

1. Übersicht über sozialwissenschaftliche Methoden; Geschichte Diekmann: 40-84

II. Probleme empirischer Sozialforschung

Forschungsmethoden

Diekmann, 40-76

Julian Brosch

071-2602641

julian.brosch@student.unisg.ch

Diekmann skizziert in diesem Kapitel, was für Probleme in der empirischen Sozialforschung auftreten. Er befasst sich mit der selektiven Wahrnehmung, Problemen bei der Prüfung von Hypothesen, sowie mit Werturteilsproblemen und Fragen der Forschungsethik. Alles in allem zeigt diese Kapitel ganz anschaulich auf, was für Fehler sich in der Forschung einschleichen können. Dabei geht es hauptsächlich um die Bewahrung von Subjektivität und das Ziel, den Einfluss von Werten, Vorurteilen, Wahrnehmungsfehlern, etc. möglichst gering zu halten. Dies ist aber vor allem in den Sozialwissenschaften kaum möglich.

1. Probleme selektiver Wahrnehmung

	<p>Geht es um die Aufdeckung und Prüfung sozialer Regelmässigkeiten oder Hypothesen, dann ist die vorurteilsgeleitete Wahrnehmung nicht selten trügerisch.</p>
Pseudo-Regelmässigkeiten	<p>Wir verspüren eine starke Neigung, Regelmässigkeiten auch dort zu erkennen, wo objektiv keine Zusammenhänge existieren. Diese Verknüpfung zufällig auftretender Ereignisse ist in der Psychologie die Grundlage „erlernter Neurosen“ und abergläubischen Denkens. Beispiel hierfür ist Wright und Bavelas' Experiment des „vielarmigen Banditen“</p>
Erwartungsabhängige Beobachtung	<p>In Polen wurde bei einer Umfrage festgestellt, dass die Befragten den ehemaligen General Jaruzelski immer mit einer dunkelbraunen Brille in Erinnerung haben. An diese Sonnenbrille erinnern sie sich auch im Zusammenhang mit Ereignissen, bei denen er gar keine trug. Dies zeigt, in wie starkem Mass Alltagsbeobachtungen von Vorurteilen und Erwartungshaltungen überlagert sein können. Weitere Beispiele hierfür sind die Entfernungsabschätzung von Ruhrpötlern im Bergland, oder die „hypothesengesteuerte Beobachtung“ von Bohnen (ja, so heisst der Mann... nicht die Hülsenfrucht). Soziale Wertschätzungen, Normen oder Konformitätsdruck tragen zu solchen Fehleinschätzungen bei.</p>
Selektive Wahrnehmung	<p>Friedrichs (1990) spricht im Anschluss an Untersuchungen der Massenmedien von einem „dreifachen Selektionsprozess“: (a) Stichprobenselektion – Auswahl bestimmter Selektionsobjekte; (b) Wahrnehmungsselektion – nur bestimmte Aspekte werden wahrgenommen; (c) Erinnerungsselektion – nur bestimmte Teile davon werden erinnert. Somit werden gewisse Sachverhalte nur selektiv und daher falsch wahrgenommen und ausgewertet. Wie Diekmann sagt: „Der Wunsch ist nicht nur der Vater des Gedankens; er steuert auch die Aufmerksamkeit und filtert die Beobachtungen.</p>
Wahrnehmung von Wahrscheinlichkeiten	<p>Die intuitive Wahrnehmung der Wahrscheinlichkeit von Ereignissen liefert besonders verzerrte Ergebnisse: kleine Risiken werden systematisch überschätzt und hohe Wahrscheinlichkeiten eher unterschätzt.</p>

Bestätigungsbias – kognitive Dissonanz	<p>Ausserdem werden bevorzugt Wahrnehmungen registriert, die liebgewonnene Vorurteile und Hypothesen bestätigen. Man spricht hier von einem Bestätigungsbias infolge gefilterter Wahrnehmung. Die Theorie der kognitiven Dissonanz kann dies zum Teil erklären [siehe Diekmann S. 25: „die Dissonanz zwischen zwei ‚kognitiven‘ Elementen (Vorstellungsinhalte wie Wahrnehmungen, normative Forderungen, etc.) eine Spannung, die im allgemeinen als unangenehm empfunden wird. ... Die Theorie prognostiziert nun, dass bei dissonanten Wahrnehmungen durch Umdeutung von Kognitionen versucht wird, die Dissonanz zu reduzieren.“]</p>
Bestätigungsbias – self-fulfilling prophecy	<p>Die self-fulfilling prophecy spielt ebenfalls eine wichtige Rolle bei der Bestätigung von Vorurteilen und Alltagshypothesen. Hier entwickeln Vorurteile und Diskriminierung oft eine Eigendynamik nach dem Muster „sich selbst erfüllender Prognosen“.</p>
Kasten II.1: Kreuztabelle und Bestätigungsbias	<p>Diese Probleme der selektiven Wahrnehmung können jedoch mit wissenschaftlichen Mitteln reduziert werden. Das einfachste Verfahren zur Prüfung einer Hypothese ist die Kreuztabelle. Ich werde hier nur schnell die Hauptaussagen von Kasten II.1 zusammenfassen. Grundsätzlich wird eine Hypothese dann bestätigt, wenn das Gewicht der Fälle in der Hauptdiagonalen die Nebendiagonale überwiegt. Darüber hinaus muss noch ein Schwellenwert definiert werden, der überschritten werden muss, um die Hypothese als bestätigt anzuerkennen. Diese Schwelle wird mit Signifikanztests festgelegt. Indem man eine Hypothese anhand einer Kreuztabelle untersucht, wird schon bei der Aufteilung der Daten (z.B. auf eine 2x2 Kreuztabelle) sichergestellt, dass keine mögliche Kombination ausgeschlossen wird. Alle Felder (in diesem Falle 4) sind eingeschlossen und können daher zur Überprüfung und Erklärung der Hypothese beitragen. Bei der selektiven Alltagswahrnehmung ist dies nicht der Fall: gewisse Felder werden von vornherein ausgeschlossen.</p>
Deduktionsfehler	<p>Ein AIDS Test habe eine „Falsch-negativ“ Quote von null und eine „Falsch-positiv“ Fehlerquote von 1%. Der Test wird demnach mit Sicherheit „HIV-positiv“ anzeigen, wenn eine Person infiziert ist, und mit 99% Sicherheit „HIV-negativ“, wenn eine Person nicht infiziert ist. Wenn jeder Tausendste tatsächlich infiziert ist, und Herr X bekommt die Nachricht, er sei HIV-positiv, dann nimmt er intuitiv an, dass die Nachricht mit 1% Wahrscheinlichkeit falsch ist. Er nimmt also an, dass er zu 1% wohl nicht infiziert ist. Diese Intuition ist jedoch falsch, da die tatsächliche Wahrscheinlichkeit 90,9% ist. Wenn z.B. 1 Million Leute getestet wurden, von denen 1000 infiziert sind, dann wird bei den restlichen 999 000 Personen 1% (9990 Personen) ein positives Testergebnis registriert. Das heisst, dass aus dem Kontingent der Nichtinfizierten 9990 Personen dennoch als infiziert diagnostiziert werden. Insgesamt werden dann also 10990 Briefe mit der Nachricht „HIV-positiv“ verschickt. Nur 1000 davon sind aber tatsächlich infiziert. Die Wahrscheinlichkeit, tatsächlich „negativ“ zu sein ist dann 90,9% (=9990/10990). Dieses Beispiel zeigt, dass intuitives Denken oft zu irrtümlichen Schlussfolgerungen führt.</p>
Empirische Forschung ist Detektivarbeit	<p>Kontrollierbarkeit in der empirischen Sozialforschung –im Gegensatz zur Alltagswahrnehmung- heisst, dass Verzerrungseffekte</p>

identifizierbar sind, ausgeblendet oder wenigstens vermindert werden können oder im Hinblick auf Richtung und Ausmass der Verzerrung berechenbar sind.

2. Probleme der Überprüfung von Hypothesen

Diekmann führt hier das Beispiel der vermuteten Effekte von Gewaltdarstellungen auf jugendliche Zuschauer ein. In diesem Zusammenhang diskutiert er eine Reihe von Problemen, die sich bei der Untersuchung verschiedener Hypothesen zu diesem Thema auf tun. Die Frage ist: Was für einen Effekt haben Gewaltdarstellung in Filmen auf die Gewaltbereitschaft von jungen Leuten. Ein Problem, das Diekmann aufführt ist das der Selbstselektion, die häufig als Kausaleffekte missdeutet werden. In diesem Beispiel hiesse dies, dass gewaltbereite Jugendliche bevorzugt Horrorfilme schauen; die Gewaltbereiten treffen also die Wahl, Gewaltfilme zu schauen. Dies steht im krassen Gegensatz zu der These, dass Gewaltbereitschaft erst durch das Anschauen von Horrorfilmen ausgelöst wird.

Eine strikte Versuchsanordnung mit einer Experimental- und einer Kontrollgruppe, deren Mitglieder zufällig zugeteilt werden, kann mögliche Selektionsverzerrungen evtl. verringern. 100%ige Gewissheit kann es jedoch auch hier nicht geben... So ist es nun einmal in der Sozialforschung.

Scheinkorrelation

Bei nicht-experimentellen Studien stellt sich das Problem möglicher Einflüsse von „Drittvariablen“. Ein extremer Fall ist die „Scheinkorrelation“, bei der ein nicht-kausaler Zusammenhang zwischen X und Y durch eine dritte Variable Z kausal produziert wird. Ein Beispiel hierfür ist die „Storchkorrelation“ (siehe Kasten II.2). Man kann in multivariaten statistischen Verfahren den Einfluss von Drittvariablen kontrollieren, indem man sie miterhebt sie bei der Datenanalyse rechnerisch konstant hält.

Bei experimentellen Studien wird durch den Prozess der Randomisierung (der zufälligen Aufteilung von Probanden auf die Versuchsgruppen) der störende Einfluss von Drittvariablen praktisch ausgeschlossen.

Weitere Probleme

Es kann ebenso zu Versuchsleitereffekten, Interviewereffekten, Problemen bei Stichproben, etc. kommen, welche zu Verzerrungen und Fehlern führen. Es ist praktisch unmöglich, sämtliche Fehlerquellen unter Kontrolle zu haben.

Erst mehrere Studien mit möglichst unterschiedlichen Methoden können etwas mehr Gewissheit geben. „Erst viele Mosaiksteine formen ein Gesamtbild“. Das Problem hier ist, dass zu selten Replikationen von Studien und Experimenten durchgeführt werden. Oft wird solch eine Arbeit einfach als zu wenig originell und kreativ empfunden. Daher werden lieber einzelne, unzusammenhängende Mosaiksteine produziert.

3. Werturteilsproblem und Forschungsethik

Auftragegeberinteressen und Forschungsergebnisse disharmonieren nur

Das Beispiel der Anthropometrie	<p>selten – daran haben wir uns gewöhnt. Interessen, Ideologien und Wertvorstellungen von Auftraggebern und Sozialforschern können einen erheblichen Einfluss auf die Forschungsergebnisse ausüben. Diekmann führt hier das Beispiel dieser anthropologischen Theorie an, welche besagt, dass ein positiver Zusammenhang zwischen Hirnmasse (gemessen an den Schädelhöhlen) und der geistigen Leistungsfähigkeit von Menschen besteht. Hiermit liessen sich dann die Vorurteile über die Rangordnung von Rassen und Geschlechtern untermauern.</p>
Werturteilsprobleme	<p>Werturteile sind präskriptive Sätze, die eine Handlung vorschreiben oder einen Sachverhalt bewerten. Sie können praktisch immer klar und transparent als Soll-Sätze formuliert werden.</p>
Vier Aspekte des Werturteilsproblems: 1. als Forschungsgegenstand 2. Die Wertbasis der Wissenschaften	<p>Werturteile können Gegenstand sozialwissenschaftlicher Forschung sein. Z.B. kann man den Wandel von Arbeitswerten oder Umweltwerten genauer unter die Lupe nehmen.</p> <p>Jede Wissenschaft kennt ethische Verpflichtungen, wie z.B. der Eid des Hippokrates in der antiken Humanmedizin. Dennoch gibt es im Bereich solcher forschungsethischen Verpflichtungen drei Probleme: (1) ohne Kontrolle und Sanktionsinstanzen ist eine Garantie der Einhaltung der Regeln nicht gegeben; (2) sind die Anwendungen wissenschaftlicher Forschungsergebnisse nicht immer vorhersehbar; und (3) sind die Möglichkeiten der Nutzung von Forschungsergebnissen vielfach ambivalent. „Eine Theorie der Revolution kann sowohl den Revolutionären als auch den Tyrannen nützlich sein, um ihre jeweils entgegengesetzten Ziele zu befördern“.</p>
3. Das Relevanzproblem	<p>Eine direkte Rolle spielen Wertentscheidungen bei der Auswahl von Forschungsproblemen (Relevanzproblem). Dies ist oft eine politische Entscheidung: „Die marxistische Sozialforschung in Osteuropa hat keine anderen Methoden verwendet, als sie in der westlichen ... Forschung im Gebrauch waren. Man hat allerdings das Forschungsinteresse stärker auf die Probleme der Arbeiter gerichtet“.</p>
4. Werturteile in wissenschaftlichen Aussagen	<p>Max Weber löste mit seiner Forderung, Wissenschaftler sollen auf wertende Stellungnahmen im Rahmen ihrer wissenschaftlichen Arbeit verzichten, den bis heute andauernden Werturteilsstreit aus. Für Weber sind Werturteile nicht objektiv begründbar und stellen nur die private Überzeugung des Wissenschaftlers dar. Dieser sollte aber nicht durch das Heranziehen wissenschaftlich-empirischer Aussagen eine Schein-Objektivität verliehen werden.</p> <p>Es ist wichtig, dass Werturteile nicht verschleiert werden und explizit erkennbar sind. Sie können in jeden Schritt einer empirischen Sozialforschung einfließen. Daher muss der Forscher die einzelnen methodischen Schritte genau dokumentieren.</p>
Persönlichkeitsschutz von Versuchspersonen	<p>Für Versuchspersonen muss in vielen Fällen Anonymität zugesichert werden und das Datenschutzgesetz respektiert werden. Ebenso sollte man Versuchspersonen im Nachhinein über die Ziele und Hypothesen des Versuchs aufklären. Ein gravierendes Problem stellen Versuche wie die sogenannte „Milgram-Studie“ dar, auch das „KZ Experiment“ genannt. Hier wurde die Gehorsamkeitsbereitschaft der Versuchsobjekte – sie mussten Elektroschocks verteilen – getestet. Die Effekte auf die Versuchspersonen können hier durchaus</p>

schwerwiegend sein.

Viele Wissenschaftsvereinigungen haben daher einen Ethik-Kodex, nach dem Personen keinen Nachteilen oder Gefahren ausgesetzt werden dürfen.

Oft ist es unmöglich, die Versuchspersonen im Voraus über Sinn und Zweck eines Experiments zu informieren, da dies die Ergebnisse verfälschen würde. In solchen Fällen sieht der Kodex der Deutschen Gesellschaft für Soziologie (DGS) vor, „andere Möglichkeiten der informierten Einwilligung zu nutzen“. Dies ist mit Sicherheit problematisch und durchaus verbesserungswürdig.

2. Wissenschaft; Wissenschaftstheorie

Diekmann: 150

Falsifikationismus und Konkurrenz von Forschungsprogrammen

Forschungsmethoden (Hug)

Diekmann, S. 150-159

Thomas Geissmann

071/223'84'15

thomas.geissmann@student...

Popper wie auch Lakatos und Kuhn betonen den Wert und die Notwendigkeit empirischer Forschung, sei es zur Prüfung und Falsifikation von Theorien (Popper) oder zur Aufdeckung von Anomalien von Forschungsprogrammen (Lakatos) und Paradigmen (Kuhn). Ein Ausgangspunkt der Kritik an Popper war das **Basissatzproblem** bzw. das **doppelte Risiko von Fehlentscheidungen**: falsche Basissätze zu akzeptieren oder wahre Basissätze zu verwerfen.

1. Ausgangsdilemma

Dilemma:

„**sichere Deduktion** (Herleitung des Besonderen aus dem Allgemeinen) **ohne empirischen Gehalt vs. unsichere, aber informative Induktion** (Herleitung von allgemeinen Regeln aus Einzelfällen)“

2. Poppers Falsifikationismus

Die **Kritik am Induktionsprinzip** diente als Ausgangslage für Poppers „Falsifikationismus“. **Aus einer endlichen Menge von Beobachtungen könne nicht auf einen allgemeinen Satz**, der sich auf eine unendliche Menge potentieller Beobachtungen bezieht, **geschlossen werden** (Bsp. „weisse Schwäne“: ich kann noch so viele weisse Schwäne beobachten, doch folgt daraus nicht, dass alle Schwäne weiss sind). Da **Hypothesen, Gesetze und Theorien** Aussagen über eine unendliche Menge von Objekten machen, sind diese grundsätzlich **nicht** anhand einer endlichen Menge von Beobachtungen **verifizierbar**.

Hingegen sind sie falsifizierbar: Steht auch nur ein wahrer Satz im Widerspruch zu einer (deterministischen) Hypothese, so gilt die Hypothese als falsifiziert. (falls bspw. irgendwo

ein schwarzer Schwan gesichtet wird, gilt die Hypothese „Alle Schwäne sind weiss“ als falsifiziert). Und: selbst wenn eine Hypothese von allen gegenwärtigen Beobachtungen „verifiziert“ (noch nie schwarze Schwäne entdeckt) wird, so könnte in Zukunft auch noch ein falsifizierendes Ereignis eintreten!

→ **Asymmetrie:**

Hypothesen und Gesetze sind nur (potentiell) falsifizierbar, nicht aber verifizierbar!

Das Interesse der Methodologie von Popper richtet sich allein auf die Begründung von Hypothesen!

Deduktiv-empirisches Modell Poppers (Ausdruck eines „kritischen Rationalismus“)

1. Wissenschaftler sollen („kühne“) Hypothesen mit möglichst **hohem Informationsgehalt** konstruieren und
2. diese einer Vielzahl empirischer Prüfungen in unterschiedlichen Situationen unterziehen („**Bewährungsproben**“).
3. Diejenigen gehaltvollen Hypothesen, die Falsifikationsversuchen widerstanden haben, werden (vorläufig) beibehalten; falsifizierte Hypothesen werden ausgesondert.

Man kann nie wissen, ob eine Hypothese tatsächlich und endgültig wahr ist. **Die Menge der potentiellen Falsifikationen ist immer unendlich** und daher alles Wissen unsicher und vorläufig: „Ich weiss, dass ich nichts weiss!“ (Sokrates) Das Wissen kann aber vermehrt werden, indem man Hypothesen harten Bewährungsproben unterzieht und diese so lange akzeptiert, wie sie allen Falsifikationsversuchen widerstehen.

3. Kritik an Poppers Falsifikationismus

3.1. Basissatzproblem: (Basissätze = singuläre Aussagen, die raum-zeitlich fixierte Beobachtungen beschreiben: „Der hier und heute beobachtete Schwan ist schwarz.“)
(→ Je nach Ausgang eines Experimentes stehen Hypothesen im Einklang (Bewährung) oder Widerspruch (Falsifikation) zu Basissätzen.)

Nun: selbst eine scheinbar unproblematische Beobachtung wie „vor mir steht ein Glas Wasser“ (Popper) erfordert stillschweigend die **Annahme einer Beobachtungstheorie**. Insbesondere bei historischen, nicht-reproduzierbaren Ergebnissen ist der Nachweis der Wahrheit von Basissätzen häufig schwer zu erbringen. Wenn nun die Akzeptanz von Basissätzen theorieabhängig ist, dann können die Basissätze natürlich auch falsch sein, denn die angewandte Beobachtungstheorie könnte sich in Zukunft als falsch herausstellen. Dann ist aber auch die Falsifikation einer Hypothese keine sichere Angelegenheit, denn der der Hypothese widersprechende Basissatz könnte ja falsch sein. **Mithin sind Theorien, Hypothesen und Gesetze weder mit Sicherheit verifizierbar noch falsifizierbar**. Bei der Prüfung von Theorien könnte somit der **Fall einer irrtümlichen Bewährung als auch der Fall einer irrtümlichen Falsifikation** auftreten → eine Hypothese sollte nicht schon bei einer einmaligen Falsifikation verworfen werden. (Bsp. mit Planet Neptun).

Die rigorose Position einer Verwerfung von Hypothesen schon nach einmaliger Falsifikation nennt Lakatos „**naiver Falsifikationismus**“ Seine Theorie des „verfeinerten Falsifikationismus“ hingegen richtet die Aufmerksamkeit auf die **Konkurrenz zwischen „degenerativen“ und „progressiven“ Forschungsprogrammen:**

Erweist sich eine Theorie in einer Reihe von Situationen als leistungsfähig, dann wird man die Theorie bzw. das Forschungsprogramm beim Auftreten einer Anomalie (=falsifizierende Beobachtung) nicht gleich aufgeben. Kritisch wird die Situation, wenn sich die Anomalien häufen. Dann wird versucht, die Anomalien durch Ad-hoc-Hypothesen „wegzuerklären“. **Das alte „degenerative“ Forschungsprogramm wird nach Lakatos erst dann nach und**

nach aufgegeben, wenn ein neues, „progressives“ Forschungsprogramm verfügbar ist.

3.2. Kuhn (1967)

Der **Wechsel wissenschaftlicher Paradigmen**: hier toleriert die „etablierte“ Wissenschaft Anomalien und Ad-hoc-Hypothesen so lange, bis ein neues Paradigma auftaucht und die **Grundannahmen** der alten Theorie revidiert (siehe zur Illustration die beiden anschaulichen Rätsel auf S. 157 und 159) → Kuhn: **Wissenschaft als „Puzzle-Solving“**)

Bsp.:

- Kopernikanische Wende: Übergang vom geozentrischen zum heliozentrischen Weltbild
- Ökonomie: der Weg vom Keynesianismus zur heute dominierenden Neoklassik

Unterschiede in den Positionen von Popper und Kuhn betreffen die Zielsetzung: Popper entwirft eine normative Methodologie, während Kuhn eher den faktischen Ablauf der Wissenschaftsgeschichte zu charakterisiert. Aber: was nützen strikte normative Postulate, wenn sich in der Wissenschaftsgeschichte häufig gezeigt hat, dass gerade deren Verletzung die Konstruktion leistungsfähiger Theorien stimuliert hat? Es macht wenig Sinn, Normen losgelöst von ihren empirischen Konsequenzen zu diskutieren! Lakatos' Konzeption der Konkurrenz von Forschungsprogrammen trägt den wissenschaftstheoretischen Bedenken eher Rechnung als Poppers „Logik der Forschung“.

„Mit einer fachwissenschaftlichen Methodenlehre, in unserem Fall mit den Methoden der empirischen Sozialforschung, wird insbesondere das Ziel verfolgt, durch die Identifizierung von Fehlerquellen und die Entwicklung geeigneter Methoden das Risiko irrtümlicher Entscheidungen über die „Wahrheitsqualität“ von Basisätzen nach Möglichkeit zu vermindern. „

3. Theorie

Diekmann: 122

4. Hypothesen

Diekmann: 107

Typen von Variablen: Definition	<p>Am Anfang dieses Kapitels wird eine Definition des Begriffs der Variable gegeben: Eine Variable ist ein Merkmal oder eine Eigenschaft eines Merkmalsträgers. Beispiele sind das Geschlecht einer Person, die Dauer einer Ehe oder die Regierungsform von Staaten.</p> <p>Grundsätzlich sind bei Variablen drei Dinge zu unterscheiden, nämlich die Variable (das Merkmal, z.B. das Geschlecht), die Ausprägung der Variable (Kategorien, Merkmalsausprägungen, z.B. männlich/weiblich) und die Merkmalsträger.</p>
Variablen und ihre Ausprägungen	<p>Anschließend werden Variablen und die Ausprägung von Variablen besprochen. Die Variablen können - wie erwähnt - in verschiedene Kategorien eingeteilt werden, z.B. Einkommensklassen oder eben Mann/Frau. Eine Variable muss mindestens zwei Ausprägungen aufweisen. Die Festlegung der Variablen und Kategorien ist dabei eine Frage der jeweiligen Definition, es gibt keine Variablen oder Kategorien an sich. Bei der Konstruktion von Kategorien ist aber drauf zu achten, dass sich die Kategorien nicht überlappen (disjunkt) und dass jeder Merkmalsträger einer Kategorie zugewiesen werden kann (erschöpfend). Anders formuliert: Jede denkbare Beobachtung muss genau einer Kategorie zugewiesen werden können.</p> <p>Weiter gibt es diskrete und kontinuierliche Kategorien. Die Geschwindigkeit wäre eine kontinuierliche, die verschiedenen Schulbildungsniveaus eine diskrete Merkmalsausprägung.</p> <p>Variablen mit zwei Ausprägungen werden als dichotom, Variablen mit mehreren Ausprägungen als polytom bezeichnet. Die Variablen können zudem absolut oder relational sein. Wird eine Variable durch die Beziehung zu anderen Variablen definiert, ist sie relational. Ein Beispiel wäre die Intensität von Freundschaftsbeziehungen.</p>
Merkmalsträger und -hierarchien	<p>Im folgenden Abschnitt wird auf die Merkmalsträger und Merkmals-hierarchien eingegangen. Zuerst werden individuelle und kollektive Merkmale unterschieden. Kollektivmerkmale zeichnen sich dadurch aus, dass der Merkmalsträger eine Personenmehrheit ist. Ein individuelles Merkmal wäre das persönliche, ein Kollektives das Durchschnittseinkommen der OECD. Bei Kollektivmerkmalen kann zwischen verschiedenen Ebenen unterschieden werden, z.B. zwischen Angestellten einer Firma, der Firma selber, der Menge aller Firmen eines Landes etc.</p> <p>Weiter wird zwischen analytischen, strukturellen und globalen Kollektivmerkmalen unterschieden. Analytische Kollektivmerkmale lassen sich durch Rechenoperationen aus absoluten Individualmerkmalen ermitteln. Strukturelle Kollektivmerkmale ergeben sich durch Rechenoperationen mit relationalen Individualmerkmalen. Beide gehen somit aus der Aggregation von Individualmerkmalen hervor. Globale Kollektivmerkmale sind dagegen nicht aus Individualmerkmalen ableitbar, sie können nicht durch die Aggregation von Merkmalen der Kollektivmitglieder definiert werden.</p>
Hypothesen: Definition	<p>In diesem Kapitel wird zuerst der Begriff der Hypothese definiert: Eine Hypothese ist demnach eine Vermutung über einen bestehenden Sachverhalt. In der Sozialwissenschaft wird unter Hypothesen im allgemeinen ein Zusammenhang zwischen zwei oder mehr Variablen verstanden.</p>

<p>Klassifizierung von Hypothesen</p>	<p>Im Folgenden werden Hypothesen dann anhand von verschiedenen Kriterien klassifiziert: Zuerst werden deterministische und probabilistische Hypothesen unterschieden. Deterministische Hypothesen sind strikte Zusammenhänge, z.B. das Fallgesetz der Physik. Solche Hypothesen sind in der Sozialwissenschaft selten, Zusammenhänge sind probabilistisch, d.h. ein vorhergesagter Zusammenhang wird nur mit einer gewissen Wahrscheinlichkeit auftreten. Anschließend werden zwei Grundtypen von Hypothesen beschrieben. Die erste ist die wenn-dann-Hypothese. Eine solche besagt, dass wenn eine Ausprägung A auftritt, dann eine andere Ausprägung B zu erwarten ist (implizite wenn-dann-Hypothese). Die äquivalente wenn-dann-Hypothese besagt, dass B erwartet wird, wenn und nur wenn A auftritt, während bei der impliziten Version auch andere Ausprägungen zu B führen können. Oft werden in der Sozialwissenschaft probabilistische wenn-dann-Hypothesen formuliert, d.h. wenn die Ausprägung A auftritt, dann wird die Ausprägung B mit höherer Wahrscheinlichkeit zu beobachten sein als bei anderen Ausprägungen. Als zweites wird die je-desto-Hypothese erwähnt. Die Kategorien der beiden beteiligten Variablen müssen bei dieser Art von Hypothese als Rangfolge interpretierbar sein, d.h. sie müssen zumindest ordinal sein. Bei einem Zusammenhang steigt z.B. bei einem Anstieg von A die Variable B ebenfalls durchgehend oder nimmt durchgehend ab. Jedoch muss der Zusammenhang nicht linear sein, er kann z.B. auch u-förmig oder umgekehrt u-förmig sein. Bei Variablen, die mindestens Intervallskalenniveau aufweisen, kann der Zusammenhang mit einer mathematischen Funktion beschrieben werden. Weiter wird bei Hypothesen eine Unterscheidung zwischen Kausalhypothesen und Merkmalsassoziationen vorgenommen. Bei letzterem wird ein Zusammenhang nicht als kausal interpretiert. Als Trendhypothesen werden Hypothesen bezeichnet, bei denen die Zeit die unabhängige Variable darstellt.</p>
<p>Individual-, Kollektiv- und Kontexthypothesen</p>	<p>Zu guter letzt werden Individual-, Kollektiv- und Kontexthypothesen unterschieden. Wenn es bei beiden Variablen um Individualmerkmale geht, handelt es sich um eine Individualhypothese, z.B. Je höher der Bildungsabschluss einer Person, desto höher ihr Einkommen. Kollektivhypothesen beziehen sich dagegen auf Zusammenhänge zwischen Kollektivmerkmalen, z.B. einen umgekehrt u-förmigen Zusammenhang zwischen den Mobilitätschancen und der Zufriedenheit in einem Unternehmen. Kontexthypothesen beziehen sich auf einen Zusammenhang zwischen einem Kollektiv- und einem Individualmerkmal, z.B. je höher die soziale Integration einer Gruppe, desto kleiner die Wahrscheinlichkeit, dass sich eine einzelne Person abweichend verhält. Jedoch folgt aus Kollektivhypothesen nicht automatisch die korrespondierende Individualhypothese, z.B. kann man, wenn in einer Region die Prozentzahl Katholiken und CDU-Wähler gleich ist (Kollektivmerkmal) nicht gleich davon ausgehen, dass ein Katholik die CDU bevorzugt (Individualhypothese). Last but not least gibt es noch die Aggregationsregel, bei der Kollektivmerkmale durch Individualmerkmale erklärt werden. Mit den Zusammenhängen zwischen Kollektiv- und Individualmerkmalen beschäftigt sich die Mehrebenenanalyse. Dabei ist es wichtig, die Merkmale von verschiedenen und nicht nur einem Kollektiv zu erheben. Das Kollektivmerkmal muss bei den einzelnen Beobachtungen in der Stichprobe verschiedene Ausprägungen aufweisen. Oft geht es in den Sozialwissenschaften um Kollektivmerkmale, z.B. um Kriminalitätsraten. Zur Erklärung von Kollektivmerkmalen können einerseits andere Kollektivmerkmale benützt werden (Kollektivhypothese), andererseits können aber auch die Kontext- und Individualhypothese sowie die Aggregationsregel dazu benutzt werden, Kollektivhypothesen vertiefter und genauer zu erklären. Hierbei wird von Kollektivmerkmalen auf Individualmerkmale (Kontexthypothese), von Individualmerkmalen auf Individualmerkmale (Individualhypothese) und dann von</p>

	Individualmerkmalen wieder auf Kollektivmerkmale geschlossen (Aggregationsregel).
--	---

5. Variablen: Qualität, Skalenniveaus
Diekmann: 216, 244

Forschungsmethoden

1. Der BIG-MAC-Index

2. Einstellungsmessung mit der Likert-Technik

3. Gütekriterien der Messung: Objektivität, Reliabilität und Validität

4. Testtheorie

Patrik Hugelshofer

Diekmann 200-234

patrik.hugelshofer@unisg.ch

Big-Mac-Index als Indikator der Kaufkraft. Operationalisierung ist der Oberbegriff für Messung, Skalierung und Indexbildung. Die Likert-Technik ist ein beliebtes Instrument der Einstellungsmessung. Die Gütekriterien der Messung sind Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit). Zur Messung der Reliabilität gibt es 3 Methoden: Paralleltest, Test-Retest, Testhalbierung. Die Validität kann in 3 Methoden unterteilt werden: Inhalts-, Kriteriums- und Konstruktionsvalidität.

Big Mac Index

Seit 1986 vom Economist erhoben. Der lokale Preis des Big Mac in einem Land dient als Indikator der Kaufkraft der Landeswährung. Auf Basis der Big-Mac-Preise wird die reale Kaufkraft einer Währung sowie die Kaufkraftparität zum Dollar geschätzt. Bsp. Daten 1994: CHF um 72 % zum USD überbewertet.

Kaufkraft einer Währung ist ein theoretisches Konstrukt deren Werte durch Indikatoren indirekt gemessen werden. Big Mac=Indikator, zu messende Grösse=Kaufkraft. Gefahr hier: nur einzelner Indikator wird gemessen. Die Messung mit einem Indikator reicht nur aus, wenn der Big-Mac-Preis mit sämtlichen anderen relevanten Güterpreisen perfekt korreliert. Big-Mac=Repräsentant für alle anderen Güterpreise. Möglichkeit zur Prüfung des Indikators ist ein Validierungstest. Stärke des linearen Zusammenhangs ist Korrelationskoeffizient der im Bereich +1 (perfekt positiv) bzw. -1 (perfekt negativ) ist. 0 bedeutet kein Zusammenhang. Der Big Mac-Index korreliert sehr stark mit dem Human-Development-Index (0.84). Die Korrelation von HDI und BSP hingegen beträgt nur 0.71. Dies zeigt, dass die Messung eines theoretischen Konstruktes durch einen ungewöhnlichen Indikator nicht von vornherein ausgeschlossen werden sollte. Andere Beispiele: Mit Wasserverbrauch (WC-Spül-Vorgang) die Sehbeteiligung von Fernsehsender ableiten.

Operationalisierung

Oberbegriff für Messung, Skalierung und Indexbildung

Messung

Zuordnung von Zahlen zu Objekten nach bestimmten Regeln

Skalierung

Spezialfall der Messung. Messung auf Basis einer Skalierung

Indizes

Variable deren Werte sich aus einer Rechenoperation ergeben

Likert-Skala

Verfahren der Einstellungsmessung. Einfach und praktisch gut brauchbar, deshalb sehr beliebt in der Sozialforschung. Bsp. Umfrage zum Umweltbewusstsein. Befragte Person soll zu jedem Item den Grad ihrer Zustimmung auf einer 5-Punkte-Antwortskala zum Ausdruck bringen → stimmte überhaupt nicht zu (=1), stimme nicht zu (=2), usw. Es werden verschiedene Aussagen (Items) aufgenommen. Jeder dieser Items gilt als Indikator der Einstellung „Umweltbewusstsein“. Von den Items wird das Summscore errechnet (Antwort stimme überhaupt nicht zu=1 plus stimme nicht zu=2, etc). Das Summscore ist ein Mass des Umweltbewusstseins, dass die gesamten Infos der einzelnen Indikatoren ausschöpft. Falls die Antworten der einzelnen Antworten untereinander konsistent sind (Bsp. in einer Antwort zeigt sich Person sehr umweltfreundlich, in einer Anderen wiederum nicht= inkonsistent), werden die einzelnen Items hoch mit dem Summscore korrelieren. Diese Korrelation nennt man

	<p>Trennschärfekoeffizient. Ist Koeffizient bei einigen Items hoch, bei anderen niedrig, dann sind letztere zur Messung der Zieldimension weniger geeignet. Die tiefe Korrelation kann darauf zurückzuführen sein, dass die Antwortreaktion auf das Item möglicherweise durch eine Fremddimension verzerrt wird. Bsp.: Frage Umweltbewusstsein → „Die Wissenschaft und Technik werden in der Lage sein, die Umweltprobleme zu lösen“. Hier tiefe Trennschärfekorrelation. Die Antwort wurde allenfalls durch die Fremddimension „Technikeinschätzung“ verzerrt.</p>
Korrelationskoeffizient	Formel, Erklärung und Beispiel sind auf S. 212 zu finden. (schon gut zusammengefasst, könnte das nicht besser)
Gütekriterien	Objektivität, Reliabilität (Zuverlässigkeit), Validität (Gültigkeit)
Messung	
Objektivität	Ausmass dafür, wie unabhängig die Ergebnisse von der Person sind, die das Messinstrument anwendet. Vollständige Objektivität: Anwender A und B erzielen das gleiche Resultat. Mass dafür ist der Korrelationskoeffizient. Noch differenzierter kann man von Durchführungs- und Auswertungsobjektivität unterscheiden.
Reliabilität	<p>Mass für die Reproduzierbarkeit von Messergebnissen. Kann durch Korrelationskoeffizient ausgedrückt werden. 3 Methoden:</p> <ol style="list-style-type: none"> 1. Paralleltest-Methode → Messung erfolgt mit zwei vergleichbaren Messinstrumenten. Korrelation Instrument A mit Instrument B informiert über Reliabilität. 2. Test-Retest-Methode → Messinstrument wird nach einem Zeitintervall wiederholt angewendet. Messung Korrelation der Zeitpunkte. Problem hier: die zeitliche Stabilität muss gegeben sein. Bsp. Berner Umweltstudie, Retest ½ Jahr nach Test. Allenfalls hat sich Umweltbewusstsein der Befragten zwischenzeitlich verändert. 3. Methode Testhalbierung → Korrelation zwischen zwei Testhälften. Gebräuchlichste Form. Problem: durch Halbierung Unterschätzung Reliabilität, da diese mit Anzahl multipler Indikatoren anwächst. Verkürzung kann aber rechnerisch kompensiert werden. Formel von Spearman/Brown. Weiter ist darauf zu achten, wie Itematterie halbiert wird. Je nach Halbierung unterschiedliche Resultate. Deshalb Berechnung Durchschnittsreliabilität für alle Halbierungsvarianten.
Beispiel	Allgemein wird Reliabilitätsschätzung eines Messinstrumentes von $r_s = 0.8$ verlangt. Mit Spearman und Brown-Formel kann Unterschätzung (Testhalbierung) korrigiert werden. Formel S. 220. Mit Cronbachs α kann die interne Konsistenz der Items gemessen werden, Formel S. 221. Hier ist die Reliabilität von der Testlänge und der Trennschärfekorrelation abhängig. Deshalb: Höhere Reliabilität falls verlängerte Skala und Interkorrelationen Items mit mittlerer Streuung. Wieso mittlere Streuung? Items die von allen Befragten befürwortet werden, leisten in unserem Beispiel nur einen geringen Betrag zur Trennung der Umweltbewussten von den weniger Umweltbewussten.
Reliabilitätsschätzung	
Validität	<p>Gültigkeit von Messinstrumenten. „Intelligenz ist, was Intelligenztests“ messen. Objektive und zuverlässige Messinstrumente sind nicht zwingend valide. Validität gibt Grad der Genauigkeit an, mit dem der Test dasjenige Merkmal misst, das es eigentlich messen sollte. Arten der Validität:</p> <ul style="list-style-type: none"> ▪ Inhaltsvalidität: In welchem Grad misst die Auswahl von Items die zu messende Eigenschaft. Bsp. Wortschatztest: Zufallsstichprobe von 50 Vokabeln aus den 2000 meistgebrauchten Wörtern → inhaltsvalide. Auswahl von 50 exotischen Vokabeln ist nicht inhaltsvalide. ▪ Kriteriumsvalidität: In welchem Grad sind die mit einem Messinstrument erzielten Resultate mit anderen relevanten Merkmalen empirisch korreliert. Eindeutig definierte Kennziffer: Korrelationskoeffizient zwischen Messergebnis und Aussenkriterium. Bsp. Aussenkriterium der Umweltskala ist Mitgliedschaft Umweltorganisation. ▪ Konstruktionsvalidität: Brauchbarkeit von Messinstrumenten für die Entwicklung von Theorien. Verlangt, dass das von einem Messinstrument erfasste Konstrukt mit möglichst vielen anderen Variablen in theoretisch begründbaren Zusammenhängen steht. Diese Validitätsart kann nicht durch ein einfaches Prüfverfahren ermittelt werden. Cronbach und Meehl (1995) und Campbell und Fiske (1959) mit „Multitrait-
Arten der Validität:	
Inhaltsvalidität	
Kriteriumsvalidität	
Konstruktionsvalidität	

Multimethod-Matrix“ haben Verfahren entwickelt (Details S. 226)

Testtheorie
(nicht auf Vorlesungs-
folien)

Behandelt Zusammenhang zwischen Messinstrumenten und theoretischen Konstrukten im Rahmen eines mathematisch-statistischen Modells. 5 Annahmen im Modell:

die 5 Annahmen des Modells

1. X (Messinstrument) = T (wahrer Wert) + E (Messfehler). Was bedeutet diese Annahme? Bei genügend of wiederholten Messungen ist Messfehler bei Null, bzw. die Messfehler heben sich gegenseitig auf.
2. Erwartungswert Messfehler ist null: $\mu(E)=0$.
3. Wahrer Wert korreliert nicht mit dem Messfehler: $\rho_{TE} = 0$
4. Messfehler von zwei wiederholten Messungen korrelieren nicht miteinander: $\rho_{E_1, E_2} = 0$
5. Korrelation zwischen Messfehler einer ersten Messung und wahrer Wert einer zweiten Messung ist Null: $\rho_{E_1 T_2} = 0$

Erklärung Annahmen

Annahmen 1+2 = präzise Definition theoretisches Konstrukt

Annahmen 2-5 = schliessen verschiedene Arten systematischer Messfehler aus.

- Annahme 3: schliesst systematische Verzerrung Korrelation wahrer Wert – Messfehler aus. Bsp: Messfehler bei Umweltskala bei umweltbewussten Personen höher als bei weniger umweltbewussten Personen → Verletzung
- Annahme 4: keine systematischen Abhängigkeiten zwischen zwei Messungen. Bsp. Messfehler in 1. Messung hoch, aber kein Einfluss auf 2. Messung.
- Annahme 5: siehe oben

Mit diesem Modell können wir nun einige Begriffe genauer definieren:

Reliabilitätskoeffizient: Quadrat der Korrelation zwischen Messwertvariable X und dem Konstrukt $T = \rho_{XT}^2$

Validitätskoeffizient = ρ_{XY} . Korrelation von X mit Messung Aussenkriterium Y .

Das Modell kann erweitert werden: Faktorenanalyse und allgemeines lineares Strukturgleichungsmodell. Sind auf S. 232-34 beschrieben. Meines Erachtens (Patrik) geht das aber zu weit.

Empirische Sozialforschung

7. Kapitel Messtheorie Diekmann, S. 244-265

Sandra Jost

079 283 30 79

sandra.jost@student.unisg.ch

Einleitung zur Messtheorie: Es stellt sich v.a. die Frage, welche Kriterien (Axiome) eine Messung erfüllen muss, um einer gew. Skalierung zugeordnet werden zu können. Genannt werden im Folgenden ein paar Messmethoden.

Skalierungen: N(ominal), O(rdinal), I(ntervall) und R(atio).

7. Einleitung zur Messtheorie Welche Kriterien muss eine Messung gemäss der Messtheorie erfüllen? Nach Stevens ist eine Messung eine "Zuordnung von Zahlen zu Objekten nach bestimmten Regeln." Die Regeln können unterschiedlich sein. Sie sind nach der "axiomatischen Messtheorie" bestimmt. Der Output einer Messung sind Zahlen, der Input sind Objekte und die zwischen Objekten bestimmten Beziehungen. Der Input wird in Zahlen widergespiegelt. Diekmann merkt immer wieder an: Wenn eine Rangordnung (resp. eine gewisse Abbildungsvorschrift) vorherrscht (z.B. 4,3,2,1), dann gilt diese Ordnung auch noch nach einer mathematischen Transformation (z.B. /5 oder

Die Methode der
Paarvergleichung

$\times 10$), solange die Rangordnung bewahrt wird. Die ursprünglichen Messwerte o_1, o_2 , usw. werden dann zu $\varphi(o_1), \varphi(o_2)$, usw.

Die Methode wird auf S. 245/246 anhand eines Beispiels beschrieben. Die Testperson hat vier Bücher zur Auswahl. Sie präferiert die Bücher in einer bestimmten Reihenfolge: $s > k > d > e$. Ordnet man den Präferenzen (s, k, d, e) Zahlen zu, erhält man z.B. $4 > 3 > 2 > 1$. Damit wird auch mathematisch veranschaulicht, dass die Testperson das Buch s bevorzugt. Diese Präferenz ist transitiv, d.h. sie ist widerspruchsfrei. Man spricht dann auch von einem Isomorphismus, welcher durch Paarvergleiche verifiziert werden kann: $s > k, s > d$, usw. Die Messungen entsprechen einer Ordinalskala (sinnvolle Reihenfolge ohne messbare, vergleichbare Abstände)

Intransitiv ist eine Beziehung, wenn im Paarvergleich ein "Widerspruch" festgestellt wird, also z.B. $a > b, b > c$ aber $c > a$. Dies kann auch numerisch bewiesen werden. Wenn $a=3, b=2$ und $c=1$ ist (wie es bei transitiven Präferenzen sein müsste), ergibt sich eine falsche Gleichung, nämlich $1 > 3$. Es ist dann keine Messung auf dem Niveau der Ordinalskala möglich.

Die genaue Definition von
"Messen"

Die Menge A der empirischen Objekte (z.B. Präferenzen, Größen,...) o_1, o_2 usw. wird als *empirisches Relativ* bezeichnet. Es kann sich auch um mehrere Relationen handeln (R_1, R_2 , usw.). Ausserdem können die Relationen mehr als binär (zweistellig) sein. Z.B. ist die Relation " o_1 ist o_2 ähnlicher als o_3 gegenüber o_4 " vierstellig. Mathematisch wird das empirische Relativ fogendermassen ausgedrückt: $\alpha = \langle A, R_1, \dots, R_n \rangle$. Falls es sich um eine Menge von Zahlen handelt, wird es nicht empirisches, sondern numerisches Relativ genannt. Wenn für jede empirische Relation eine korrespondierende numerische existiert, spricht man von einem Isomorphismus (siehe oben). Die Abbildungsvorschrift ist beim Iso. eindeutig: jedem Objekt wird eine Zahl zugeordnet und umgekehrt. Diese Bedingung ist im allgemeinen jedoch zu restriktiv. Es genügt, wenn man eine eindeutige Abbildung fordert. D.h. jedes Objekt muss genau einer Zahl zugeordnet werden können, aber nicht umgekehrt. Man spricht dann von einem Homomorphismus. Die Relation zwischen zwei Objekten entspricht dann derselben numerischen Relation. Eine Messung im Sinne der Messtheorie liegt vor, wenn ein Isomorphismus oder ein Homomorphismus zwischen einem empirischen und einem numerischen Relativ existiert.

Repräsentation, Eindeutigkeit
und Bedeutsamkeit

Was sind die Bedingungen (=Axiome) für die Existenz eines Homomorphismus? Es kommt auf die Art der Ordnung an (d.h. nominal, ordinal, ...) welche Axiome erfüllt sein müssen. Die richtigen Axiome mathematisch zu finden, nennt man Repräsentationsproblem. Kennt man die Axiome der Repräsentation, können jene empirisch überprüft werden. Nur wenn alle notwendigen Axiome erfüllt sind, kann eine Messung in der entsprechenden Skale vorgenommen werden.

Die Eindeutigkeitsproblem beschäftigt sich mit der Frage, welche mathematischen Transformationen vorgenommen werden dürfen, damit die Rangordnung bewahrt wird (siehe oben). Es kommt dabei auf das Skalenniveau (=Messniveau) an.

Die Frage nach der Bedeutsamkeit von Messungen bezieht sich auf Schlussfolgerungen, die bei einer bestimmten Art der Messung als zulässig gelten (z.B. ist die Aussage "10 Grad ist doppelt so warm wie 20 Grad falsch, da kein definierter Nullpunkt vorhanden ist. Richtig ist aber, dass die Luft bei 20 Grad wärmer ist als 10 Grad.)

Nur solche statistischen Kennziffern (Mittelwert, Modus, Median, Varianz,...) sind bezüglich eines Skalentyps bedeutsam, wenn die hiermit formulierten Aussagen bei allen zulässigen Transformationen (siehe Eindeutigkeit) invariant bleiben. Dabei müssen die Kennziffern selbst nicht die gleichen bleiben. Das Verhältnis zwischen den Werten darf sich aber nicht verändern.

Wir unterscheiden in vier Skalenniveaus: Nominal-, Ordinal-, Intervall, und Ratioskala (Eselsbrücke: NOIR).

Nominalskala: Klassifikation von Objekten nach der Relation der Gleichheit oder der Verschiedenheit. Es kann keine logische Rangfolge gebildet werden. Der einzig logische Messwert ist der Modalwert, welcher die Frage beantwortet, welche Klasse die höchste Besetzungszahl aufweist.

Ordinalskala: Hier wird die Rangordnung der Objekte bezüglich einer Eigenschaft vorausgesetzt (strenge Ordnungsrelation, wie z.B. $>$ oder $<$). Auch Indifferenz (=Gleiche Wertschätzung, gleiche Größe, usw.) ist zulässig. Empirisch überprüft werden kann dieses Axiom durch Paarvergleiche (siehe oben). Ordinalskalen sind eindeutig im Bezug auf alle positiven monotonen Transformationen. Die Wert können also z.B. logarithmiert werden. Die Rangfolge der Werte wird dadurch beibehalten. Neben dem Modus ist hier auch die Berechnung des Median (Unterteilung in 50% unten und 50% oben) sinnvoll. Die Berechnung des arith. Mittels ist aber nicht zulässig, da die Abstände zwischen den Werten in keinem bestimmbar Verhältnis stehen.

Intervallskala: Neben der Rangfolge informiert diese Skala auch über die Abstände zwischen den Messwerten. Es existiert aber kein klar definierter Nullpunkt, d.h. er kann willkürlich festgelegt werden, wie auch die Skaleneinheit. Das obige Temperaturbeispiel wäre auch hier nicht zulässig, weil die Werte in Fahrenheit nicht mehr im gleichen Verhältnis stehen wie in Grad Celsius. Bsp. für eine zulässige Intervallskalierung wären die Auswertung von Intelligenztests, Berufsprestigeumfragen, usw.. Abstände können hier bestimmt werden, es kann aber keine Aussage über das Verhältnis dieser Abstände gemacht werden. Die liegt am nicht klar definierten Nullpunkt. Sinnvoll erweist sich bei dieser Skalierung jedoch die Berechnung des arith. Mittels. Zeugnisnoten sind z.B. das arith. Mittel von den zuvor geschriebenen Prüfungen (=Durschnittswert). Achtung: der Nachweis der Normalverteilung von Messwerten gibt noch nicht über das Messniveau Auskunft. Empirisch prüfbar ist die Intervallskalierung erst durch die axiom. Messtheorie, und zwar durch eine empirische Prüfung der im Repräsentationsproblem aufgeführten Axiome. Oftmals werden bei einer Messung Mittelwerte und Korrelation berechnet, d.h. es wird unterstellt, dass die Abstände messbar sind und die Variablenausprägungen somit intervallskaliert sind. Eine häufig anwendbare Messstruktur ist beispielsweise die additiv-verbundene Messung, welche zu einer Intervallskalierung führt. Mit den Axiomen dieser Messstruktur kann überprüft werden, ob die Unterstellung, eine Messung sei intervallskaliert, zutrifft. Ein Vorzug dieses Modells ist die Mehrdimensionalität. Die Grundidee des Modells ist, dass sich eine Eigenschaft (z.B. Sozialprestige von Berufen oder Personen) additiv aus einzelnen Teildimensionen (z.B. Einkommen oder Bildungsgrad) zusammensetzt. Es kann also durch die Addition der "Punkte" im Bereich Einkommen, Bildung und Beruf das Sozialprestige einer Person ermittelt werden.

Ratioskala: Ein natürlicher Nullpunkt lässt sich bei dieser Skalierung feststellen. Deswegen können alle Werte in ein math. definierbares Verhältnis gesetzt werden. Als Beispiele können Schuljahre,

Vermögen, Preise, Zinsen, usw. genannt werden. Auch Verhältnisskalen sind eindeutig bis auf positive proportionale Transformationen $\varphi' = a\varphi$ (für $a = >0$). Neben dem Modus, Median und arith. Mittel ist hier auch die Berechnung des geom. und das harmonische Mittel möglich.

Absolutskala: Diese Skalen sind eindeutig festgelegt und Skalentransformationen sind nicht erlaubt. (Oben kann z.B. eine Intervall- in eine Ratioskala transformiert werden, wenn ein Nullpunkt bestimmt wird.) Bsp. sind Wahrscheinlichkeitswerte (z.B. zu erwartende Demonstranten) und Häufigkeiten (z.B. Anzahl Fam.mitglieder).

Während die Klasse der zulässigen Transformationen mit steigendem Messniveau eingeschränkt wird, verhält es sich mit der Bedeutsamkeit genau umgekehrt. Die Zahl der zuläss. Rechenoperationen und erlaubter Aussagen (z.B. oben genannte Messwerte) steigt mit zunehmendem Skalenniveau.

Skalentypen und ihre Eigenschaften: Auf der S. 255 befindet sich eine Tabelle, welche Eindeutigkeit, Bedeutsamkeit und Beispiele für jeden Skalentyp angibt.

Framing

Je nach Formulierung der Fragen können Testpersonen divergierende Antworten geben, auch wenn es sich inhaltlich um dasselbe handelt. Es kommt dabei sehr darauf an, ob die jeweilige Frage positiv oder negativ formuliert ist. Laut der Nutzentheorie dürfte dies aber nicht so sein, da der Nutzen derselbe bleibt. Deswegen muss man in einen Verlust-Frame und in einen Gewinn-Frame unterscheiden, wobei Testpersonen bei ersterem risikobereiter sind als bei zweitem.

8. Hinweise und nützliche Tipps

Messverfahren können nach div. Kriterien klassifiziert werden. 1. Repräsentationsmessung oder Annahme des Skalenniveaus, 2. deterministischer oder probabilistischer Charakter des Skalierungsmodells, 3. eindimensional (eine Variable) oder mehrdimensional (zwei oder mehr Variablen), 4. Dominanzrelation (Präferenz) vs. Ähnlichkeitsrelation, 5. Zugehörigkeit der zu skalierenden Objekte zu *einer* Menge oder zu *zwei oder mehreren* Mengen.

Vom Skalenniveau der Var. hängt ab, welche statistischen Analyseverfahren später bei der Datenauswertung verwendet werden können.

Es gibt sog. Einstellungsskalen, die ihre eigenen Regeln haben und nicht den obigen Skalen entsprechen. Im Forschungsfall ist aber die Abstützung der Daten auf eine bewährte und math. überprüfte Skala zu bevorzugen.

6. Forschungsdesign: Experimente, Erhebungsebene Diekmann: 266

Querschnitt- und Längsschnitterhebungen

Forschungsmethoden
Diekmann, S. 266-285

Kurzzusammenfassung (nur ein paar Sätze)
--

Bei längeren Texten Inhaltsverzeichnis

Einführungsbeispiele

Aussagen wie „lieber 5% Inflation als 5% Arbeitslosigkeit“ verkennen, dass Querschnittsdaten statische Momentaufnahmen sind und beispielsweise nichts aus über die (wechselnde) Zusammensetzung der Arbeitslosen und damit die effektive Anzahl Betroffener aussagen. Um zeitbezogene Hypothesen über den sozialen Wandel untersuchen zu können, müssen deswegen Längsschnittdaten herangezogen werden.

1. Querschnitt-, Trend- und Paneldesign Definitionen

Es gibt drei Arten von Erhebungsdesign, die von der Methodik unabhängig sind und sich durch den einbezogenen Zeitraum voneinander unterscheiden:

1. *Querschnittsdaten*: Die Querschnitterhebung bezieht sich auf einen Zeitpunkt oder eine kurze Zeitspanne, in der eine einmalige Erhebung der Eigenschaften bei N Untersuchungseinheiten vorgenommen wird.
2. *Trenddesign*: Bei der Trenderhebung werden (a) die Werte der gleichen Variablen zu (b) mehreren Zeitpunkten mit (c) jeweils unterschiedlichen Stichproben erhoben. Damit können Veränderungen auf der Aggregatebene verfolgt werden. Das Trenddesign entspricht einer Abfolge von Querschnitterhebungen zum selben Thema. Bsp: Trendanalyse der Armutsdynamik
3. *Paneldesign*: Bei der Panelerhebung werden (a) die Werte der gleichen Variablen zu (b) mehreren Zeitpunkten, jedoch (c) auf der Grundlage einer identischen Stichprobe erhoben. Dadurch können Veränderungen auf der individuellen Ebene verfolgt werden. Bsp: Analyse der Armutsdynamik. Die einzelnen Erhebungen heissen Panelwellen. Das Paneldesign ist in der Informationshierarchie zuoberst: Es liefert Informationen über Querschnitts- und Trenddaten, umgekehrt ist dies nicht der Fall.

Fehlerquellen

- Stichproben- und

Beim Trenddesign können Stichproben- und Messfehler auftreten. Beim Paneldesign entfallen Stichprobenfehler, da sich die Erhebungsbasis nicht

Messfehler	verändert. Messfehler können aber auftreten wegen dem Problem zwischen Stabilität und Realität. Kleine Schwankungen sind deswegen bei beiden Designs mit Vorsicht zu interpretieren.
• Veränderungen der Ausgangspopulation	<p>Problematisch sind logischerweise Trend- und Panelaussagen, die sich auf nur zwei Erhebungswellen abstützen.</p> <p>Wenn sich die Zusammensetzung der Population im Beobachtungszeitraum stark verändert, stellt eine Panelerhebung nicht mehr eine repräsentative Stichprobenerhebung dar. Die Informationshierarchie zwischen Trend-design und Paneldesign ist in diesem Fall aufgehoben.</p>
• Systematische Fehler beim Paneldesign wegen systematischen Ausfällen (Panelmortalität)	<p>In der Praxis ist die Stichprobenbasis für Panelerhebungen keineswegs konstant. Unter dem Begriff „Panelmortalität“ werden Todesfälle, Wegzug und sonstige Unerreichbarkeit zusammengefasst. Die Panelmortalität ist oft systematisch (Junge ziehen häufiger weg etc.) und beträgt von Welle zu Welle ca. 10 Prozent. Dies kann zu systematischen Verzerrungen führen, die je länger je stärker werden. Beispielsweise kann in der Realität die Armutsquote konstant sein, aber aufgrund der schwindenden Basis sinkend erscheinen. Um die Panelmortalität möglichst gering zu halten, ist ein erheblicher Aufwand bei der „Adressenpflege“ notwendig.</p>
Wichtigste Panlerhebungen	<p>Die umfassendste Panelerhebung im deutschsprachigen Raum ist das Sozioökonomische Panel (SÖP) seit 1984 mit einer Basis von 12'000 Leuten. Daneben gibt es die Allgemeine Bevölkerungsumfrage in den Sozialwissenschaften (Allbus) mit einer Basis von 3000 Befragten. Sie erlaubt durch ihre Verlinkung mit dem International Social Survey Program (ISSP) internationale Vergleiche.</p>
2. Daten und Designs	<p>Die Übereinstimmung von Daten- und Designhierarchie ist nur gegeben, wenn der Zeitpunkt der Messung mit dem der Erhebung übereinstimmt (bspw. gegenwärtiges Einkommen). Häufig kommen aber Retrospektiv-fragen vor (Ausbildung der Eltern, Zahl der Zahnarztbesuche im letzten Jahr), bei denen Paneldaten auch ohne Paneldesign erhoben werden können. Das Erhebungsdesign darf deswegen nicht mit dem Datentyp verwechselt werden. In der Realität empfiehlt sich dennoch eine Panelerhebung für Paneldaten, da das Erinnerungsvermögen ausser bei harten Fakten oft gering ist.</p>
Erhebungstyp≠Datentyp bei Retrospektivfragen	<p>Die Übereinstimmung von Daten- und Designhierarchie ist nur gegeben, wenn der Zeitpunkt der Messung mit dem der Erhebung übereinstimmt (bspw. gegenwärtiges Einkommen). Häufig kommen aber Retrospektiv-fragen vor (Ausbildung der Eltern, Zahl der Zahnarztbesuche im letzten Jahr), bei denen Paneldaten auch ohne Paneldesign erhoben werden können. Das Erhebungsdesign darf deswegen nicht mit dem Datentyp verwechselt werden. In der Realität empfiehlt sich dennoch eine Panelerhebung für Paneldaten, da das Erinnerungsvermögen ausser bei harten Fakten oft gering ist.</p>
Definition der verschiedenen Datentypen nach Zeitbezug	<p>Die verschiedenen Datentypen lassen sich nach ihrem Zeitbezug unterscheiden:</p> <ol style="list-style-type: none"> <li data-bbox="552 2007 1406 2038">1. <i>Querschnittsdaten</i>: bezogen auf einen Zeitpunkt,

normalerweise werden die Werte der Variablen bei einer grösseren Anzahl N von Untersuchungseinheiten erhoben. ($i=1,2,\dots,N; t=1$)

2. *Zeitreihendaten*: eine Sequenz von Variablenwerten bei einer Untersuchungseinheit zu T verschiedenen Zeitpunkten. Bsp: Aggregierte Armutsquoten Tabelle VII.1. ($i=1; t=1,2,\dots,T$)

3. *Paneldaten*: jeder der Werte von N Untersuchungseinheiten muss zu mindestens zwei Zeitpunkten ($T>1$) vorliegen, wobei das Zeitintervall bei allen Untersuchungseinheiten identisch ist. Paneldaten können damit als Zeitreihendaten für mehrere Untersuchungseinheiten aufgefasst werden, wobei bei Paneldaten T meistens grösser und N geringer ist. ($i=1,2,\dots,N; t=1,2,\dots,T, N>T$)

Datenquader

Die Kombination von M Variablen für N Untersuchungen zu T Zeitpunkten kann zu einem Datenquader wie auf S. 278 führen.

4. *Verlaufs- oder Ereignisdaten*: Zeitintervalle zwischen zwei Ereignissen. Bsp. Dauer einer Ehe, Zeitspanne bis zum Wohnortswechsel. Der Informationsgehalt ist höher als von Paneldaten, weil bspw. auch Heirat, Scheidung und Wiederheirat erfasst werden. Zur Erhebung kommen Querschnitt- oder Paneldesigns in Frage.

4. Kohortendesign

Das Kohortendesign ist ein Sonderfall der Erhebung von Längsschnittdaten. Untersucht wird dabei eine Bevölkerungsgruppe, die durch ein zeitlich ge-meinsames, längerfristig prägendes Startereignis (Geburt, Eheschliessung etc.) definiert wird. Geburtskohorten sind bspw. denselben sozial-ökonomischen Ereignissen ausgesetzt, die so genannte Kohorteneffekte bewirken können. Das Kohortenkonzept ist deswegen grundlegend für die Untersuchung sozialen Wandels.

Startereignis

Kohortenuntersuchung

Kohortendaten werden durch Querschnitts- oder Paneldesigns erhoben und ex ante oder ex post berücksichtigt.

In der Kohortenuntersuchung gibt es drei typische Effekte:

1. *Kohorteneffekte*: systematische Unterschiede zwischen den

Kohorten

2. *Lebenszykluseffekte*: systematische Zusammenhänge zwischen einem interessierenden Merkmal und der seit dem Startereignis verstrichenen Zeit.

3. *Periodeneffekte*: bei historisch einmaligen Ereignissen

Lebenszyklus-Fehlschluss

Kohortenfehlschluss

Die Unterscheidung zwischen den drei Typen erfolgt am besten durch Längsschnittdaten. Wenn nur Querschnittdaten vorliegen und bestehende Kohorteneffekte nicht beachtet werden, kommt es bei Schätzungen von Lebenszykluseffekten zum „Lebenszyklus-Fehlschluss“ (Beispiele vgl. S. 283).

Es gibt auch den Kohortenfehlschluss, bei dem die Schätzung von Querschnittsdaten unterstellt, dass die betrachteten Merkmale im Lebenslauf konstant sind. So können Lebenslaufeffekte fälschlicherweise als Kohorteneffekte interpretiert werden. Dies zeigt, dass Kohorten- und Lebenszykluseffekte oft sehr schwer zu trennen sind.

Eine wichtige Anwendung von Kohorten- und Alterseffekten ist die Lebenserwartung.

Experimentelle und quasi-experimentelle Designs

Forschungsmethoden

Diekmann, S. 289-304

Eliane Hess

076 373 81 65

Eliane.Hess@student.unisg.ch

Neben der zeitlichen Dimension (Entscheid, ob Querschnitt- oder Längsschnittdesign) ist bei der Forschungsplanung auch die **Varianzkontrolle** zu berücksichtigen. Es handelt sich dabei um die Bestimmung von *Vergleichsgruppen* und den *Modus der Aufteilung* von Untersuchungspersonen auf die Vergleichsgruppen. (In dieser Zusammenfassung werden die Abschnitte 1 und 2 des Kapitels VIII behandelt.)

Einleitung

Die Bildung von Vergleichsgruppen, d.h. die Zuweisung von Untersuchungseinheiten zu den Kategorien der unabhängigen Variable einer zu prüfenden Hypothese, kann *vor* der Erhebung nach dem Prinzip der

Zufallsaufteilung oder *nach* der Erhebung der Daten (**Ex-post-facto-Design**) erfolgen. Werden die Vergleichsgruppen ex-ante bestimmt, wird die Varianz der unabhängigen Variable bereits vor der Datenerhebung kontrolliert. Bei einem **experimentellen Design** werden die Probanden ferner per Zufall auf die Vergleichsgruppen aufgeteilt (= Randomisierung). Dadurch können eventuell verzerrende Einflüsse von Drittvariablen neutralisiert werden. Der Prüfung von Hypothesen anhand von Daten aus Bevölkerungsumfragen liegt im allgemeinen ein Ex-post-facto-Design zugrunde. Der Versuch der Neutralisierung eventuell verzerrender Effekte (Scheinkorrelation!) erfolgt also erst im nachhinein und ist nur näherungsweise möglich. Sofern realisierbar, ist ein **experimentelles Design das ideale Versuchsarrangement zum Test von Kausalhypothesen.**

Unter dem Gesichtspunkt der Varianzkontrolle sind drei Klassen von Designs unterscheidbar:

- **Experimentelle Designs**
- **Quasi-experimentelle Designs** (eine Art Zwischentyp): Vergleichsgruppen werden explizit und i.d.R. im Vorhinein festgelegt, die Untersuchungseinheiten werden den Vergleichsgruppen aber nicht zufällig zugewiesen \Rightarrow Experimente ohne Randomisierung
- **Ex-post-facto-Designs:** Varianzkontrolle, Auswahl von Vergleichsgruppen und Ausblendung verzerrender Störfaktoren erfolgt erst im Zuge der Datenanalyse nach der Datenerhebung.

Campbell und Stanley erwähnen weiter noch **vorexperimentelle Designs**, die wissenschaftlichen Anforderungen an die Methoden der Hypothesenprüfung *nicht* genügen. Anhand von ihnen lässt sich aber eine Reihe von Fehlerquellen gut erkennen.

Vorexperimentelle Designs

Bsp.: Ein in die Hände klatschender Mann sagt: „Ich vertreibe durch das Klatschen Elefanten.“ (X) Ein anderer: „Aber es sind doch gar keine Elefanten hier!“ (O) Der Mann: „Sehen sie, das kommt daher, weil ich in die Hände klatsche.“ In der Sprache experimenteller Versuchspläne gibt das ein Design **X – O** (X = experimenteller Stimulus, O = Beobachtung, Messung einer abhängigen Variable). Diesem sowie weiteren Beispielen (S. 291) ist gemeinsam, dass sie keine relevante Vergleichs- oder Kontrollgruppe berücksichtigen (bzw. unvollständige Angaben zu diesen geben). Es fehlt somit jegliche Vergleichsmöglichkeit! Bei einem **vorexperimentellen XO-Design** werden die Beobachtungen (O) nur für eine Kategorie der unabhängigen Variable registriert (nur die Werte einer Spalte, Bsp. S. 292 f.). Werte zu den Vergleichsgruppen werden keine gemacht, so dass die Ermittlung eines Effekts der Variablen mit einem XO-Design nicht möglich ist. Auch das **Design falscher Vergleichswerte** liefert unvollständige Tabelleninformationen (nur die Werte einer Zeile, Bsp. S. 294). Beim XO-Design ist die Varianz der unabhängigen Variable null, beim Design falscher Vergleichswerte ist die Varianz der abhängigen Variable null. Ohne Varianz von Variablen gibt es aber keine Vergleichsmöglichkeiten – es lassen sich keine Zusammenhangshypothesen

überprüfen und keine Effekte abschätzen.

Ein weiteres vorexperimentelles Design wird angesprochen: Bei einem Versuchsplan nach dem Muster $O_1 - X - O_2$ ist ein Vergleich immerhin durch die Vorher-nachher-Messung gewährleistet. Problem (insbesondere für Experimente in den Sozialwissenschaften, Medizin und Psychologie): Zwischen den beiden Beobachtungszeitpunkten können Reifungsprozesse (z.B. Besserung nach einer Grippe) stattfinden. Liegt Reifung vor, kann sich der Schluss von den Beobachtungen auf den Effekt einer Massnahme X (Bsp. Medikament gegen Grippe) somit als Fehlschluss erweisen. Mit einem experimentellen oder quasi-experimentellen Kontrollgruppendesign sind dagegen Reifungsprozesse identifizierbar.

Experimentelle Designs

Von einem experimentellen Design spricht man, wenn drei Bedingungen vorliegen:

1. Es werden mindestens zwei experimentelle Gruppen gebildet.
2. Randomisierung (R): Die Versuchspersonen werden experimentellen Gruppen nach einem Zufallsverfahren zugewiesen.
3. Die unabhängige Variable wird vom Forscher „manipuliert“ (Forscher präsentiert einen oder mehrere „experimentelle Stimuli“: Gruppe, die neues Medikament bekommt = Stimulus- bzw. Versuchsgruppe, Vergleichsgruppe = Kontrollgruppe).

Standarddesign

Für das **klassische Versuchs- und Kontrollgruppen-Design** ergibt sich folgende Form: **R – X – O Versuchsgruppe; R – O Kontrollgruppe.**

Der experimentelle Faktor hat zwei Ausprägungen: Anwesenheit (X) und Abwesenheit eines experimentellen Stimulus. Ist den Probanden nicht bekannt, in welcher Gruppe sie sind bzw. ist ihnen die zu prüfende Hypothese nicht bekannt, so handelt es sich um einen *Blindversuch* (Ziel: Selbstsuggestion, Verhaltensänderungen aufgrund der Kenntnis ausschliessen). Gilt das gleiche auch für den Versuchsleiter, dann spricht man von einem *Doppelblindversuch* (Ziel: zusätzlich auch Beeinflussungsmechanismen durch den Versuchsleiter verhindern).

Nicht bei allen Experimenten ist die Bezeichnung Kontroll- und Versuchsgruppe sinnvoll. Werden verschiedene experimentelle Stimuli (Bsp.: verschiedene Unterrichtsmethoden und deren Wirkung auf die Lernleistungen) getestet, ist jede Gruppe gleichermassen Kontroll- und Versuchsgruppe: **R – X₁ – O₁ Versuchsgruppe 1; R – X₂ – O₂ Versuchsgruppe 2, usw.**

Der **Hauptvorteil des experimentellen Designs** ist die Zufallsaufteilung der Probanden auf die Versuchsgruppe (Randomisierung). Damit wird der Einfluss sämtlicher, auch unbekannter Drittvariablen neutralisiert (Drittvariablen haben – von Zufallsfehlern abgesehen – die gleiche Verteilung in allen Versuchsgruppen) \Rightarrow Problem der Scheinkorrelation kann nicht auftreten! Es folgt ein Beispiel ohne (ergibt Scheinkorrelation) und mit Randomisierung (S. 298). Wird eine Studie mit Randomisierung

durchgeführt, gibt es grössere Gewissheit, aber noch nicht Sicherheit über die Korrelation zwischen X und O. Folgende *Fehlerquellen* können auftreten:

1. Mit dem Faktor X ist eine Variable Y konfundiert, die auch durch Randomisierung nicht neutralisiert werden konnte. Lösungsansätze: sorgfältig kontrollierte Laborexperimente, weitere Experimente
2. Ist den Teilnehmern in der Versuchsgruppe bewusst, dass sie Gegenstand wissenschaftlicher Forschung sind, könnte allein die Tatsache wissenschaftlicher Aufmerksamkeit zu Einstellungs- und Verhaltensänderungen führen (sog. **Hawthorne-Effekt** oder Verzerrung durch **Reaktivität**). Lösungsansätze: Blind- oder Doppelblindversuche, Feldexperimente in natürlichen Situationen
3. Die Randomisierung könnte paradoxerweise einen verzerrenden Effekt haben. Dieses Problem ist eher die Ausnahme. Bei der Evaluation sozialer Programme ist auch nur selten eine Randomisierung möglich \Rightarrow Grund, weshalb oft quasi-experimentelle Designs angewendet.
4. Vor allem bei geringen Fallzahlen kann die Zufallsaufteilung missglücken. Lösungsansatz: Kombination von Randomisierung und Matching. Könnte beispielsweise das Geschlecht eine relevante Drittvariable sein, so können die weiblichen und männlichen Versuchspersonen separat randomisiert werden. Von **Gruppen-matching** spricht man, wenn sichergestellt wird, dass die Verteilung bekannter Drittvariablen in den Versuchsgruppen gleich ist. **Paarweises Matching** liegt vor, wenn auch Kombinationen von Merkmalen (z.B. Geschlecht, Bildung, Alter) berücksichtigt werden (jede Versuchsperson hat „Zwillingspartner“ in der Kontrollgruppe). Im allgemeinen ist Randomisierung besser als Matching ohne Randomisierung, noch besser ist die Kombination der beiden.

Weitere experimentelle Designs

Andere experimentelle Designs neben dem Standarddesign sind die **Vorher-Nachher-Messung bei der Versuchs- und Kontrollgruppe** und der **Vier-Gruppen-Versuchsplan von Solomon**.

Vorher-Nachher-Messung: $R - O_1 - X - O_2$ bzw. $R - O_3 - O_4$. Es werden zusätzlich die Ausgangsniveaus O_1 und O_3 kontrolliert. Problem: Die Vorher-Messung könnte die Nachher-Messung beeinflussen.

Vier-Gruppen-Versuchsplan: $R - O_1 - X - O_2$; $R - O_3 - O_4$; $R - X - O_5$; $R - O_6$. Effekte des Messinstruments können isoliert werden und der vermutete Einfluss der experimentellen Variablen X wird strenger und mehrfach geprüft.

Faktorielles Design

Mit experimentellen Designs kann auch die simultane Wirkung von zwei und mehr Experimentalvariablen erforscht werden. Bsp.: Untersuchung des Effekts von zwei Unterrichtsmethoden bei drei Grössen von Schulklassen

⇒ 2x3-**faktorielles Design** mit sechs Versuchsgruppen. Faktorielle Designs ermöglichen auch die Untersuchung der Wechselwirkungen oder Interaktionseffekte der Faktoren. Zudem kann das simultane Zusammenspiel mehrerer Variablen im Experiment studiert werden, dies ist aber mit grossem Aufwand verbunden. Dies ist ein Grund für die Verwendung von ex-post-facto-Designs! Der Hauptgrund für die Dominanz nicht-experimenteller Designs in den Sozialwissenschaften ist aber der Umstand, dass zur Prüfung vieler sozialer Zusammenhänge eine Randomisierung praktisch gar nicht möglich ist.

Vor- und Nachteile der verschiedenen experimentellen und quasi-experimentellen Designs

Campbell und Stanley unterscheiden zwischen der **internen** (Ausblendung von Störvariablen; durch 8 Fehlerquellen gefährdet, z.B. Reifung) und **externen Validität** (Generalisierbarkeit experimenteller Effekte; durch vier Fehlerquellen gefährdet) von Versuchsplänen. Das Standarddesign, Vorher-Nachher-Design und Solomon-Vier-Gruppen-Design (Annahme: mit Randomisierung) tragen bei sorgfältiger Anwendung den acht Fehlerquellen bezüglich der internen Validität Rechnung. Die **Kontrolle von Störfaktoren** ist eben die **wesentliche Leistung experimenteller Versuchspläne** mit einer Zufallsaufteilung der Probanden auf die Versuchsgruppen! Die externe Validität wird mit experimentellen Designs aber nicht unbedingt garantiert. Probleme: Künstlichkeit der Laborsituation; die meisten (sozial-) psychologischen Experimente werden mit Studenten, v.a. Psychologie-studenten durchgeführt. Die Versuchspersonen stammen somit mehrheitlich aus einer speziellen sozialen Gruppe – damit sind eventuell wichtige Einflussfaktoren (Bildung etc.) durchgehend konstant und Interaktionseffekte oder Wechselwirkungen können übersehen werden. Lösungsansätze: Versuchspersonen aus verschiedenen sozialen Gruppen, Feldexperimente, Befunde mit anderen Designs und Methoden reproduzieren (Triangulation).

Zusammenfassend nochmals die wichtigsten Vor- und Nachteile experimenteller Designs:

Vorteile

- Der experimentelle Stimulus wird im Experiment „produziert“ und geht der vermuteten Wirkung zeitlich voraus.
- Durch die Randomisierung werden verzerrende Effekte durch Drittvariablen neutralisiert.

⇒ experimentelle Versuchspläne sind ideale Designs zum Test von Kausalhypothesen!

Probleme

- Problem der externen Validität der geprüften Zusammenhänge.
- Das Problem der Reaktivität.

- Der hohe Aufwand bei der simultanen Prüfung komplexer Zusammenhänge.
- Bei der Untersuchung vieler sozialer Zusammenhänge ist die Randomisierung aus praktischen oder ethischen Gründen nicht möglich.

Empirische Sozialforschung

Diekmann, A. S. 304-326

Rico Hofmann

076 577 88 54

Rico@Hofmann.net

Zusammenfassung: Der Abschnitt 3 im Kapitel VIII des Buches (Experimentelle und quasi-experimentelle Designs) illustriert die praktische Anwendung von Experimenten und quasi-Experimenten an einem bekannten und verständlichen Beispiel. Im Abschnitt 4 wird dann auf die Möglichkeiten, Probleme und Bedeutung solcher Quasi-Experimente in der Evaluationsforschung, d.h. in der Untersuchung von Effekten bestimmter Massnahmen, eingegangen. Die beiden Abschnitte sind relativ einfach formuliert und verständlich. Insbesondere werden viele illustrative Beispiele angeführt, auf die ich im Folgenden nicht detailliert eingehe. Wenn die Zusammenfassung des theoretischen Hintergrunds also nicht klar genug sein sollte, wäre es u.U. angebracht, zur Erläuterung kurz die Beispiele im Buch zu überfliegen.

<p>3. Ein Beispiel: Von der Verantwortungsdiffusion zur experimentellen Spieltheorie</p>	<p>Das „Diffusion-von-Verantwortung-Prinzip“ besagt, dass z.B. bei einem Unfall jeder der Zeugen von einem anderen Beobachter erwartet, dass dieser die Hilfeleistung erbringen wird. Da aber jeder Betroffene so denkt, führt das dazu, dass die Hilfeleistung gänzlich unterbleibt. „Je grösser die Zahl der Zuschauer (n) in einer Hilfeleistungssituation ist, desto geringer ist die (individuelle) Wahrscheinlichkeit (p), dass eine <i>bestimmte</i> Person Hilfe leistet.“ Mit Umfragen kann eine solche Hypothese nicht geprüft werden, mit Experimenten zur Simulation solcher Hilfeleistungssituationen kann die Hypothese aber eindrucksvoll bestätigt werden. (Beispiel auf Seite 305)</p> <p>Diese Situation kann in einem spieltheoretischen Modell formalisiert werden: Der Wert der Hilfeleistung sei für alle Beteiligten ein (kollektives) Gut mit dem Nutzen U. Die Hilfeleistung selbst ist aber mit Kosten K verbunden. Ein Trittbrettfahrer, der keine Hilfeleistung erbringt, schliesslich aber auch davon profitiert, erhält dann eine Auszahlung U, während kooperative Personen nur U-K erhalten. (Entscheidungsmatrix siehe Seite 307)</p> <p>Die Lösung dieses „Freiwilligendilemmas“ hängt natürlich davon ab, wie hoch U, K und n sind. Die Gleichung zur Nash-Gleichgewichtslösung lautet:</p> $p = 1 - \sqrt[n]{\frac{K}{U}}$ <p>Die Wahrscheinlichkeit p der Kooperation steigt also mit dem Wert des Kollektivguts U, sinkt mit den Kosten K und – wie von der Hypothese der Verantwortungsdiffusion behauptet – mit der Gruppengrösse n.</p>
<p>4. Quasi-Experimente und Evaluationsforschung</p>	<p>„Quasi-Experimente kann man definieren als Versuchsanordnungen, die dem Vorbild des Experiments nahe kommen und der experimentellen Logik folgen, jedoch nicht die strengen Anforderungen an experimentelle Designs erfüllen.“ In der Regel das zentrale Kriterium der Randomisierung verletzt. Beispielsweise ist bei der Untersuchung und Evaluation von (rechtlichen, wirtschaftlichen oder sozialen) Massnahmen eine Zufallsaufteilung von Untersuchungspersonen auf die einzelnen Versuchsgruppen oftmals nicht möglich. Man greift deshalb auf zwei verschiedene quasi-experimentelle Designs zurück: Versuchsanordnungen</p>

	mit nicht gleichartiger Kontrollgruppe und Zeitreihen-Experimente.
Versuchsordnungen mit nicht gleichartiger Kontrollgruppe	<p>Die erste Möglichkeit entspricht einem Experiment mit einer Vorher-nachher-Messung bei einer Versuchs- und einer Kontrollgruppe, nur dass eben keine Zufallsaufteilung erfolgt:</p> <p>O_1 X O_2 Massnahmegruppe (X: Massnahme) O_3 O_4 Kontrollgruppe</p> <p>Durch die Vorher-nachher-Messung kann ein allfälliger Effekt von X auf die Massnahmegruppe (O_2 verglichen mit O_1) mit der Situation der Kontrollgruppe (O_4 verglichen mit O_3) verglichen werden, z.B. die Wirkung von Förderunterricht (eine Gruppe von Schülern erhält diesen Zusatzunterricht, eine andere nicht).</p> <p>Bei dieser Methode treten allerdings Probleme auf, die zu Verzerrungseffekten führen können:</p> <ol style="list-style-type: none"> 1. Nicht-Vergleichbarkeit der Gruppen, z.B. infolge von Selbstselektion (nur speziell motivierte Kinder melden sich zum Förderunterricht an) 2. Systematischer Ausfall von Probanden (leistungsschwächere Schüler brechen den Förderunterricht eher ab) <p>Beide Faktoren verzerren O_2 „nach oben“, so dass die Massnahme X überschätzt wird.</p> <p>Für diese zwei Probleme bieten sich zwei Lösungen (mit begrenzter Wirkung) an:</p> <p>Beim „paarweisen Matching“ wird jeder Person in der Massnahmegruppe bezüglich spezifischer Kriterien (z.B. Geschlecht, bisherige Schulnoten etc.) ein „Zwilling“ in der Kontrollgruppe zugeordnet.</p> <p>Bei der „nachträglichen Kontrolle von Drittvariablen mit multivariaten statistischen Verfahren“ werden ausgewählte Kontrollvariablen (dummy-Variablen) nachträglich bei der Datenauswertung eingefügt und rechnerisch konstant gehalten.</p> <p>Es bleibt aber das Problem, dass wenn keine Zufallsaufteilung erfolgt, immer gewisse Drittvariablen den Effekt einer Massnahme verzerren können, welche nicht gemessen und explizit berücksichtigt werden können.</p> <p>Eine weitere Fehlerquelle ist der Regressionseffekt (Definition siehe Seite 313), welcher dann Verzerrungen hervorruft, wenn sich die Werte der Vorhermessung in der Versuchs- und der Kontrollgruppe stark unterscheiden. In der Evaluationsforschung sind Fehlinterpretationen aufgrund des Regressionseffekts nicht selten, da Massnahmen häufig erst in extremen Situationen ergriffen werden. Beispielsweise werden gesetzliche Massnahmen zur Kriminalitätsbekämpfung oft erst bei steigender Kriminalitätsrate realisiert. Nimmt die Kriminalität danach ab, wird dieser Effekt den gesetzlichen Massnahmen zugeschrieben, obwohl die Abnahme (z.B. nach einem „Ausnahme-Jahr“) vielleicht ohnehin stattgefunden hätte (Regression zur Mitte).</p>
Zeitreihen-Design	<p>Mit einem Zeitreihen-Design können Regressionseffekte zum Teil kontrolliert werden: Zur Prüfung eines Effekts von X und zur Schätzung der Stärke dieses Effekts werden statistische Methoden der Zeitreihenanalyse verwendet. Statt nur zwei Situationen (vorher und nachher) zu vergleichen, wird ein längerfristiger Trendvergleich angestellt. Hält der Trend nach der Massnahme nachhaltig an, ist der Effekt eher X zuzuschreiben, als wenn er ein Jahr danach gleich wieder abfallen würde. Allerdings besteht auch hier ein Problem, dass nicht genau zwischen einem allgemeinen längerfristigen Trend (der ohnehin besteht) und dem effektiven oder eventuellen Effekt der Massnahme unterschieden werden kann.</p> <p>Um noch ein genaueres Bild zu erhalten, können auch Designs mehrfacher Zeitreihen angewendet werden (d.h. zwei oder mehr Vergleichsgruppen werden über einen längeren Zeitraum hinweg beobachtet), nur ist es in der Regel schwierig, einigermaßen geeignete Vergleichsgruppen zu finden.</p>
Evaluationsforschung	Die quasi-experimentelle Versuchsanordnung ist das typische Design einer Evaluationsstudie im Sinne wissenschaftlicher Untersuchung von

	<p>Massnahmen: Eine Massnahme wird von einem Akteur (Behörde, Legislative, Unternehmen) versuchsweise oder auf Dauer eingeführt. Ziel der sozialwissenschaftlichen Forschung ist die Erfolgskontrolle und die Untersuchung eventueller Nebenwirkungen. Mit der Evaluationsforschung kann z.B. eine „experimentelle Reformpolitik“ betrieben werden: Gesetze, Programme und Massnahmen werden nach Möglichkeit zunächst befristet; erst im Erfolgsfall wird die Massnahme dauerhaft implementiert.</p> <p>Um zu korrekten Aussagen bezüglich der (Un-)Wirksamkeit von Massnahmen zu kommen, ist aber eventuellen Fehlerquellen wie Drittvariableneffekte, Regressionseffekt usw. besondere Aufmerksamkeit zu widmen. Vor allem der Zeitrahmen einer Evaluation ist von besonderer Wichtigkeit. Massnahmen mit einem „Strohfeuereffekt“ (d.h. zuerst grosse Wirkung, die aber mit der Zeit stark nachlässt), werden bei Kurzzeit-Evaluationen überschätzt. Massnahmen mit „Anlaufschwierigkeiten“ (d.h. es ist ein Anpassungs- und Lernprozess nötig) hingegen werden bei zu kurzen Untersuchungszeiträumen unterschätzt. Auch ist Vorsicht geboten bei der Übertragung von Massnahmen, die in einem Kontext erfolgreich waren, auf einen anderen Kontext oder eine andere, (z.B. in kultureller Hinsicht) unterschiedliche Situation.</p> <p>Schliesslich kann auch die Definition des gewünschten Ziels Probleme bereiten. Die Verringerung von CO₂-Ausstoss kann z.B. relativ einfach gemessen und verfolgt werden. Meist ist aber eine Vielzahl von Leistungskriterien zu berücksichtigen. Die Zielkriterien „zufriedene Mitarbeiter“ oder „glückliche Bürger“ wären da als Beispiele wesentlich komplexer...</p>
--	--

7. Stichproben
Diekmann 325, 371

Empirische Sozialforschung
Diekmann, A. S. 325-346

Stefanie Kothmiller

076 566 88 54

Stefanie.Kothmiller@web.de

Zusammenfassung: Im Kapitel IX wird auf die verschiedenen Methoden von Stichproben Auswahlmethoden, ihre Art, die Anwendbarkeit, Aufwendigkeit, Effizienz und empirische Relevanz näher eingegangen. Genauer gesagt werden die Grundbegriffe (Abschnitt 2) vorerst besprochen und dann wird ein Vergleich zwischen Zufallsstichproben (Abschnitt 3) und Quotenauswahl (Abschnitt 4) aufgestellt. In Abschnitt 5 werden Stichproben aus speziellen Populationen näher besprochen. Ich habe versucht, in der Zusammenfassung die Hauptpunkte der Methoden und die theoretischen Aspekte kurz aufzuzeigen. Die Abschnitte im Buch sind aber sehr verständlich formuliert und auch noch mit sehr guten Beispielen untermalt, die man aber am besten zum besseren Verständnis im Buch nachliest.

Kapitel IX: Stichproben

<p>1. Gallup gegen „Literary Digest“</p>	<p>Die Zeitschrift „Literary Digest“ hatte schon öfters so genannte „polls“ (Wahlumfragen) in großem Stil durchgeführt. Dazu wurden zehn Millionen Probestimmzettel an Amerikaner verschickt, deren Adressen im Verzeichnis „Telefon und Auto“ eingetragen waren. Bei der Wahl Landon versus Rosevelt brachte diese Methode allerdings ein Ergebnis zugunsten Landon.</p> <p>Georg Gallup wählte dagegen eine andere Methode, nämlich die der Quotenbefragung, die in wesentlichen Merkmalen einem verkleinerten Abbild der amerikanischen Wählerschaft entsprach, und kam damit mit sehr viel weniger Aufwand zu einer richtigen</p>
---	---

	<p>Einschätzung, nämlich für Rosevelt. Dies bedeutete den Niedergang der Zeitschrift „Literary Digest“ wohingegen die Gallup-Technik und der Name Gallup zum Inbegriff von Meinungsumfragen wurden. Und obwohl später auch einmal Gallup das Wahlergebnis falsch voraussagte, schadete dies seinen Instituten nur wenig und sie sind auch heute noch weltweit tätig.</p> <p>Allerdings ist auch die Quotenmethode noch immer eine umstrittene, vor allem unter Statistikern, weil eigentlich Fehlerparameter dafür nicht wirklich berechnet werden können. Was uns allerdings der Erfolg von Gallup lehrt, ist: Im Allgemeinen gilt das statistische Prinzip, dass größere Stichproben unter sonst gleichen Bedingungen genauer sind als kleinere Stichproben. Diese Regel trifft aber dann nicht zu, wenn eine Stichprobe krass verzerrt ist. Sind große Stichproben stark selektiv, dann werden kleinere, unverzerrte Stichproben in der Regel genauere Schätzungen liefern. Wählt man aber den Weg der Stichprobenziehung, dann besteht immer die Gefahr eines Selektionsbias.</p>
<p>2. Grundbegriffe</p>	<p>Grundgesamtheit: Die Gesamtheit aller zu untersuchenden Elemente (Bsp.: alle wahlberechtigten Personen)</p> <p>Stichprobe: Auswahl von Elementen aus der Grundgesamtheit.</p> <p>Erhebungseinheit: Ein Element der Grundgesamtheit (Bsp.: ein Wähler)</p> <p>Stichprobenverfahren: Ist charakterisierbar durch eine explizite Vorschrift, die festlegt, in welcher Weise Elemente der Grundgesamtheit ausgewählt werden:</p> <ol style="list-style-type: none"> 1. Wahrscheinlichkeitsauswahl 2. bewusste Auswahl 3. willkürliche Auswahl
<p>3. Verschiedene Arten der Wahrscheinlichkeitsauswahl</p>	<p>Einfache Zufallsstichproben: Bei der einfachen Zufallsstichprobe ist</p> <ul style="list-style-type: none"> • die Auswahlwahrscheinlichkeit für alle Elemente der Grundgesamtheit identisch und grösser als null • und die Auswahl erfolgt direkt in einem einstufigen Auswahlvorgang. <p>Dabei wird entweder eine Liste sämtlicher Elemente der Grundgesamtheit herangezogen (Listenauswahl, Karteiauswahl), oder aber die Auswahl basiert auf der Zeichung von Elementen aus einem Gebiet (Gebietsauswahl, Flächenstichprobe).</p> <p>Mögliche Formen:</p> <ul style="list-style-type: none"> • Listenauswahl • Random-Digit-Dialing • Lotteriewahl • Random-Route-Verfahren • Adressrandom • Schwedenschlüssel • Geburtsmethode <p>Mehrstufige Zufallsauswahl Erstreckt sich die Zufallsauswahl über mehrere Ebenen, spricht man von einer mehrstufigen Zufallsauswahl. Zu beachten ist dabei, dass die Auswahlseinheiten auf den einzelnen Stufen meist ungleich gross</p>

	<p>sind, d.h. eine unterschiedliche Anzahl von Elementen der Grundgesamtheit enthalten. Die Elemente müssen daher je nach der Grösse bzw. Häufigkeit gewichtet werden.</p> <p>PPS-Samples Ein wichtiger Spezialfall zweistufiger Zufallsstichproben sind so genannte PPS Samples. Hier ist die Auswahlwahrscheinlichkeit auf der ersten Stufe proportional zur Grösse der Auswahleinheiten. Dabei ist die Auswahl der Elemente in der ersten Stufe proportional zu ihrer Grösse. In der zweiten Stufe wird dann die gleiche Anzahl von Personen aus den Elementen zufällig ausgewählt. Eine nachträgliche Gewichtung dieser ist nicht mehr erforderlich.</p> <p>Klumpenstichproben Eine Klumpenauswahl ist eigentlich ein Spezialfall einer mehrstufigen Zufallsauswahl. Auf der ersten Stufen werden dabei „Klumpen“(Cluster, Bsp.: Schulklassen) zufällig ausgewählt. Auf der zweiten Stufe der Ziehung werden dabei allerdings sämtliche Klumpenelemente berücksichtigt (Bsp.: Schüler), d.h. die Elemente eines Klumpens werden mit Wahrscheinlichkeit eins in die Stichprobe aufgenommen. Die Fehlerbreite ist dabei</p> <ul style="list-style-type: none"> • invers zur Anzahl der Klumpen • wächst mit der Grösse der Klumpen relativ zur Population • wächst mit der Homogenität der Elemente zwischen den Klumpen. <p>Schichtung Ist die Heterogenität eines Merkmals in der Grundgesamtheit sehr hoch, dann kann es lohnenswert sein, das Fehlerintervall der Schätzung mit einer geschichteten Stichprobe zu reduzieren (Bsp. Einkommensklassen). Eine notwendige Voraussetzung der Schichtung ist aber, dass wir über Vorwissen bezüglich der Merkmalverteilung in der Population verfügen. Ausserdem muss die Möglichkeit bestehen, Elemente der Grundgesamtheit separat nach Schichten auszuwählen. Für jede Erhebungseinheit muss die Schichtzugehörigkeit bekannt sein. Je homogener die Elemente innerhalb der Schichten bezüglich des untersuchten Merkmals und je grösser die Heterogenität zwischen den Schichten, desto ausgeprägter ist der Schichtungseffekt.</p>
<p>4. Quotenauswahl</p>	<p>Der Grund für das Vorherrschen und die Beliebtheit von Quotenbefragungen ist nicht unbedingt ihre Effizienz und Genauigkeit, die noch immer sehr umstritten ist, sondern vielmehr, dass sie wesentlich kostengünstiger und weniger aufwendiger sind, vor allem bei persönlichen Interviews. Bei Telefoninterviews wird häufiger eine Zufallsstichprobenauswahl getroffen.</p> <p>Bei der Quotenauswahl wird insofern eine bewusste Auswahl von Elementen aus der Grundgesamtheit vorgenommen, indem gewisse Auswahlkriterien festgelegt werden. Diese sollen garantieren, dass die befragte Stichprobe in einer gewissen Weise die Grundgesamtheit widerspiegelt und somit bei kleinen Stichproben</p>

	<p>bessere Schätzungen getroffen werden können wie bei einer zufälligen Auswahl.</p> <p>Quoten können einfach oder kombiniert sein. Bei kombinierten Quoten wird die gemeinsame Merkmalsverteilung von zwei oder auch mehr Merkmalen vorgegeben. Kombinierte Quoten erfordern ein noch grösseres Vorwissen und stellen erhöhte Anforderungen bei der praktischen Realisierung der Quotenauswahl.</p> <p>Argumente für die Quotenauswahl:</p> <ol style="list-style-type: none"> 1. Die Quotenmerkmale sind mit anderen, eigentlich interessierenden Merkmalen korreliert. Die Stichprobe ist daher nicht nur repräsentativ für die kontrollierten Quotenmerkmale, sondern auch bezüglich der unbekanntem Merkmalsverteilungen in der Population. 2. Wird der Entscheidungsspielraum der Interviewer durch präzise Quoten begrenzt, dann wird der Interviewer innerhalb der Quoten praktisch oder wenigstens näherungsweise eine Zufallsauswahl treffen. Mit einer Quotenauswahl wird mithin eine Zufallsauswahl angenähert. 3. Ein Quotenplan entspricht dem Design einer geschichteten Stichprobe mit proportionaler Schichtung. <p>Diese Argumente sind allerdings sehr umstritten weil eine Korrelation der Merkmale nicht von vornherein angenommen werden kann und auch nicht immer zu finden ist. Man müsste immer vorher die Merkmale untersuchen und dann eine individuelle, auf das jeweilige Thema der Untersuchung abgestimmte Quoten festlegen, was praktisch aber kaum möglich ist.</p> <p>Mit zu komplizierten Vorschriften und Quotenanweisungen wird der Interviewer leicht zum Fälschen erzogen, indem die Daten der Interviewten einfach an die Quotenvorgaben angepasst werden. Auch die These der Schichtenhomogenität konnte empirisch eigentlich widerlegt werden.</p> <p>Die „Theorie“ der Quotenauswahl steht also auf sehr schwachem Fundament und ist somit sehr fraglich. Eine statistische Theorie der Stichprobenziehung existiert nur für Zufallsstichproben, was bedeutet, dass die Anwendung von Formeln zur Fehlerabschätzung nur dafür zulässig ist.</p>
<p>5. Stichproben aus speziellen Populationen</p>	<p>Sind seltene Populationen in einer Liste verzeichnet, dann ist eine Zufallsauswahl kein Problem. Ist dies aber nicht der Fall, was insbesondere auftritt, wenn diese Population durch ein abweichendes Verhalten definiert wird, ist eine Listenauswahl nicht möglich, in diesem Fall können spezielle Sampling-Methoden angewendet werden.</p> <p>Capture-Recapture-Methode:</p> <p>Mit dieser Methode kann unter gewissen Voraussetzungen der Umfang „verborgener“ Populationen ermittelt werden. Dafür werden für eine erste Stichprobe die Elemente zufällig ausgewählt und mit Namen oder einer Kennziffer notiert. Nach einem bestimmten zeitlichen Abstand wird dann eine zweite zufällige Stichprobe gezogen. Danach wird ermittelt, wie viele von den</p>

	<p>Elementen in der zweiten Stichprobe schon in der ersten vorhanden waren.</p> <p>Voraussetzung ist allerdings, dass es sich bei beiden Stichproben um voneinander unabhängige Zufallsstichproben handelt und die Population zwischenzeitlich konstant bleibt. Auch die Wiedererkennung ist ein heikler Punkt dieser Methode.</p> <p>Schneeballtechnik</p> <p>Die Fragebögen werden dafür z.B. an Freunde und Bekannte verteilt, die das entsprechende Verhalten aufweisen, welche wiederum weitere Fragebögen in ihrem Freundes- und Bekanntenkreis weiterreichen.</p> <p>Damit eng verwandt ist die <i>Nominationsmethode</i>, bei der von anonym befragten Personen weitere Personen angegeben werden, die den Vorgaben entsprechen würden.</p>
--	--

Empirische Sozialforschung

Diekmann, A. S. 347-369

Maila Schärer

079 699 14 26

maila.schaerer@unisg.ch

<p>6. Theorie der Zufallsstichproben</p>	<p>Fall 1: dichotome Variablen: Mit der Informationen einer Zufallsstichprobe wollen wir die jeweiligen Anteilswerte in der Population schätzen. N= Stichprobenumfang, n= die Anzahl Personen in der interessierenden Merkmalskategorie in der Stichprobe, p= der unbekannte Populationsparameter „Anteil der Population“, p^{\wedge}= der Stichprobenschätzwert des Parameters. <i>Zur Schätzung des p: $p^{\wedge} = n/N$</i></p> <p>Bsp: $N=120$ Pers.= Zufallsstichprobe aus einem Stimmenbezirk, $n=48$ Pers die angeben SP-Wähler zu sein. Der Schätzwert p^{\wedge} für den Anteil der SP-Wähler in dem Stimmenbezirk beträgt dann $0,40 = 40\%$.</p> <p>Wie gross ist der <i>Stichprobenfehler</i>? Dazu muss man die Stichprobenverteilung von p^{\wedge} berechnen. Man stellt sich vor der wahre Wert von p ist $=40\%$, es werden jetzt sehr viele hypothetische Stichproben von $N=120$ gezogen, für jede von denen wird p^{\wedge} berechnet. Die meisten p^{\wedge}s werden in der Nähe von $0,40$ liegen, also liegt eine Binomialverteilung vor. Von dieser weiss man, dass sie sich sehr rasch der Normalverteilung annähert, ist p nicht sehr klein oder sehr gross, so ist die Normalverteilung schon ab $N \geq 30$ eine gute Approximation für die Binomialverteilung.</p> <p>Wir ziehen aber im Normalfall nur eine Stichprobe und berechnen nur einen Wert von p^{\wedge}. mit einem statistischen Modell können wir aber die hypothetische Frage beantworten, welche p^{\wedge}- Werte mit welcher Häufigkeit wir erhalten würden, wenn eine sehr grosse Anzahl von Stichproben mit gleichem N aus einer Grundgesamtheit mit dem wahren Parameterwert p gezogen würden. Hierüber informiert die Stichprobenverteilung. Anhand der Streuung der Stichprobenverteilung(=Standardfehler) kann dann ein Fehlerintervall der konkreten Stichprobenschätzung bestimmt werden. Die Stichprobenverteilung ist für kleine N eine Binomialverteilung, für $N \geq 30$ eine Normalverteilung. P^{\wedge} ist normalverteilt mit:</p> <ol style="list-style-type: none"> 1. Erwartungswert (Mittelwert) p 2. Standardfehler: $\sqrt{(p(1-p)/N)}$ <p>Den Standardfehler können wir durch ersetzen von p mit p^{\wedge} schätzen. Die Stichprobenanteile streuen sich also entsprechend einer Gausschen Glockenkurve um den Mittelwert p, den wahren Anteilswert in der Grundgesamtheit.</p> <p>Man möchte meistens das Fehlerintervall auf dem 95%- oder 99%-Niveau berechnen. Mit einer Wkeit von $0,95$ liegt der Wert einer normalverteilten Variable im Bereich von $\pm 1,96$ Standardabweichungen um den Mittelwert, bei einer Wkeit von $0,99$ sind es $\pm 2,58$ Standardabweichungen. Die kritischen Grenzen der Normalverteilung für den 95% bzw. 99% oder generell $w\%$- Bereich bezeichnen wir mit $z(95)$, $z(99)$, generell $z(w)$. die <i>Grenzen des Konfidenzintervalls</i> sind I_1 und I_2. damit erhalten wir die Formel für die Berechnung des Vertrauensbereich:</p> $I_{1,2} = p^{\wedge} \pm z(w) * \sqrt{(p^{\wedge}(1-p^{\wedge})/N)}$ <p>Bsp. von oben: $I_{1,2} = 0,40 \pm 1,96 * \sqrt{(0,40(1-0,40)/120)} = 0,40 \pm 0,088$</p> <p>d.h. mit einer Wkeit von $0,95$ liegt der geschätzte Wähleranteil der SP-Wähler in dem Stimmenbezirk im Bereich von $31,2\%$ bis $48,8\%$.</p> <p><i>\sqrt{N}-Gesetz:</i> Der Vertrauensbereich ist umgekehrt proportional zur Wurzel aus dem Stichprobenumfang. D.h. will man die Präzision der Schätzung verbessern, z.B. den</p>
--	---

	<p><i>Vertrauensbereich halbieren, dann muss der Umfang der Stichprobe vervierfacht werden.</i> Mit dieser Formel kann auch der Stichprobenumfang berechnet werden, der erforderlich wäre, um eine gewünschte Präzision zu erzielen. Voraussetzung ist, dass wir eine Vermutung über die Grössenordnung von p^{\wedge} haben. $N = p^{\wedge}(1-p^{\wedge}) * (z(w)/h)^2$, $h =$ die Länge des Vertrauensbereichs.</p> <p>Fall 2: metrische Variablen: Wir gehen davon aus, dass eine metrische variable X in der Population normalverteilt ist mit dem Mittelwert $\mu(x)$ und Standardabweichung $\sigma(x)$. $\mu(x)$ wird durch \bar{x} geschätzt. Dann sind die \bar{x}-Werte aller hypothetischen Stichproben des Umfangs N normalverteilt um den Mittelwert $\mu(x)$ und Standardabweichung. <i>Standardfehler = $\sigma(x) / \sqrt{N}$</i> Man wird i.d.R. die Standardabweichung in der Population nicht kennen, deshalb wird geschätzt: $s(x) = \sigma^{\wedge}$: $s(x) = \sqrt{(1/N-1) * \sum (\bar{x} - x)^2}$</p> <p>Für das Konfidenzintervall gilt wiederum die Formel: $\bar{I}_{1,2} = \text{Stichprobenschätzwert} \pm t(w) * \text{Standardfehler}$: $\bar{I}_{1,2} = \bar{x} \pm t(w) * s(x) / \sqrt{N}$ Oder für $N < 30$ und einen 95%-Vertrauensbereich: $\bar{I}_{1,2} = \bar{x} \pm 2 * s(x) / \sqrt{N}$ Das “\sqrt{N}- Gesetz“ gilt auch hier, das heisst die Präzision der Schätzung wächst mit \sqrt{N}.</p>
7.Zufallsstichproben in der Praxis	Bei allgemeinen Bevölkerungsumfragen verliert die Theorie der Zufallsstichprobe in der Praxis schnell an Bedeutung. Warum das so ist werden wir v.a. anhand der Fehlerquellen feststellen, ausserdem befassen wir uns mit einem typischen Zufallsstichprobendesign bei Bevölkerungsumfragen, dem ADM-Design.
ADM-Design	<p>Beim ADM-Design handelt es sich um eine 3-stufige Zufallsstichprobe mit Gebietsauswahl, also eine Flächenstichprobe. Als Grundgesamtheit gelten alle Privathaushalte in Deutschland.</p> <ol style="list-style-type: none"> 1. Stufe: die Stimmbezirke werden mit einer Wahrscheinlichkeit proportional zur Grösse ausgewählt. 2. Stufe: Mit dem Random-Route-Methode wird eine Zufallsstichprobe von Adressen ermittelt. 3. Stufe: die zu befragende Person wird per Schwedenschlüssel ausgewählt.
Fehlerquellen	<p>In der Survey-Forschung unterscheidet man 3 Fehlerquellen, der Gesamtfehler setzt sich aus diesen zusammen:</p> <ol style="list-style-type: none"> 1. Zufallsfehler der Stichprobe (sampling variability) 2. Systematische Fehler aufgrund des Verfahrens der Stichprobenauswahl 3. Verzerrungen, die nicht direkt durch das auswahlverfahren produziert wurden (non sampling bias). <p>(Die 1. Fehlerquelle schon behandelt), 2. Systematische Fehler treten auf, wenn Elemente der Population mit grösserer oder geringerer W'keit in das Sample eingehen. Sofern die unterschiedlichen W'keiten bekannt sind, kann der Bias durch entsprechende Gewichtung korrigiert werden. Problematischer sind systematische Fehler in der Auswahl aufgrund von praktischen Problemen mit Random-Route und Schwedenschlüssel. Bei beiden Verfahren besteht keine Garantie, dass die festgelegten Regeln auch eingehalten werden. Bsp.: Schwedenschlüsselproblem: Interviewer werden pro Interview bezahlt. Gemäss Regel muss Person X befragt werden, aber nur Person Y ist anwesend und auskunftsbereit. Interviewer befragt Y und gibt an X befragt zu haben. So werden die Auswahlregeln verletzt und leichterreichbare Personen werden im Sample überrepräsentiert sein. Diese W'keiten sind nicht bestimmbar und können somit auch nicht durch nachträgliche Gewichtung korrigiert werden.</p> <p>Das Hauptproblem bei Umfragen ist aber Fehlerquelle 3: Messfehler, Fehler im Interviews, Diskrepanz zwischen Zielpopulation und Surveypopulation (Unterdeckung= Zielpopulation grösser als Surveypopulation) und Non-response.</p>

<p>Non-Response</p>	<p>Eines der grössten Probleme bei Umfragen ist die erhebliche Zahl von Ausfällen durch Verweigerung oder Nicht-Erreichbarkeit.</p> <p><i>Ausschöpfungsquotient</i> $A = \frac{\text{Anzahl ausgewerteter Interviews}}{\text{Bereinigter Stichprobenumfang}} * 100$</p> <p>Bsp.: $A = \frac{2955}{4620-367} * 100 = 69.5\%$, d.h. 4620= Bruttostichprobe, 2955 auswertbare Interviews, 367= stichprobenneutrale Ausfälle, d.h. gewisse Kriterien nicht erfüllt, z.B nichtdeutscher Haushalt usw.</p> <p>Die <i>Ausfallquote</i> beträgt also in diesem Bsp. $100-A = 29.5\%$</p> <p>Es zeigt sich relativ deutlich ein steigender Trend bei der Nichterreichbarkeit, während die Anzahl nicht-kooperativer Personen stark vom befragenden Institut abhängt (Institutsfaktor). Nicht nur bei den Interviewern besteht Raum zur Interpretation, auch auf der Institutsebene kann manches getan werden um die Ausschöpfungsquote besser darzustellen, v.a. die „stichprobenneutralen Ausfälle“ sind Dreh- und Angelpunkt. Wer viele Ausfälle in diese Kategorie packt wird eine höhere Ausschöpfungsquote erreichen.</p> <p>Welche Probleme bereitet denn eine niedrige Ausschöpfungsquote bei der inhaltlichen Interpretation der Ergebnisse? Es ist logisch, dass Ausfälle in der Grössenordnung von 40% nicht zufällig erfolgen, sondern es sind v.a. Singels und mobile Personen die schwer erreichbar und deshalb im Sample unterrepräsentiert sind. Deshalb ist die resultierende Stichprobe verzerrt.</p> <p>Ein besonderes Problem sind die Ausfallquoten bei replikativen Umfragen. Hier bereitet die <i>Veränderung der Ausschöpfungsquoten</i> Probleme. Da die Ausschöpfungsquoten institutsabhängig sind empfiehlt es sich dasselbe Institut mit den replikativen Umfragen zu befassen. Auch wenn es schlechte Arbeit leistet, sind die Verzerrungen geringer als wenn man es wechselt.</p> <p>Was tun um die Ausfallquote zu senken? Mehr Kontaktversuche, Einsatz von telefonischer und schriftlicher Befragung, Kooperationsbereitschaft erhöhen mit monetären Anreizen usw. Gute Bevölkerungsumfragen sind dann eben teurer als manche „quick and dirty“- Studien.</p> <p>Es ist darauf hinzuweisen, dass das Non-Response- Problem v.a. ein Problem bei Schätzungen von Mittelwerten und Anteilswerten vieler Merkmale der Population ist. Zusammenhänge werden durch Non-Response weniger verzerrt. Die Schätzung von Korrelationen ist gegenüber systematischen Stichprobenfehlern i.d.R. robuster.</p> <p>Das <i>Item-non-Response-Problem</i> tritt v.a. bei Einkommenschätzungen auf, d.h. der Befragte erklärt sich einverstanden sich interviewen zu lassen, aber verweigert dann die Beantwortung einzelner Fragen, eben z.B. über sein Einkommen. Bsp.: 70% beteiligen sich an Umfrage und 70% von diesen beantworten auch die Einkommensfrage, somit liegt die Ausschöpfungsquote dieser Frage bei gerade $0,7 * 0,7 = 0,49$, also 49%.</p>
<p>Gewichtung</p>	<p>Zur Schätzung der Populationsparametern werden die Rohdaten der Stichprobe häufig gewichtet. Es gibt 3 verschiedene Arten der Gewichtung. Die 2. und die 3. Art sind umstritten:</p> <ol style="list-style-type: none"> 1. theoretisch-statistische Gewichtung aufgrund der bekannten Auswahlw'keiten gemäss Stichprobenplan (Designgewicht). 2. Nachgewichtung 3. Gewichtung aufgrund einer empirischen Hypothese <p>1. Beim ADM-Design werden Zielpersonen mit dem Schwedenschlüssel ausgewählt, dann ist die Auswahlw'keit $1/m$, wobei m die Haushaltsgrösse bezeichnet. Ohne Gewichtung wären Personen aus kleinen Haushalten (bei gleicher Erreichbarkeit) in der Stichprobe überproportional vertreten. Durch die Gewichtung mit der Haushaltsgrösse wird diese Verzerrung ausgeglichen.</p> <p>2. Bei der Nachgewichtung wird die Verteilungen der Stichprobe nachträglich an bekannte Verteilungen der Grundgesamtheit angepasst. Bps.: 30% Männer, 70% Frauen von $N=1000$, weil Frauen tagsüber eher erreichbar, da diese Verteilung unschön aussieht wird nachgewichtet. Aus amtlichen Statistiken ist bekannt, dass der Anteil der Männer in der Population 53:47 beträgt, also werden alle männlichen Interviews hochgewichtet. Dies ist allerdings nicht unbedingt gerechtfertigt.</p> <p>3. Es wird eine empirische Hypothese zugrundegelegt. Ob diese Gewichtung das Ergebnis verbessert hängt ab inwieweit diese Hypothese empirisch zutreffend ist. Es ist aber sehr schwierig eine Hypothese objektiv und unabhängig zu überprüfen.</p>

<p>Repräsentative Stichprobe?</p>	<p>Problem: Eine Stichprobe repräsentiert niemals sämtliche Merkmalsverteilungen einer Population. Deshalb ist der Ausdruck „repräsentative Stichprobe“ mit Vorsicht zu verwenden. Bei einer Untersuchung sollen genaue Angaben über ihre Charakteristika gemacht werden :</p> <ol style="list-style-type: none"> 1. Angabe der Stichprobentechnik (Quoten- oder Zufallsstichprobe) 2. Quotenmerkmale oder Ausschöpfungsquote (je nach Stichprobentechnik) 3. Anzahl der realisierten Interviews 4. Erhebungsmethode (Telefonisch, schriftlich usw.) 5. Gewichtungungsverfahren 6. Angabe der genauen Fragetexte mit Antwortkategorien <p>Angesichts der zahlreichen Fehlerquellen ist unbedingt Transparenz nötig.</p> <p>Bei Hypothesentests denkt man weniger über die „Repräsentativität“ nach, sondern vielmehr über adäquate Designs der Varianzkontrolle und die Ausschaltung von Störfaktoren. Diese Überlegung bezieht sich demnach v.a. auf Schätzungen von Verteilungen durch Zufallsstichproben oder Quotenstichproben.</p>
-----------------------------------	---

8. Befragung
Diekmann: 373, 439

Empirische Sozialforschung
Diekmann, A. S. 371 - 399

Sara Schuler

071 / 220 61 17

sara-christina.schuler@unisg.ch

Zusammenfassung:
 In diesem Kapitel geht es um die Arten von Interviewtechniken, die Art der Fragestellungen und das Antwortverhalten der Befragten. Sehr zentral ist in diesem Teil der sogenannte „Ort der sozialen Erwünschtheit“, welcher bei der Antwortreaktionen der Befragten erhebliche Verzerrungen hervorrufen kann.

<p>1. Stellenwert in der Sozialforschung</p>	<p>Das Interview ist eine ausgesprochen reaktive Erhebungsmethode, was heisst, dass die Interviewsituation, das Interviewverhalten und die Art und Weise der Konstruktion des Messinstruments „Fragebogens“, nicht ohne Einfluss auf die Antwortreaktionen sind. Zur Erhebung sozial- und wirtschaftsstatistischer Daten (allgemeine Bevölkerungserhebung, Meinungsforschung etc.) ist die Methode der Befragung aber trotz aller Kritik unverzichtbar. Die heutigen Kenntnisse der Sozialstruktur und sozialen Schichtung, der Mobilität und Bildungschancen und anderer sozialstruktureller Merkmale und Zusammenhänge sind hauptsächlich das Ergebnis quantitativer Bevölkerungsbefragungen.</p>
---	---

<p>2. Formen der Befragung</p>	<p>Es können 3 Typen der Befragung unterschieden werden:</p> <ol style="list-style-type: none"> 1) das persönliche Interview (Face-to-Face) 2) das telefonische Interview 3) die schriftliche Befragung (questionnaire) <p>Befragungen können aber auch nach dem <i>Grad der Strukturierung oder der Standardisierung</i> unterschieden werden. Dabei handelt es sich um ein Kontinuum mit den Polen „vollständig strukturiert“ auf der einen und „unstrukturiert, offen“ auf der anderen Seite. Bei einem vollständig strukturierten Interview werden:</p> <ol style="list-style-type: none"> a) alle Fragen mit b) vorgegebenen Antwortkategorien in c) festgelegter Reihenfolge gestellt. <p>Offene Interviews dagegen erfordern nur minimale Vorgaben, im Extremfall nur die Vorgabe eines Themas der Befragung.</p> <p>Objektivität, Reliabilität und Validität sind Ansprüche, denen das Messinstrument „Fragebogen“ in möglichst hohem Grad gerecht werden sollte. Allem voran sollte die Objektivität einen möglichst hohen Wert aufweisen. Die Standardisierung fordert aber auch einen Preis: so erhält man bei geschlossenen Fragen keine Informationen jenseits des Spektrums der vorgelegten Antwortkategorien. Häufig verwendet man deshalb Mischformen, die sowohl standardisierte als auch weniger strukturierte Fragen beinhalten.</p> <p>Bei stark strukturierten Interviews sprechen wir auch von <i>quantitativen</i> Befragungen. Dagegen zählen weniger strukturierte Interviewtechniken (Leitfadeninterview, fokussiertes oder narratives Interview) zu den <i>qualitativen</i> Methoden der Befragung.</p> <p>Bei der <i>Gruppendiskussion</i> (ein Interviewer als Moderator, mehrere Befragte) liegt das Ziel in der Erforschung von Meinungsbildungsprozessen.</p>
<p>3. Theorie des Interviews</p>	<p>Definition Interview: “Ein planmässiges Vorgehen mit wissenschaftlicher Zielsetzung, bei dem die Versuchsperson durch eine Reihe gezielter Fragen oder mitgeteilter Stimuli zu verbalen Informationen veranlasst werden soll.“</p> <p>Ein Interview</p> <ul style="list-style-type: none"> - knüpft an die alltägliche Situation des Fragestellens an. - besteht aus einem Gespräch. - ist eine künstliche, asymmetrische Interaktion zwischen Fremden mit der stillschweigenden Vereinbarung, dass keine dauerhafte Beziehung eingegangen wird. <p>Die <i>neutrale</i> Interviewtechnik (im Gegensatz dazu die asymmetrische Interaktion wie z.B. Verhör oder Prüfung) versucht der befragten Person das Gefühl zu vermitteln, dass ihre Antwortreaktionen weder während des Interviews noch im nachhinein in irgendeiner Weise positiv oder negativ sanktioniert werden. Denn normalerweise werden die Reaktionen in einem Alltagsgespräch von Zustimmung oder Missbilligung. Bei der neutralen Interviewtechnik erhofft man sich jedoch, dass die Chancen gross ist, unverfälschte Antworten zu erhalten (da der Sanktionsmechanismus wegfällt). In der Praxis ist der Neutralitätsanspruch allerdings ein unerreichbares Ideal, denn Körpersprache, Mimik, Stimmlage etc. des Interviewers werden von der befragten Person immer gewertet.</p> <p>Anders verhält es sich bei der „weichen“ und bei der „harten“ Interviewtechnik. Hier wird die Neutralitätsforderung bewusst aufgegeben. In „weichen“, nicht-direktiven Interviews soll der Interviewer durch zustimmenden Reaktionen Hemmungen bewusst abbauen, das Gespräch unterstützen und so zu weiteren (möglichst unverfälschten) Antworten ermuntern.</p> <p>Bei der „harten“ Interviewtechnik tritt der Interviewer autoritär auf, macht auf Widersprüche aufmerksam und versucht (wie bei einem Verhör), „dem</p>

	<p>Befragten das Gefühl der Zwecklosigkeit unvollkommener Aufgaben zu vermitteln“.</p> <p>Die Methode des Interviews ist nur anwendbar, wenn die folgenden Bedingungen gegeben sind:</p> <ol style="list-style-type: none"> 1) Kooperation der Befragten im Regelfall, 2) Die Existenz „Aufrichtigkeit“ in Gesprächen mit Fremden, 3) Eine „gemeinsame Sprache“ zwischen Interviewer und Befragtem. <p>Die Methode der Erhebung und speziell bei Interviews die Art der Fragestellung sollte jeweils dem kulturellen Kontext angepasst werden. Die Bedeutung von Fragen und Antworten sollte vom Interviewer als auch vom Befragten in gleicher Weise interpretiert werden. Insbesondere in Randgruppen und speziellen Kulturen kann diese Bedingung keineswegs immer erfüllt werden. Probleme der Verständigung kann man in Pre-Tests herausfinden. Auch durch die Auswahl von Interviewern, die mit der entsprechenden Subkultur vertraut sind, lassen sich manche Verständigungsprobleme umgehen.</p>
<p>Die soziale Erwünschtheit</p>	<p>Mit Theorien des Interviews wird beabsichtigt, die Antwortreaktionen von Befragten systematisch zu erklären. Es gibt dabei <i>drei Perspektiven</i>:</p> <ol style="list-style-type: none"> a) Die faktorenanalytische „Theorie der Frage“ (nach HOLM) b) Die Anwendung der rationalen Entscheidungstheorie auf das Befragtenverhalten. c) Die Anwendung von Hypothesen zur Informationsverarbeitung aus der kognitiven Psychologie. <p>Bei HOLM's „Theorie der Frage“ setzt die konkrete Anwendung voraus, dass eine Variable im Fragebogen mit einer Fragebatterie, d.h. mit mehreren Items, operationalisiert wurde. Die Antwortreaktionen auf die einzelnen Fragen (Items) der Fragebatterie sind nun nicht nur von der eigentlich interessierenden Zieldimension abhängig. Die Antwortreaktionen können verzerrt werden zum einen durch eine oder mehrere <i>Fremddimensionen</i>, zum anderen durch den Faktor der <i>sozialen Erwünschtheit</i>.</p> <p>Mit der <i>sozialen Erwünschtheit</i> wird die Tendenz umschrieben, dass jemand sein tatsächliches Verhalten oder die eigene Meinung in Richtung auf das sozial erwartete Verhalten oder die sozial erwartete Meinung verzerrt. Dies kann bewusst oder unbewusst geschehen. (wird z.B. eine Person nach ihrem Umweltverhalten befragt, so ist anzunehmen, dass die sie eigenen Aktivitäten eher „schönfärberisch“ nach oben verzerrt). HOLM formulierte dafür ein linear additives Modell (siehe S. 379).</p> <p>Wird einer Person eine Frage gestellt, so muss sie eine Entscheidung treffen, welche Antwortalternativen sie mitteilt oder in einem schriftlichen Antwortfragebogen ankreuzt. Es liegt daher nahe, mit allgemeinen Entscheidungstheorien auch das Antwortverhalten der Befragten in Interviews näher zu untersuchen: Gleich ob bewusst oder unbewusst, bei der Entscheidung für eine bestimmte Antwortreaktion spielen unterschiedliche Motive eine Rolle. Mit der Anwendung von Entscheidungstheorien wird das Verhalten des Befragten aus einer Kosten-Nutzen-Perspektive analysiert. Nach ESSER sind es vor allem zwei</p>

	<p>motivale Grundbedürfnisse, die das Handeln der Befragten bestimmen:</p> <ul style="list-style-type: none"> - soziale Anerkennung - Vermeidung von Missbilligung. <p>An einem anderen Punkt setzt die Hypothese zur Informationsverarbeitung an. So richten sich in einem Interview i.d.R. zahlreiche Fragen auf die Ermittlung von Ereignissen des Befragten, die in der Vergangenheit liegen (Retroperspektive). Bekanntlich sind diese nur unscharf im Gedächtnis gespeichert. Die Abrufung von Informationen aus dem Gedächtnis erfolgt nun nicht nur passiv, sondern unterliegt psychologischen Gesetzmässigkeiten aktiver Informationsverarbeitung (er die Verknüpfung mit Hinweisen [clues] und parallelen Ereignissen [Anker] ermöglicht das Abrufen).</p> <p>Fazit: Systematische Verzerrungen der Antwortreaktionen sind nach dieser Theorie durch die Eigentümlichkeiten des psychologischen Prozesses der Informationsverarbeitung erklärbar. Im Gegensatz zum Modell der Entscheidungstheorie sind verfälschte Antworten nicht allein auf motivationale Faktoren (u.a. Vermeidung Missbilligung und Wunsch nach sozialer Anerkennung) zurückzuführen.</p>
<p>4. Fehlerquellen im Interview</p>	<p>Eine wichtige Aufgabe der Methodenforschung ist die Untersuchung systematischer Antwortverzerrungen im Interview. Dabei bedient man sich häufig der Technik des „Fragesplits“ (Frageversion A oder B) oder des „Methodensplits“ (Interviewsituation, Reihenfolge der Fragen etc.). Wir unterscheiden drei Kategorien von Fehlerquellen:</p> <ol style="list-style-type: none"> 1) Befragtenmerkmale (soziale Erwünschtheit, Response-Set, „Meinungslose“) 2) Fragemerkmale (Frageformulierung, Frageposition, Effekt von Antwortkategorien) 3) Merkmale des Interviewers und der Interviewsituation (Interviewermerkmale, Anwesenheit Dritter, Interviewsituation)
<p>1) Befragtenmerkmale</p>	<p>Fast alle sozialen Aktivitäten und Eigenschaften werden Bewertungen unterzogen. Je nach Schicht oder sozialer Klassenzugehörigkeit, Bildungsgrad und weiteren Merkmalen werden die Bewertungen individuell variieren. Mit dem <i>Ort der sozialen Erwünschtheit</i> bezeichnen wir die Lage des subjektiv wahrgenommenen Maximums einer positiven Bewertung von Handlungen, Meinungen oder anderen Eigenschaften. Je weiter nun der tatsächliche Variablenwert von diesem Ort abweicht, desto unangenehmer wird die Angabe der entsprechenden Antwortkategorie empfunden.</p> <p>Nach KOOLWIJK kann man das „Ausmass der Unangenehmheit“ (Y) in Abhängigkeit von dem im Interview offenbarten „Variablenwert“ (X) als Kurve in einem Diagramm darstellen (siehe S. 384). Die Personen wurden</p>

	<p>in diesem Bsp. nach ihrem täglichen Fernsehkonsum befragt. Die befragten Personen machten in diesem Bsp. einen Kompromiss zwischen Wahrheitsliebe und dem (nach ihrer Ansicht) sozial erwünschten Verhalten. Die Angabe von zwei Stunden täglichem Fernsehkonsum ist in der Kurve also durch den Effekt der sozialen Erwünschtheit verzerrt.</p> <p>Je heikler eine Frage ist – die Kurve würde sich dann entsprechend nach oben verschieben – desto stärker wird sich aber der Effekt der sozialen Wünschbarkeit bemerkbar machen. Des weiteren ist der „wahre“ Wert der befragten Person selbst keineswegs immer genau bekannt. Bei retrospektiven Verhaltensangaben z.B. kommen noch Erinnerungsfehler hinzu. Manchmal wird auch einfach geschätzt und innerhalb dieses Schätzbereichs wird eher der Ort der sozialen Wünschbarkeit mitgeteilt, als der „wahre“ Schätzwert.</p> <p>Wir können daraus also die Hypothese ableiten: <i>Je heikler eine Frage aus der Sicht einer Person ist und je grösser der Unsicherheitsbereich bezüglich des wahren Variablenwertes, desto stärker wird sich der Effekt der sozialen Erwünschtheit bemerkbar machen.</i></p> <p>Gegenmassnahmen dazu wären:</p> <ul style="list-style-type: none"> - möglichst neutrale Frageformulierungen wählen - die gegenteilige Strategie suggestiver Fragen, die einleitend im Fragetext ein abweichendes Verhalten als normales Verhalten darstellt, wird in Ausnahmefällen angewendet, jedoch nicht empfohlen. (z.B. „viele Menschen lassen gelegentlich in einem Geschäft etwas mitgehen, ohne die Ware zu bezahlen. Ist Ihnen das auch schon passiert?“) - bei persönlichen Interviews die Antwort in einem Kuvert verschliessen. - Die vollständige Anonymisierbarkeit mittels der Random-Response-Technik garantieren (siehe Punkt 7). - Skalen zur Messung der sozialen Erwünschtheit ermitteln. <p>Fragen nach der Kenntnis oder Bewertung von Personen, Produkten oder Sachverhalten setzen Vorwissen voraus. Zur Kategorie des Effekts der sozialen Erwünschtheit kann auch gezählt werden, dass die meisten Personen <i>ungern ihr Nichtwissen eingestehen</i>, selbst wenn die Antwortkategorie „weiss nicht“ vorgegeben ist.</p> <p>Ein weiterer typischer Spezialfall ist die <i>Akquieszenz</i> (Ja-Sage-Tendenz) von gewissen Befragten (kann auch aufgrund eines kulturellen Höflichkeitkontextes geschehen) oder dass es nicht genügt, Fragebogen einfach in eine andere Sprache zu übersetzen.</p> <p>Ein weitere Problemkreise wird <i>Meinungslosigkeit, Pseudo-Opinions oder „Non-Attitude-Problem“</i> genannt. Die übliche Praxis besteht darin, meinungslose Personen durch einen Filter auszuschliessen. Eine mögliche Filterfrage kann z.B. sein: „Haben Sie zum Thema X eine Meinung?“ Falls nicht, dann wird diese Person nicht mehr weiter befragt. Filterfragen müssen jedoch gut überdacht werden, denn sie können für die Befragten auch eine Signalwirkung haben und abschrecken wirken. Manche Befragten glauben dann z.B., die darauffolgenden Fragen würden besonders komplex oder schwierig werden. Auf diese Weise werden auch Personen ausgefiltert, die durchaus etwas zum Befragungsthema zu sagen hätten. So ist es im Regelfall besser, auf globale Filterfragen zur Meinungslosigkeit zu verzichten. Statt dessen empfiehlt sich die Alternative, Personen nach ihrer Meinung und einem eventuellen Meinungsbildungsprozess oder Kenntnisse zum Thema zu befragen.</p>
<p>Wahlprognosen und Wahlbörsen – Beispiel einer Alternative zur Umfrageforschung</p>	<p>Soziale Erwünschtheit, Meinungslosigkeit und darüber hinaus Stichprobenprobleme verzerren auch die Ergebnisse von politischen Wahlumfragen. Eine interessante Alternative zur beliebten „Sonntagsfrage“ (Wen würden Sie wählen, wenn nächsten SO Wahlen wären?) ist daher die Wahlbörse. Die Teilnehmer der Wahlbörse haben ein Eigeninteresse daran, das Ergebnis möglichst exakt zu prognostizieren und</p>

	entsprechend zu handeln. Prognosefehler werden nämlich vom Markt durch Gewinneinbussen bestraft (für genauere Angaben, lese Kasten S. 390).
2) Frageeffekte	<p>Die Art der Frageformulierung kann die Antwortreaktionen erheblich beeinflussen. Z.B. gibt eine wesentliche Differenz von ca. 20% zwischen „verbieten“ oder „nicht erlauben“ innerhalb einer Fragestellung.</p> <p>Auch <i>Rating und Ranking</i> sind zwei verschiedene Fragetechniken zur Ermittlung der Prioritäten von politischen Themen oder Werten.</p> <p>Beim <i>Rating</i>-Verfahren wird die Bedeutung jeweils separat auf einer Skala mit den Polen „sehr wichtig“ bis „überhaupt nicht wichtig“ eingestuft.</p> <p>Das <i>Ranking</i>-Verfahren sieht dagegen vor, dass die Befragten die einzelnen Themen nach ihrer Wichtigkeit in eine Rangfolge bringen.</p> <p>Es ist leicht einsehbar, dass ein Thema X (z.B. Was empfinden Sie als die grösste politische Herausforderung? - Umweltschutz) beim Rating höhere Prozentwerte der Wichtigkeit erhält als beim Ranking, da beim Ranking ja jeweils auch noch die Alternativen berücksichtigt werden müssen (z.B. Umweltschutz, Arbeitslosigkeit, Wirtschaftswachstum in eine Reihenfolge bringen). Die beiden Frageversionen produzieren höchst unterschiedliche Rangfolgen.</p> <p>Des weiteren wird der <i>Einfluss von vorgegebenen Antwortkategorien</i> auf die Häufigkeit/Dauer von selbstberichtetem Verhalten untersucht. D.h. dass Leute bezüglich ihres TV-Konsums befragt wurden. In der 1. Frageversion wurde eine Kategorie „bis 2.5 h“ genannt, in der 2. Frageversion wurde diese Kategorie in 5 Unterkategorien (von „1/2 h“... bis „2.5 h“) unterteilt, was dazu führt, dass sich eine Person mit 2.5 h TV-Konsum bei der Version 1 im unteren Bereich der Skala befindet, bei Version 2 jedoch im oberen Extrembereich. Kaum einer wird jedoch seinen täglichen TV-Konsum kennen, also orientiert er sich bei seiner Schätzung an „Referenzpunkten“. Warum? – weil alltägliche Routinehandlungen von geringer subjektiver Bedeutung im episodischen Gedächtnis nur unscharf gespeichert werden. Wer also glaubt, sein TV-Konsum liege im Durchschnitt, wird bei Version 1 eine weit höhere Schätzung angeben als bei Version 2.</p> <p>Solche <i>Retroperspektivfragen</i> zu selbstberichteten Verhaltensaktivitäten kommen in Umfragen sehr häufig vor. Wird nach der Dauer/Häufigkeit eines Verhaltens gefragt und sind dabei Verzerrungen zu erwarten (Erinnerungsprobleme, Schätzstrategien), gibt es eine einfache Alternative: man stellt eine „offene Frage“ (ohne vorgegebene Antwortkategorien!!).</p> <p>„<i>Verankerungen</i>“ sind bei der Fragestellung sehr wichtig (besser: „Haben Sie zwischen Jan 03 und Jan 04 ein Auto gekauft?“ anstelle von „Haben Sie in den letzten 12 Mt. ein Auto gekauft?“), denn bei einer unpräzisen Referenzzeit werden emotional hoch bewertete Handlungen häufig auch dann berichtet, wenn sie vor diesem Zeitraum liegen. Gut erinnerte Ereignisse erscheinen zeitlich näher zur Gegenwart, als sie tatsächlich stattgefunden haben (Teleskopeffekt: Das berichtete Ereignis wird sozusagen in die Referenzperiode „hineingeschoben“).</p> <p>Neben Effekten der Frageformulierung und Antwortkategorien ist die Positionierung der Fragen von Bedeutung, den Fragen können auch auf andere Fragen „ausstrahlen“ (sog. Halo-Effekt). Z.B. wird die Frage nach der allgemeinen Lebenszufriedenheit entschieden positiver beantwortet, wenn vorangehend nach dem Dating/Flirt-Verhalten gefragt wird.</p>
3) Interviewer und Interview-situation	Im Allgemeinen sind die Ergebnisse bei „sensiblen“ Fragen umso weniger verzerrt, je geringer die soziale Distanz zwischen Interviewer und Befragten ist. Bei telefonischen Interviews entfallen die äusseren Interviewmerkmale (Kleidung, Alter, Mimik...), allerdings nie ganz. Nur bei postalischen, schriftlichen Befragungen werden Interviewereinflüsse vollständig ausgeblendet.

Forschungsmethoden

Diekmann, S. 400 - 428

Thomas Meier

071 223 17 06

tmeier@abstrakt.ch

Der Text ist sehr schlecht strukturiert, deshalb mag auch die Zusammenfassung ein wenig chaotisch erscheinen. Ich habe mich aber bemüht, Klarheit in diesen Diekmannschen Mischmasch zu bringen. Ich habe ihn in Abschnitte unterteilt, diese Abschnitte sind in der linken Spalte fett gedruckt. Das Thema ist *Datenerhebung* und die *Theorie des Interviews*. Im ersten Abschnitt (399 – 403) geht es vor allem um Fehler, die durch den Interviewer und/oder die Situation des Interviews verursacht werden. Im zweiten Abschnitt (404 – 418) werden verschiedene Fragetypen und -ausprägungen vorgestellt. Am Schluss werden einige Grundregeln zur Befragung, der Konstruktion von Fragebogen und zur Frageformulierung aufgestellt. Im dritten Abschnitt (419 – 428) wird die Randomized-Response-Technik erläutert, welche man bei heiklen Fragen benutzen kann, und dann zeigt Diekmann auch noch, was Fragebogenbefragungen zur sozialen Kontextanalyse beitragen können.

Interviewer und Interviewsituation

Persönliche (äussere und innere) Merkmale des Interviewers wie Geschlecht, Sprache, nonverbale Kommunikation etc. beeinflussen die Antwortreaktion. Durch telefonische Befragung können diese zwar teilweise reduziert werden, es gibt jedoch immer Fehlerquellen. Persönliche Befragungen haben auch ihre Vorteile.

Einflussnahme durch Interviewer

Fehlerquelle: Bewusste Fälschung durch den Interviewer (z. B. selber ausfüllen um die Bearbeitungszeit zu verkürzen) Strategie: „Kernfragen stellen, Kontakte sichern und die Bögen selber vervollständigen“
Gegenmassnahmen: (manchmal beeinflussen Fälschungen die Resultate nicht schwerwiegend), Feldkontrollen, Art und Höhe der Bezahlung, selber Befragung durchführen

Einflüsse durch Interviewsituation

- Atmosphäre des Raumes, wo die Befragung stattfindet
- Anwesenheit dritter (z. B. Ehepartner, S. 402)
- Sponsorship – Effekt (Auftraggeber der Studie ist bekannt)

Antwortverzerrung im Interview (im Sinn einer Zusammenfassung der vorangegangenen Abschnitte, siehe Abb. S. 403)

- Soziale Erwünschtheit
- Response Set
- Meinungslosigkeit
- Situationale Effekte
- Fälschungen

- Anwesenheit Dritter
- Interviewermerkmale
- Fragekontext
- Antwortkategorien
- Frageformulierung
- Sponsorship – Effekt

Fragetypen

Es gibt verschiedene Arten der Fragetypen, für die sich verschiedene Arten von Fragestellungen unterschiedlich eignen:

1. Einstellungsfragen

Zur Erhebung von Einstellungen werden den Befragten häufig Aussagen vorgelegt (statements, items), die auf einer Rating- oder Likert – Skala eingestuft werden sollen. In der Regel empfiehlt es sich, eine Einstellungsdimension mit multiplen Items zu messen (für Konsistenzanalyse etc.)

2. Überzeugungsfragen

Überzeugungen sind subjektive Aussagen über Fakten. Hier bieten sich Multiple-Choice, dichotome Fragen, oder Alternativfragen an. Zu bemerken ist, dass offene Fragen besser sind als Kategorien vorzugeben, denn Kategorien können ungewollt Informationen vermitteln.

3. Verhaltensfragen

Das sind i. d. R. Retrospektivfragen, z. B. nach Art, Häufigkeit und Dauer von Handlungen. Zu beachten ist, dass das Erinnerungsvermögen nicht immer ganz genau die Realität wiedergibt.

4. Sozialstatistische Merkmale

Für solche statistische Daten gibt es verschiedene Erhebungsmethoden, die alle ihre Vor- und Nachteile aufweisen.

Bezug von Fragen

- Selbstauskünfte(zur eigenen Person)
- Fremdauskünfte (zu einer Fremdperson)

Fragestrategien

- Direkt
- indirekt (um Verzerrung durch soziale Erwünschtheit zu vermeiden)

Spezialfall: projektive Fragen → Wünsche oder Einstellungen

aufdecken, die der Befragte nicht direkt angeben will/kann.

Im Allgemeinfall ist von ungeprüften indirekten oder projektiven Fragen abzuraten.

Form von Fragen

1. geschlossene, offene, halboffene
 - a. *geschlossene* vs. *offene*: vorgegebene Antwortkategorien sind leichter und schneller zu beantworten, aber man erhält nur Informationen im Rahmen der vorgegebenen Kategorien.
 - b. halboffene (Hybridfragen) sind ein Kompromiss, zwei Antwortmöglichkeiten
2. Filterfragen, Gabel, Fragetrichter
→ Fragetechnische "Kunstkniffe"
 - a. Filterfragen werden nur der relevanten Personengruppe gestellt → Einengung (z. B. nur verheiratete werden nach Ehezufriedenheit befragt)
 - b. Gabel: Erweiterung. (wie die Weichen bei Eisenbahnschienen)
 - c. Trichterfragen: vom Allgemeinen zum Besonderen (für heikle Fragen), → geringere Verweigerungsquote; aber Achtung: Kontexteffekte!

Grundregeln der Frageformulierung

3. Eisbrecherfragen: am Anfang von Fragebögen
 1. Kurz, verständlich und hinreichend präzise
 2. Keine platten Anbieterungen (also nicht cool, geil, wow oder so!)
 3. Keine doppelte Verneinung
 4. Antwortkategorien disjunkt, erschöpfend und präzise
 5. Vorsicht bei stark wertbesetzten Begriffen
 6. Keine mehrdimensionale Fragen
 7. Normalerweise keine indirekten Fragen
 8. Normalerweise keine Suggestivfragen
 9. Befragte sollten nicht überfordert werden

Grundregeln der Fragebogengestaltung

Fragebogen in „thematische Blöcke“ aufteilen. Ein neuer Fragebogen sollte unbedingt einem oder mehreren Pretests unterzogen werden

1. Eröffnungsfragen (Eisbrecherfragen)
2. Wichtigste Fragen im zweiten Drittel platzieren (wegen der abnehmenden Spannungskurve)
3. Filterfragen und Gabeln helfen, überflüssige Fragen zu vermeiden.
4. Überleitungsformulierungen bei Themenwechseln
5. Sozialstatistische Fragen am Ende des Fragebogens
6. Interviews nicht zu lang gestalten
7. Kontaktprotokoll dazuheften

**Grundregeln der Befragung
(Durchführung)**

1. Interviewer: Schulung, zeitliche Belastung, Bezahlung, Kontrolle, etc.
2. Kontaktaufnahme bei Befragten:
 - a. Thema nicht erwähnen (nur allgemein)
 - b. Anonymität versichern, aber nicht allzu stark, denn das macht misstrauisch
 - c. Omnibus- oder Einzelbefragung?

Detailplanung einer Durchführung ist nicht zu unterschätzen oder zu vernachlässigen!

**Randomized-Response-
Technik**

Garantiert in Interviews vollständige Anonymität und ist darum für heikle Fragen anwendbar. Auch nachträglich sind die individuellen Antworten nicht identifizierbar.

Ablauf: Am besten selber im Buch nachlesen, S. 419 – 424

**Soziometrie und soziale
Netzwerke**

Soziometrie: Analyse sozialer Beziehungen zwischen Gruppenmitgliedern.

Kontextanalysen (Untersuchungen der Zusammenhänge zwischen dem sozialen Kontext und individuellen Merkmalen) basieren häufig auf „atomisierten“ Individualbefragungen. Auf der Basis von erhobenen Informationen kann eine Beziehungsmatrix aufgestellt werden, die Aufschluss gibt über die Struktur der sozialen Gruppe. So kann man untersuchen, ob die Einstellungen von Kontaktpersonen Einfluss haben auf die Einstellung der Befragten.

Empirische Sozialforschung

Forschungsmethoden

Andreas Diekmann, S. 429-455, Kapitel C , X. : 9. -11.

Marion Meister

0786606060

marion.meister@student.unisg.
ch

Dank der technologischen Entwicklung hat sich das Telephoninterview als Methode der Markt- und Meinungsforschung stark verbreitet, v.a. das computerassistierte Telefoninterview CATI ermöglichte Effizienzsteigerungen. Die Vorteile des Telephoninterviews sind die hohe Standardisierung, geringe Kosten, rel. hohe Datenqualität und Ausschöpfung. Das CATI kann komplex ausgestaltet sein und die Auswertung ist nicht aufwendig. Falls visuelles Material gezeigt oder Rankings gemacht werden sollen ist das PAPI (Paper and Pencil Interview) jedoch geeigneter. Es gibt 3 Arten von schriftlichen Interviews: (1) die postalische Befragung, (2) die Gruppenbefragung und (3) die Kombination von telephon./persönl. Interview mit nachträglicher schriftlicher Befragung. Positiv daran ist, dass der Befragte die Fragen besser durchdenken kann, oder dass Merkmale des Interviewers keinen Einfluss haben. Das Interview darf nicht komplex sein. Problematisch ist auch die geringe Rücklaufquote, welche aber mit bestimmten Massnahmen (TDM) gesteigert werden kann.

Anders als die quantitative Sozialforschung, betont die qualitative Forschung die **Subjektbezogenheit** und will mehr in die Tiefe gehen um nicht nur zu statistischen Kennwerten zu kommen. Es ist umstritten, ob die qualitative Forschung nur für explorative Untersuchungen (Entwicklung von Typologien, Kategoriensystemen, Generierung von Forschungshypothesen) geeignet ist, oder auch für die *Prüfung* von Hypothesen. Es gibt 3 qualitative Interviewtypen, alle mit nicht-direktivem Fragestil: (1) das fokussierte Interview, welchem einen Stimulus (Werbepot, Film...) vorausgeht, dessen Wirkung untersucht wird, (2) das narrative Interview, in welchem auf noch unstrukturierter Weise Erfahrungen/ biographische Abläufe erzählt werden und (3) das problemzentrierte Interview, bei welchem der Interviewer eine aktivere Rolle wahrnimmt, falls der Befragte über geringe narrative Kompetenz verfügt. Die Nachteile der qualitativen Methode sind: Stichprobenauswahl, Validität und Reliabilität und der grosse Aufwand bei der Datenauswertung.

Diekmann betont, dass es oft sinnvoll ist qualitative und quantitative Erhebungs- und Auswertungsmethoden zu kombinieren.

Inhaltsverzeichnis:

C Datenerhebung

X. Befragung

9. Telephonische Befragung

10. Schriftliche Befragung

11. Qualitative Methoden der Befragung

9. Telephonische Befragung

Entwicklung	In den letzten 10 Jahren hat sich das telephonische Interview als Methode der Markt- und Meinungsforschung enorm rapide verbreitet. 2/3 aller Umfragen von Meinungsforschungsinstituten in der Schweiz werden am Telefon durchgeführt. Die Verbreitung dieser Methode geht v.a. auf technologische Entwicklungen zurück: Die Netzdichte der Telephonanschlüssen hat stark zugenommen, aktuelle Verzeichnisse mit allen Privatanschlüssen sind auf CD-Rom erhältlich, das computerassistierte Telefoninterview - CATI ¹ – ermöglichte eine Effizienzsteigerung. Telephoninterviews sind viel billiger als persönliche oder schriftliche Befragungen.
Ergebnisse der Methodenforschung	Die frühere Kritik an der telephonischen Befragungsmethode wurde von der methodischen Forschung der vergangenen zwei Jahre widerlegt.

¹ Bei CATI-Interviews liest der Interviewer die Fragen vom Bildschirm eines Computers ab und tippt die Antworten zu jeder Frage umgehend ein. Das CATI-System automatisiert Stichprobenziehung, Adressenverwaltung und Terminvereinbarungen, steuert die Befragung und speichert die eingetippten Daten.

1. Mit entsprechenden Massnahmen (vorheriges Anschreiben, mehrere Kontaktversuche in nicht zu geringer Zeitspanne) lässt sich in allgemeinen Bevölkerungsumfragen eine Ausschöpfung von 70% realisieren.
2. Möglich sind Telephoninterviews, die 1 bis 1½ Stunden dauern. Keine ernsthaften Probleme bereiten Umfragen bis 45 Minuten. In der Praxis dauern die meisten Interviews von 10 bis 30 Minuten.
3. Auch komplexe Fragen wie Ranking, Listenabfragen, Rating bei Itembatterien sind ohne Qualitätsverlust möglich, wenn die Fragen entsprechend angepasst werden.
4. Datenqualität ist nicht geringer als bei persönlichen Interviews. Die Antwortreaktionen sind sehr ähnlich, erkannte Rogers als er den Vergleich einer telephonischen mit einer persönlichen Umfrage zum Wahlverhalten machte. Bei beiden Methoden weisen die Angaben eine deutliche systematische Verzerrung in Richtung soziale Erwünschtheit auf: Viele Personen gaben fälschlicherweise eine Wahlteilnahme an.

Telephonische Bevölkerungsumfragen sind, falls der Untersuchungszweck eine Repräsentativerhebung ist, natürlich nur möglich in Ländern, in denen mehr als 90% der privaten Haushalte mit einem Telefon ausgestattet sind. In Ostdeutschland z.B. müsste also mit einem Stichproben-Selektionsfehler („undercoverage“) gerechnet werden.

Stichproben

Ein Vorteil telephonischer Bevölkerungsumfragen ist die einfache Prozedur der Ziehung von Zufallsstichproben: RDD-Methode (Random Digit Dialing) oder Listenauswahl aus einem Telefonverzeichnis. Die Stichprobenziehung wird heute mit den CATI-Systemen automatisiert. Von der einstufigen Zufallsstichprobe von Haushalten wird in einem zweiten Schritt die Zielperson zufällig – mit dem Schwedenschlüssel oder der „Geburtstagsmethode (Zielperson ist, wer zuletzt Geburtstag hatte) - ausgewählt.

Der Fragebogen

Die Fragen, Antwortkategorien und Skalen müssen meistens kaum von denen eines persönlichen Interviews modifiziert werden. Nur Rankings von mehr als 4 Items sind am Telefon schwierig vorzunehmen. Da müssen Anpassungen vorgenommen werden. Will man z.B. eine Rangordnung von 10 Berufen nach dem Sozialprestige erhalten, kann man den Befragten bitten jedem Beruf eine Zahl zwischen 0 und 100 zu nennen oder Paare zu vergleichen.

Bei telephonischen Interviews mit CATI Technik können Filter, Gabeln und Konsistenzprüfungen in beliebiger Komplexität programmiert werden. Verzweigungsanweisungen und Filterführungen für den Interviewer fallen weg, da das System automatisch die nächste Frage in Abhängigkeit von den zuvor eingetippten Antworten steuert (= CAPI computerassistiertes persönliches Interview). Ein weiterer Vorteil des CATI ist, dass die Phase der nachträglichen Kodierung und Datenübertragung wegfällt. Die Daten stehen sofort zur Verfügung.

Durchführung

Beim CATI muss der Fragebogen zuerst programmiert werden. Dann sollten mehrere Pretests und eine Interviewschulung gemacht werden.

Ein Telephonlabor verfügt über 20 bis 50 Plätze. Die Supervisors kontrollieren, indem sie sich von einem Gerät aus in die laufenden Interviews einschalten können, und leisten Hilfe. Fälschungen sind kaum möglich, die Interviews sind sehr standardisiert. Die Interviews werden zu einem grösseren Teil abends an Werktagen (19-20h) und 14/15h an Wochenenden durchgeführt. Dies sind die ertragreichsten Perioden. Im Normalfall werden am Anfang des Interviews die Haushaltsmitglieder ermittelt und dann die Zielperson ausgewählt, was wegen der Gefahr eines Interviewabbruchs Geschick verlangt. Hat die ausgewählte Person keine Zeit müssen Termine vereinbart werden, meistens auch bei längeren Interviews. Eine seriöse Umfrage von etwa 1000 Interviews und einer Ausschöpfung von 60 bis 70% dauert zwei bis vier Wochen.

→ Kasten X.7:

Befragungen in Zeitschriften etc. sind oft schlecht dokumentiert. Es fehlen Angaben wie Fragetext, Anzahl der Befragten, Befragungszeitraum, Befragungsmethode (persönlich oder telephonisch), Verfahren der Stichprobenwahl, Stufigkeit der Auswahl, Ausschöpfungsquote. Ebenfalls kritisch ist der Autor gegenüber sog. „Blitzumfragen“, in welchen lediglich 500 Personen an einem Abend befragt werden. Solche Stichproben seien meistens verzerrt.

CATI oder PAPI?

PAPI ist das Etikett für das herkömmliche persönliche Interview „Paper-and-Pencil-Interview“. Das Telephoninterview hat gegenüber dem PAPI Vorteile: ebenso hohe Datenqualität, grössere Anonymität, Interviewmerkmale weniger relevant, fälschungssicherer. Dazu kommen noch Vorteile durch die Verwendung des CATI-Systems. Es gibt aber auch Interviews, die besser persönlich gemacht werden, z.B. wenn visuelles Material präsentiert werden soll oder bei bestimmten Zielgruppen und Fragestellungen. Oft ist auch ein Methodenmix (telephon., persönlich, schriftl. Interviews) sinnvoll, da viele Leute nicht gerne übers Telephon Auskunft geben.

10. Schriftliche Befragung

Arten	1. Postalische Befragung, 2. Gruppenbefragungen (in Anwesenheit eines Versuchsleiters, z.B. Kursevaluation in der Schule, in Institutionen), 3. Kombination von telephonischer oder persönlicher Befragung mit schriftlicher Nachbefragung
Vorteile	1. Befragte können die Fragen besser durchdenken. 2. Merkmale und Verhalten von Interviewern haben keinen Einfluss. 3. geringere Kosten und Aufwand
Nachteile	1. Bei Verständnisproblemen keine Hilfe durch Interviewer (ausser bei Gruppenbefragungen) 2. Fragebogen muss einfach und selbsterklärend sein, wenig Filter und Verzweigungen. Rankings von Kärtchen, etc. nicht möglich. Wissensfragen nicht möglich, da im Lexikon nachgeschaut werden kann (ausser bei Gruppenbefragungen). 3. Ungewiss, ob Zielperson oder eine andere die Befragung ausfüllt. 4. Adressenliste der Grundgesamtheit erforderlich um Zufallsstichprobe zu machen 5. geringe Rücklaufquote (ausser bei Gruppenbefragungen) → Selektionsbias
Systematische Steigerung der Rücklaufquote	Die Rücklaufquote hängt von Merkmalen des Fragebogens ab (siehe Kasten X.8). Bei einem Fragebogen mit freundlichem Anschreiben ohne weitere Massnahmen beträgt die Rücklaufquote <20%. Dillmann hat die „Total-Design-Methode“ (TDM) entwickelt, gemäss welcher die Qualität jedes Aspekts der schriftlichen Befragung maximiert werden soll: <ul style="list-style-type: none">- Fragebogen: Broschürenform, weisses Papier- Fragenanordnung: mit leichten, interessanten Fragen beginnen, Demographie am Ende- Fragendesign: übersichtlich, Vermeidung von Frageteilungen, visuelle Hilfen- Anschreiben: offizielles Briefpapier, Nützlichkeit der Studie, Dank, Unterschrift – „blue ball point“ (Aussehen wie persönliche Unterschrift)- Verpackung/Versand: mit Rückantwortcouvert, Mitte der Woche

(Wochenende dazwischen), nach 1 Woche Postkarte als freundliche Erinnerung, 3 Wochen später nochmals Fragebogen, 7 Wochen später nochmals aber als „certified mail“
Damit erreichte Dillmann durchschnittlich eine Rücklaufquote von 74%. Besonders lässt sich die Quote durch Nachfassaktionen oder kleinen Belohnungen steigern.

11. Qualitative Methoden der Befragung

quantitative vs.
qualitative Methode

Die quantitative Sozialforschung will möglichst maximale Standardisierung der Interviews erreichen, was die qualitative Forschung kritisch betrachtet. Soziale Phänomene die nicht in den Frageraster passen werden ausgeblendet, anders als in Alltagsgesprächen gibt es Hemmschwellen und unfreie Formulierung. Die qualitative Forschung ist an den „Sinndeutungen“ des Befragten interessiert und betont die **Subjektbezogenheit**, die Offenheit bezüglich der Fragen, Antworten und Methoden und Untersuchungen in Alltagssituationen.

Aus Sicht der quantitativen Forschung ist die qualitative nur für explorative Untersuchungen (Entwicklung von Typologien, Kategoriensystemen, Generierung von Forschungshypothesen) anwendbar, nicht aber für die *Prüfung* von Hypothesen. Für Vertreter qualitativer Methoden gehen die Forschungsziele aber weiter:

- Deskription empirischer Sachverhalte und sozialer Prozesse
- Aufstellung von Klassifikationen oder Typologien
- Gewinnung von Hypothesen am empirischen Material

- Prüfung von Forschungshypothesen
Bei der qualitativen Forschung arbeitet man mit Stichproben von wesentlich kleinerem Umfang (Fallstudien), man versucht aber mehr in die Tiefe zu gehen und das Material intensiver auszuwerten, nicht nur zu statistischen Kennwerten.

Die Kombination von quantitativer und qualitativer Methoden ist oft sinnvoll.

Das fokussierte Interview

Das fokussierte Interview ist geeignet zur Untersuchung der Wirksamkeit von Massenmedien und Propaganda. Es hat folgende Merkmale:

1. Alle Personen haben eine konkrete Situation/„Stimulus“ erlebt, z.B. ein Film, Rundfunksendung, Werbespot, Bild, Geschichte, Buch...
2. Die Situation wird vor dem Interview von einem Forscher analysiert und einige Hypothesen über die Wirkung formuliert.
3. Er entwickelt einen Interviewleitfaden, der die anzusprechenden Themen enthält.
4. Es erfolgt ein Interview, das offen ist für neue Gesichtspunkte, die sich aus dem Gespräch ergeben. Ziel des Interviews ist es die Hypothesen zu überprüfen.

Für die Interviewphase gelten 4 Prinzipien:

1. Nicht-Beeinflussung, d.h. einen „weichen“, nicht-direktiven Interviewstil
2. Spezifität, d.h. Reaktionen auf Details der erlebten Situation herausfinden
3. Erfassung eines breiten Spektrums thematischer Aspekte, unerwarteter Reaktionen
4. Tiefgründigkeit und personaler Bezugsrahmen, z.B. werden emotionale Worte wie „gut“, „hässlich“ usw. durch nachfragen tiefer ausgelotet

Das narrative Interview

Das narrative Interview ist noch weniger strukturiert als das fokussierte Interview. Der Interviewer gibt eine Themenstellung vor und ermuntert in nichtdirektiver Weise zur Erzählung. Es werden damit erfahrungsnahe, subjektive Aussagen über Ereignisse oder biographische Abläufe/Phasen (Ehescheidung, Ausbildungs- und

Berufsverläufe, Arbeitslosigkeit...) gewonnen.

Mit der Auslösung des Erzählflusses wird quasi automatisch gewissen Strukturprinzipien Rechnung getragen, der Befragte befolgt unbewusst 3 Darstellungszwängen: *Zwang zur Gestalterschliessung* (Erzählung muss vollständig, verständlich und ausgewogen sein), *Kondensierungszwang* (Befragte muss Schwerpunkte setzen) und *Detaillierungszwang* (Motive und Zusammenhänge müssen verständlich gemacht werden). Der Interviewer hat die Rolle des Zuhörers bis er in einer zweiten Phase wieder Nachfragen stellen darf („Wie kam es dazu?“...) um evtl. neue narrative Sequenzen auszulösen. Die 3 Phasen des Interviews (Stimulierung der Erzählung – Erzählung – Nachfragephase) dauern bis zu 3 Stunden und werden auf Tonband aufgezeichnet. Gesprächspausen, Lachen etc. geben ein authentisches Bild des Interviews wieder.

Problemzentriertes
Interview

Da nicht alle Interviewpartner über „narrative Kompetenz“ verfügen, spielt der Interviewer beim Problemzentrierten Interview eine aktivere Rolle. Gestützt auf einen Leitfaden bringt er situationsbedingt auch während der Erzählphase Fragen vor. Zusätzlich werden mit einem quantitativen Kurzfragebogen die sozialstatistischen Daten erhoben. Viele vertreten diese Richtung des additiven Zusammenspiels qualitativer und quantitativer Methoden um ein verlässliches Gesamtbild eines sozialen Gegenstandes zu erstellen (*Triangulation*).

Probleme qualitativer

1. Auswahl einer Stichprobe

Die Stichprobe wird in der Regel „bewusst“ ausgewählt, z.B nach einem 2x2 Desing², wenn man gewisse Typen vergleichen will. Es liegt dabei eine explizite oder implizite Hypothese zugrunde.

2. Reliabilität und Validität qualitativ gewonnener Daten
Wie auch bei standardisierten Interview gibt es Fehlerquellen wie soziale Erwünschtheit, Interviewereinflüsse, Fragereiheneffekte u.a.m.
3. Datenauswertung
Die Transkription und Auswertung von qualitativen Interviews erfordert einen erheblichen Aufwand. Da es oft keine eindeutig Interpretation gibt, sollten mind. 2 Personen das Material auswerten (→ *Durchführungsobjektivität* Kapitel VI).

Es besteht auch die Möglichkeit qualitative Erhebungsverfahren mit quantitativer Auswertung zu verbinden. Z.B. wurden Kinder nach ihren Weihnachtswünschen befragt (qualitativ) und diesen dann Zahlen – die Preise – zugeordnet (quantitative Auswertung). Dies würde in der Tabelle X.10 Typ II entsprechen. Der in der empirischen Sozialforschung häufigste Typ ist I, in der qualitativen Sozialforschung dominiert Typ IV. In letzter Zeit haben sich die Frontstellungen zw. „qualitativ“ und „quantitativ“ gelockert.

Tabelle X.10 Kombination von Erhebungs- und Auswertungsmethode

		Erhebungsmethode	
		<i>quantitativ</i>	<i>qualitativ</i>
Auswertungsmethode	<i>quantitativ</i>	I	II
	<i>qualitativ</i>	III	IV

² Bsp: man befragt arbeitslose Jugendliche mit niedriger und hoher Qualifikation aus Grosstädten und Landgemeinen = 4 Gruppen.

9. Weitere Methoden der Datenerhebung: nicht-reaktive, Beobachtung
Diekmann: 456, 517

Empirische Sozialforschung

Forschungsmethoden
Diekmann, A. S. 456 - 480

Stefan Rüegg

071 220 92 37

stefan.rueegg@student.unisg.
ch

Dieses Kapitel dient der genauen Untersuchung und Systematisierung von Beobachtungsmethoden. Insbesondere die systematisierte Aufstellung und Nennung der Probleme unter **4. Verschiedene Arten der Beobachtungstechnik und ihre Probleme** scheint mir **beachtenswert**.

XI. Beobachtung

**1. Beobachtung als
Methode der
Sozialforschung**

Beobachtungsverfahren im allgemeinen Sinn: sämtliche empirischen Methoden.

Beobachtungsverfahren im spezifischen Sinn der Sozialforschung: direkte Beobachtung menschlicher Handlungen, sprachlicher Äusserungen, nonverbaler Reaktionen (Mimik, Gestik, Körpersprache) und anderer sozialer Merkmale (Kleidung, Symbole, Gebräuche, Wohnformen, usw.)

Abgrenzung zwischen Sozialreportagen und wissenschaftlichen Beobachtungsstudien (obwohl Grenzen fließend):

1. Bezug auf Forschungshypothese; 2. stärkere Kontrolle und Systematik der Beobachtung

Probleme, denen systematische Beobachtungstechniken Rechnung tragen müssen:

1. Problem der Verzerrung durch selektive Wahrnehmung; 2. Problem der (Fehl-)Interpretation des beobachteten sozialen Geschehens.

Lösung: Schulung von Beobachtern, Einsatz mehrerer Beobachter, Verwendung von Leitfäden, etc. → häufig auch Kombination

**2. Die Arbeitslosen
von Marienthal**

Inhalt der Studie von Jahoda et al.:

Marienthal (kleiner Ort östlich von Wien) 1930 besonders hart von Arbeitslosigkeit betroffen (durch Schliessung Textilfabrik): von total 478 Familien hatten 367 Familien kein einziges Familienmitglied mit Job.

2 widersprüchliche Hypothesen werden getestet:

1. Massenarbeitslosigkeit führt zu Radikalisierung der Arbeiter

2. Massenarbeitslosigkeit führt zu Resignation und Apathie bei Arbeitslosen

Verwendet wurde ein Multi-Methoden-Mix:

Lebensgeschichten = *qualitative Interviews*; Zeitverwendungsbogen = *quantitative Zeitbudgeterhebung*; Schulaufsätze = *Inhaltsanalyse*; ...

Die Mitarbeiter der Studie nahmen auch aktiv am Geschehen teil;

Merkmale und Charakterisierung der Beobachtungstechniken:

- Beobachtung in natürlicher Situation (*Feldbeobachtung*)
- Sozialforscher nimmt an den Interaktionen in der sozialen Situation teil (*teilnehmende Beobachtung*)
 - Ermöglicht Zugang zum sozialen Feld
 - Protokollierung erst nachträglich erfolgt

Bsp: Einteilung der Familien in „ungebrochen“, „resigniert“, „verzweifelt“ und „apathisch“.

Diese Typologie hat zudem einen dynamischen Aspekt: mit zunehmender Dauer der Arbeitslosigkeit werden anfangs ungebrochene Familien auf der Stufenleiter der einzelnen Phasen bis zu apathisch absteigen.

Zeitstruktur wurde mit Hilfe *quantitativer Daten* erhoben: „Häufigkeit des Stehenbleibens“ und „Gehgeschwindigkeit“

Männer (Fabrikarbeiter) verloren Zeitgefühl viel schneller als Frauen (Hausarbeit) nach Verlust des Arbeitsplatzes

- **Studie hilft nachträglich zu verstehen, wieso Ideologie des Nationalsozialismus so erfolgreich war**

3. Wenn Prophezeiungen fehlschlagen

Was reagieren gläubige Mitglieder einer „Ufo-Sekte“, die Leben auf Weltuntergangs-Prophezeiung ausgerichtet haben, auf das Nichteintreffen der Prophezeiung?

Hypothesen (*ex ante* formuliert) aus der Theorie der kognitiven Dissonanz:

- Kognitive Dissonanz (nach Nichteintreffen) wird reduziert, indem
 1. der Glaube aufgegeben wird
 2. Glaube uminterpretiert wird, so dass Fehlschlag der Prophezeiung subjektiv plausibel erscheint
 - Wird nur geschehen, wenn die Interpretationen soziale Unterstützung erhalten (Gläubige untereinander in engem Kontakt)

2 Gruppen wurden anhand *verdeckter Beobachter* untersucht:

1. Hauptquartier um Sektenchefin (Gruppe mit engem Zusammenhalt)
2. Filiale an College (Mitglieder stärker voneinander isoliert)

3 Probleme mit verdeckt ermittelnder Beobachtung:

1. der Zugang zum sozialen Feld
2. Problem der Beeinflussung des soz. Geschehens durch Beobachter
3. Protokollierung der Beobachtungen

Zusätzlich: forschungsethisches Problem der Ausspähung einer Gruppe gegen deren Willen!

Ergebnis der Studie: Hypothesen wurden bestätigt

In Filiale: Abwendung der Mitglieder von Sekte; Im Hauptquartier:

Uminterpretation und Festhalten der Mitglieder an Glauben

4. **Verschiedene Arten der Beobachtungstechnik und ihre Probleme**

Systematisierung der einzelnen Beobachtungsverfahren (Wahl der Methode hängt wesentlich von Untersuchungsgegenstand und –ziel ab):

1. *teilnehmende* vs. *nichtteilnehmende* Beobachtung

Problem der Identifikation (Trübung des Blicks; „going native“); bei Nicht-Teilnahme (Vorteil): muss nicht gleichzeitig interagieren und beobachten

2. *offene* vs. *verdeckte* Beobachtung

verdeckt teilnehmend: „Spionagemethode mit Undercover-Agenten“

verdeckt nichtteilnehmend: „Schlüssellochmethode“ → nicht reaktiv

Problem verdeckt allgemein: ethisch?? → Güterabwägung zwischen Schutz der Intimsphäre und dem öffentlichen Interesse

3. *Feldbeobachtung* vs. *Beobachtung im Labor*

Labor erlaubt gezielte Vorgabe experimenteller Stimuli und Kontrolle von Störfaktoren; Feldbeobachtung besser für „externe Validität“

→ oft Kombination der Instrumente (z.B. 1. Labor; 2. externe Überprüfung)

4. *unstrukturierte* vs. *strukturierte* Beobachtung

strukturiert: erhöht Objektivität und Zuverlässigkeit (kleinere Gefahr der selektiven Wahrnehmung)

5. *Fremdbeobachtung* vs. *Selbstbeobachtung*

Selbstbeobachtung (=Introspektion): erfüllt nicht das Kriterium

intersubjektiver Nachprüfbarkeit, kann aber der Hypothesen-Gewinnung dienen; Prüfung der Geltung dann anhand von Fremdbeobachtung

5. **Strukturierte Beobachtung**

„unstrukturiert“ → „Leitfaden“ → „hoch strukturiert“

Beobachtungsleitfaden: Liste von Gesichtspunkten, auf die die Aufmerksamkeit des Beobachters gelenkt werden soll

Je strukturierter, desto stärker eingeschränkt der Spielraum des Beobachters.

Ziel der Strukturierung: Gewährleistung einer möglichst hohen Objektivität

Möglichkeiten zur Strukturierung: z.B. 2 Beobachter, Schulung, Testlauf, detailliert vorgefertigtes Protokollformular, ...

Unstrukturierte Beobachtung kann verwendet werden, um Vorwissen über ein Milieu zu erlangen. In der 2. Phase erfolgt dann strukturierte Beobachtung.

Inhaltsanalyse

*Forschungsmethoden
Diekmann, Seiten 481-516*

Jens Wiesenhütter

076 374 8088

Jens.wiesenhuetter@unisg.ch

Die Inhaltsanalyse (= Textanalyse, Dokumentanalyse, Bedeutungsanalyse) befasst sich mit der Erhebung und Auswertung von Texten, Bildern oder Filmen. Inhaltsanalyse greift zu kurz, weil auch formale Aspekte von Texten usw. erhoben werden können. Im Regfall ist es aber der Inhalt. Wichtig ist wie immer, die intersubjektive Nachvollziehbarkeit und Systematik.

Bei längeren Texten Inhaltsverzeichnis

Definition

Die Inhaltsanalyse ist eine empirische Methode zur systematischen, intersubjektiv nachvollziehbaren Beschreibung inhaltlicher und formaler Merkmale von Mitteilungen.

Anwendungsbereiche (Beispiele): Studien zu Kriegspropaganda, Massenkommunikation, in der klinischen Psychologie, von Annoncen, literarische Texte, Träume, Musikstücke etc.

Ziel ist es über z.B. einen Text oder Film (**Mitteilung**) Rückschlüsse auf den Autor, Produzenten (**Sender**) oder den Leser, Zuschauer (**Empfänger**) zu ziehen. Der Sender kodiert seine Aussage in Mitteilung (Bilder oder Buchstaben) und der Empfänger dekodiert sie wieder. Allerdings muss man die Bedeutung (inhaltlich) der Zeichen verstehen (Bedeutungsproblem).

Aspekte und Zielsetzung

Man muss drei Aspekte der Zeichenfolge unterscheiden:

- Syntaktik: formale Regeln der Verknüpfung von Zeichen (Zeichen-Zeichen-Relation)
- Semantik: Bedeutung von Zeichen, d.h. Zuordnung der Zeichen zu Objekten oder Eigenschaften (Zeichen-Bedeutungs-Relation)
- Pragmatik: Herkunft und Art der Verwendung und Wirkung von

Zeichen in einer spezifischen Situation

Zudem gibt es drei Zielsetzungen der Inhaltsanalyse:

- Formal-deskriptive Analyse: formale Aspekte des Texts, z.B. wie oft werden Verben verwendet
- Diagnostische Analyse: Sender-Mitteilungs-Beziehung. Was will der Sender mitteilen, welche Werte fließen ein.
- Prognostische Inhaltsanalysen: Erforschung der Wirkung von Mitteilungen beim Empfänger.

Vorteile der Inhaltsanalyse:

- Man kann weit in die Vergangenheit blicken (Archive)
- Nicht reaktiv, wird nicht durch Forscher beeinflusst
- Der soziale Wandel ist erforschbar

Beispiel

Im folgenden wird eine Inhaltsanalyse anhand des Beispiels einer schwedischen Studie zum Verhältnis der Werte Freiheit und Gleichheit in Leitartikeln wichtiger schwedischer Zeitungen gezeigt. Ziel der Studie, den Wertewandel zu beschreiben: Bürgerlich Freiheit vor Gleichheit, Sozial umgekehrt.

Grundgesamtheit:

5 grösste Zeitungen in Schweden und von denen die Editorials.

Stichproben:

Jede achte Ausgabe, für alle Zeitungen am gleichen Tag. (Achtung, keine Periodizitäten, wie z.B. immer am Samstag – könnte verzerren)

Analyseeinheit:

Ausmass in dem von Freiheit/Gleichheit gesprochen wird.

Kategorien:

Müssen disjunkt, erschöpfend, präzise und eindeutig zuzuordnen sein.

Im Beispiel: Die Variablen Freiheit und Gleichheit haben je nur zwei Ausprägungen. Die Analyseeinheit nimmt entweder auf Freiheit (gilt immer auch für Gleichheit) bezug (Kategorie 1) oder nicht (Kategorie 2).

Dazu kamen dann Subkategorien für jeweilige Kat 1 (bezug gegeben).

z.B. (1) Freiheit allgemein, (2) nationale Unabhängigkeit, (3) persönliche Freiheit, (3) Pressefreiheit, etc.

und (1) Gleichheit allg. (2) Fairness, etc.

(Nicht vergessen, die Ausprägungen müssen zu einem und nur einem zuzuordnen sein. Wichtig ist auch, dass diese Kategorien gut gewählt sind,

wegen der Reliabilität und Validität.)

Kodierung

Zunächst Kodierungsregeln festlegen. Man könnte alle Worte auflisten, welche in die jeweilige Kategorie gehören (ist sehr aufwendig). Man kann die Kodierer auf die Eigenschaften der Kategorie schulen (intensionale Definition). Bsp: Wirtschaftsfreiheit: Alle Einzelworte und Wortverbindungen, die sich auf die Handel-, Unternehmens- und Wirtschaftsfreiheit beziehen, wie Wegfall von Handelsschranken, freies Unternehmertum, Lockerung der Ladenöffnungszeiten und vieles mehr). Dann macht man Pretests, um zu überprüfen ob die Kodierer zu ähnlichen Resultaten bei gleicher Stichprobe (Reliabilität).

Aggregationsregel

In diesem Beispiel könnte man eine Aggregationsregel formulieren: z.B. Der Kommentar bringt den Wert „Freiheit“ zum Ausdruck, wenn mindestens eine Analyseeinheit (hier: Wort oder Wortverbindung) des Kommentars auf eine Unterkategorie von Freiheit verweist. (Gleichheit analog).

Spezielle Formen der
Inhaltsanalyse

Frequenzanalyse

Häufigkeiten von Worten, Themen, Begriffen, Bildern,...

Die Frequenzanalyse kann sich auf Inhalt und Form beziehen. Beispiel aus Schweden ist Frequenzanalyse.

Kontingenzanalyse

Ausgangspunkt sind auch Häufigkeitszählungen. Ziel ist aber die Ermittlung von Assoziationsstrukturen.

Z.B: Sind Freiheit und Gleichheit positiv assoziiert in Arbeiterzeitungen?

Wäre nun die Auftretenswahrscheinlichkeit für Freiheit 30% und Gleichheit 20%, dann wäre die Wahrscheinlichkeit, dass beide (wenn nicht assoziiert)

in einem Text auftreten: $P(F \text{ und } G) = P(F) \cdot P(G) = 0.3 \times 0.2 = 0.06$

Wären die Worte nun positiv assoziiert müssten sie mehr als in 6% der Fälle gemeinsam in einem Text auftauchen.

D.h. man vergleicht die erwartete Häufigkeit, wenn keine Verknüpfung besteht mit der beobachteten Häufigkeit. Man kann auch

Kontingenztabellen machen (S. 499). Dazu muss man, da es immer noch eine Stichprobe ist noch Signifikanztests machen.

Bewertungsanalyse:

Messung der Intensität von Bewertungen, die ein Sender in einem Text gegenüber Objekten oder Ereignissen zum Ausdruck bringt. Z.B.: Hat sich

die Bewertung der NATO seit dem Ende des Kalten Kriegs geändert?
 Man unterscheidet Einstellungsobjekte (AO) und allgemeines Material (cm) und Ausdrücken für Handlungen zum Beispiel Verben (c). AO's werden von unterschiedlichen Sendern unterschiedlich bewertet, cm von allen Sendern gleich, die Verben verknüpfen beide logisch.
 Zuerst werden alle AO's im Text markiert und verschlüsselt (z.B. Atomkraftwerke, CDU), d.h. durch Codebuchstaben ersetzt, damit der nachfolgende Kodierer das Resultat nicht beeinflussen kann. Dann werden die Sätze umformuliert, dass sie nur noch aus (1) Handlungsursprung (actor), (2) Handlung (action), (3) Handlungsobjekt (complement) bestehen. Jetzt gibt es nur noch 4 Möglichkeiten von Aussagen:

1. AO c cm z.B. Atomkraftwerke (AO) bergen (c) ein hohes Risiko (cm)
2. AO₁ c AO₂
3. AO c cm₀ (cm₀ ist ein als neutral bewertetes cm)
4. cm₁ c cm₂

3 und 4 sind neutral und bleiben deshalb unberücksichtigt.
 c und cm bekommen ein Rating zwischen -3 (stark negativer Zusammenhang) und +3 (stark positiver Zusammenhang). 0 gibt es nicht, da der Fall vorhin schon eliminiert wurde.

Bsp: XY (Code für Atomkraftwerke) bergen (+2) ein hohes Risiko (-3). Die Resultate werden dann multipliziert (-6) und dann werden verschiedene Aussagen verknüpft, so dass man für jedes AO eine Bewertung bekommt. (Formeln findest Du auf S. 502f.)

Das ganze ist sehr aufwendig und man kann es eigentlich nur mit kleinen Stichproben machen.

Qualitative Inhaltsanalysen

Eigentlich sind auch quantitative Analysen qualitativ, z.B. bei der Bedeutung von Worten. Aber es gibt da wohl verschiedene Schulen. 3 grundlegende Varianten:

- Zusammenfassung: wichtige Aussage herausarbeiten
- Explikation: Bedeutungsanalyse problematischer Textstellen unter Rückgriff auf erklärende Verbindungen im selben Text (enge Kontextanalyse), oder auf externes Material (weite Kontextanalyse, z.B. Berichte über Autor, etc.)
- Strukturierung: Strukturmerkmale anhand eines Kategoriensystems herausarbeiten. (Kategorien,

Ankerbeispiele (damit man weiss, was in welche Kategorie kommt), Kodierregeln