebefragung etwa muss klar bestimmbar sein, wer als 'Mitglied' gezählt wird, usw.

Zu berücksichtigen ist auch, dass verschiedene Bevölkerungsgruppen nicht oder nur unter erschwerten Bedingungen befragbar sind (wie etwa Gefängnisinsassen, kranke und behinderte Menschen u.a.). Kinder und minderjährige Personen dürfen im Allgemeinen häufig nur mit Einwilligung der Eltern befragt werden, abgesehen davon, dass die Befragung von Kindern andere Befragungsinstrumente einschliesst als die Befragung von Erwachsenen. Umfragen bei betagten Menschen stehen vor dem Problem, dass Personen in privaten und institutionellen Haushaltungen unterschiedlichen Befragungssituationen unterliegen (und etwa demenzkranke Betagte höchstens via 'proxy-Interviews' (= Befragung von Angehörigen oder Pflegepersonen) erfasst werden können).

Je heterogener die Untersuchungspopulation in geographischer, sozialer, altersmässiger oder kultureller Hinsicht ist, desto aufwändiger wird eine Befragung, abgesehen davon, dass heterogene Untersuchungspopulationen die Bereitstellung eines für alle gut verständlichen Fragebogens erschweren. Fragen, welche für Jugendliche wichtig sind, sind für ältere Menschen unangebracht, und umgekehrt.

Speziell dort, wo von vornherein mit eingeschränkten finanziellen Mitteln gearbeitet wird, ist in vielen Fällen eine geographische, soziale, altersmässige und kulturelle Homogenisierung der Untersuchungspopulation zu empfehlen (z.B. Beschränkung auf Agglomerationsgebiete, altersmässige Limitierung der Stichprobe).

3) Zur Wahl der Stichprobengrösse

Eine zweite zentrale Entscheidung bei jeder (repräsentativen) Befragung ist die Wahl der Stichprobengrösse bzw. Samplegrösse (= Zahl der zu befragenden Personen). Sollen 500, 1000 oder 2000 Personen befragt werden?

Die Stichprobengrösse wird zum einen von finanziellen Überlegungen beeinflusst: Je mehr Interviews, desto kost- und zeitaufwendiger ist Befragung, namentlich bei kostenintensiven Umfrageformen (wie face-to-face-Interviews). Aber auch bei kostengünstigen Verfahren - wie telefonische oder schriftliche Befragung - nehmen die Kosten mit steigender Stichprobengrösse linear zu. Einzig bei elektronischen Befragungen (Internet-Umfrage) ergeben sich

höchstens indirekte Mehrkosten (z.B. Nachkontrollaufwendungen). Aus finanziellen Gründen wird somit namentlich bei mündlichen, teilweise aber auch bei telefonischen Befragungen eine möglichst geringe Zahl von Befragungen gewünscht.

Die Wahl der Stichprobengrösse wird zum anderen durch statistische Überlegungen bestimmt. Das Gesetz der grossen Zahl besagt, dass ein Stichprobenwert umso eher mit dem echten Wert der Grundgesamtheit übereinstimmt, wenn die Stichprobengrösse sich der Grösse der Grundgesamtheit nähert. Daraus ergibt sich, dass die aus einer beschränkten Stichprobe gewonnenen Werte mehr oder minder stark von ihrem wahren Wert abweichen müssen (= Stichprobenfehler).

Alle aus Stichproben gewonnenen Ergebnisse können daher lediglich wahrscheinlichkeitstheoretische Aussagen sein, deren Genauigkeit vom Stichprobenfehler abhängt. Der Stichprobenfehler seinerseits hängt mit der Zahl an Befragten ab, und er reduziert sich mit zunehmender Stichprobengrösse. Je genauer die gemachten Aussagen sein sollen, desto mehr Personen müssen befragt werden.

Die "Kunst" liegt darin namentlich bei kostenintensiven Befragungsformen zwischen finanziellen Grenzen (möglichst wenig Interviews) und statistischer Genauigkeit (möglichst viele Interviews) zu optimieren.

Bei reinen Zufallsstichproben lässt sich der Stichprobenfehler (= wahrscheinliche Abweichung des Stichprobenwertes vom wahren Wert) ohne grosse Probleme errechnen. Der Stichprobenfehler bei einer reinen Zufallsstichprobe bezüglich eines gemessenen Merkmal hängt von zwei Grössen ab:

a) von der Variabilität des Merkmals in der Grundgesamtheit, die durch die Stichprobenvarianz geschätzt wird. Je grösser die Streuung (s) eines Merkmals, desto grösser der Schätzfehler. Bei einer Streuung von $p = 50\%$ (bzw. 50% Ja, 50% Nein) ist der Schätzfehler am höchsten.

b) von der Grösse der Stichprobe. Der Stichprobenfehler verändert sich umgekehrt proportional zur Quadratwurzel aus n (= Stichprobengrösse). In anderen Worten: Um den Stichprobenfehler zu halbieren, muss man einen viermal so grossen Personenkreis befragen. Wegen dieser nichtlinearen Relation zwischen Sample-Umfang und Fehler verzichtet man in der Praxis oft auf eine minimale Fehlereinengung zugunsten der niedrigeren Kosten einer kleineren Stichprobengrösse.

Konkrete Beispiele: Bei einer Befragung von 1000 Personen hat eine Frage einen Prozentsatz positiver Antworten von $p = 80\%$ erbracht. Als Stichprobenfehler ergibt sich folgendes:

$$sp = \sqrt{p(100-p)/N} = \sqrt{80 * 20/1000} = 1.265.$$

$\sqrt{}$ = Quadratwurzel

Bei Verdoppelung der Stichprobengrösse auf 2000 reduziert sich der Stichprobenfehler bei dieser Antwortverteilung auf 0.89. Bei 500 Befragten erhöht er sich auf 1.79.

Variiert die Verteilung zwischen positiven und negativen Antworten stärker, im Verhältnis von 50% zu 50%, ergibt sich bei einer Stichprobengrösse von 1000 gemäss obiger Formel ein Stichprobenfehler von 1.58. Bei $N = 500$ ist der Fehler schon 2.24.

Bei nicht allzu kleinen Stichproben kann von normal verteilten Grössen ausgegangen werden, und der Stichprobenfehler kann zur Berechnung von Vertrauensintervallen (oder Konfidenzintervallen) benützt werden. Nehmen wir etwa eine Sicherheitswahrscheinlichkeit von 95%, ergibt sich aus der statistischen Tafel zur Normalverteilung eine Standardabweichung (SD)

von 1.96. Eine Multiplikation der SD mit dem oben errechneten Stichprobenfehler zeigt uns, in welchem Bereich sich der effektive Wert befindet.

Beispiel: N: 1000, 80% positive Antworten. Stichprobenfehler: $1.265 * 1.96 = 2.48$. D.h. Aufgrund unserer Stichprobe können wir davon ausgehen, dass in der Gesamtbevölkerung zwischen 77.5% und 82.5% eine positive Antwort geben. Hätten wir jedoch nur 500 Leute befragt, könnten wir nur sagen, dass sich der wahre Wert zwischen 76.5% und 83.5% bewegt. Und bei einem Sample von nur 100 Personen wird die Unsicherheit noch grösser (Stichprobenfehler: $4.0 * 1.96 = \pm 7.84$. Das heisst der tatsächliche Wert bewegt sich zwischen 72% und 88%.

Zu beachten: Die oben angeführten Formel gilt nur für einfache Zufallsstichproben. Bei mehrstufigem Sampling gelten komplexere Formel zur Berechnung des Stichprobenfehlers bzw. der Vertrauensintervalle. Zudem gilt die obige Formel unkorrigiert nur, wenn die Stichprobe nicht mehr als 20% der Untersuchungspopulation umfasst (vgl. Henry, Gary T. (1990) Practical Sampling, Sage, Newbury Park.

Es ist zudem wichtig zu verstehen, dass Umfrageergebnisse nur dann auf die Grundgesamtheit verallgemeinert werden dürfen, wenn die Stichprobe nach den Regeln der Zufälligkeit gezogen wurde. Nur eine statistische Zufallsauswahl gewährleistet eine unverzerrte Widerspiegelung der Gesamtbevölkerung. Verzerrungen einer Stichprobe - durch verzerrte Adressenwahl, durch Verweigerungen - schränken ihre "Repräsentativität" ein.

Faktisch ist bei allen Umfragen mit Ausfällen zu rechnen; sei es dass bestimmte Zielpersonen nicht erreicht werden können (z.B. wegen Auslandsaufenthalt, ungenügenden Adressangaben usw.). Zudem ist immer auch mit Verweigerungen zu rechnen.

Generell gilt: Ein frühzeitiges Abklären der Möglichkeiten und Probleme, welche sich bei der Auswahl von Befragten aus einer definierten Untersuchungspopulation ergeben können, erspart später viele Probleme. So ist es etwa in der Schweiz durchaus möglich, aktualisierte Adressen der Einwohnergemeinden zu erhalten, bei zunehmend mehr Gemeinden braucht es dazu allerdings eine Genehmigung durch die politischen Oberbehörden (was sachgemäss Zeit braucht). Bei spezifischen sozialen Gruppen (wie z.B. Alleinerziehenden, 'Singles', 'working poors' usw.) liegen zudem keine brauchbaren Adressen vor, und teilweise ist eine repräsentative Stichprobe nur via telefonischen Suchinterviews überhaupt erreichbar (was entsprechende zeitliche und finanzielle Aufwendungen beim Sampling einschliesst).

4) Kurzbeschreibung unterschiedlicher Samplingverfahren

Es gibt unterschiedliche Samplingverfahren und -strategien. Im Folgenden werden wichtige Verfahren kurz aufgeführt:

A) Einfache Zufallsstichprobe (simple random sampling)

Diese statistisch optimalste und theoretisch klare Auswahlmethode bringt in der Praxis oft Schwierigkeiten mit sich. Die einfache Zufallsstichprobe ist deshalb in der Forschungspraxis weniger verbreitet als in statistischen Lehrbüchern, da die einfache Zufallsstichprobe ein guter Ausgangspunkt zur Darstellung einer Zufallsauswahl ist. Die einfache Zufallsauswahl setzt im Wesentlichen voraus, dass von der Grundgesamtheit eine vollständige (Adressen)kartei vorliegt, in der jedes Mitglied einmal vertreten ist. Aus dieser Kartei werden rein zufällig so viele Adressen gezogen wie benötigt werden. Die Auswahl geschieht etwa dadurch, dass jede Kartei eine Nummer erhält, und Zufallszahlen bestimmen, welche Kartei gezogen werden. Durch diese Methode hat jede Person die gleiche Chance, ausgewählt zu werden.

Eine einfache Zufallsauswahl funktioniert somit am besten, wenn aktualisierte und vollständige Adressenlisten vorliegen (z.B. Studentenverzeichnis, Mitgliederlisten, Einwohnerkarten). Dort wo – wie etwa in den USA oder Grossbritannien – keine aktualisierten Einwohnerregister vorhanden sind, wird häufig das sogenannte ‚random-route-Verfahren‘ gewählt, d.h. nach einer zufallsgesteuerten Auswahl werden innerhalb einer gegebenen Region Haushalte ausgewählt (und innerhalb des Haushalts zufallsgesteuert Haushaltsbewohner ausgewählt).

B) Systematische Zufallsstichprobe (systematic sampling)

Bei grösseren Grundgesamtheiten - z.B. Einwohner einer Stadt - ist die einfache Zufallsauswahl (mit ihrer Durchnummerierung aller Personen und Zuordnung von Zufallszahlen) faktisch zu aufwändig. In der Praxis hat sich deshalb vielfach die systematische Zufallsstichprobe durchgesetzt. Man wählt die Adressen aus einer Kartei nicht mit Hilfe von Zufallszahlen aus, sondern zieht systematisch jede x -te Adresse. Der Abstand zwischen den gezogenen Adressen bleibt fix und berechnet sich gemäss der Formel $k = N/n$. Z.B.: In einer Gemeinde von 10'000 Stimmbürger/innen wird bei einer Stichprobe von 500 Adressen jede 20 Einwohnerkarte ausgewählt. Die Zufälligkeit der systematischen Auswahl ist dadurch gegeben, dass die erste auszuwählende Adresse mit Hilfe einer Zufallszahl aus den ersten k -Adressen gezogen wird.

C) Geschichtete Stichprobenauswahl (stratified sampling)

Das Ziel einer Stichprobenauswahl ist es, möglichst genaue Ergebnisse zu erzielen bzw. den Stichprobenfehler möglichst gering zu halten, ohne die Kosten allzu sehr in die Höhe zu treiben. Ein einfaches Mittel, um bei gleicher Stichprobengrösse eine höhere Genauigkeit bzw. einen geringeren Stichprobenfehler zu erzielen, ist eine geschichtete Stichprobenauswahl. Bei diesem Verfahren werden einzelne bedeutsame Merkmale der Zielgruppe (wie Geschlecht, Alter, Zivilstand, Wohnort usw.) schon von vornherein in ihrem richtigen Verhältnis im Sample repräsentiert.

Beispiel: Bei einer Befragung von Studierenden an einer gegebenen Universität weiss man, dass sich die Studierendenpopulation zu 55% aus Männern und zu 45% aus Frauen zusammensetzt. Der Stichprobenplan wird so organisiert, dass die Stichprobe von vornherein 55% Studenten und 45% Studentinnen enthält, z.B. dadurch dass die (systematische) Zufallsauswahl je nach Geschlecht getrennt durchgeführt wird. Das Merkmal Geschlecht entfällt bei der eigentlichen Zufallsauswahl. Damit wird auch der Stichprobenfehler reduziert, da ja der Stichprobenfehler nichts anderes ist als der Einfluss der Zufälligkeit auf die Zusammensetzung der Stichprobe. (Bei geschichteten Stichproben ist der Stichprobenfehler die gewichtete Durchschnittsvarianz der Standardabweichungen innerhalb der einzelnen Schichten).

Je mehr Stichprobenschichten (sample stratifications) gebildet werden, desto mehr wird der Stichprobenfehler bei sonst gleicher Samplegrösse verringert. Eine geschichtete Stichprobe bleibt dennoch eine zufällige Stichprobe, da die Auswahl der Zielpersonen innerhalb jeder Schicht nach den Regeln statistischer Zufälligkeit erfolgt.

Die faktisch am häufigsten benützten Schichtungsmerkmale sind Geschlecht, Alter und Wohnort. Sofern die vorliegenden Adressen dies erlauben, wird unter Umständen auch nach Zivilstand oder Berufskategorie geschichtet (wobei solche Kriterien teilweise darunter leiden, dass etwa Zivilstand und notierter Beruf nicht (mehr) den tatsächlichen Lebens- bzw. Berufsverhältnissen entsprechen). Bei Telefonbefragungen ist schon aufgrund des Charakters des Telefons (haushaltsgebundener Telefonanschluss kombiniert mit individualisierenden Mobiltelefons) oft eine sorgfältige Schichtung nach Haushaltsgrösse bzw. Benutzerzahl pro Telefonanschluss notwendig, um nicht etwa Ein-Personen-Haushalte gegenüber Mehr-Personen-Haushalte zu privilegieren bzw. zu verhindern, dass Personen mit mehr als ein Mobiltelefon höhere Befragungschancen aufweisen als Personen mit nur einem Telefon.

Werden die einzelnen Stichprobenschichten proportional zur ihrer Verteilung in der Grundgesamtheit ausgewählt, spricht man von einer **proportional geschichteten Stichprobe** (propor-

tionate stratification). In anderen Fällen kann jedoch eine **disproportional geschichtete Stichprobe** (disproportionate stratification) sinnvoll sein. Eine bestimmte Bevölkerungsgruppe kann in der Stichprobe gezielt übervertreten sein (oversampling), um genügend Interviews für diese spezifische Gruppe zu erhalten. Beispiel: Umfrage bei Müttern. Es ist klar, dass gegenwärtig in der Schweiz die meisten Mütter in einer Partnerbeziehung leben. Daneben gibt es aber auch ledige, verwitwete und geschiedene Mütter. Werden etwa 600 Mütter interviewt, und die Gesamtpopulation enthält nur 6% ledige Mütter ohne Partner, werden demgemäss nur etwa 36 ledige Mütter befragt. Dies ist für Aussagen zur Situation lediger Mütter sicherlich zu wenig. Es kann daher sinnvoll sein, ledige Mütter schon bei der Stichprobenauswahl stärker zu berücksichtigen (so dass z.B. 100 Interviews mit ledigen Müttern vorliegen). Um bei proportionaler Schichtung 100 ledige Mütter zu erhalten, müssten sonst 1667 Interviews durchgeführt werden. Bei gerontologischen Studien werden nach Alter disproportional geschichtete Stichproben oft auch benützt, um genügend Hochbetagte zu erhalten (vgl. Betémps et al. 1997). Ein disproportionaler Sampling (mit oversampling) ist vor allem sinnvoll, wenn neben einem repräsentativen Querschnitt der Bevölkerung auch Minderheiten oder zahlenmässig schwach vertretene Gruppen speziell analysiert werden sollen.

D) Klumpen und Mehrstufenverfahren (cluster and multistage sampling)

Sofern eine Befragung eine umfassende Region oder gar ein ganzes Land abdeckt, ergeben sich zusätzliche Probleme: Zum einen sind in der Schweiz die Einwohnerkarteien kommunal dezentralisiert. Zum anderen setzt eine grossflächige Stichprobe ein fein verteiltes Interview-Netz voraus. Namentlich bei mündlichen Befragungen entstehen umso mehr Reisekosten, je dezentralisierter die Befragung ist.

Aus diesen Gründen wird bei mündlichen Interviews (face-to-face-interviews) teilweise ein Klumpenverfahren (cluster) benützt, um die regionale Verteilung von Interviews einzugrenzen. Anstatt z.B. aus 3000 Schweizer Gemeinden eine Adressenauswahl zu treffen, beschränkt man sich auf z.B. auf 60 Gemeinden, in denen etwa je 10 Interviews durchgeführt werden. Das heisst die Stichprobe wird in 60 Klumpen à je 10 Interviews aufgegliedert (vgl. dazu etwa Jann 2006 (Berner Stichprobenplan).

Ein Klumpenverfahren reduziert den Aufwand für Adressenbeschaffung und Interviewspesen (vor allem innerhalb dünn besiedelter Gebiete). Andererseits erhöht jedes Klumpenverfahren den Stichprobenfehler. Die Cluster sollten demzufolge nicht zu gross sein (um z.B. die Gefahr zu verhindern, dass ganze Gebiete wegfallen). (Faustregel: höchstens 20-30, eventuell auch nur 10 Interviews pro Cluster).

Bei schriftlichen, telefonischen oder elektronischen Befragungen sind Klumpenverfahren sachgemäss weniger nützlich, weil die Reisekosten wegfallen.

Die einzelnen Klumpen (cluster) (z.B. Befragungsgemeinden) können ebenfalls via Zufallsverfahren ausgewählt werden. Wird innerhalb einzelner Klumpen eine weitere Auswahl getroffen, spricht man von einer **mehrstufigen Stichprobe** (multistage sampling). Zum Beispiel wird zuerst eine (geschichtete) Auswahl von Bezirken oder Gemeinden gewählt. Innerhalb jeder Gemeinde wird die vorgesehene Adressenauswahl ebenfalls gemäss Zufallsprinzip gewählt. Oder bei einer Befragung von Pflegefachpersonen wird zuerst eine (eventuell nach Spitalgrösse geschichtete) Auswahl von Spitälern getroffen, und innerhalb jedes Spitals wird eine Abteilung gewählt, um danach eine Auswahl von Krankenschwestern zu treffen.

In Ländern, wo keine brauchbaren Bevölkerungskarteien oder Einwohnerregister vorliegen - wie z.B. den USA oder Grossbritannien - wird oft ein Klumpenverfahren verwendet, das unter dem Begriff der **Flächenstichprobe** bekannt ist: Auf einer Landkarte oder einem Stadtplan wird das in Frage kommende Gebiet in Einzelflächen oder Quadrate aufgeteilt. Nach Zufallsprinzip wird eine bestimmte Zahl von Planquadraten ausgewählt, und in jedem ausgewählten Planquadrat werden entweder alle Personen oder eine Zufallsauswahl davon befragt. Sofern die Einwohnerzahl der einzelnen Planquadrate allerdings unbekannt ist, können sich mit die-

sem Verfahren regionale Verzerrungen ergeben (z.B. in dem ländliche, grossflächige Gegenden überpräsentiert sind).

Ein anderes Verfahren, welches bei fehlender Adresseninformation – wie angeführt - ebenfalls häufig benützt wird, ist ein 'random-route'-Verfahren: InterviewerInnen müssen - ausgehend von einem gegebenen Standort - gemäss einfachen Zufallsregeln (drittes Haus von links, oberste Klingel) zu einem Hauseingang geleitet, um dort ein Interview durchzuführen. Ein solches Verfahren ist allerdings relativ aufwändig, und zudem muss gewährleistet werden, dass die Interviewer tatsächlich auch eine zufällige Route wählen.

E) Zwei-Phasen-Verfahren (two-phase-sampling)

Sofern kein geeignetes Adressenmaterial für die Personenauswahl zur Verfügung steht oder unklar ist, wie gross die Zielgruppe ist, sind die vorher diskutierten Sampling-Verfahren unbrauchbar. Dies gilt etwa für Befragung spezieller Gruppen (z.B. Befragung von Alleinerzieherinnen, Leuten, die letztes Jahr eine Wohnung gewechselt haben usw.). In solchen Fällen kann bzw. muss ein Zwei-Phasen-Verfahren (two-phase sampling) verwendet werden: In einer ersten Erhebungsstufe sammelt man aus einer grossen Zufallsstichprobe die notwendigen Informationen, etwa via telefonischen Suchinterviews. In der zweiten Stufe wird aus der ersten Stichprobe eine neue Stichprobe von Personen gezogen, welche die gewünschten Merkmale aufweisen. Z.B. sollen 300 Frauen befragt werden, die in feministischen Organisationen aktiv sind. In diesem Fall muss zuerst eine grosse Stichprobe von etwa 2000-3000 Frauen danach gefragt werden, ob sie politisch aktiv sind (z.B. via Telefon-Interviews). Aus der Zahl an erfassten aktiven Frauen werden 300 Frauen für das eigentliche Interview gezogen.

F) Panelstudie und Längsschnittuntersuchung (panel design/longitudinal study)

Die bisher vorgestellten Verfahren bezogen sich faktisch auf einmalige Querschnittsbefragungen (cross-sectional study). Geht es jedoch um Kausalzusammenhänge ist vielfach eine Längsschnittbetrachtung sinnvoll bzw. unumgänglich. In einem solchen Fall wird am besten eine Panelstudie durchgeführt, bei welcher die gleichen Personen zwei- oder mehrmals befragt werden, z.B. im Abstand von zwei Jahren. Die Stichprobeauswahl der ersten Panel-Befragung kann nachden oben angeführten Samplingverfahren durchgeführt werden. Bei der Zweitbefragung ergeben sich allerdings zusätzliche Probleme, und die Erhaltung der Stichprobe kann schwierig sein, z.B. müssen einige der Befragten neu aufgesucht und ihre Adresse muss verfolgt werden (etwa bei Wegzug). Zudem können sich Ausfälle durch Todesfälle oder Verweigerung des Zweitinterviews ergeben. Bei langjährigen Längsschnitt-Studien werden Ausfälle etwa durch Wegzug teilweise ersetzt, um die Stichprobengrösse zu behalten. Dabei wird nach sogenannten "statistischen Zwillingen" gesucht, d.h. es werden Personen neu ausgewählt werden, die weitgehend dieselben Charakteristika haben wie die ausgefallenen Personen. In anderen Panelstudien werden neue Altersgruppen an das Panel "angehängt", um z.B. auch die nachwachsende Generation zu berücksichtigen und zusätzlich zu Alterseffekten auch Kohorteneffekte zu überprüfen.

Die unter A) bis F) aufgeführten Verfahren sind alles Samplingverfahren, welche in irgendeiner Weise ein Zufallsverfahren einschliessen. Es sind somit Verfahren, welche die Repräsentativität zumindest der Stichprobenziehung gewährleisten. Es ist allerdings nochmals zu erwähnen, dass auch eine Zufallsstichprobe von Adressen noch keineswegs die Repräsentativität einer Umfrage gewährleistet. Verweigerungen können etwa zu systematischen Verzerrungen der Ergebnisse führen.

Neben den eigentlichen Random-Verfahren bestehen zusätzliche Auswahlverfahren für Umfragen, die je nach Forschungszwecken ebenfalls sinnvoll sein können. Zu erwähnen sind als "nonprobability sampling" namentlich:

G) Pragmatische Auswahl (haphazard, convenience or accidental sampling)

Es werden die Leute befragt, die gerade zur Verfügung stehen oder die man gerade auf der Strasse antrifft (z.B. Strassenbefragung). Oder es werden Fragebogen aufgelegt oder einer Zeitung bzw. dem Internet beigelegt und es werden diejenigen berücksichtigt, die antworten (z.B. Leserumfrage einer Zeitung). Wichtig ist die Tatsache, dass auch eine sehr hohe Zahl von Rückantworten - etwa auf einen aufgelegten Fragebogen oder eine Web-Befragung - keine Repräsentativität gewährleistet.

In einigen Fällen wird auch ein "Schneeballprinzip" verwendet: Befragte werden um Adressen von Freunden/Bekanntem gebeten. Das "Schneeballprinzip" ist etwa zur Untersuchung sozialer Netzwerke/ "peer-groups" usw. nützlich. Aber auch hier ist streng genommen keine Zufallsauswahl vorhanden.

H) Gesteuerte Auswahl (purposive sampling, expert choice)

Dieses Verfahren wird vor allem bei Experteninterviews benützt. Um beispielsweise Kenntnisse über eine bestimmte Unternehmung zu erhalten, ist es oft nicht sinnvoll, eine Zufallsauswahl aller Beschäftigten zu befragen, sondern für interne, organisatorische Dimensionen genügt die Befragung einiger weniger ausgewählter Personen, die sich auskennen. Expertenbefragungen werden oft auch für Delphi-Erhebungen – zur Erfassung von Zukunftstrends eingesetzt, z.B. Entwicklung der Gesellschaft, der Robotertechnik: Experten werden in einer ersten Runde über die Einschätzung bestimmter Entwicklungspfade befragt, das Ergebnis der ersten Runde wird den befragten Experten mitgeteilt und ist die Grundlage für eine zweite Befragungsrunde.

I) Quotensampling (quota sampling)

Eine Form nicht-randomisierter Befragung, die praktisch oft benützt wird, ist das Quota-Verfahren. Es findet etwa in Marktumfragen, aber auch bei vielen Umfrageinstituten regen Gebrauch. Es ist häufig, wenn ein Personenkreis befragt werden soll, für den kein (aktualisiertes) Adressenmaterial vorliegt. Der Unterschied zur Random-Stichprobe ist, dass hier keine Adressen zugrunde liegen, sondern die Zielpersonen werden nach "Quoten" ausgewählt: Der Interviewer hat keine Adresse zur Verfügung, sondern er sucht sich die Zielpersonen selber aus, allerdings nach genauen Vorgaben von bestimmten Merkmalen, eben den Quoten. Diese Quoten schreiben ihm etwa vor, wie viele Frauen und Männer er mündlich oder telefonisch zu befragen hat, wobei neben dem Geschlecht meist auch noch alters- und berufsbezogene Quoten einbezogen sind (Z.B. eine Interviewerin erhält den Auftrag, 3 Hausfrauen unter 30, 2 erwerbstätige Frauen unter 30 J., die in Büroberufen arbeiten, usw. zu interviewen). Auch bei elektronischen Befragungen – wo die Grundgesamtheit unbekannt ist – können Quotensampling-Verfahren eingesetzt werden.

Die Quoten entsprechen oft einer geschichteten Auswahl, welche eine Repräsentativität der Befragten zumindest gemäss den Quotenmerkmalen gewährleisten sollen. Um zu vermeiden, dass die Erreichbarkeit zu allzu starken Verzerrungen führt, sollten die Quoten auch Merkmale umfassen, die mit der Erreichbarkeit stark korrelieren.

Eine breite Streuung der Stichprobe erhält man bei einem Quoten-Sample dadurch, dass verschiedene Quoten-Merkmale miteinander gekoppelt werden (z.B. Geschlecht mit Alter mit Beruf). Je mehr Koppelungen vorgesehen sind, desto besser wird die Stichprobe, aber desto schwerer werden es die Interviewer finden, eine geeignete Person zu finden.

Die Quotenmethode ist seit Jahrzehnten Gegenstand von methodischen Diskussionen. Statistiker und Theoretiker werfen ihr mit Recht vor, dass sie letztlich keine Zufallsauswahl sei (womit man auch den Stichprobenfehler nicht abschätzen könne). Auch ergeben sich stille Verzerrungen dadurch, dass Interviewer sachgemäss primär jene Personen mündlich oder telefonisch befragen, welche leicht erreichbar und zugänglich sind. Dies gilt sowohl für mündliche wie telefonische Befragungen. Dennoch wird ein Quotensampling in der Markt-

und Umfrageforschung vor allem wegen seiner pragmatischen Vorteile oft verwendet, namentlich auch bei Telefonumfragen, welche innert kurzer Zeit durchgeführt werden müssen).

Benützte Literatur:

- Bétemps, Christine; Bickel, Jean-François; Brunner, Matthias; Hummel, Cornelia (1997) Journal d'une enquête: La récolte des données d'une recherche transversale par échantillon aléatoire stratifié, Lausanne: Réalités Sociales.
- Gabler, Siegfried; Hoffmeyer-Zlotnik, Jürgen H.P. (Hrsg.) (1997) Stichproben in der Umfragepraxis, Opladen: Westdeutscher Verlag.
- Gabler, S.; Häder, S. (Hrsg.) (2002): Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland. Münster u.a.: Waxmann.
- Faulbaum, Frank; Wolf, Christof (Hrsg.) (2006) Stichprobenqualität in Bevölkerungsumfragen, Bonn: Informationszentrum Sozialwissenschaften.
- Henry, Gary T. (1990) Practical Sampling, Sage, Newbury Park.
- Häder, Michael; Häder, Sabine (Hrsg.) (2009) Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung, Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hoffmeyer-Zlotnik, Jürgen H.P. (2006) Stichprobenziehung in der Umfragepraxis. Die unterschiedlichen Ergebnisse von Zufallsstichproben in face-to-face-Umfragen, in: Frank Faulbaum, Christof Wolf (Hrsg.) Stichprobenqualität in Bevölkerungsumfragen, Bonn: Informationszentrum Sozialwissenschaften: 19-36.
- Holm, Kurt (Hrsg.) (1986) Die Befragung 1. Der Fragebogen - Die Stichprobe, Tübingen: Francke-Verlag (3. Auflage).
- Jann, Ben (2006) Der Berner Stichprobenplan. Ein Vorschlag für eine effiziente Klumpenstichprobe am Beispiel der Schweiz, in: Frank Faulbaum, Christof Wolf (Hrsg.) Stichprobenqualität in Bevölkerungsumfragen, Bonn: Informationszentrum Sozialwissenschaften: 63-83
- Kalton, Graham (1983) Introduction to Survey Sampling, Beverly Hills: Sage.
- Neubäumer, Renate (1982) Die Eigenschaften verschiedener Stichprobenverfahren bei wirtschafts- und sozialwissenschaftlichen Untersuchungen, Bern/Frankfurt: Lang Verlag.
- Sudman, Seymour (1976) Applied Sampling, New York.

Ausschöpfungs- und Verweigerungsraten: Berechnung und Konsequenzen von Ausfällen bei Repräsentativumfragen

Nicht-Erreichbarkeit, ungültige Adressen und Verweigerungen erweisen sich bei allgemeinen Bevölkerungsbefragungen als immer schwerwiegenderes Problem. Aufgrund der höheren Mobilität sind mehr Adressen der Einwohnerämter nicht mehr gültig, und die Erreichbarkeit verschiedener Gruppen der Bevölkerung hat sich verschlechtert. Gleichzeitig kann die Bereitschaft, an einer Befragung teilzunehmen, nicht mehr vorausgesetzt werden, und die Verweigerungsraten haben sich in den letzten Jahrzehnten erhöht.

Nachfolgend stehen zwei Fragen im Vordergrund:

- a) Welche Formen von Ausfällen treten auf, und welches sind die wichtigsten Punkte einer sorgfältigen Ausschöpfungsanalyse?
- b) Welche Folgen für die Repräsentativität einer Befragung haben die verschiedenen Formen von Ausfällen bzw. Verweigerungen? Mit welchen Verzerrungen der Aussagekraft von Umfragen durch Ausfällen und Verweigerungen ist zu rechnen, und wie kann mit Ausfällen umgegangen werden?

Durch die Analyse und Diskussion dieser Fragen kann gewährleistet werden, dass grobe Fehlinterpretationen von Befragungsdaten vermieden werden.

1) Formen von Ausfällen und ihre Konsequenzen

1.1 Ausfälle aufgrund ungültiger Adressen

Ausfälle bei den Adressen ergeben sich durch Adressfehler bzw. ungültige Telefonnummer, nicht gemeldete Wohnortsänderungen oder noch nicht registrierte Todesfälle. Diese Ausfälle sind umso höher, je langsamer Adressänderungen verarbeitet werden, je unwirksamer die Einwohnerkontrolle funktioniert bzw. je mehr Leute sich nicht um eine korrekte An- und Abmeldung kümmern.

Vielfach werden diese Ausfälle als 'qualitätsneutrale' Ausfälle bezeichnet. Damit wird suggeriert, dass sich dabei keine sozialen Verzerrungen ergeben und dass diese Ausfälle ohne weiteres durch Ersatz-Adressen ersetzt werden können. Dies ist allerdings nur bedingt der Fall.

An und für sich muss bei adressbedingten Ausfällen zwischen zwei Formen unterschieden werden:

- a) Ausfälle aufgrund der Tatsache, dass die entsprechende Person nicht oder nicht mehr der definierten Grundgesamtheit angehört (z.B. verstorbene oder aus der Untersuchungsregion weggezogene Personen). Dazu gehören unter Umständen auch Personen, die nicht den von der Untersuchung verlangten Alterskategorien usw. angehören. Diese Ausfälle führen zu keinerlei Verzerrungen und können ohne Probleme durch Ersatzadressen ersetzt werden.

Sofern Einwohneradressen nicht mehr voll aktuell sind, kann die Zahl von Weggezogenen recht hoch sein, vor allem bei regionalen oder lokalen Studien. Ob Personen aus der Untersuchungsregion weggezogen sind - und deshalb nicht mehr zur Untersuchungspopulation gehören - oder ob sie innerhalb der Untersuchungsregion gewandert sind, ist im Einzelnen oft schwierig nachprüfbar. Im Fall veralteter Adressen bei hoher interregionaler Wanderung können sich auch insofern (kleine) Verzerrungen der Befragungspopulation ergeben, als Neuzuzüger/innen - die neu zur Untersuchungspopulation gehören - nicht berücksichtigt werden.

- b) Ausfälle aufgrund falscher oder nicht mehr gültiger Adressangaben von Personen der untersuchten Grundgesamtheit.. Hier handelt es sich primär um Personen, die innerhalb der Untersuchungsregion ihren Wohnort gewechselt, aber nicht gemeldet haben (z.B. bei gesamtschweizerischer Befragung: (noch) unangemeldete interne Migration, minus Auswanderung

ins Ausland). Bei E-Erhebungen können auch veraltete e-Mail-Adressen zu Ausfällen führen. Da geographische Mobilität lebenszyklisch variiert, ergeben sich durch solche Ausfälle insofern soziale Verzerrungen, als mobile Gruppen tendenziell unterrepräsentiert bleiben. Konkret handelt es sich vielfach um Jugendliche und junge Erwachsene (im Übergang zur Familiengründung). Bei Beginn einer Lehre oder eines Studiums bleibt etwa der Wohnort der Eltern weiterhin formeller Wohnort, ungeachtet der faktischen Wohnsituation. Junge Frauen ziehen zu ihrem Freund, ohne dies offiziell anzugeben usw. Offene und transitorische Lebensformen junger Erwachsener (nichteheliche Lebensgemeinschaften, Wohngemeinschaften, Alleinleben nach Scheidung usw.) sind deshalb in vielen Fällen eher unterrepräsentiert (verglichen mit stabilen Lebensformen (Ehe). Zunehmend häufiger, vor allem bei gehobenen Bildungsschichten, finden sich auch Muster von 'living apart together' bzw. von Wochenpendler/innen usw. Der faktische Wohnsitz stimmt nur noch zeitweise mit dem legalen Wohnsitz überein, was zumindest die Erreichbarkeit erschwert. Auch bei anderen Situationen - Spitalaufenthalt, Heimaufenthalt - stimmt legaler und faktischer Wohnort nicht überein.

Faktisch sind die beiden erwähnten Formen von Ausfällen deshalb nur bedingt 'qualitätsneutral'. Vor allem bei 'veralteten' Adressen ist die Wahrscheinlichkeit relativ hoch, dass hochmobile Gruppen (junge Erwachsene im Übergang, marginale Gruppen im Sinne von 'Landstreicher, Obdachlose) unterrepräsentiert bleiben. Besonders betroffen werden durch solche Ausfälle Schätzungen von Migrationsraten jeder Art (Wohnortwechsel, Wohnungswechsel, Übergang in Heim- und Spital usw.). Dabei besteht die Gefahr einer Unterschätzung sowohl von Zuwanderungsraten als auch von Abwanderungsraten.

Die Stichprobenausfälle aufgrund ungültiger oder falscher Adressen können von den Brutto-Ausgangsadressen abgezählt werden, um die bereinigte Ausgangsstichprobe (Netto-Ausgangsadressen) zu errechnen. Die nachfolgenden Ausfallraten (Anteil an nicht-erreichten Personen, Verweigerungsraten) können - sofern keine Ersatzadressen benützt werden - sowohl von der Brutto-Ausgangsstichprobe oder der Netto-Ausgangsstichprobe berechnet werden. In der entsprechenden englischsprachigen Literatur wird deshalb unterschieden zwischen:

- a) 'completion rate' (number of completed interviews or questionnaires divided by the total sample),
- b) 'response rate' (number of completed interviews divided by the effective sample (total sample minus the ineligible) (vgl. Goyder 1987: 9).

1.2 Ausfälle aufgrund von Nichterreichbarkeit

Was die Ausfallrate aufgrund von Nichterreichbarkeit betrifft, ist diese Rate einerseits von den gewählten Forschungsstrategien abhängig und andererseits mit sozialen Merkmalen von Befragten assoziiert.

Es ist offensichtlich, dass die Häufigkeit nichterreichbarer Personen von folgenden Aspekten des Untersuchungsdesign bzw. der Befragungsstrategie beeinflusst wird:

- a) Kontrolle der Ausgangsadressen auf Gültigkeit. Es ist naheliegend, dass eine gründliche Kontrolle und Bereinigung der Ausgangsadressen spätere Probleme der Nichterreichbarkeit bzw. die Häufigkeit vergeblicher Kontaktversuche reduziert. Faktisch wird eine Adresse oftmals erst nach diversen vergeblichen Kontaktversuchen als 'ungültig' erkannt.
- b) Dauer und Zeitraum einer Erhebungs- bzw. Interviewphase: Je länger die Interviewphase, desto höher ist die Wahrscheinlichkeit, dass Befragte erreicht werden. Bei mehrmonatiger Feldarbeit reduziert sich die Zahl von Befragten, die aus Urlaubsgründen, wegen kürzerem Spitalaufenthalt usw. nicht kontaktiert werden können. Die Nachteile einer längeren Erhebungsdauer liegen nicht allein in erhöhten Kosten, sondern möglicherweise auch in kon-

textspezifischen Veränderungen. So können bestimmte politische Ereignisse die Einstellungen von Befragten beeinflussen, so dass zu Beginn und gegen Ende befragte Personen systematisch anderes Antwortverhalten zeigen. Ebenso können konjunkturelle Einflüsse (Beginn einer Krise, Wiederaufschwung) bei längerer Befragungsdauer zu verstärkten periodenspezifischen Effekten führen. Befragungen, die sich über mehrere Monate hinziehen, können zudem saisonalen Effekten unterliegen, was namentlich bei Konsum- und Freizeitverhalten zu bedeutsamen Periodeneffekten führt.

Sachgemäss erweisen sich Ferienzeiten (Schul- und Betriebsferien) als ungünstige Befragungsperioden. Namentlich in den Sommermonaten sind viele Personen urlaubsbedingt abwesend, und grössere Bevölkerungsumfragen sollten wenn möglich ausserhalb der Haupturlaubssaison unternommen werden. Der zunehmende Reise- und Urlaubstourismus auch in den übrigen Jahreszeiten hat allerdings generell dazu geführt, dass die Erreichbarkeit namentlich junger mobiler Erwachsener sowie kaufkräftiger Bevölkerungsgruppen (z.B. reiche Rentner/innen) heute generell auf mehr Schwierigkeiten stösst, als vor zwei oder drei Jahrzehnten. Schönwetter- oder Schlechtwetterperioden beeinflussen die Erreichbarkeit ebenfalls, und vor allem bei kurzfristig durchgeführten Telefoninterviews, die auf Quoten-Sample basieren, können sich dabei subtile Verzerrungen ergeben (etwa wenn bei schönem Reisewetter primär wenig mobile, behinderte Rentner/innen Zuhause erreicht werden).

- c) Häufigkeit, Form und Zeitpunkt von Kontaktversuchen: Es ist naheliegend, dass die Häufigkeit von Kontaktversuchen die Erreichbarkeit beeinflusst. Allerdings gibt es auch in diesem Fall eine Optimierung zwischen Aufwand und Ertrag zu beachten (und eine genaue Protokollierung der Kontaktversuche erweist sich dafür als sinnvoll). Zunehmend häufig sind heute (schon aus Kostengründen) telefonische Vorkontakte. Die Beschränkung auf telefonische Kontaktversuche hat allerdings den Nachteil, dass einige Gruppen (z.B. ärmere Personen ohne eigenes Telefon oder Personen, die keinen eigenen Telefonanschluss haben) von vornherein ausgeschlossen werden. Der Zeitpunkt von Kontaktversuchen (auch abends oder am Wochenende) ist ebenfalls relevant, ansonsten Personen mit speziellen Arbeitsbedingungen (Schichtarbeit, Wochenendaufenthalter usw.) unberücksichtigt werden.

Erreichbarkeit bzw. Nichterreichbarkeit (an einem gültigen Wohnort) ist mit verschiedenen sozialen Merkmalen assoziiert. Zuhause leicht erreichbar, namentlich unter Tag sind etwa Hausfrauen oder andere nichterwerbstätige Personen (Betagte), aber auch Heimarbeiter/innen sowie Personen, die krankheitsbedingt zuhause bleiben usw. Speziell bei Umfragen auf der Basis von Quotensamples sind Hausfrauen häufig überrepräsentiert, vor allem im Vergleich zu vollerwerbstätigen Frauen. Schwieriger zu erreichen sind vor allem mobile Gruppen (sei es aufgrund ihrer beruflichen Mobilität (z.B. Handelsreisende, Pendler); sei es aufgrund ihrer ausserberuflichen Mobilität (z.B. reise- und urlaubsfreudige junge Erwachsene oder Rentner/innen). Abwesenheit ist im Allgemeinen speziell bei jüngeren, unverheirateten Personen vergleichsweise häufig, und sie nimmt nach einer Familiengründung bzw. Geburt von Kindern eher ab. Deutliche Ausfallraten aufgrund Abwesenheit sind deshalb primär in jüngeren Altersgruppen zu verzeichnen.

Daneben gibt es spezifische Bevölkerungsgruppen, die aus institutionellen Gründen nicht erreichbar sind, zumindest nicht für ein Interview. Dazu gehören Gefängnisinsassen (meist jüngere Männer) oder Personen, die sich längerfristig in Spitälern, Anstalten oder Heimen aufhalten. Selbst bei vorhandener Kommunikationsfähigkeit sind Personen in institutionellen Haushaltungen im Allgemeinen schlecht erreichbar, was dazu führt, dass sich die meisten Umfragen explizit oder implizit auf Personen in Privathaushalten beschränken. Auch in manchen Erhebungen wird die Stichprobe oft auf 'Zuhause lebenden Personen' limitiert. Dies führt mit steigendem Alter zu einer zunehmend stärkeren Selektivität der untersuchten Population

(z.B. in Hinblick auf Gesundheitszustand, Selbständigkeit, soziale Ressourcen usw.). So leben in der Schweiz (2008) fast 20% der über 80-jährigen Bevölkerung in sozio-medizinischen Einrichtungen.

Ganz allgemein führt Nichterreichbarkeit zu zwei sozialen Verzerrungen: Zum einen ergibt sich ein Bias in Richtung 'Häuslichkeit', in dem relativ häuslich ausgerichtete Personengruppen leichter erreichbar sind als mobile Gruppen. Dadurch werden namentlich berufliche und freizeitbezogene Mobilität und die Häufigkeit ausserhäuslicher Aktivitäten tendenziell unterschätzt. Zum anderen beschränken sich viele Befragungen explizit oder implizit auf Personen in Privathaushaltungen. Institutionelle Haushaltungen bleiben unberücksichtigt. Dies führt zu Bias namentlich bei Hochbetagten, aber auch bei der Untersuchung marginaler Subgruppen.

1.3 Ausfälle wegen Kommunikationsproblemen (Nicht-Befragbarkeit)

Nicht-Erreichbarkeit und Nicht-Befragbarkeit werden häufig nicht gesondert ausgewiesen. Es handelt sich jedoch um zwei grundsätzlich andere Ausfälle, die in weiten Teilen einer unterschiedlichen sozialen Verteilung unterliegen.

Ausfälle wegen Kommunikationsproblemen können sich namentlich aus den folgenden Gründen ergeben:

a) Sprachprobleme: Ein in der Landessprache durchgeführtes Interview kann nicht durchgeführt werden, sofern die Person der entsprechenden Sprache nicht mächtig ist. Dies betrifft einerseits ausländische Personen sowie andererseits Personen, die einer Sprachminderheit angehören (z.B. Rätoromanen). Mit zunehmender multikultureller Ausrichtung einer Gesellschaft verstärken sich solche Probleme (und entsprechende Übersetzungen des Fragebogens bzw. Mitarbeit von Übersetzern stossen auf finanzielle Grenzen). In der Schweiz ergeben sich auch bei einigen ländlichen Bevölkerungsgruppen einige Kommunikationsprobleme, als der Fragebogen nicht dem ortsüblichen Dialekt entspricht. Dies wirkt sich vor allem bei telefonischen und schriftlichen Befragungen nachteilig aus. Bei schriftlichen Befragungen kommt hinzu, dass ein nicht unbeträchtlicher Teil der erwachsenen Bevölkerung mehr oder weniger lese- und schreibunkundig ist. Speziell bei schriftlichen Befragungen ergeben sich deshalb meist namhafte bildungsbezogene Verzerrungen.

b) Hör- und Sehbehinderungen: Hör- und Sehbehinderungen können ebenfalls zu Ausfällen - sowie zu missverstandenen Antworten - führen. Da Hör- und Sehbehinderungen mit steigendem Lebensalter häufiger werden, ergeben sich diesbezügliche Kommunikationsprobleme namentlich bei gerontologischen Studien.

Bei der Befragung von betagten Personen sollte deshalb schon beim Aufbau des Fragebogens auf mögliche Kommunikationsprobleme geachtet werden. Dies schliesst zum einen gut lesbare Schriftgrösse (bei schriftlichen Interviews oder bei Antwortblättern und Karten mit Antwortvorgaben) ein. Zum anderen sollte bei mündlichen oder telefonischen Interviews auf eine gute Artikulation geachtet werden, und Fragen sollten nicht zu rasch vorgelesen werden. Dies schliesst eine entsprechende Schulung der Interviewer ein. Generell sollten Fragen zudem so kurz sein, dass sie das Kurzzeitgedächtnis nicht überfordern. Dies gilt vor allem für mündliche und telefonische Befragungen. Die allgemeine Regel, Fragen möglichst kurz und unkompliziert zu gestalten, impliziert auch, dass auf komplexe Formulierungen oder doppelte Negation verzichtet wird. In vielen Fällen scheint es besser zu sein, auf interne Antwortvorgaben zu verzichten, sondern auf externe Antwortvorgaben (etwa via Antwortkarten) zurückzugreifen.

William Foddy (1993: 99) schlägt generell vor, komplexere Fragen sowohl mündlich als auch schriftlich zu präsentieren, d.h. den Befragten die Möglichkeit zu geben, die entsprechende Frage nicht nur zu hören, sondern auch zu lesen. Dies gibt gerade älteren Menschen mit Hör- oder Sehproblemen sowie fremdsprachig aufgewachsenen Personen eine bessere

Chance, die Fragen zu verstehen (und die Interviewsituation zu bestehen, ohne dass indirekt ihre funktionalen Einbussen offen gelegt werden). Nach William Foddy (1993: 99) erweist es sich generell als vorteilhaft, den Befragten eine Kopie des Fragebogens zu übergeben: 'The author has found that this procedure has the added benefit of making respondents more comfortable (perhaps because it gives them something to do). It also eliminates the need for 'show' cards etc. While traditionalists might object to this procedure on the grounds that some respondents are likely to skip ahead and think about questions before the interviewer gets to them, the author has never found this to be a problem.' (Foddy, 1993: 99).

c) Kommunikationsunfähigkeit wegen Krankheit: Dazu gehören nicht nur Kranke, die aus ethischen Gründen nicht einem Interview zu unterwerfen sind, sondern auch Personen, die nicht ansprechbar sind. Dazu gehören namentlich stark depressive Personen sowie Personen mit Hirnleistungsstörungen (Alzheimer-Patienten). Sachgemäss nimmt der Anteil der Personen, die krankheitsbedingt nicht befragbar sind, in den höheren Altersgruppen deutlich zu.

Nicht-Befragbarkeit - aufgrund von Sprachproblemen oder Kommunikationsunfähigkeit - betrifft einerseits vor allem kürzlich eingewanderte Ausländer/innen aus fremden Kulturen, aber auch Angehörige kultureller Minderheiten (mit eigenem Dialekt). Zum anderen betrifft es Kranke und Behinderte, was primär bei der Befragung älterer Bevölkerungsgruppen zu namhaften Ausfällen führen kann. Aufgrund solcher Probleme nimmt gerade bei schriftlichen Befragungen die Rücklaufquote (und teilweise auch die Antwortqualität) mit dem Alter tendenziell ab (vgl. Kaldenberg/Koenig/Becker 1994). **Insgesamt wird dank solchen Ausfällen die kulturelle und physisch-psychische 'Normalität' einer Bevölkerung überbetont** (und kommunikative Probleme tendenziell unterschätzt). Aufgrund solcher Ausfälle geben Befragungsstudien häufig eine verbesserte Realität (**a slightly improved version of reality**) wieder (vgl. Lindström 1983: 81, Schnell 1991).

Eine Lösung, um dennoch Informationen etwa über hör-, sehbehinderte oder demenzerkrankte Personen zu erhalten, liegt in der Durchführung von Proxi-Interviews (Befragung von Angehörigen oder von Betreuungspersonen). Sachgemäss können damit höchstens soziodemographische und ausgewählte biographische Hintergrundvariablen einer Person erfasst werden.

Sowohl Probleme der Nicht-Erreichbarkeit als auch Probleme der Nicht-Befragbarkeit lassen sich durch eine vorgängige Spezifikation der zu untersuchenden Grundgesamtheit reduzieren, z.B. indem explizit nur Personen in Privathaushalten, nur Bürger/innen befragt werden oder eine obere Altersgrenze (nur Personen bis 75 J.) gezogen wird. Dadurch reduzieren sich die entsprechenden Ausfallquoten, allerdings sind die Einschränkungen der Studie bei der Interpretation der Ergebnisse zu beachten. So ergeben Studien, die sich ausschliesslich auf zu Hause lebende Betagte beziehen, ein sozial selektives Bild zur Situation von alten Menschen.

1.4 Ausfälle aufgrund von Verweigerungen

Die bisher diskutierten Ausfälle implizieren keine gezielte Entscheidung der zu befragenden Personen. Dies hat zur Folge, dass solche Ausfälle zumindest nicht direkt mit Einstellungen assoziiert sind. Dies ist bei Ausfällen aufgrund einer Verweigerung nicht der Fall, namentlich nicht bei Forschungsfragen, die mit der Teilnahme an bzw. Verweigerung einer Befragung assoziiert sind. Im Unterschied zur Nichterreichbarkeit ist die Verweigerung der Befragung eine unmittelbare Reaktion des Befragten vor Beginn des "Messvorgangs". Dabei wirken eine Anzahl von Faktoren mit, die auch beim Interviewvorgang von Bedeutung sind (Motivation zur Teilnahme, Probleme der Rollenfindung, Misstrauen gegenüber Fremden usw.).

Da ein Interview auf Freiwilligkeit beruht, ist in allen Studien mit einer geringeren oder größeren Zahl von Verweigerungen zu rechnen, wobei die angegebenen Gründe (z.B. Zeitmangel) nicht mit den effektiven Motiven, eine Befragung zu verweigern, übereinstimmen. Ausfälle durch Verweigerung gehören zu den üblichen Feldproblemen der Interviewforschung, wobei die Verweigerungsraten bei allgemeinen Bevölkerungsumfragen seit den 1950er Jahren insgesamt angestiegen sind (vgl. Steeh 1981). Eine ganze Reihe von Studien deutet darauf hin, dass die Antwortquoten (response rates) bei mündlichen und telefonischen Befragungen in den letzten Jahrzehnten in diversen Ländern gesunken sind (vgl. Goyder 1987 (USA, Kanada); Betlehem/ Kersten 1981 (Niederlande)). Eine Meta-Analyse von 45 Methodenstudien bestätigt den Trend zu sinkenden Antwortquoten in der Zeitperiode 1947 bis 1992 für mündliche und telefonische Befragungen (vgl. Hox/De Leeuw 1994: 338). Dieser Trend hat sich in urbanen Regionen auch in den letzten Jahrzehnten teilweise fortgesetzt (allerdings sind bei den Verweigerungsraten grosse soziale Unterschiede feststellbar).

Für mehr Verweigerungen scheinen zum einen strukturelle Wandlungen (wie zunehmende Urbanisierung) verantwortlich zu sein. Zum anderen führen sozio-kulturelle Wandlungen - in Richtung einer Individualisierung und Pluralisierung von Verhaltensmuster - dazu, dass die Teilnahme an vorgegebenen Rollen (wie Befragter) nicht mehr als selbstverständlich vorausgesetzt werden kann. Skepsis gegenüber Wissenschaft, aber auch temporäre Effekte - wie Verunsicherung und Anomie während Krisen- und Umbruchzeiten - können die Teilnahme an Befragungen ebenso negativ beeinflussen, wie etwa die Teilnahme an Abstimmungen und Wahlen. Dabei ist anzunehmen, dass die sinkende politische Partizipation (geringere Teilnahme an Abstimmungen) und die geringere Bereitschaft zu Interviews von ähnlichen Faktoren beeinflusst werden. Gleichzeitig ist - wie früher erwähnt - aufgrund höherer Mobilität mit verstärkten Problemen der Nichterreichbarkeit zu rechnen.

Substanzielle Ausfälle aufgrund von Verweigerungen sind heute deshalb in den Sozialwissenschaften keine Ausnahme, sondern die Norm. Damit stellt sich grundsätzlich auch die Frage, welche Befragungen heute tatsächlich noch das Prädikat 'repräsentativ' verdienen.

2 Verweigerungen: Raten, Strategien und mögliche Effekte

Im Folgenden sollen kurz drei Aspekte diskutiert und analysiert werden: a) die Berechnung von Verweigerungsraten, b) mögliche Strategien zur Reduktion von Verweigerungsraten, und c) mögliche Effekte von Verweigerungen auf die Aussagekraft von Umfragen.

2.1 Zur Berechnung der Verweigerungsraten

Am saubersten ist die Berechnung der Verweigerungen als Prozentsatz der Ausgangsadressen (Brutto-Verweigerungsrate und/oder als Prozentsatz der bereinigten Ausgangsadressen (Netto-Verweigerungsrate)). Dies ist jedoch nur möglich, wenn klar zwischen nicht-erreichten Adressen, nicht-befragbaren Personen und verweigerten Interviews unterschieden werden kann. Zudem lassen sich Verweigerungsraten im obigen Sinn nur errechnen, wenn Ausfälle aufgrund von Nichterreichbarkeit oder Nicht-Befragbarkeit nicht durch andere Adressen ersetzt wurden. Nach Benützung von Ersatzadressen lässt sich hingegen nur das Verhältnis von angefragten Personen und realisierten Interviews errechnen (Realisierungsquote), da die ursprüngliche Zufallsstichprobe (Ausgangsadressen) verändert wurde. Dabei ist zu beachten, dass das Verhältnis von angefragten Personen und realisierten Interviews nicht nur von der Häufigkeit von Verweigerungen bestimmt wird, sondern auch davon beeinflusst ist, in welchem Masse Ersatzadressen zur Verfügung standen bzw. benützt wurden (z.B. um eine vorher festgelegte Mindestzahl von Interviews zu erreichen).

Auch bei schriftlichen Befragungen sind Verweigerungsraten nicht zu berechnen (da nicht zwischen Falsch-Adressen, Nicht-Erreichbarkeit, Nicht-Befragbarkeit und Verweigerungen unterschieden werden kann). Bei schriftlichen Befragungen kann nur die sogenannte 'Rücklaufquote' berechnet werden (als Verhältnis zwischen versandten und ausgefüllt zurückgeschickten Fragebogen). Im übrigen gilt, dass auch beim Quoten-Sample - das streng genommen keine Zufallsauswahl einschliesst - keine Ausfall- oder Verweigerungsrate berechnet werden können (da keine Ausgangsadressen vorliegen). Es können allenfalls Zahl vergeblicher Kontakte bzw. verweigerter Interviews mit der Zahl realisierter Interviews verglichen werden.

2.2. Strategien zur Reduktion von Verweigerungen

Die Strategien zur Verringerung der Verweigerungen bzw. zur Erhöhung der Interviewbereitschaft sind je nach Form einer Befragung unterschiedlich.

Bei schriftlichen Befragungen variiert die Rücklaufquote enorm. Je nach Befragung kann sie 10% bis 70% erreichen. Bei allgemeinen Bevölkerungsbefragungen via schriftlichen Fragebogen sind Rücklaufquoten von über 50% als Erfolg zu verzeichnen. Höhere Rücklaufquoten ergeben sich nur bei spezifischen Befragungsgruppen, die gewohnt sind, Formulare zuverlässig auszufüllen (wie z.B. Beamte, Professionelle usw.). So ergab sich bei einer schriftlichen Befragung von Gemeindeforschern eine Rücklaufquote von über 75%. Schriftliche Befragungen sind deshalb - als allgemeine Regel - umso erfolgreicher, je gezielter und professioneller die angefragte Zielgruppe ist.

Daneben sind Faktoren wie Aufbau, Layout und Länge eines Fragebogens bedeutsam. Gut lesbare, einfach und klar aufgebaute Fragebögen, die nicht zu lang erscheinen, haben eine höhere Rücklaufquote. Ein portofreies Rückantwortcouvert erhöht die Chance einer Rückantwort ebenfalls. Mahnbriefe oder ein Zweitversand des Fragebogens - bei Personen, die noch nicht geantwortet haben - können die Rücklaufquote gleichfalls erhöhen. Inwiefern monetäre Anreize (Teilnahme an einem Wettbewerb, Gutschein) vertretbar sind, ist umstritten, aber es scheint, dass die Beilage einer kleinen Entschädigung oder eines Gutscheins die Antwortbereitschaft stärkt.

Die Meta-Analyse von Hox/De Leeuw (1994) weist darauf hin, dass die Rücklaufquote schriftlicher Befragungen im Zeitvergleich (1947-1992) eher angestiegen sind. Dafür lassen sich zwei Hauptfaktoren anführen: Zum einen hat sich das Ausbildungsniveau der Bevölkerung eher erhöht, und zum anderen ist die Durchführung schriftlicher Befragungen professioneller geworden.

Elektronische Erhebungen unterliegen den gleichen sozialen Gesetzmäßigkeiten wie eine schriftliche Befragung (Teilnahmequote abhängig von Vertrautheit mit schriftlich-elektronischer Kommunikation). Zusätzlich wird eine E-Befragung erleichtert, wenn Aufbau und Steuerung klar aufgebaut sind, und primär geschlossene Fragen (zum Ankreuzen) gestellt werden. Offene Fragen - wo ganze Textantworten erwartet werden - nerven.

Bei telefonischen Interviews ist die Schulung der Interviewer in Bezug auf Kontaktnahme ein entscheidender Faktor. Eine freundliche Stimme oder eine ansprechende Argumentation sind - wie bei anderen 'Verkaufsgesprächen' - wichtige Elemente. Daneben spielt auch der Zeitpunkt eine bedeutsame Rolle, und die Antwortbereitschaft sinkt während Essenszeiten, während populären Fernsehsendungen oder wichtigen Sportübertragungen. Insgesamt gesehen ist die Antwortbereitschaft bei (professionell durchgeführten) Telefoninterviews ebenso hoch wie bei mündlicher Befragung (vgl. Farago/Zeugin 1993: 448ff.). In einigen Fällen (z.B. bei relativ heiklen Themen) oder bei spezifischen Gruppen (z.B. Personen, die Mühe haben, jemand zu treffen oder in ihre Wohnung zu lassen) ist die Antwortbereitschaft bei Telefoninterviews sogar leicht höher als bei mündlicher Befragung (auch weil kein 'face-to-face'-Kontakt notwendig ist).

Im Allgemeinen ist die Verweigerungsquote bei mündlichen Interviews - aufgrund des persönlichen Kontakts und des "Zwangs zur Interaktion" häufig deutlich geringer als bei schriftlichen Befragungen. Allerdings stellen sich bei mündlichen Interviews eher Probleme der Erreichbarkeit. Dennoch sind Antwortquoten bei mündlichen Befragungen vielfach höher als bei schriftlichen Befragungen, vor allem bei breit angelegten Stichproben.

Die Interviewbereitschaft bzw. Verweigerungsrate bei mündlichen Befragungen ist durch eine Vielzahl oft spezifischer Faktoren bestimmt, so dass allgemeine Regeln zur Verminderung der Verweigerung einer mündlichen Befragung nur schlecht vermittelt werden können. Dies ist umso mehr der Fall, als die Interviewbereitschaft teilweise von kontextspezifischen Faktoren beeinflusst wird, auf die Forscher/innen wenig Einfluss nehmen können. So sind wirtschaftliche Krisenzeiten mit starker Verunsicherung weiter Bevölkerungskreise für Befragungen weniger günstig als Phasen allgemeinen Wohlbefindens. Wissenschaftskritische Auseinandersetzungen können die Bereitschaft zum Interview unter Umständen ebenfalls reduzieren, ebenso wie ein genereller Autoritätsverlust von Institutionen die Stellung der Wissenschaft (und damit wissenschaftlicher Studien) beeinträchtigen kann. Dennoch zeigt sich, dass wissenschaftliche Institute im Allgemeinen auf eine höhere Antwortbereitschaft stossen als Marketingorganisationen. Zudem stossen Forschungsthemen, die von den Befragten als wichtig angesehen werden, auf eine höhere Antwortbereitschaft (sofern das Thema nicht als zu eng mit dem Intimbereich assoziiert wird) (vgl. Hox/De Leeuw 1994: 337).

Trotz des Einflusses kontextueller und periodenspezifischer Faktoren können erfahrungsgemäss einige organisatorische Faktoren aufgeführt werden, die bei mündlichen Befragungen zur Reduktion der Verweigerungsquoten führen:

So ist die Erfahrung eines Interviewers ein bedeutsamer Faktor, und erfahrene Interviewer weisen deutlich weniger Verweigerungen auf. Es empfiehlt sich daher, grössere Umfragen nicht durch unerfahrene Student/innen durchzuführen. Eine gute Schulung und Begleitung des Interviewer-Stabes kann nicht nur die Interviewqualität erhöhen, sondern sie erlaubt es auch frühzeitig, spezifischen Problemsituationen zu begegnen (wodurch die Verweigerungsquote reduziert wird).

Interviewverweigerungen - oft durch Zeitmangel oder 'kein Interesse' rationalisiert - lassen sich zudem durch eine flexible Gestaltung des Interviewzeitpunktes oder -ortes (leicht) reduzieren. Eine freie Wahl des Zeitpunktes erhöht die Bereitschaft vor allem vielbeschäftigter Personen. Die Möglichkeit, ein Interview ausserhalb der Privatwohnung durchzuführen (z.B. in einem Restaurant), ist angebracht für Gruppen, die Mühe haben, fremde Personen in ihre Privatwohnung zu lassen (z.B. ältere, alleinlebende Frauen und Männer).

Inwiefern eine Voranmeldung und Vorankündigung die Interviewbereitschaft erhöht, ist hingegen weniger klar. Der 'Zwang zur Interaktion' kann bei unangekündigtem Erscheinen grösser sein, wogegen ein Ankündigungsbrief Personen die Möglichkeit öffnet, ein Nein zu formulieren. Eine Vorankündigung (schriftlich oder telefonisch) ist allerdings aus forschungsethischen Gründen angebracht. Unangekündigt erscheinende Interviewer/innen sollten zumindest einen offiziellen Ausweis tragen.

Die Wirkung eines unterstützenden Ankündigungsbriefs ist ebenfalls umstritten. Eine prestigereiche Unterschrift (Hochschulprofessoren, Regierungsmitglieder) kann einerseits die Bedeutung der Studie hervorheben, andererseits zu Abwehrreaktionen führen (vor allem bei oppositionellen Gruppen oder bei generellem Autoritätsverlust des 'Establishments'). Ein pointiertes Einführungsschreiben oder eine öffentliche Kampagne vor Beginn der Interviewphase kann ähnliche 'Backlash'-Effekte auslösen und unter Umständen eine verstärkte Selektion von Antwortenden und Verweigerern zur Folge haben. Ein neutrales Ankündigungsschreiben hingegen hat kaum negative, aber auch wenig positive Konsequenzen, zumindest für die Antwortbereitschaft. Eine telefonische Voranmeldung scheint besser zu sein, da auf Gegenargumente (namentlich: keine Zeit) gezielt Rücksicht genommen werden kann.

2.3 Auswirkungen von Verweigerungen auf die Aussagekraft von Umfragen

Auch bei der Verweigerung - wie bei der Nichterreichbarkeit - ist die Hauptfrage, ob sie ein mehr zufälliges Ereignis darstellt, das mit dem Inhalt der Befragung nicht korreliert, oder ob bestimmte Gruppen in feststellbarem Ausmass eher verweigern als andere Gruppen, wodurch die Repräsentativität der Befragung in Frage gestellt wird. Bei Praktikern der Umfrageforschung besteht die Tendenz, Verweigerung - wie Nichterreichbarkeit - als zufällige Ereignisse zu betrachten. **Allerdings gibt es viele Studien, die zeigen dass Interviewverweigerung gerade nicht zufällig geschieht, und dass Verweigerungen zu Verzerrungen der Aussagen bzw. zu Einschränkungen der Repräsentativität führen können.**

Die erste Frage ist, ob es eine Kerngruppe ('Hardcore') von Personen gibt, die Interviews prinzipiell und unabhängig von Thema oder Kontaktart ablehnen. Diese Frage ist nicht eindeutig zu beantworten, aber soweit ersichtlich gibt es kaum klare empirische Belege dafür, dass eine konstante Gruppe von Verweigerer existiert (vgl. Schnell, Hill, Esser, 1989: 287). Durchgehend schwierig für standardisierte Interviews zu gewinnen sind einzig 'Eliteangehörige' (oberste Kader, Regierungsmitglieder usw.). Diese Gruppe ist jedoch zahlenmässig relativ gering.

Selbst wenn der 'Hardcore' von Verweigern (momentan noch) gering ist, schliesst dies systematische und signifikante Beziehungen zwischen Interviewverweigerung und sozialen Merkmalen keineswegs aus. Einerseits hängt die Kooperationsbereitschaft zum Interview davon ab, inwieweit die Rolle des Informanten auch im Alltag eingeübt ist und allgemein Kontakte zu Fremden nicht ungewöhnlich sind. Andererseits spielen soziale Integration und soziale Teilnahme eine Rolle. Es fallen in erster Linie solche Bevölkerungsgruppen durch überdurchschnittliche Verweigerungsquoten auf, deren Beziehung zu ihrer sozialen Umwelt mehr oder weniger eindeutig mit "Disengagement" umschrieben werden kann. Bei den Verweigerern sind durchweg Personen mit niedrigem sozio-ökonomischem Status und geringer Schulbildung, Personen mit geringem sozialem oder politischem Engagement sowie Personen mit "Randpersönlichkeiten" übervertreten. Interviewverweigerung kann soziologisch als Ausdruck einer geringen Intensität der sozialen Verflechtung mit der dominanten Kultur im Alltag interpretiert werden. **In Bevölkerungsbefragungen sind damit in der Regel Personen umso weniger vertreten, je deutlicher sie von der vorherrschenden Mittelschichtkultur abweichen, und die meisten Befragungen weisen deshalb einen klaren "Mittelschicht-Bias" auf. Dieser Mittelschicht-Bias ist bei Quotensampling - aufgrund des Auswahlprozesses - und bei schriftlichen Befragungen - aufgrund der erforderlich hohen schriftlichen Kommunikationsfähigkeit - ausgeprägter als bei mündlich durchgeführten Interviews gemäss Randomverfahren.**

Durch Ausfälle (namentlich Nicht-Erreichbarkeit und Verweigerungen) kann die Verteilung von Merkmalen in bedeutsamer Weise verzerrt werden. Die Frage, ob zudem auch die Beziehungen zwischen Variablen verzerrt werden, ist nicht eindeutig beantwortbar (sofern die Ausfälle nicht zu einer starken Homogenisierung der befragten Bevölkerung führen). Es ist aber generell zu befürchten, dass tendenziell solche Personen befragt werden können, die:

- a) den Verhaltensanforderungen eines Interviews (als komplexer und geregelter sprachlicher Interaktionsvorgang) am ehesten entsprechen.
- b) von ihren Eigenschaftskonstellationen die von Sozialforschern - des gleichen sozio-kulturellen Kontextes - formulierten Hypothesen stärker stützen als andere Populationsteile.

Dies hat zwei Konsequenzen: Eine Befragung wird erstens aufgrund der Verweigerungen als unproblematischer erfahren als sie tatsächlich ist. Zweitens ergibt sich aufgrund der Resultate

eine Standardisierung und Normalisierung von Ergebnissen, wodurch standardisierte Befragungen das Bild einer (gutbürgerlichen) Gesellschaft eher stärken.

Diese 'Normalisierungstendenz' wird durch standardisierte Kategorisierungen und Auswertungen noch verstärkt (da spezielle, seltene Konstellationen bei der Auswertung meist zusammengefasst werden). Es ist in diesem Rahmen auch darauf hinzuweisen, dass grössere Bevölkerungsbefragungen soziale und politische Ungleichheiten auch deshalb abschwächen, weil Befragte ungeachtet ihres wirtschaftlichen, sozialen und politischen Einflusses gleichgewichtet werden. Befragungen verstärken damit demokratische Gesellschaftsstereotype, die von den bestehenden sozialen Machtungleichgewichten absehen.

2.4 Zum Umgang mit Ausfällen und Verweigerungen

Selbst bei sorgfältigster Planung und Durchführung einer Befragung lassen sich Ausfälle und Verweigerungen kaum vermeiden, und wie erwähnt ist unter heutigen gesellschaftlichen Bedingungen meist mit recht substantiellen Ausfällen und Interviewverweigerungen zu rechnen. Umso bedeutsamer sind möglichst frühzeitige Überlegungen, zum späteren Umgang mit Ausfällen und Verweigerungen. In diesem Zusammenhang existieren diverse, mehr oder weniger sinnvolle Strategien, die im Folgenden kurz aufgeführt werden:

a) Redefinition der Untersuchungspopulation: Nach Abschluss der Feldarbeit (und entsprechender Kontrolle von Ausschöpfungsraten für verschiedene Gruppen oder Regionen kann die Grundgesamtheit bzw. Untersuchungspopulation redefiniert werden. Beispielsweise wird bei einer Befragung der Wohnbevölkerung festgestellt, dass Personen in Kollektivhaushaltungen krass untervertreten sind (z.B. wegen Nichterreichbarkeit). Durch eine Einschränkung der Untersuchungspopulation (Personen in privaten Haushaltungen) kann die Repräsentativität erhöht werden (allerdings eben mit der Einschränkung, dass die Aussagen nur für Personen in privaten Haushaltungen gelten). Oder wenn festgestellt wird, dass fremdsprachige Personen krass untervertreten sind, kann es angebracht sein, die Auswertung nur auf Personen zu begrenzen, die die einheimische Sprache beherrschen. Auch bei ausgeprägten regionalen Differenzen der Ausschöpfung kann eine nachträgliche Limitierung der Untersuchungspopulation sinnvoll sein. Die Nachteile sind Informationsverluste bzw. unnütze Ausgaben für unausgewertete Interviews, die unter Umständen durch eine frühzeitige Limitierung der Befragung vermeidbar gewesen wären. Eine nachträgliche Redefinition bzw. Einschränkung der Grundgesamtheit ist zudem nur möglich, wenn klar abgrenzbare Populationsteile ausgeschlossen werden, die nur einen vergleichsweise geringen Anteil der Interviews ausmachen.

b) Verwendung von Ersatzadressen: Heute werden Ausfälle und zum Teil Verweigerungen häufig durch neue Interviewadressen ersetzt. Dementsprechend werden in den meisten Befragungen, die auf Adressen beruhen, schon von vornherein mehr Adressen 'gezogen'. Heute hat es sich eingebürgert, gut 40% bis 60% Ersatzadressen zu ziehen (auch um eine nachträgliche Belästigung von Einwohnerämtern zu vermeiden). Das explizite Ziel von Ersatzadressen liegt meist darin, eine bestimmte (vertraglich vereinbarte) Stichprobenzahl zu erreichen.

Damit wird zwar dem Gesetz der grossen Zahl Rechnung getragen, indem auch bei hohen Ausfall- und Verweigerungsraten die gewünschte Stichprobengrösse erreicht wird. Die Stichprobengrösse ist jedoch nur eine Voraussetzung für induktive Schlüsse. Teilweise besteht insofern Verwirrung, als eine grosse Stichprobenzahl an sich als genügende Voraussetzung für Repräsentativität angenommen wird. Es wird oft einzig auf die statistische Signifikanz geachtet, und je mehr 'signifikante Beziehungen' gefunden werden, desto besser. Es ist klar, dass signifikante Beziehungen bzw. Gruppendifferenzen umso häufiger auftreten, je grösser die Stichprobe ist, und bei genügend vielen Fällen ist faktisch alles signifikant. Die statistische Signifikanz wird nicht selten mit dem (parametrischen oder nicht-parametrischen) Erklärungsgrad (erklärte Varianz) vermischt. Bei grossen Stichproben können auch hochsignifikante Beziehungen wenig oder nichts erklären. Wie häufig Signifikanztests missverstanden wer-

den, zeigt sich darin, dass Signifikanztests oft auch bei Vollerhebungen, Zensusdaten oder internationalen Vergleichen, die alle Nationen einer Region einschliessen, benützt werden. Dabei wird vergessen, dass Signifikanztests nur sinnvoll sind, wenn nicht allein eine bestimmte Stichprobengrösse vorhanden ist, sondern es sich gleichzeitig um eine Zufallsauswahl handelt. Auch eine Befragung von dreissigtausend Personen ist nicht repräsentativ, sofern sie nicht nach einem Zufallsverfahren durchgeführt wurde (ein Punkt, der etwa bei grossen Leserumfragen oft vergessen geht).

Die Benützung von Ersatzadressen erhöht zwar die Zahl realisierter Interviews, löst jedoch das Problem der Ausfälle und Verweigerungen - und die damit verbundenen Verzerrungen nicht.

Eine höhere Zahl von Interviews durch Ersatzadressen erhöht die Repräsentativität somit nicht, sie führt bestenfalls zu einer höheren internen Heterogenität der Stichprobe. Mit höherer Zahl an realisierten Interviews nimmt die Zahl von Interviews auch aus schlecht erreichbaren Gruppen mit geringer Interviewbereitschaft zu. Dies führt zwar nicht zu einer Reduktion der Verzerrungen, aber verbessert die Möglichkeit von Gruppenvergleichen.

Durch zusätzliche Interviews via Ersatzadressen wird somit nicht die Repräsentativität der Verteilungswerte (Mittelwerte, Standardabweichungen usw.) erhöht, jedoch die Stabilität von Gruppenunterschieden. Die erhöhte interne Heterogenität bei grösserer Interviewzahl erklärt, wieso auch bei wenig repräsentativen Befragungen viele sozialen Gruppenunterschiede ähnlich sind wie bei Repräsentativbefragungen. Generell scheinen soziale Beziehungen (vor allem die Richtung von Gruppenunterschieden) bei hoher Interviewzahl von sozialen Verzerrungen durch Ausfälle und Verweigerungen weniger betroffen zu sein als die Verteilungswerte (allerdings gilt auch dies nur unter Bedingung, dass die Gruppenunterschiede der Ausschöpfungsrate nicht zu krass sind).

Relativ unproblematisch ist die Benützung von Ersatzadressen nur, wenn Personen aus der ursprünglichen Stichprobe nicht oder nicht mehr zur definierten Grundgesamtheit gehören (wie etwa verstorbene Personen). Wie früher erwähnt, führt der Ersatz 'ungültiger Adressen' höchstens dann zu leichten Verzerrungen, wenn auch die Ersatzadressen nicht aktuell sind (und damit beispielsweise Zuzüge weiterhin untervertreten sind).

Die Benützung von Ersatzadressen bei Ausfällen aufgrund von Nichterreichbarkeit führt theoretisch dann zu einer Verbesserung der Stichprobenqualität, wenn die Nichterreichbarkeit in Bezug auf zentrale Messvariablen gewissermassen 'zufälligen Charakter' hat. Dies ist allerdings faktisch kaum der Fall, da berufliche Mobilität, Ferienabwesenheit, aber auch Krankheiten bekanntlich sozial selektive Prozesse sind. Bei Ausfällen wegen Nicht-Befragbarkeit und Verweigerungen lässt sich das Problem der damit verbundenen Verzerrungen durch Ersatzadressen nicht lösen.

Auch die hie und da vorgeschlagene Verwendung statistischer Zwillinge löst das Problem nicht, da die zur Nichtbefragbarkeit und Verweigerung führenden Faktoren nicht korrigiert werden können. Die Verwendung statistischer Zwillinge - wie auch ein gezieltes Oversampling schlecht erreichbarer Gruppen - hat dennoch seine Vorteile: Werden nicht-erreichte Personen oder Verweigerer durch Personen ersetzt, die sozio-demographisch die gleichen Merkmale aufweisen (statistische Zwillinge) werden zumindest die sozio-demographischen Verzerrungen (nach Geschlecht, Alter, Haushaltszugehörigkeit, sozialer Schicht) reduziert. Damit kann zumindest die sozio-demographische Repräsentativität der Stichprobe verbessert werden. Dies ist vor allem wichtig, wenn die erhobenen Befragungsdaten auf die Gesamtpopulation hochgerechnet werden sollen und wenn sozio-demographische Variablen (wie Alter, Geschlecht) zentrale Erklärungs- oder Kontrollvariablen darstellen. Dies kann allerdings auch vorgängig, durch eine gezielte Schichtung der Stichprobe erreicht werden.

Inwiefern soziale Verzerrungen damit eliminiert werden, ist weniger klar. Nehmen wir beispielsweise an, dass die erste Befragungsrunde (ohne Ersatzadressen) einen klaren 'Mittel-

schicht-Bias' zeigt. Werden die Ersatzadressen - sofern möglich - schichtspezifisch geschichtet, kann der Mittelschicht-Bias korrigiert werden. Unterschicht und Mittelschicht wären zum Schluss entsprechend ihrer Verteilung repräsentiert. Nicht gelöst wird damit allerdings das Problem, dass innerhalb der Unterschicht und - weniger ausgeprägt - der Mittelschicht primär die besser integrierten, besser erreichbaren Segmente weiterhin übervertreten sind. Anstelle eines allgemeinen 'Mittelschichtsbias' haben wir zwei gruppenspezifische Verzerrungen. Die Repräsentativität in Bezug auf allgemeine Merkmale wird verbessert, die spezifischen, gruppeninternen Verzerrungen bleiben.

c) Nachträgliche Gewichtung von Gruppen: Eine Möglichkeit, um soziale Verzerrungen auszugleichen, besteht darin, die Gruppen mit geringerer Ausschöpfungsrate entsprechend stärker zu gewichten. Dies setzt detaillierte Informationen über die Ausschöpfungsraten nach Merkmalen voraus (z.B. Alter, Geschlecht usw.). Zumindest wird die sozio-demographische Repräsentativität verbessert, was namentlich bei Hochrechnungen bedeutsam ist. Das Problem sozialer Verzerrungen - etwa aufgrund von Verweigerungen - wird damit nicht gelöst. Im schlimmsten Fall kann es sich durch die Gewichtung noch verstärken. Nehmen wir beispielsweise an, dass in einer Bevölkerungsbefragung die Gruppe der 75 und mehr jährigen Personen untervertreten ist (weil nur gesunde, zu Hause lebende Betagte geantwortet haben). Durch eine Übergewichtung der 75 und mehr jährigen Befragten wird zwar die Altersstruktur 'repräsentativ', aber die Gesundheitsindikatoren werden zusätzlich verzerrt. Gewichtung kann somit nur eine Teillösung sein (deren Konsequenzen genau überprüft werden müssen).

d) Beibezug von zusätzlichen Informationen: Es ist sicherlich ideal, wenn Befragungsdaten durch andere Daten (statistische Daten, institutionelle Informationen) ergänzt werden können. Im Minimalfall können aggregierte Verteilungen verglichen werden, um sozio-demographische Verzerrungen der befragten Population zu überprüfen. Im Maximalfall können individuelle Daten zur Überprüfung der Validität von Aussagen und zur Analyse von Ausfällen bzw. Verweigerungen benützt werden.

Namentlich bei Kommunikationsproblemen kann auch das Verfahren von Proxi-Interviews benützt werden: an Stelle der ausgewählten Person wird eine ihr nachstehende Person (Hausangehörige, Nachbarn, Familienangehörige) über diese Person befragt. Sachgemäß können damit nur Tatbestände (und nicht Einstellungen oder subjektive Aspekte) erfasst werden, und die Qualität von Proxi-Interviews variiert je nach Familienzugehörigkeit. Proxi-Interviews bei Verweigerungen sind zwar in einigen Fällen theoretisch möglich, faktisch aber heikel. So ist es beispielsweise nicht ratsam, Informationen über erwerbstätige Verweigerer via Arbeitgeber einzuholen. Tatsächlich werden Proxi-Interviews primär bei Nicht-Befragbarkeit (z.B. im Rahmen sozio-medizinischer und gerontologischer Studien) eingesetzt.

e) Analyse der Ausfälle und Verweigerungen: Eine genaue Dokumentation der Ausfälle und Verweigerungen gehört zu jeder sorgfältig durchgeführten Befragungsstudie. Nur so lassen sich Verzerrungen - z.B. durch Vergleich mit Zensus-Daten usw. - und Ausschöpfungsraten überhaupt festhalten. Inwieweit eine Analyse von Verweigerungsgründe möglich ist, ist offen, da die angegebenen Gründe (keine Zeit, kein Interesse) vielfach nicht mit den tatsächlichen Gründen übereinstimmen.

Bei schriftlichen Befragungen hat es sich bewährt, Frühantwortende und Spätantwortende miteinander zu vergleichen. Da Spätantwortende im allgemeinen Nicht-Antwortenden ähnlicher sind als Frühantwortende, kann der entsprechende Gruppenvergleich zumindest einige Hinweise auf die Struktur der Nicht-Antwortenden vermitteln.

Bei mündlichen oder telefonischen Interviews kann ein Vergleich zwischen leicht und schwererreichbaren Personen ebenfalls nützlich sein, um Hinweise auf die Struktur der nicht-erreichbaren Personen zu erhalten.

Eine weitere Methode, zumindest zu einigen Informationen zu gelangen, besteht darin, die jeweilige Befragung durch eine Zweitmethode zu ergänzen. Zum Beispiel kann eine schriftliche Befragung durch eine telefonische Kurzbefragung ergänzt werden, um zumindest einige Merkmale bei Nichtantwortenden zu erhalten. Ebenso kann Personen, die im Rahmen einer mündlichen oder telefonischen Befragung nicht erreicht wurden oder die eine Befragung verweigert werden, ein schriftlicher Kurzfragebogen zugeschickt werden. Damit können zumindest einige sozio-demographische Grundmerkmale erfasst werden. Selbst wenn nur ein Teil der Verweigerer den Ergänzungsfragebogen zurücksenden, vermittelt dies einige Hinweise auf die sozio-demographischen Merkmale der Personen, die das mündliche Interview verweigerten. Zudem lässt sich der Datensatz ergänzen (was z.B. Hochrechnungen auf die Gesamtpopulation erleichtert).

f) Annahmen über die Verweigerer bei der Auswertung: Wenn ein Verdacht auf Verzerrungen der Daten durch Verweigerungen besteht, kann dies bei der Datenauswertung mitberücksichtigt werden. Denkbar sind verschiedene Verfahren. Ein radikales Verfahren besteht darin, Hypothesen unter der Annahme zu testen, alle Verweigerer hätten Antworten gegeben, die der Hypothese widersprechen würde. Sofern die Hypothese auch unter dieser Bedingung Bestand hat bzw. eine Beziehung auch bei dieser Annahme signifikant ist, kann sie durchaus als erhärtet gelten. Diese radikale Strategie führt allerdings bei hohen Verweigerungsquoten zur Relativierung aller Aussagen (was methodisch zwar adäquat ist, forschungspraktisch jedoch an die Grenzen der Selbstverleugnung stösst). Ein verfeinertes Verfahren besteht darin, abzuklären, wie hoch der Anteil der Verweigerer mit dem Merkmal x oder der Einstellung y sein müsste, damit die gefundene Beziehung insignifikant wird. Der Realitätsgehalt einer solchen Merkmalsverteilung bei den Verweigern - verglichen mit der Verteilung bei den Befragten - lässt sich zwar nicht beweisen, aber immerhin unter soziologischen Gesichtspunkten diskutieren. In jedem Fall lässt sich die Verallgemeinbarkeit von Aussagen durch solche Kontrollen auch bei Verdacht auf verzerrte Stichproben klar begründen.¹

Literaturhinweise

- Armstrong, J.Scott; Lusk, Edward J. (1987) Return Postage in Mail Surveys: A Meta-Analysis, *Public Opinion Quarterly*, 51: 233-248.
- Bétemps, C.; Bickel, J.-F.; Brunner, M.; Hummel, C. (1997) *Journal d'une enquête: La récolte des données d'une recherche transversale par échantillon aléatoire stratifié*, Lausanne: Réalités Sociales.
- Betlehem, J.G.; Kersten, H.M.P. (1981) The nonresponse problem, *Survey Methodology* 7: 130-156.
- Blass-Wilhelms, Wolfgang (1982) Der Einfluss der Frankierungsart auf den Rücklauf von Antwortkarten, *Zeitschrift für Soziologie*, 11: 64-68.
- Farago, Peter; Zeugin, Peter (1993) The Telephone Survey: A valid and reliable research method for the social sciences, *Schweiz. Zeitschrift für Soziologie*, 19: 445-462.
- Foddy, William (1993) *Constructing Questions for Interviews and Questionnaires. Theory and Practice in Social Research*, Cambridge: University Press 1993.
- Fox, Richard J.; Crash, Melvin R.; Kim, Jonghoon (1988) Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response, *Public Opinion Quarterly*, 52: 467-491.
- Frasch, Gerhard (1987) *Der Rücklaufprozess bei schriftlichen Befragungen: formale Modelle zur Analyse kollektiver Regelmässigkeiten*, Bern: Lang.
- Fuchs, Marek (1994) *Umfrageforschung mit Telefon und Computer. Einführung in die computergestützte telefonische Befragung*, Weinheim: Beltz.

- Gabler, Siegfried; Hoffmeyer-Zlotnik, Jürgen H.P. (Hrsg.) (1997) *Stichproben in der Umfragepraxis*, Opladen: Westdeutscher Verlag.
- Goyder, John (1985) *Face-To-Face Interviews and Mailed Questionnaires: The Net Difference in Response Rates*, *Public Opinion Quarterly*, 49: 234-252.
- Goyder, J. (1987) *The Silent Majority*, Cambridge: Blackwell.
- Häder, Michael; Häder, Sabine (Hrsg.) (2009) *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*, Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hoffmeyer-Zlotnik, Jürgen H.P. (2006) *Stichprobenziehung in der Umfragepraxis. Die unterschiedlichen Ergebnisse von Zufallsstichproben in face-to-face-Umfragen*, in: Frank Faulbaum, Christof Wolf (Hrsg.) *Stichprobenqualität in Bevölkerungsumfragen*, Bonn: Informationszentrum Sozialwissenschaften: 19-36.
- Hox, Joop. J.; De Leeuw, Edith D. (1994) *A Comparison of Nonresponse in Mail, Telephone, and Face-to-face Surveys. Applying multilevel modeling to meta-analysis*, *Quality & Quantity*, 28: 329-344.
- Kaldenberg, Dennis O.; Koenig, Harold F.; Becker, Boris W. (1994) *Mail Survey Response Rate Patterns in a Population of the Elderly*, *Public Opinion Quarterly*, 58: 68-76.
- Kalton, Graham (1983) *Introduction to Survey Sampling*, Beverly Hills: Sage.
- Lindström, Hakan (1983) *Non-Response Errors in Sample Surveys*, Oerebrö: Statistiska Centralbyran.
- Reubaud, Karl-Heinz; Blasius, Jörg (1996) *Face-to-face, telefonische und postalische Befragungen. Ausschöpfungsquoten und Antwortmuster in einer Grosstadt-Studie*, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 48,2: 296-318.
- Schnell, Rainer (1991) *Wer ist das Volk? Zur faktischen Grundgesamtheit bei 'allgemeinen Bevölkerungsumfragen': Undercoverage, Schwererreichbare und Nichtbefragbare*, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 43: 106-137.
- Schnell, Rainer (1997) *Nonresponse in Bevölkerungsumfragen. Ausmass, Entwicklung und Ursachen*, Opladen: Leske & Budrich.
- Steeh, Charlotte G. (1981) *Trends in Nonresponse Rates, 1952-1979*, *Public Opinion Quarterly*, 45: 40-57

Interviewer und befragte Personen – Interviews als soziale Situation und Interaktion

Das (mündliche Interview als künstliche dyadische Sozialbeziehung

Jedes Interview bildet einen sozialen Prozess, der mit allen Merkmalen dyadischen Verhaltens verknüpft ist, wie etwa Hang zur Selbstpräsentation, Verminderung von Dissonanzen, Wahrnehmung in Stereotypen usw. Beim mündlichen Interview handelt es sich allerdings um eine spezifische dyadische Sozialbeziehung. Vor allem drei soziale Aspekte kennzeichnen ein (mündliches oder auch telefonisches) Interview:

A) Auf beiden Seiten stehen sich relativ Fremde gegenüber. Sowohl Interviewer wie Befragter sind einander zu Beginn des Interviews fremd, und die Begegnung mit Fremden ist selten spannungslos; es muss Misstrauen überwunden werden; Befragte und Interviewer werden sich zuerst sozial einschätzen (auch was sozialen Status betrifft). Der Befragte wird versuchen, sich dem fremden Interviewer möglichst günstig darzustellen, was die Antworten beeinflussen kann.

B) Beim Interview handelt es sich um eine künstliche Sozialbeziehung, die vielen alltäglichen Begegnungen nicht entspricht: es werden vorgegebene Fragen gestellt und entsprechend klare Antworten verlangt; die Antworten werden notiert, das Interview ist von vornherein zeitlich begrenzt usw. Kurz und gut: ein Interview ist für die meisten Personen eine ungewohnte soziale Situation und diesbezüglich herrscht oft Verhaltensunsicherheit, vor allem bei Menschen, die nicht gewohnt sind, ‚ausgefragt‘ zu werden. Entsprechende Beobachtungen zeigen, dass spannungsreduzierendes Verhalten - wie z.B. Lachen, Kichern - bei Befragten unerwartet häufig auftreten, was auf ein gewisses Mass an Stress für die Befragten schliessen lässt.

C) Beim Interview handelt es sich um eine ausgeprägt asymmetrische soziale Situation: der Interviewer stellt Fragen und der Befragte hat sie zu beantworten. Es ist gerade kein Austausch von Höflichkeiten und kein gegenseitiges Gespräch. Ein Interview ist im Gegenteil eine einseitige kommunikative Angelegenheit: der Befragte muss über sich Auskunft geben, und zwar nur zu den Punkten, die explizit befragt werden, ohne dass er das Recht hat, auch den Interviewer auszufragen. Auf eine so offensichtlich einseitige Austauschbeziehung tendieren manche Befragte mit Ausgleichstrategien, etwa durch Gegenfragen: "Was würden Sie dazu sagen?", "Wie meinen Sie soll ich darauf antworten?"

Befragungsergebnisse sind nicht nur ein Produkt des Fragebogens, sondern sie werden auch durch Befragten- und Interviewerverhalten sowie die Interviewsituation beeinflusst oder im Extremfall verzerrt. Ein mündliches Interview bringt nicht einfach gewünschte "Reize" (= vorgelesene Fragen), sondern auch unerwünschte, das Untersuchungsergebnis möglicherweise verzerrende Reize. Verzerrungen durch soziale Intervieweffekte sind bei mündlichen und telefonischen Befragungen nicht auszumerzen, da sie unabdingbar mit dem Forschungsinstrument verbunden sind.

Interviewer-Effekte - im Sinne von unerwünschten Verzerrungen der Befragungsergebnisse - können sich vor allem in drei Richtungen ergeben:

- a) Interview-Effekte können das Ergebnis intentionaler Handlungen des Interviewers selbst darstellen; z.B. erwartungsbezogene Vercodungen, Beeinflussung des Befragten, Protokollierungsfehler. Interviewer-Erwartungen, wie sie sich aus dem Befragungskontext ergeben (z.B. bezüglich Geschlechtsrollen oder Statuserwartungen) können sich in unterschiedli-

chem Vercodungsverhalten niederschlagen. Dies geschieht umso eher und umso stärker, je weniger eindeutig die Reaktion des Befragten ist (d.h. in Abhängigkeit von Frage-Ambiguitäten). Generell kann man davon ausgehen, dass eine bewusste Einflussnahme des Interviewers auf den Befragten eher die Ausnahme als die Regel darstellt. Allerdings sind Probleme der nicht-bewussten Beeinflussung, aufgrund non-verbaler Kommunikation oder Verstärkung bestimmter Befragten-Reaktionen, damit nicht ausgeschlossen.

- b) Interview-Effekte können die Folge von Reaktionen des Befragten auf sichtbare Eigenschaften des Interviewers sein, z.B. aufgrund von Alter, Geschlecht, Kleidung, Auftreten). Die Anwesenheit eines Interviewers führt oft zur Antwortanpassung in Richtung auf allgemeine kulturelle Normen bzw. Stereotypen. Bei speziellen Eigenschaften des Interviewers wird jeweils ein situationspezifisches "looking good" aktualisiert. Zu erwähnen ist in diesem Zusammenhang, dass unter den Interviewern Frauen, Jüngere und Ledige oftmals stärker vertreten sind als bei den Befragten. Interviewer zeichnen sich oftmals durch eine höhere Schulbildung und höhere berufliche Stellung aus als Befragte (was Rückwirkungen auf Interviewer-Befragten-Verhältnis hat).
- c) Interview-Effekte können aber auch das Ergebnis eines Prozesses der wechselseitigen Beeinflussung von Interviewer und Befragten sein, z.B. Übertragung von Interviewer-Erwartungen). Ein guter Fragebogen für eine mündliche Erhebung sollte bekanntlich so aufgebaut sein, dass er - z.B. via natürlichen Gesprächsablauf - eine "gemütliche" und entspannte Interviewsituation herbeiführt. Damit erhöht sich die Antwortbereitschaft, aber gleichzeitig wird aber auch die soziale Dyade "Interviewer/Befragten" gestärkt, mit denkbaren Nebenfolgen, wie z.B. zunehmend kritischere Einstellung beider Personen zum Interview, abnehmende Ernsthaftigkeit, unnötige Verlängerung des Interviews, Bestärkung der Ja-Sage-Tendenz, immer stärker werdende nicht-verbale Kommunikation).

(Verzerrende) Effekte des Interviewers bzw. der Interviewsituation sind in ihrer Stärke auch von Fragedimensionen abhängig:

- a) Einstellungsfragen sind eher oder stärker anfällig für Interviewer-Effekte als Verhaltensfragen.
- b) Heikle bzw. unangenehme Fragen sind eher anfällig als nicht-heikle Fragen.
- c) schwierige Fragen sind eher anfällig als leichte Fragen (da schwierige Fragen des öfters ein Eingreifen bzw. Nachhacken des Interviewers erfordern und damit interviewspezifische Bedingungssetzungen wahrscheinlicher werden).
- d) offene Fragen sind eher anfällig als geschlossene Fragen, da bei offenen Fragen mehr Raum für Interviewer-Effekte vorliegt. Zudem können Interviewer die Antworten unterschiedlich genau notieren.
- e) durch den Interviewer nicht überprüfbare Antworten sind stärker anfällig als überprüfbare Antworten.

In welchem Masse Einstellungen und Erwartungen des Interviewers einen verzerrenden Einfluss haben, ist situationsabhängig. Die Einstellungen des Interviewers können etwa bei unklaren Antworten des Befragten die Codierung beeinflussen. In verschiedenen Studien wurden signifikante Korrelationen zwischen spezifischen Interviewer-Einstellungen und Antworten der Befragten festgestellt. Beim deutschen Allbus 1980 wurde z.B. festgestellt, dass die Einstellung der Interviewer zu "liberalen Erziehungszielen" die Wahrscheinlichkeit erhöhte, dass auch der Befragte "liberale Erziehungsziele" vertrat (Gamma: .32) (Grund: "liberale Erziehungsziele" galten 1980 als sozial erwünscht). Auch die Partesympathie von Interviewer und Befragten erwiesen sich auch nach Kontrolle von Drittfaktoren als signifikant korreliert. Dagegen hatten bei dieser Studie soziale Interviewer-Merkmale - wie Geschlecht, Alter, berufliche Stellung usw. - wenig Einfluss auf die Reaktion der Befragten (vgl. K.-U. Mayer, P.

Schmidt (Hrsg.), Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Beiträge zu methodischen Problemen des Allbus 1980, Frankfurt 1984: 72-113).

Bei schriftlichen oder elektronischen Befragungen ergeben sich Verzerrungen eher dadurch, dass die Befragungssituation nicht kontrollierbar ist, und bei einer schriftlichen oder elektronischen Erhebung ist unbekannt, ob der Fragebogen allein oder zusammen mit anderen Bezugspersonen ausgefüllt wird, oder ob die Befragung zu einem Zeitpunkt stattfindet, als sich Menschen sich aus irgendwelchen Gründen gestresst fühlten, ob der Fragebogen aufmerksam oder unaufmerksam beantwortet wurde usw. Noch stärker als bei mündlichen oder telefonischen Interviews ist bei schriftlicher oder elektronischer Befragung völlig unbekannt, ob die befragte Person die angeführten Fragen oder Antwortkategorien überhaupt verstanden hat oder nicht (und ein funktionaler Analphabetismus ist weiter verbreitet als häufig angenommen).

Die Befragten: Anforderungen an Befragten und Reaktionen:

In der Literatur zur Interviewmethode wird immer wieder festgestellt, dass die Lehre vom Befragten im Vergleich zum Wissen über andere Elemente der Befragung am geringsten entwickelt sei. Der Befragte wird als "Objekt" der Untersuchung wahrgenommen, und es werden teilweise primitive Stimulus-Reaktions-Modelle unterstellt (ein Denkverfahren, das empirischen "Verobjektivierungstendenzen" der Sozialwissenschaften entgegenkommt). Die Vernachlässigung der Verhaltensweisen des Befragten im gesamten Interviewprozess hängt auch damit zusammen, dass den Forschungsleitern nur geringe Möglichkeiten offen stehen, auf den Befragten im Sinne eines "optimalen" Interviewverhaltens einzuwirken.

Das Prinzip der Zufallsauswahl verbietet es, nur "qualifizierte" Befragte auszuwählen, und im Grunde muss ein Fragebogen auf die "dümmsten" Befragten ausgerichtet sein. Die Verwendung standardisierter Verfahren und die Anwendung sozialwissenschaftlicher Erhebungsinstrumente auf sehr heterogene Populationen setzt ein "standardisiertes", vereinfacht modelliertes Befragtenverhalten voraus. Schliesslich wird bei der statistischen Auswertung der Umfragedaten explizit von der prinzipiellen Vergleichbarkeit der Antworten auf vorliegende Fragestimuli ausgegangen.

Interaktionstheoretische Erklärungsmodelle zum Verhalten des Befragten wurden erst in den 70er Jahren systematisch entwickelt, namentlich von Hartmut Esser.

In mündlichen und telefonischen Interviews ergeben sich einige Anforderungen an den Befragten, die sich unmittelbar aus der typischen Interaktionsbeziehung als kurzfristige, konsequenzlose Beziehung einander fremder Personen in dyadischen Form ableiten lassen, z.B. Begegnung von zwei Fremden, die vorher und nachher keine Beziehung pflegen, Vorherrschen kognitiver Kommunikationsinhalte, Selbstdefinition des Befragten als "Auskunftsperson" und Datenträger usw. Die Reaktionen eines Befragten haben sich auf die Stimuli zu beschränken, die der Forscher in Gestalt von Fragen vorsetzt (und auch der Interviewer muss Fragen stellen, unabhängig davon, ob er sie gut oder peinlich findet). Eine weitergehende Reaktivität, etwa auf den Vorgang der Befragung selbst ("Meerschweinchen-Effekte, Interviewer-Effekte usw.) gelten als fehlererzeugendes Verhalten.

Bei standardisierten Interviews müssen befragte Personen an und für sich eine sozial eingeübte Verhaltensweise verinnerlicht haben, in einer Beziehung mit Fremden mitteilbare Meinungen und Informationen zu besitzen und sich zu ihrer Äusserung fähig und kompetent zu fühlen. Ein Interview erfordert vom Befragten ein hohes Mass der sprachlichen Kommunikationsfähigkeit (und es ist deshalb nicht verwunderlich, dass die Verweigerung eines Interviews stark schichtspezifisch variiert (in Form eines "Mittelschicht-Bias"). Ein Interview setzt auch

eine hohe semantische Einheit und Übereinstimmung innerhalb verschiedener Gruppen von Befragten voraus (so dass gleiche Begriffe inhaltlich für verschiedene Befragte gleiches bedeuten). Die Rolle des Befragten ist so gesehen recht anspruchsvoll (was einschliesst, dass Interviews nur in demokratischen Gesellschaften, die durch starke Mittelschichts-Normen durchtränkt sind und eine hohe Offenheit besitzen, sinnvoll durchgeführt werden können. Im allgemein lässt sich bei manchen Befragten eine Tendenz zur Antwortanpassung in Richtung auf allgemeine kulturelle Normen (social desirability) feststellen. Bei speziellen Eigenschaften des Interviewers wird zudem ein situationsspezifisches "looking good" aktualisiert.

Interviewer-Effekte in Telefonumfragen

Interviewer-Effekte bei Telefonumfragen zeigen sich einerseits bei der Gesprächseröffnung: Dass verbale Kommunikation während der Gesprächseröffnung Einfluss auf die Teilnahme in Telefonumfragen ausübt, wird durch zahlreiche Arbeiten belegt. Nach einer deutschen Studie (Hüfken, Schäfer 2003) wird die Realisierung eines Telefoninterviews begünstigt, wenn die Interviewer sich eher enthusiastisch und eher freundlich gaben; wenn sie sich in einer eher heiteren Stimmung präsentieren und sich interessiert zeigen. Als entscheidend erwies sich zudem auch eine persönliche Anrede der Zielperson. Personen, die mit ihrem Namen angesprochen werden, fühlen sich nicht nur als blosses Informationsobjekt unter vielen, sondern als Person, die wichtig ist und ernst genommen wird. Stimmliche Merkmale an sich erweisen sich nicht in allen Studien als bedeutsam, aber tendenziell zeigt sich, dass Gesprächseröffnungen, die freier und weniger abgelesen klingen sowie eine gute und passende Betonung aufweisen eher zu einem Interview führen.

Wenn befragte Personen Vorbehalte gegenüber einer Teilnahme am Interview geltend machen, zeigt sich ein positiver Effekt von Tailoring (= Informationen der Kooperationsablehnung des Befragten werden vom Interviewer als Basis für Gegenargumente benützt). Noch erfolgreicher ist eine Erweiterung der Argumentation für eine Beteiligung, die über die ablehnenden Argumente des Befragten hinausgehen.

Nach der Gesprächseröffnung ist vor allem eine akustisch klare und nicht zu rasche Befragung zentral, da sonst Missverständnisse auftreten können. Davon abgesehen sind die Interviewer-Effekte bei telefonischen Umfragen eher geringer als bei face-to-face-Umfragen, weil körperliche Signale weitgehend wegfallen. Alter und ethnische Zugehörigkeit fallen als Einflussfaktoren eher weg, und auch geschlechtsspezifische Reaktionsmuster sind bei telefonischen Umfragen weniger klar. Gleichzeitig erhält aber ein Interviewer auch weniger Signale bzw. Informationen darüber, ob Fragen verstanden wurden oder nicht.

Bei computerunterstützten telefonischen Bevölkerungsumfragen lassen sich auch die Antwortreaktionszeiten auswerten und etwa zur Aufdeckung von Response-Effekten verwenden, vgl. dazu: Mayerl, Jochen, Urban, Dieter (2008) Antwortreaktionszeiten in Survey-Analysen. Messung, Auswertung und Anwendungen, Wiesbaden: VS Verlag für Sozialwissenschaften.

Benützte Literatur:

- Erbslöh, Eberhard; Wiendieck, Gerd (1975) Der Interviewer, in: Van Koolwijk, Jürgen; Wieken-Mayser, Maria (Hrsg.), Techniken der empirischen Sozialforschung, Pp. 83-103, München: Oldenburg.
- Esser, Hartmut (1975) Der Befragte, in: Van Koolwijk, Jürgen; Wieken-Mayser, Maria (Hrsg.), Techniken der empirischen Sozialforschung, Pp. 107-129, München: Oldenburg.
- Esser, Hartmut (1986) Können Befragte lügen? Kölner Zeitschrift für Soziologie und Sozialpsychologie, 38, 1986: 314-336.
- Fowler, Floyd J.; Mangione, Thomas W. (1990) Standardized Survey Interviewing. Minimizing Interviewer-related Error, Newbury Park: Sage.
- Hüfken, Volker; Schäfer, Anne (2003) Zum Einfluss stimmlicher Merkmale und Ueberzeugungsstrategien der Interviewer auf die Teilnahme in Telefonumfragen, Kölner Zeitschrift für Soziologie und Sozialpsychologie, 55,2: 321-339.
- Koop-Steenstra, Hanneke Hout (2000) Interaction and the Standardized Survey Interview. The living questionnaire, Cambridge: University Press (M 1277).
- Lueger, Manfred (1989) Die soziale Situation im Interview, Österreichische Zeitschrift für Soziologie 14,3: 22-36.
- Mayerl. Jochen, Urban, Dieter (2008) Antwortreaktionszeiten in Survey-Analysen. Messung, Auswertung und Anwendungen, Wiesbaden: VS Verlag für Sozialwissenschaften.
- Reinecke, Jost (1991) Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte, Opladen: Westdeutscher Verlag.
- Scholl, Armin (1993) Die Befragung als Kommunikationssituation: zur Reaktivität im Forschungsinterview, Opladen: Westdeutscher Verlag.

Integrative Gesamtschau: Vor- und Nachteile einer standardisierten Erhebung

Vorteile einer standardisierten Erhebung

Forschungsmethode:

- Grosse Zahl von vergleichbaren Befragungen möglich, und damit Möglichkeit, repräsentative Ergebnisse zu erhalten.
- Von vornherein definierte Operationalisierung und Kategorienbildung, und damit Möglichkeit, vergleichbare Informationen zu erhalten.
- Standardisierte Daten, und damit Möglichkeit zur einfachen Codierung und raschen statistischen Auswertung der Daten.
- Standardisierung der Fragen und Antwortkategorien erleichtert Vergleichbarkeit von Antworten zwischen Befragten, aber auch zwischen verschiedenen Studien.
- Möglichkeit einer Wiederholbarkeit gleicher Fragesets, wodurch das Testen von Hypothesen erleichtert wird.

Forschungshintergrund:

- Es besteht ein breiter, ausdifferenzierter methodischer Erfahrungshorizont, und damit auch viel methodisches Wissen über die Optimierung einer Befragung.
- Es besteht eine Vielfalt an statistischen Auswertungsverfahren um standardisierte Befragungsdaten auszuwerten.
- Es ist ein polyvalentes Forschungsinstrument, das für viele Themen in analoger Weise einsetzbar ist (universalistisches Forschungsinstrument, das auch interdisziplinär eingesetzt werden kann).

Forschungsplan:

- Standardisierte Befragungen sind vergleichsweise gut planbar und ihr Aufwand ist gut abschätzbar.
- Eine Delegation der Befragungen an professionelle Instanzen ist möglich, und es bestehen spezielle Umfrageinstitute mit entsprechender Infrastruktur.
- Ansprüche an Interviewer/innen können gering gehalten werden bzw. es ist möglich, auch Nicht-Fachleute für die Befragung einzusetzen.
- Es besteht die Möglichkeit, Erhebung und Auswertung zu differenzieren (was für Forscher/innen eine wesentliche Zeitersparnis bringen kann).
- Aufwand und Anforderungen für Befragte kann tief gehalten werden.

Pragmatische Gesichtspunkte:

- Es ist eine wissenschaftlich gut akzeptierte Forschungsmethode, die bei Forschungsstellen und Geldgebern eine vergleichsweise hohe Akzeptanz genießt.
- Gute Publikationsmöglichkeiten dank empirischer Datenbasis.
- Es besteht die Möglichkeit einer Sekundärauswertung (d.h. die Daten können auch von 'Aussenstehenden' ausgewertet werden).

Nachteile einer standardisierten Erhebung

Forschungsmethode:

- Lässt wenig Spielraum für individuelle und subjektive Aspekte der Befragten und ihrer Situation offen.
- Es erfasst nur bestimmte Formen von verbaler Kommunikation, und die Methode ist für nicht-verbale Kommunikation oder unbewusste Sachverhalte nicht geeignet.
- Die Gefahr von Artefakten ist hoch, da Fehler (Missverständnisse, Falschantworten) nicht immer sichtbar werden.
- Es werden nur 'oberflächliche' Einstellungen/Meinungen erfasst, da eine Vertiefung und ein Nachfragen nicht möglich sind.
- Spezielle Einzelsituationen und individuelle Einzelschicksale gehen in der (massenstatistischen) Auswertung verloren (Tendenz zur 'Normalität')
- Es kann höchstens subjektiv wahrgenommene und verbal formulierbare 'Realität' erfasst werden.
- Eine standardisierte Befragung muss in ihrem Aufbau, ihrer Kategorienbildung, ihren Fragen usw. immer auf den kleinsten gemeinsamen Nenner ausgerichtet sein.

Forschungsplan:

- Es braucht eine aufwendige Vorbereitung und viel Vorwissen (und ist für explorative Studien deshalb ungeeignet).
- Möglichkeit der Routinisierung und der vorgegebenen Antwortkategorien reduziert die Chancen für neue Ideen, und die Methode wirkt eher konservativ.
- Vor allem bei Trennung von Erhebung und Auswertung können Fehlinterpretationen entstehen und Informationen verloren gehen.

Pragmatisch

- Eine standardisierte Umfrage kann auch ohne Forschungskonzept oder Theorie durchgeführt werden (Theoriedefizit).

Letzte Aktualisierung: November 2011.