

ZUMA

Zentrum für Umfragen,
Methoden und Analysen

ZUMA-NACHRICHTEN

38

Mai 1996

AMUS

Zentrum für Umfragen,
Methoden und Analysen**Herausgeber**

Zentrum für Umfragen, Methoden und Analysen (ZUMA)
ZUMA ist Mitglied der Gesellschaft Sozialwissenschaftlicher
Infrastruktureinrichtungen e.V. (GESIS)

Vorsitzender des Trägervereins ZUMA e.V.: Prof. Dr. Max Kaase**Direktor:** Prof. Dr. Peter Ph. Mohler**Hausanschrift**

B 2, 1
68 159 Mannheim

Postanschrift

Postfach 12 21 55
68 072 Mannheim

Telefon 0621/1246 - 0
Fax 0621/1246 - 100
Redaktion 0621/1246 - 268

E-Mail zuma@zuma-mannheim.de**Internet** <http://www.social-science-geis.de/>
<gopher://gopher.social-science-geis.de/>**NSD-HOTLINE** 0621/1246 - 111.**Redaktion:** Dr. Paul Lüttinger**ISSN 0721-8516** 20. Jahrgang

Die ZUMA-Nachrichten erscheinen im Mai und November eines Jahres.
Sie werden Interessenten auf Anforderung kostenlos zugesandt.

Namentlich gekennzeichnete Beiträge geben die Meinung der Autoren wieder.
Der Nachdruck von Beiträgen ist nach Absprache möglich.

Druck: Verlag Pfälzische Post GmbH, Neustadt/Weinstraße.
Gedruckt auf chlorfrei gebleichtem Papier.

© ZUMA

In eigener Sache.....	5
-----------------------	---

FORSCHUNGSBERICHTE

Zeit und Geld: Empirische Verhaltensklärung mittels Restriktionen am Beispiel der Verkehrsmittelwahl. <i>Sebastian Bamberg</i>	7
Unimodalität und Unimodalitätstests. <i>Siegfried Gabler und Ingwer Borg</i>	33
Gastarbeiter oder Ausländer? Ergebnisse des Splits mit den reformulierten Gastarbeiterfragen im ALLBUS 1994. <i>Thomas Blank und Martina Wasmer</i>	45
Vertrauen in Institutionen in Deutschland zwischen 1991 und 1995. <i>Dieter Walz</i>	70

PROJEKTBERICHTE

Fragen zu Fragen: Instrumentbedingte Antwortverzerrungen? <i>Willy H. Eirmbter und Rüdiger Jacob</i>	90
Der Altenheimsurvey: Durchführung und Repräsentativität einer Befragung in den Einrichtungen der stationären Altenhilfe. <i>Thomas Klein und Siegfried Gabler</i>	112

MITTEILUNGEN

Sozialwissenschaftenbus 1996/97: Termine und Preise.....	135
--	-----

BUCHBESPRECHUNGEN

KOTZ, S./JOHNSON, N. L. (Eds.): Breakthroughs in Statistics. Vol. 1 and Vol. 2. (<i>Thomas Rahlf</i>)	138
BROSIUS, G./BROSIUS, F.: SPSS. Base System and Professional Statistics. (<i>Michaela Thoma/Cornelia Züll</i>).....	147
KELLE, U.: Computer-Aided Qualitative Data Analysis. Theory, Methods and Practice. (<i>Jürgen H.P. Hoffmeyer-Zlotnik</i>).....	149
SUDMAN, S./BRADBURN, N./SCHWARZ, N.: Thinking about Answers. The Application of Cognitive Processes to Survey Methodology. (<i>Michaela Thoma/Michael Braun</i>) ...	151
ASSENMACHER, W.: Deskriptive Statistik. (<i>Siegfried Gabler</i>)	154

PUBLIKATIONEN

Buchhinweise	157
ZUMA-Arbeitsberichte	164

ZUMA-TAGUNGEN

Workshop: „Simulation for the Social Sciences“, 3.-5. September“	167
Workshop: „Einführung in die computerunterstützte Inhaltsanalyse (CUI) mit TEXTPACK PC“, 24.-25. September	167
Symposium: „Measurement in Comparative Research“, 15.-16. und 17.-18. Oktober	168
Workshop: „Einführung in die Korrespondenzanalyse“, 5.-7. November	169
Workshop: „Delphi-Methode“, 14. November.....	169
Workshop: „Multi-Method-Pretesting“, 3.-4. Dezember	170
Workshop: „Einführung in das Internet“, 11.-13. Dezember	171

ADRESSENPFLEGE**BESTELLUNGEN****DURCHWAHL-RUFNUMMERN**

IN EIGENER SACHE

In dieser Ausgabe der ZUMA-Nachrichten beschäftigt sich *Sebastian Bamberg* am Beispiel der Wahl von Verkehrsmitteln mit der Frage, wie umweltrelevante Verhaltensweisen erklärt werden können. Er geht davon aus, daß hierbei nicht nur einschlägige Einstellungen eine Rolle spielen, sondern auch Restriktionen im Sinne von monetären und nicht-monetären Kosten von Bedeutung sind.

Im folgenden Beitrag steht eine der am meisten verwendeten statistischen Kennziffern zur Beschreibung von Daten im Blickpunkt: der Mittelwert. Er kann bekanntlich nur dann sinnvoll interpretiert werden, wenn die Verteilung auch tatsächlich eine zentrale Tendenz aufweist und beispielsweise nicht U-förmig verläuft. *Siegfried Gabler und Ingwer Borg* beschreiben den DIP-Test, mit dem die Unimodalität von Verteilungen überprüft werden kann.

Thomas Blank und Martina Wasmer berichten über die Ergebnisse eines Methodenexperimentes im ALLBUS 1994. Bei der zufällig ausgewählten Hälfte der Befragten wurde bei den sogenannten „Gastarbeiter“-Items der Begriff „Gastarbeiter“ durch „in Deutschland lebende Ausländer“ ersetzt und beide Formulierungen auf ihre Äquivalenz untersucht.

Mit einem Aspekt der politischen Integration West- und Ostdeutschlands beschäftigt sich *Dieter Walz*. Er untersucht, wie sich dort das Vertrauen in zentrale politische Institutionen, darunter das Bundesverfassungsgericht, der Bundestag und die Bundesregierung, zwischen 1991 und 1995 entwickelt hat.

Willy H. Eirmbter und *Rüdiger Jakob* erörtern in ihrem Beitrag „Fragen zu Fragen: Instrumentbedingte Antwortverzerrungen?“ die Ergebnisse eines Methodensplits im Sozialwissenschaftenbus II/1995. Um den Einfluß der Reihenfolge von Antwortvorgaben zu untersuchen, wurden bei mehreren Skalen die unterschiedlichen Antwortmöglichkeiten systematisch variiert.

Der Altenheimsurvey ist eine Befragung unter Bewohnern von Altenheimen mit dem Ziel, z.B. Näheres über die Gründe des Heimeintritts oder die Lebensumstände vor dem Eintritt zu erfahren. Die Durchführung der Befragung solcher Spezialpopulationen ist mit besonderen Schwierigkeiten verbunden. *Thomas Klein* und *Siegfried Gabler* berichten

über die „Durchführung und Repräsentativität einer Befragung in den Einrichtungen der stationären Altenhilfe“.

Im personellen Bereich haben sich bei ZUMA folgende Veränderungen ergeben. Ausgeschieden sind Carmen Eilinghoff (ALLBUS), Kirsten Schiller (Verwaltung) und Rita Kley (Mikrodaten). In den Ruhestand verabschiedet wurde Lisbeth Koch, die in den letzten Jahren das Sekretariat des Direktors von ZUMA engagiert und kompetent leitete. Ihre Nachfolgerin ist Birgit Lenz. Ebenfalls neu bei ZUMA sind Irene Fischer (Mikrodaten) und Matthias Engelhard (Mikrodaten) als Vertretung von Achim Wackerow, der zwei Jahre für das GESIS-Gemeinschaftsprojekt „Internet“ freigestellt wurde. Alle früheren und neuen Mitarbeiter begleiten unsere besten Wünsche.

Max Kaase
Vorsitzender des ZUMA e.V.

ZEIT UND GELD: EMPIRISCHE VERHALTENSERKLÄRUNG MITTELS RESTRIKTIONEN AM BEISPIEL DER VERKEHRSMITTELWAHL¹⁾

SEBASTIAN BAMBERG

Während „Präferenzen“ im Kontext sozialwissenschaftlicher Modelle ein zentrales Konstrukt zur Erklärung menschlichen Verhaltens darstellt, bemühen sich Ökonomen, Verhalten möglichst ohne Bezug auf dieses Konstrukt zu erklären. Im Mittelpunkt ökonomischer Analysen steht hingegen der Versuch, Verhaltensveränderungen bzw. Unterschiede im Verhalten auf Veränderungen bzw. Unterschiede in externen Restriktionen zurückzuführen. Während die traditionelle Ökonomie sich dabei vor allem auf die Analyse monetärer Restriktionen konzentriert, hat Gerry Becker den ökonomischen Ansatz um die nicht-monetäre Restriktion „Zeit“ erweitert. In der vorliegenden Arbeit wird versucht, aus Beckers Ansatz empirisch testbare Hypothesen abzuleiten, wie die Restriktionen „Geld“ und besonders „Zeit“ die individuelle Verkehrsmittelwahl determinieren. Diese Hypothesen werden an Daten aus einer empirischen Untersuchung getestet, in der für eine spezifische Fahrstrecke die mit der Bus- bzw. der PKW-Nutzung verbundenen Zeit- und Geldkosten erhoben wurden. Wie aus Beckers Ansatz theoretisch abgeleitet, übt die wahrgenommene Zeitkostendifferenz der Bus- oder PKW-Nutzung einen deutlichen Einfluß auf die tatsächliche Bus- bzw. PKW-Nutzung aus, während der Einfluß der wahrgenommenen Geldkostendifferenz statistisch nicht signifikant ist.

In sociological or socialpsychological models „preferences“ are used as a central construct explaining human behavior, whereas economic analysis are trying to explain behavior without reference to this construct. The economic-approach tries to explain behavioral changes and differences by changes and differences in external restrictions. While the traditional consumption-theory analyses mainly the effect of monetary restriction, Gerry Becker has enlarged the economic approach by the non-monetary restriction „time“. After a short description of Becker’s approach this article tries to use Becker’s

approach deriving hypotheses how the restrictions „money“ and „time“ determine the individual transportation mode choice. To test these hypotheses empirically a field-study was conducted, in which for a specific route the time- and monetary-costs are raised, connected with the usage of the transportation means „bus“ and „car“. As expected the results of the data-analysis show that the perceived differences of the time-costs between the transportation means „bus“ and „car“ exert a significant impact onto the actual usage of the bus or car, while the impact of the perceived differences of the monetary-costs are insignificant.

1. Einleitung

In den Sozialwissenschaften, besonders der empirischen Sozialforschung, stellen individuelle Präferenzen das zentrale theoretische Konstrukt zur Erklärung von Verhalten dar. Sozialpsychologen verwenden anstelle des Begriffs „Präferenz“ traditionell den Begriff „Einstellung“. Unter Einstellung wird die globale positive bzw. negative Bewertung eines Objekts durch eine Person verstanden. Eine relativ präzise Operationalisierung des Einstellungskonzepts erfolgt in multiattributiven Einstellungsmodellen, wie z.B. dem von Ajzen/Fishbein (1980). Eine Anwendung dieses Modells auf die Erklärung individueller Verkehrsmittelwahl findet sich in Bamberg/Schmidt (1994) oder Bamberg/Bien (1995).

Auch das ökonomische Verhaltensmodell geht davon aus, daß Präferenzen prinzipiell eine Verhaltensdeterminante darstellen. Das Individuum bewertet die ihm zur Verfügung stehenden Verhaltensalternativen entsprechend seinen Präferenzen, d.h. es wägt Vor- und Nachteile, Nutzen und Kosten gegeneinander ab. Schließlich entscheidet es sich für die Alternative, die seinen Präferenzen am ehesten entspricht.

Für die meisten Ökonomen stellen jedoch Präferenzen nicht beobachtbare und damit nicht meßbare Größen dar. Nach Ansicht von Ökonomen haftet daher dem Präferenz-Konstrukt im Kontext empirischer Forschung etwas tautologisches an. Da sich Präferenzen nicht unabhängig erfassen lassen, stellt die Erklärung mittels Präferenzen nur eine Umschreibung der Verhaltensänderung unter Verwendung eines anderen Wortes dar: Eine Präferenzänderung muß stattgefunden haben, weil sich das Verhalten verändert hat. (Hätte sich das Verhalten nicht geändert, hätte auch keine Präferenzänderung stattgefunden.) Eine empirische Überprüfung einer derartigen Umschreibung ist natürlich nicht möglich, da sie immer wahr sein muß (Frey 1990).

Stigler/Becker (1977) schlagen daher in ihrem berühmt gewordenen Aufsatz „De Gustibus Non Est Disputandum“ vor, bei der Analyse von Verhaltensveränderungen bzw. Ver-

haltensunterschieden völlig auf das Präferenz-Konstrukt zu verzichten. Präferenzen sollen statt dessen als stabile Größen behandelt werden. Nach Stigler und Becker ist davon auszugehen, daß sich die Präferenzen von Arm und Reich, von Menschen in unterschiedlichen Gesellschaften und Kulturen nicht sehr voneinander unterscheiden. Dabei beziehen sich die als stabil vorausgesetzten Präferenzen nicht auf Güter und Dienstleistungen, wie z.B. Orangen und Autos, sondern auf grundlegende Wahlobjekte, wie Gesundheit, Prestige, Sinnenfreuden, Wohlwollen oder Neid. Die Annahme stabiler Präferenzen bietet nach Stigler und Becker eine Grundlage, um Voraussagen über Verhaltensveränderungen bzw. Verhaltensunterschiede vollständig auf Veränderungen in beobachtbaren und meßbaren externen Restriktionen zurückzuführen. Restriktionen, die den Spielraum festlegen, innerhalb dessen das Individuum die Auswahl aus verschiedenen Handlungsmöglichkeiten hat, stellen im Rahmen der ökonomischen Verhaltenstheorie das zentrale Erklärungs-konstrukt für Verhalten dar.

Neben der Frage, ob Präferenzen meßbar sind, kritisieren Ökonomen wie Frey (1990) an den meisten sozialwissenschaftlichen Ansätzen, daß Restriktionen entweder gar nicht berücksichtigt werden oder daß die Trennung zwischen dem Präferenz- und Restriktions-konzept unzureichend ist. So spielen die in den ökonomischen Verhaltensmodellen im Vordergrund stehenden monetären Restriktionen „Einkommen“ und „Preise“ in vielen sozialwissenschaftlichen Ansätzen kaum eine Rolle.

In der vorliegenden Arbeit soll die sicherlich provozierende Kritik der ökonomischen Verhaltenstheorie an sozialwissenschaftlichen Erklärungsmodellen ernst genommen werden. Am Beispiel einer umweltpolitisch bedeutsamen Verhaltensweise, der Wahl zwischen Verkehrsmitteln, soll die empirische Fruchtbarkeit des Ansatzes untersucht werden, Verhaltensunterschiede bzw. Verhaltensveränderungen ausschließlich auf die Wirkung von Restriktionen zurückzuführen. Nach einer kurzen Skizzierung der theoretischen Überlegungen zur Verhaltenswirksamkeit von Restriktionen wird versucht, daraus empirisch testbare Hypothesen über den Einfluß monetärer und nicht-monetärer Restriktionen auf die individuelle Verkehrsmittelwahl abzuleiten. In einem zweiten Teil werden diese Hypothesen an empirischen Daten überprüft.

2. Theorie

Die ökonomische Theorie beschäftigt sich traditionell besonders intensiv mit der Verhaltenswirksamkeit von Güterpreisen. Preis ist die monetäre Gegenleistung eines Käufers, den dieser für ein Wirtschaftsgut bestimmter Qualität aufbringen muß. In Höhe des Preises müssen beim Erwerb eines Produktes finanzielle Mittel aus dem Haushaltseinkom-

men vom Käufer bereitgestellt werden. Da alle Produkte um das gleiche Einkommen des Konsumenten konkurrieren, beeinflussen ihre Preise auch die Verteilung des Einkommens auf diese Produkte. Es können nur so viele Produkte bzw. Produktmengen gekauft werden, wie verfügbare Geldmittel vorhanden sind. Die Verhaltenswirksamkeit des Produktattributs „Preis“ beruht damit auf der Einkommensrestriktion.

Im Rahmen der traditionellen Konsumtheorie (siehe z.B. Deaton/Muellbauer 1980; Maier/Weiss 1990; Varian 1984) stellen die jeweiligen monetären Preise unterschiedlicher Transportalternativen die alleinigen Determinanten der individuellen Verkehrsmittelwahl dar. Konsequenterweise wird daher auch von wirtschaftswissenschaftlicher Seite in der Diskussion, wie bisherige PKW-Nutzer dazu gebracht werden können, stärker die umweltfreundlicheren öffentlichen Verkehrsmittel zu benutzen, die Verhaltenswirksamkeit des Preis-Instruments betont: Mittels Maßnahmen wie „Road-Prizing“ oder Benzinspreiserhöhung soll die PKW-Nutzung verteuert werden, während mittels Maßnahmen wie „Umweltticket“ oder „Job-Ticket“ die Preise für die Nutzung öffentlicher Verkehrsmittel gesenkt werden sollen.

Nun haben jedoch eine Vielzahl von Studien gezeigt, daß weder die reinen monetären Fahrkosten noch die mit dem Nettoeinkommen gewichteten Fahrkosten einen substantiellen und signifikanten Einfluß auf die individuelle Verkehrsmittelwahl ausüben (Domencich/McFadden 1975; Hensher/Dalvi 1978; Held 1982; Ben-Akiva/Lerman 1985; Erke 1990; Molt 1990; Brüderl/Preisendörfer 1994; Diekmann 1994). In diesen Studien zeigt sich hingegen, daß die individuelle Verkehrsmittelwahl stark von der wahrgenommenen Fahrtzeit sowie durch soziodemographische und -ökonomische Hintergrundvariablen wie Familiengröße, Alter, Geschlecht, Bildung und beruflicher Status beeinflusst wird. Anscheinend läßt sich die Entscheidung, ein spezifisches Verkehrsmittel zu nutzen, nur schlecht im Rahmen eines Ansatzes erklären, der den Nutzen von am Markt angebotenen Verkehrsmittelalternativen ausschließlich durch den Bezug auf monetäre Preise und monetäres Einkommen ableitet.

Kritiker des ökonomischen Ansatzes der Verhaltensklärung können mit Recht darauf hinweisen, daß sich die vorliegenden empirischen Befunde eher so interpretieren lassen, daß die individuelle Verkehrsmittelwahl weniger durch finanzielle Restriktionen als durch individuelle Unterschiede in Motiven, Einstellungen oder Normen gesteuert wird, kurz durch Größen, die als Indikatoren für zugrundeliegende individuelle Präferenz angesehen werden.

Für Becker (1965) beruht hingegen die Schwäche der traditionellen Theorie des Konsumverhaltens bei der Erklärung vieler alltäglicher Verhaltensweisen nicht auf der

mangelnden Berücksichtigung von Präferenzen, sondern darauf, daß sie mit ihrer Beschränkung auf monetäre Marktpreise und monetäres Einkommen den Einfluß nicht-monetärer Restriktionen auf Wahlentscheidungen unberücksichtigt läßt.

Nach Becker wird die Bedeutung einer zweiten, nicht-monetären Ressource, der Zeit, deutlich, wenn Haushalte nicht nur als Konsumenten, sondern auch als Produzenten betrachtet werden. Haushalte produzieren elementare Güter ("basic commodities", Becker 1965), wie „schönes Wohnen“, „Freizeiterlebnisse“, „gesunde Ernährung“, indem sie entsprechend der Kostenminimierungsregeln der traditionellen Unternehmenstheorie Marktgüter und Zeit als Inputs kombinieren. Der Preis eines im Haushalt produzierten Gutes ergibt sich danach aus der Summe der Preise der Marktgüter und dem Preis der eingesetzten Zeit.

Wie läßt sich aber der Preis der benötigten Zeit quantifizieren? Die Zeit ist begrenzt und hat für jeden Konsumenten einen Wert, der sich aus ihrer Knappheit ergibt. Anders als Geld ist Zeit aber eine Ressource, die nicht gespeichert werden kann. „Zeitsparen“ bedeutet deshalb die Reallokation der Zeit von einer Aktivität auf andere. Der Wert der für eine bestimmte Aktivität erforderlichen Zeit ist nur indirekt über den Verlust an Geld bzw. den Nutzenentgang zu ermitteln, der dadurch eintritt, daß andere Aktivitäten in dieser Zeit nicht ausgeübt werden können (Opportunitätskosten des Zeitbedarfs). Nach Beckers "full income concept" (1965: 497) fallen Opportunitätskosten der Zeit in Höhe des Verdienstausfalls ("forgone income") für diese Zeit an. Dabei geht Becker (1965: 496) von einer Zweiteilung des täglichen Zeitbudgets aus: Es gibt bezahlte Zeit (Erwerbsarbeitszeit) und unbezahlte Zeit (Konsumzeit). Da außer der zur physischen Reproduktion notwendigen Zeit prinzipiell jede Zeit zum Gelderwerb eingesetzt werden kann, entspricht der Preis einer Zeiteinheit dem dafür von einem Individuum erzielbaren Lohn. Wenn das Einkommen eines Konsumenten steigt, steigen somit auch die Opportunitätskosten der Zeit, d.h. die Zeit wird teurer. Auf diese Änderung der Opportunitätskosten reagieren die Konsumenten mit einer Substitution zeitintensiver Aktivitäten (z.B. der Zubereitung von Speisen) durch zeitsparende Aktivitäten (Einkauf von Fertiggerichten).

Becker setzt damit die gegenseitige Substituierbarkeit von Geld und Zeit voraus. Zeit ist Geld und Geld ist Zeit. Durch Einsatz von Zeit kann Geld gespart werden (eigene Durchführung einer Reparatur) und durch den Einsatz von Geld kann Zeit gespart werden (Ausführung der Reparatur durch einen Handwerker). Wenn Zeit in Geld und Geld in Zeit „umgetauscht“ werden kann, dann gibt es sowohl einen Zeitwert des Geldes als auch einen Geldwert der Zeit.

Aus dem Haushalts-Produktions-Ansatz ableitbare inhaltliche Implikationen für die Nutzung von Verkehrsmitteln

Nach Beckers Ansatz lassen sich damit Transporte aus Sicht des Haushalts als Produktionsprozeß charakterisieren, in dem der Haushalt „Marktgüter“ (deren Erwerb der Einkommensrestriktion unterliegt) und „Zeit“ (die der Restriktion eines knappen Zeitbudgets unterliegt) mit dem Ziel kombiniert, seinen Nutzen zu maximieren. Der Gesamtpreis des Gutes „Transport“ setzt sich somit aus einer direkten Kostenkomponenten (dem Fahrpreis in Geldeinheiten) und einer indirekten Kostenkomponente (dem Zeitpreis gemessen in Geldeinheiten entgangenen Lohneinkommens) zusammen.

Da es sich bei alltäglich beschafften Transport-Gütern um relativ billige, aber zeitintensive Güter handelt, ist davon auszugehen, daß der Gesamtpreis des jeweiligen Gutes „Transport“ deutlich stärker durch die Fahrzeit (d.h. den Zeitpreis gemessen in Geldeinheiten des entgangenen Lohneinkommens) bestimmt wird als durch den jeweiligen Fahrpreis (d.h. die Zeit, die pro Geldeinheit des Fahrpreises ausgegeben werden muß). Das sei an einem Beispiel demonstriert.

Angenommen, für die Zurücklegung eines 15 km langen Weges zur Arbeit benötigt eine Person mit dem Bus 50 Minuten und mit dem PKW 20 Minuten. Die Kosten für die Busnutzung belaufen sich beim Bus auf 35 Pfennig/km und beim Auto auf 72 Pfennig/km. Wenn eine Person einen Netto-Lohn von 20 DM/Stunde erzielt, würden sich nach Beckers Ansatz die Gesamtkosten dieser Person für eine Hin- und Rückfahrt beim Bus auf 44 DM belaufen und beim PKW auf 35 DM, obwohl der reine Fahrpreis des PKW doppelt so hoch ist wie der des Busses. Somit läßt sich aus der Zerlegung der Gesamtkosten des Gutes „Transport“ in eine Geld- und eine Zeitkostenkomponente eine schlüssige Erklärung für den oben berichteten empirischen Befund ableiten, daß der reine monetäre Fahrpreis kaum Einfluß auf die Entscheidung zwischen Verkehrsmitteln hat, die Fahrzeit hingegen einen sehr starken.

Wie oben berichtet, hat sich in empirischen Studien ein Einfluß soziodemografischer und sozioökonomischer Variablen wie Familiengröße, Alter, Geschlecht, Bildung, Erwerbstätigkeit und beruflicher Status auf die Verkehrsmittelwahl gezeigt. Während sozialwissenschaftliche Ansätze, wie z.B. die Lebensstilforschung (siehe dazu Labonte, 1996) diese Effekte auf Präferenzunterschiede zurückführen, führt Becker den Einfluß dieser Größen direkt auf Veränderungen bzw. Unterschiede in den Geld- und Zeitrestriktionen zurück.

1. Auswirkung real steigenden Lohneinkommens auf die Verkehrsmittelwahl

So erhöht ein gleichförmiger prozentualer Lohnanstieg die Kosten der für Konsumzwecke aufgewendeten Zeiteinheiten um den gleichen Prozentsatz. Die relativen Preise der vom Haushalt produzierten Güter steigen jedoch in Abhängigkeit von der für ihre Produktion eingesetzten Zeit unterschiedlich an: Die Preise für Haushaltsgüter, die viel Konsumzeit benötigen, steigen stärker als die Preise für Güter, die weniger Konsumzeit benötigen. Nach Becker reagieren die Haushalte auf die Veränderung der Preise für zeitintensive Güter mit zwei Verhaltensweisen: Sie verlagern den Konsum von den zeitintensiven auf den Konsum von marktgüterintensiven Güter und/oder sie versuchen, zeitintensive Produktionsmethoden von Haushaltsgütern durch weniger zeitintensive Produktionsmethoden zu substituieren.

Übertragen auf die Verkehrsmittelwahl heißt das, mit steigendem Lohneinkommen wird die Nutzung des Busses in Relation zur PKW-Nutzung um so teurer, je größer die Fahrzeitdifferenz zwischen Bus und PKW ist, auch wenn der reine Fahrpreis konstant bleibt. Damit in dieser Situation eine Person, die zwischen PKW und Bus wählen kann, auch weiterhin den Bus nutzt, muß entweder der Marktpreis für die Busnutzung oder die Fahrzeit in dem Umfang gesenkt werden, daß damit der Anstieg der teurer gewordenen Fahrzeit kompensiert werden kann. Sonst kommt es zu einer Substitution der zeitintensiveren Busfahrt durch die zeitsparendere PKW-Fahrt.

Der durch steigende Lohneinkommen ausgelöste Substitutionsprozeß zeitintensiver durch marktgüterintensive Methoden bei der Produktion von Haushaltsgütern bietet sowohl eine Erklärung für die in den letzten 35 Jahren exponentiell anwachsende PKW-Nutzung, wie auch für die immer noch zu beobachtende stärkere ÖV-Nutzung von Personen mit geringerem Einkommen im Vergleich zu Personen mit höherem Einkommen. Der Effekt von Bildung und beruflichem Status auf die Verkehrsmittelwahl läßt sich so gesehen als indirekter Lohneinkommenseffekt interpretieren: Bildung und beruflicher Status sind Indikatoren für den am Markt erzielbaren Arbeitslohn.

2. Effekt der Erwerbstätigkeit auf Verkehrsmittelwahl

Eine andere Implikationen von Beckers Ansatz ist, daß die jeweilige Einkommensquelle ebenfalls einen Einfluß auf den Konsum eines Haushalts haben kann. So werden nach Becker die jeweiligen Zeitkosten nur durch das entgangenen Lohneinkommen bestimmt. Daraus folgt, daß bei Personengruppen, wie z.B. Rentnern, deren Einkommen nicht aus Lohnarbeit, sondern aus Transferleistungen resultiert, die Zeitkosten niedriger sind als für Erwerbstätige. Damit läßt sich der empirische Befund, daß Rentner stärker öffentliche Verkehrsmittel nutzen, auch wenn sie über einen PKW verfügen, durch die Art des Einkommens erklären: Für nicht erwerbstätige Menschen spielt der Arbeitslohn

als Verdienstquelle eine untergeordnete Rolle. Damit sinken auch die relativen Kosten der für den Konsum einer Transportdienstleistung aufgewendeten Zeit. Nach Becker führt das tendenziell zu einer Substitution marktgüterintensiver Güter durch zeitintensive Güter: Sie fahren mehr Bus als PKW.

3. Einfluß des Geschlechts auf die Verkehrsmittelwahl

Der Befund, daß Frauen öfter den Bus nutzen als Männer läßt sich ebenfalls durch Unterschiede in den Zeitkosten der Haushaltsmitglieder erklären. So verwenden Haushaltsmitglieder, die bei Markt-Aktivitäten relativ effizienter sind, weniger Zeit für Konsumaktivitäten als andere Mitglieder. Daraus folgt eine Reallokation der Zeit aller anderen Haushaltsmitglieder hin zu Konsumaktivitäten, um so dem ersteren die Möglichkeit zu geben, mehr Zeit für Marktaktivitäten einzusetzen. Wenn nur ein Haushaltsmitglied berufstätig ist und nur ein PKW zur Verfügung steht, wird danach dieser PKW vor allem von dem berufstätigen Haushaltsmitglied genutzt, da die relativen Zeitkosten und damit Transportkosten für diese Person höher sind als die der nichtberufstätigen Personen. Da es derzeit in vielen Haushalten immer noch eine geschlechtsgebundene Arbeitsteilung in dem Sinne gibt, daß Männer voll berufstätig sind, während die „Erziehungs-“ und „Haushaltsarbeit“ hauptsächlich von den Frauen geleistet wird, läßt sich die häufigere PKW-Nutzung der berufstätigen Männer daraus ableiten, daß bei ihnen das Gewicht des entgangenen Verdienstes größer ist als bei ihren nichtberufstätigen Frauen. Die stärkere Bus-Nutzung der Frauen ist damit eine Folge der Wirkung von Restriktionen und nicht einer besonderen Präferenz von Frauen für das Verkehrsmittel Bus.

Weiter lassen sich Hypothesen formulieren, welchen Effekt der berufliche Wiedereinstieg von Personen, die bisher Hausarbeit geleistet haben auf die Verkehrsmittelwahl hat. Die Aufnahme einer Berufstätigkeit führt zu Veränderungen des Zeitbudgets (die von einer Person für Konsumaktivitäten aufgewendete Zeit knapper und damit teurer) und des Einkommensbudgets (das verfügbare Einkommen steigt an). Als Reaktion auf diese Veränderungen in den Budgetrestriktionen tendieren diese Personen dazu die langsamere Busfahrt durch die schnellere PKW-Fahrt zu substituieren. Ein Zweitwagen wird angeschafft.

3. Empirie

Im folgenden soll über die Ergebnisse einer Studie berichtet werden, in der versucht wird, bezogen auf eine relativ kurze, alltäglich zurückgelegte Fahrstrecke die beobachtbare individuelle Verkehrsmittelnutzung nicht auf den Einfluß von Präferenzen,

sondern ausschließlich auf die Wirksamkeit externer Restriktionen, besonders der Zeitrestriktion zurückzuführen.

Die verkehrsinfrastrukturellen Rahmenbedingungen

Untersucht wurde die Verkehrsmittelwahl auf der Strecke Wettenberg - Gießen. Wettenberg ist eine kleine, ca. 6.000 Einwohner große hessische Gemeinde, die sechs km von dem regionalen Oberzentrum Gießen entfernt liegt. Es gibt eine direkte Busanbindung nach Gießen mit insgesamt fünf Zustiegsmöglichkeiten. Die Fahrtdauer des Busses von Wettenberg zum Gießener Marktplatz variiert in Abhängigkeit von der jeweiligen Einstiegshaltestelle zwischen 21 Minuten (erste Haltestelle am Ortseingang von Gießen aus gesehen) und 26 Minuten (Haltestelle am Ortsende von Gießen aus gesehen). Weiter ist von keinem Punkt der Gemeinde eine Bushaltestelle mehr als 400 Meter entfernt. Der Fahrpreis für eine Einzelfahrt beträgt 2,70 DM. Die Fahrzeit mit dem PKW beträgt für diese Strecke rund zehn Minuten. Es ist also davon auszugehen, daß die infrastrukturellen Rahmenbedingungen bei Fahrten in die Gießener Innenstadt sowohl bei der PKW- wie bei der Bus-Nutzung für alle Wettenberger Bürgerinnen und Bürger relativ ähnlich sind.

Abbildung 1: Modal-Split der Befragten (N=200) bei Wegen von Wettenberg nach Gießen

Die Untersuchungsstichprobe

Im Juni 1994 wurden 200 zufällig aus dem Telefonbuch ausgewählten Wettenberger Bürgerinnen und Bürger über 18 Jahre telefonisch zu ihrer Verkehrsmittelwahl befragt. Von den Befragten waren 57 Prozent weiblich und 43 Prozent männlich. Das durchschnittliche Alter lag bei 43,8 Jahren. Von den Befragten gaben 59 Prozent an, erwerbstätig zu sein. Über keinen formalen Schulabschluß verfügen zwei Prozent der Befragten, 41 Prozent gaben an, einen Hauptschulabschluß zu haben, 30 Prozent einen Real/Handelschulabschluß und 27 Prozent gaben als Schulabschluß das Abitur an. Weiter wird davon ausgegangen, daß es sich bei den 24 Personen, die über 60 Jahre alt sind und angeben nicht erwerbstätig zu sein, um Rentner handelt.

Erhebung des Verkehrsmittelnutzungsverhaltens

Die abhängige Variable „Verkehrsmittelnutzungsverhalten“ wurde erhoben, indem gefragt wurde, ob die Befragten am Befragungstag einen Weg nach Gießen unternommen haben und wenn ja, welches Verkehrsmittel sie dazu genutzt haben. Wenn die Befragten angaben, am Befragungstag keinen Weg nach Gießen unternommen zu haben, wurden sie gebeten, sich an den letzten zurückliegenden Weg nach Gießen zu erinnern und anzugeben, welches Verkehrsmittel sie für diesen Weg genutzt haben. Abbildung 1 läßt sich entnehmen, wie die Verkehrsmittelnutzung der von uns Befragten bei Wegen nach Gießen aussieht.

Operationalisierung der Zeit-, Geld- und Komfort-Kosten

Ziel der vorliegenden Untersuchung ist es, die beobachtbare individuelle Verkehrsmittelwahl anhand der Restriktionen „Einkommen“ und „Zeit“, bzw. der daraus resultierenden Geld- und Zeitkosten zu erklären. Die empirische Fruchtbarkeit dieses Ansatzes hängt also von der validen und reliablen Erfassung dieser Kosten ab. Eine Möglichkeit bestünde in der objektiven Messung dieser Kosten. Dazu müßte der Forscher die Person „mit Stoppuhr und Notizblock“ begleiten. Diese Art der Datenerfassung ist jedoch sehr aufwendig und teuer. In den meisten Studien wird daher auf eine Erhebung der objektiven Kosten verzichtet. Statt dessen werden subjektive Schätzungen der Fahrzeit in Minuten und der Fahrkosten in DM durch die Befragten erhoben. Neben forschungsökonomischen Gründen wird dabei als inhaltliche Begründung für dieses Vorgehen darauf verwiesen, daß nicht die objektiven Größen „Fahrzeit“ und „Fahrpreis“ an sich entscheidungsrelevant sind, sondern deren subjektive Wahrnehmung.

Auch in der vorliegenden Studie wurden solche subjektiven Schätzungen erhoben. Eine nähere Analyse hat jedoch starke Zweifel an der Reliabilität dieser subjektiven Schätzungen aufkommen lassen. So variierte die geschätzte Dauer der Busfahrt nach Gießen in der Gruppe der PKW-Nutzer zwischen zehn Minuten und 60 Minuten, die

geschätzten Fahrpreise zwischen 2,70 DM und 10 DM. Aber auch die Schätzungen für die Fahrzeit und den Fahrpreis des von ihnen tatsächlich genutzten PKWs variieren in der Gruppe der PKW-Nutzer beträchtlich: von vier bis 35 Minuten für die Fahrzeit sowie von 0,7 bis 7 DM für den Fahrpreis. Das größte Problem bei der Erhebung der subjektiven Schätzungen der Fahrkosten stellt jedoch die hohe Rate fehlender Werte bei der Schätzung des PKW-Fahrpreises durch die PKW-Nutzer dar: Rund die Hälfte der befragten PKW-Nutzer wollten oder konnten keine Angaben zu den Kosten einer PKW-Fahrt nach Gießen machen. Im weiteren soll deshalb auf die Nutzung dieser subjektiven Zeit- und Kostenschätzungen verzichtet werden.

Eine dritte Möglichkeit der Messung von mit einem Verhalten verbundenen Kosten stellen „indirekte“ Kostenindikatoren dar. Dieser Ansatz verzichtet auf die direkte Erhebung von Kosten in Geld- oder Zeiteinheiten. Stattdessen wird versucht, die Verhaltenskosten indirekt zu messen, in dem potentiell „kostenträchtige“ Eigenschaften einer Verhaltensalternative den Befragten zur subjektiven Beurteilung vorgelegt werden. Damit setzt die Verwendung von indirekten Kostenindikatoren jedoch eine theoretische Begründung (sog. Brückenannahme) dafür voraus, worin die „Kostenrelevanz“ einer spezifischen Eigenschaft einer Verhaltensalternative besteht.

Im folgenden wird davon ausgegangen, daß sich die mit der Nutzung eines Verkehrsmittels verbundene Gesamtzeit- und Gesamtgeldkosten in einzelne Kostenkomponenten zerlegen lassen. So bestehen die mit der Busnutzung verbundenen Gesamtzeitkosten aus der Summe der Zeit, die zur Informationsbeschaffung über die angebotenen öffentlichen Personentransportdienstleistungen, der Zeit für den Zugang- und für den Abgang von der Haltestelle, der Zeit, die an der Haltestelle auf den Bus gewartet wird, der reinen Fahrzeit des Busses sowie der Zeit, die für eventuell notwendige Umstiegsvorgänge benötigt wird.

Die reinen Geldkosten der Busnutzung entsprechen dem direkt zu zahlenden Fahrpreis, brauchen also nicht weiter differenziert zu werden. Eventuell kann zwischen dem Preis eines Einzelfahrscheins und den günstigeren Dauerscheinen unterschieden werden.

Im Unterschied zur direkten Kostenmessung bewerten die Befragten diese einzelnen Kostenkomponenten nicht in Minuten oder DM, sondern z.B. mit Schulnoten von eins (sehr gut) bis sechs (ungenügend). Hinter dieser Bewertung steht die Annahme, daß die Befragten Eigenschaften einer Verhaltensalternative um so negativer bewerten, je kosten-trächtiger diese Eigenschaften von ihnen wahrgenommen werden. Tabelle 1 stellt die in der vorliegenden Untersuchung gemessenen Kostenaspekte dar.

Tabelle 1: In der Untersuchung erhobene spezifische Zeit- und Geldkosten der Busnutzung

<u>Zeit-Kosten</u>	<u>Geld-Kosten</u>
- aktuelles Buswissen	- Preis des Einzelfahrscheins
- Verständlichkeit der Fahrpläne	- Preis der Dauerkarte
- Informationsmöglichkeiten bei Verkehrsunternehmen	
- gut merkbare Abfahrtszeiten der Busse	
- Erreichbarkeit der Haltestelle	
- Dauer der Busfahrt	
- Fahrthäufigkeit der Busse	
- Pünktlichkeit der Busse	

Der aktuelle Wissensstand über das Busangebot wurde mittels folgender drei Fragen gemessen: „Wüßten Sie, wo die nächste Bushaltestelle ist?“, „Wüßten Sie, mit welcher Buslinie Sie zum Ziel kommen?“ und „Wüßten Sie, wann der nächste Bus zum Zielort fährt?“ Aus diesen, jeweils mit Ja oder Nein beantworteten drei Items wurde die Indexvariable „aktuelles Buswissen“ gebildet.

Zusätzlich zu den indirekten Indikatoren für spezifische „Kostenkomponenten“ wurden in der vorliegenden Untersuchung noch indirekte Maße für die von den Befragten mit der Nutzung der Verkehrsmittel Bus und PKW verbundenen Gesamtzeit- und Gesamtgeldkosten erhoben. Dazu wurden die Befragten gebeten, sowohl für das Verkehrsmittel Bus wie PKW die jeweilige Zutreffenswahrscheinlichkeit der Attribute „preiswert“ und „schnell“ auf einer dreistufigen Skala von „sehr zutreffend“ bis „nicht zutreffend“ einzustufen. Dem separaten Erheben globaler Maße für die mit der Nutzung eines Verkehrsmittels verbundenen Gesamtzeit- und Gesamtgeld-Kosten liegen zwei Überlegungen zugrunde. Zum einen lassen sich damit die theoretisch abgeleiteten Brückenannahmen empirisch testen. So sollen die spezifischen Zeitkosten nur mit dem globalen Maß für die Gesamtzeitkosten zusammenhängen und nicht mit dem globalen Maß der Geldkosten. Zum anderen ist es wenig wahrscheinlich, daß in einer Entscheidungssituation eine Person die spezifischen Kostenaspekte zur Beurteilung einer Verkehrsmittelalternative heranzieht. Sie wird sich vielmehr auf die im Gedächtnis gespeicherte globale Beurteilung „schnell“ eines Verkehrsmittels stützen, die ihrerseits eine Art Summe aus der Beurteilung der Einzelaspekte darstellen.

Untersuchungshypothesen

1. Unter der Restriktion begrenzter Zeit- und Einkommensbudgets versuchen Personen in einer Entscheidungssituation ihren Nutzen zu maximieren, indem sie die Verkehrsmittelalternative wählen, die am wenigsten von diesen knappen Ressourcen verbraucht. Da es sich bei dem hier untersuchten Gut um den Transport über relativ kurze Strecken handelt, soll der Gesamtpreis dieses Gutes viel stärker durch das Gewicht der indirekten Kostenkomponente „Fahrzeit“ bestimmt werden als durch die direkte Kostenkomponente „Fahrpreis“. Die subjektiv wahrgenommene Fahrzeitdifferenz zwischen den beiden Verkehrsmittelalternativen „PKW“ und „Bus“ soll die zentrale Determinante der jeweiligen Verkehrsmittelwahl sein, während die für diese Strecke relativ geringe Fahrpreisdifferenz keinen Einfluß auf die jeweilige Verkehrsmittelwahl hat.

2. Die globale Einschätzung der mit der Benutzung des Verkehrsmittels „Bus“ verbundenen Zeit- und Geldkosten beruht auf der Beurteilung der in Tabelle 1 dargestellten spezifischen „kostenträchtigen“ Eigenschaften dieses Verkehrsmittels. So soll die globale Beurteilung des Verkehrsmittels Bus als „schnell“ von der Beurteilung der Pünktlichkeit der Busse, der Fahrtdauer, der Fahrthäufigkeit, sowie dem Zeitaufwand zur Beschaffung von Informationen über das Busangebot abhängen. Die globale Beurteilung des Verkehrsmittels „Bus“ als „preiswert“ hängt von der Beurteilung der Preise für einen Einzelfahrschein und für Dauerfahrschein ab.

3. Es wird erwartet, daß die Merkmale „Rentner“, „Erwerbstätigkeit“ und „Bildung“ einen systematischen Einfluß auf die jeweilige Beurteilung der mit der Busnutzung verbundenen Zeit- und Geldkosten haben. Dabei wird davon ausgegangen, daß diese soziodemographischen Variablen indirekt Unterschiede in den jeweiligen Zeit- und Einkommensbudgets messen.

Da Rentner ihre Einkommen unabhängig von der Lohnarbeit erzielen, ist ihr „Zeitpreis“ geringer als der von Berufstätigen. Daher wird erwartet, daß Rentner die mit der Busnutzung verbundenen Zeitkosten tendenziell niedriger einschätzen als berufstätige Personen, die mit der Busnutzung verbundenen Geldkosten hingegen als tendenziell höher.

Wenn man den formalen Bildungsabschluß einer Person als Indikator für den jeweiligen Arbeitslohn nimmt, den ein Berufstätiger am Arbeitsmarkt realisieren kann, steigen mit Höhe des formalen Bildungsabschlusses auch die jeweiligen Zeitkosten. Es wird erwartet, daß Personen mit höherem Bildungsabschluß die mit der Busnutzung verbundenen Zeitkosten tendenziell höher einschätzen als Personen mit niedrigerem

Bildungsabschluß, die Geldkosten jedoch tendenziell niedriger. Abbildung 2 stellt die inhaltlichen Hypothesen grafisch dar.

Abbildung 2: Grafische Darstellung der inhaltlichen Hypothesen

Analysen

Deskriptive Analysen

Im Folgenden soll ausschließlich die Entscheidung zwischen den Verkehrsmitteln PKW und Bus für Fahrten nach Gießen analysiert werden. Deshalb wurden die Radfahrer und der Motorradfahrer von den weiteren Analysen ausgeschlossen. Damit reduziert sich die Stichprobe auf $N = 176$. Um sicherzustellen, daß die Befragten wirklich zwischen den Alternativen PKW und Bus wählen können, wurden alle Personen, die auf die Frage „Können Sie ständig über einen PKW verfügen,“ mit „Nein“ geantwortet haben, ebenfalls von der weiteren Analyse ausgeschlossen. Die folgenden Analysen beruhen auf einem $N = 115$ Personen.

Tabelle 2 läßt sich die mittlere Zutreffenswahrscheinlichkeit der Attribute „schnell“ und „preiswert“ auf die Fahrt mit dem Bus bzw. dem PKW nach Gießen für alle Befragten sowie getrennt für die Subgruppe der Rentner, aller Erwerbstätigen und der Erwerbstätigen getrennt nach der Höhe des Bildungsabschlusses „Abitur“ - „kein Abitur“ entnehmen. Weiter enthält die Tabelle Differenzwerte, die aus der Zutreffenswahrscheinlichkeit der Attribute „schnell“ und „preiswert“ auf die Busnutzung minus der Zutreffenswahrscheinlichkeit dieser Attribute auf die PKW-Nutzung gebildet wurden (negativer Wert bedeutet Attribut trifft stärker auf die PKW-Nutzung zu).

Tabelle 2: Mittlere Zutreffenswahrscheinlichkeit der Eigenschaften "schnell" und "preiswert"

	Gesamt N=115	Rentner N=13	Erwerbs- tätige N=74	Erwerbstät. ohne Abitur N=50	Erwerbstät. mit Abitur N=24
BUS schnell	1.5	2.1	1.5	1.5	1.3
PKW schnell	2.6	2.7	2.6	2.6	2.6
BUS preiswert	1.4	1.3	1.4	1.4	1.5
PKW preiswert	2.2	2.3	2.2	2.3	1.8
Differenzvariable „Schnell“	-1.1	-0.6	-1.1	-1.1	-1.3
Differenzvariable „Preis“	-0.8	-1.0	-0.7	-0.9	-0.3

Skala für "schnell" und "preiswert"; 1 (unzutreffend), 2 (teilweise zutreffend), 3 (voll und ganz zutreffend); in der Kategorie „Rentner“ werden Personen zusammengefaßt, die über 60 Jahre sind und angeben, nicht mehr erwerbstätig zu sein.

Wie sich Tabelle 2 entnehmen läßt, werden in der Gesamtgruppe die mit der PKW-Nutzung verbundenen Geld-, besonders aber die Zeitkosten, deutlich niedriger eingestuft als bei der Busnutzung. Angesichts dieser deutlichen Kostendifferenz zwischen Bus und PKW verwundert es nicht, daß von den 115 Personen, die ständig über einen PKW ver-

fügen können, nur vier Personen (3,5 Prozent) angeben, für die letzte Fahrt nach Gießen den Bus genutzt zu haben.

Weiter läßt sich Tabelle 2 entnehmen, daß sich der theoretisch erwartete Zusammenhang zwischen der Bewertung der mit der Nutzung des Verkehrsmittels Bus verbundenen Kosten und den Variablen „Rentner“, „Erwerbstätigkeit“ und „Bildung“ zeigt. So halten die Rentner nicht nur das Attribut „preiswert“ für deutlich weniger auf die Busfahrt von Wettenberg nach Gießen zutreffend als das Attribut „schnell“, sondern bei den Rentnern ist auch die Differenz der Attribute „Bus preiswert“ und „Pkw preiswert“ am größten. Die Rentner nehmen also den Bus in Relation zum PKW vor allem als deutlich teureres Verkehrsmittel wahr. Die Erwerbstätigen stufen hingegen das Attribut „schnell“ als am wenigsten auf die Busnutzung zutreffend an. Bei ihnen ist auch die Differenz der Attribute „Bus schnell“ und „PKW schnell“ am größten. Die Erwerbstätigen nehmen den Bus also in Relation zum PKW vor allem als das deutlich langsamere Verkehrsmittel wahr. Zwischen Erwerbstätigen ohne Abitur und Erwerbstätigen mit Abitur zeigen sich ebenfalls die erwarteten Unterschiede. Bei Erwerbstätige mit Abitur fällt die Differenz der Zeitkosten für die Busnutzung und der PKW-Nutzung noch einmal höher aus als bei den Erwerbstätigen ohne Abitur, während die wahrgenommenen Unterschiede in den Geldkosten zwischen Bus und PKW deutlich geringer ausfallen als bei den Erwerbstätigen ohne Abitur. Tabelle 3 stellt die mittlere Bewertung der mit der Busnutzung verbundenen spezifischen Kostenkomponenten dar.

Tabelle 3: Mittlere Beurteilung der spezifischen Kosten

	Gesamt	Rentner	Erwerbs- tätige	Erwerbst. ohne Abitur	Erwerbst. mit Abitur
Zeitkosten					
aktuelles Buswissen	1.7	1.9	1.7	1.7	1.5
Verständlichkeit Fahrpläne	3.2	2.9	3.2	3.4	3.1
Info-Möglichkeit	2.9	2.1	3.0	3.1	3.0
gut merkbare Abfahrtszeit	3.0	2.6	3.1	3.0	3.2
Erreichbarkeit Haltestelle	2.3	2.8	2.2	2.4	2.1
Dauer der Fahrt	3.3	2.6	3.4	3.4	3.8
Fahrthäufigkeit	4.7	4.4	4.7	4.7	4.9
Pünktlichkeit	2.2	2.0	2.2	2.0	2.3
Geldkosten					
Einzelfahrschein	4.7	4.1	4.7	4.7	4.8
Dauerkarte	4.0	3.8	4.0	3.9	4.1

Skala für „aktuelles Buswissen“ (0 = kein Wissen bis 3 = hohes Wissen). Alle anderen Skalen von 1 = sehr gut bis 6 = ungenügend.

Danach werden die beiden Aspekte „Erreichbarkeit der Haltestelle“ und „Pünktlichkeit der Busse“ am positivsten beurteilt, während die Aspekte „Fahrthäufigkeit der Busse“, „Preis des Einzelfahrscheins“ und „Preis der Dauerkarte“ besonders negativ beurteilt werden. Weiter zeigen sich deutliche Unterschiede zwischen den Rentnern und Erwerbstätigen bei der Beurteilung der Aspekte „Informationsmöglichkeiten“, „Erreichbarkeit der Haltestelle“, „Fahrtdauer“, „Preis des Einzelfahrscheins“ sowie „gut merkbare Abfahrtszeiten“. Tabelle 4 stellt die bivariaten Zusammenhänge (polychorische Korrelationen) zwischen den Differenzwerten der Attribute „schnell“, „preiswert“ und „bequem“ und den spezifischen Kostenaspekten sowie den Variablen „Bildung“, „Rentner“ und „Erwerbstätigkeit“ dar.

Tabelle 4: Interkorrelation der spezifischen Kostenaspekte, der Differenzvariablen sowie der soziodemografischen Merkmale

	Differenzvariable „Schnell“	Differenzvariable „Preiswert“	Bildungsabschluß	Rentner	Erwerbstätig
Verständlichkeit Fahrpläne	-0.12	0.05	0.07	0.12	-0.18
Informationsmöglichkeiten	-0.01	0.22*	0.06	0.53*	-0.31*
Pünktlichkeit	0.02	-0.03	0.28*	0.19	0.19
gut merkbare Abfahrtszeit	-0.01	-0.01	-0.27*	0.26*	-0.18
Erreichbarkeit Haltestelle	-0.10	0.10	0.11	-0.34*	0.09
Dauer der Fahrt	0.37*	-0.08	-0.19	0.43*	-0.35*
Preis Einzelfahrschein	0.09	0.36*	0.00	0.30*	-0.09
Preis Dauerkarte	0.09	0.37*	-0.01	0.18	0.07
Fahrthäufigkeit	0.08	0.06	-0.14	0.11	-0.16
aktuelles Buswissen	0.40*	0.06	-0.17	0.20*	-0.08
Bildungsabschluß	0.00	0.27*			
Rentner	0.34*	-0.13			
Erwerbstätig	-0.11	0.00			

Koeffizienten mit * sind auf .05-Niveau signifikant.

Wie aus Tabelle 4 ersichtlich, korrelieren die Zeitkostenaspekte „Fahrtdauer“ und „aktuelles Buswissen“ signifikant mit der Differenzvariablen „schnell“. Mit der Differenzvariable „preiswert“ korrelieren die Geldkostenaspekte „Preis des Einzelfahrscheins“ und „Preis der Dauerkarte“ sowie der Zeitkostenaspekt „Informationsmöglichkeiten“. Ferner zeigt sich eine substantielle positive Korrelation zwischen der Variable „Rentner“ und der Differenzvariable „schnell“ sowie den spezifischen Kostenaspekten „Informationsmöglichkeiten“, „gut merkbare Abfahrtszeiten“, „Fahrtdauer“ und „Preis des Einzelfahrscheins“ sowie eine negative Korrelation mit

„Erreichbarkeit der Haltestelle“. Die Variable „Bildungsabschluß“ korreliert positiv mit der Differenzvariable „preiswert“ und negativ mit den spezifischen Kostenkomponenten „Pünktlichkeit“ und „gut merkbare Abfahrtszeiten“. Die Variable „Erwerbstätigkeit“ korreliert negativ mit den spezifischen Kostenkomponenten „Informationsmöglichkeiten“ und „Fahrtdauer“.

*Multivariate Analysen*²⁾

Im Folgenden soll zum einen die Hypothese getestet werden, daß die Beurteilung der mit der Busnutzung verbundenen globalen Zeit- und Geldkosten ihrerseits auf der Beurteilung spezifischer Kostenaspekte beruht. Zum anderen soll die Annahme getestet werden, daß die soziodemografischen Variablen die Verkehrsmittelwahl nicht direkt beeinflussen, sondern nur indirekt über ihren Effekt auf die Beurteilung der spezifischen Kostenaspekte. Dabei wird davon ausgegangen, daß die Variable „Bildung“ und „Erwerbstätigkeit“ einen negativen Effekt auf die Beurteilung der spezifischen Zeitkostenaspekte hat und einen positiven Effekt auf die Geldkostenaspekte, während die Variable „Rentner“ einen positiven Effekt auf die Beurteilung der spezifischen Zeitkostenaspekte und einen negativen Effekt auf die Geldkosteneffekte haben soll.

Diese Annahmen lassen sich in ein System von linearen Gleichungen überführen, die sich mittels Strukturgleichungsmodellen simultan an den Daten testen lassen. Abbildung 3 stellt die Ergebnisse eines solchen Strukturgleichungsmodells grafisch dar. Die dargestellten Parameter des linearen Gleichungssystems wurden mittels ML-Schätzer durch das Programm LISREL 8 (Jöreskog/Sörbom 1993) geschätzt³⁾. Als Input-Matrix dienen polychorische Korrelationen. Die über die Fitmaße gemessene Anpassung des in Abbildung 2 dargestellten Modells an die Daten ist zufriedenstellend ($\chi^2_{(N = 115, df=24)} = 23.88$; $p = .47$; GFI = .97; RMR = .07).

Wie sich dem in Abbildung 3 dargestellten Modell entnehmen läßt, üben lediglich drei der insgesamt zehn erhobenen spezifischen Kostenaspekte einen statistisch signifikanten direkten Einfluß auf die Differenzwerte „schnell“ und „preiswert“ aus. Der Differenzwert „schnell“ wird von den beiden Zeitkostenaspekten „Dauer der Fahrt“ und „aktuelles Buswissen“ beeinflusst. Der Differenzwert „preiswert“ wird von dem Geldkostenaspekt „Preis des Einzelfahrscheins“ beeinflusst. Die Koeffizienten der anderen spezifischen Kostenaspekte sind statistisch nicht signifikant. Wie erwartet zeigt sich ein positiver Effekt der Variablen „Rentner“ auf die Beurteilung der Fahrtdauer sowie auf das aktuelle Buswissen. Die Variable „Erwerbstätigkeit“ hat wie erwartet einen negativen Effekt auf die Beurteilung der Fahrtdauer ebenso wie die Variable „Bildung“ einen negativen Effekt auf die Beurteilung der Pünktlichkeit der Busse hat. An einigen Stellen

mußte die theoretisch erwartete Modellstruktur modifiziert werden. So beeinflusst das aktuelle Buswissen direkt die Beurteilung des Einzelfahrkartenpreises.

Abbildung 3: Grafische Darstellung des akzeptierten LISREL-Modells

Weiter beeinflußt die Beurteilung des Zeitkostenaspekts „Pünktlichkeit“ nicht direkt den Differenzwert „schnell“, sondern nur indirekt über die Beurteilung der Fahrdauer. Die Variable „Bildung“ übt keinen direkten Einfluß auf den spezifischen Kostenaspekt „Preis der Einzelfahrkarte“ aus, sondern einen direkten positiven Effekt auf die Differenzvariable „preiswert“. Ferner wurden noch Korrelationen zwischen der Variablen „Rentner“ und „Bildung“ (-.34) und zwischen „Rentner“ und der Variablen „Erwerbstätigkeit“ (-.53) zugelassen.

Prädiktion der beobachteten Verkehrsmittelnutzung anhand der Zeit- und Geldkostendifferenzen zwischen den Verkehrsmitteln Bus und PKW

Als nächstes soll die Hypothese getestet werden, daß die individuelle Wahl zwischen den Verkehrsmitteln Bus und PKW bei den Wegen nach Gießen ausschließlich auf der wahrgenommenen Zeitkostendifferenz der beiden Verkehrsmittel beruht. Dazu wird angenommen, daß sich die Erhöhung der Zeit- und der Geldkosten in gleicher Weise auf die Nutzung der Verkehrsmittel Bus und PKW auswirkt. Diese theoretischen Annahmen lassen sich auf elegante Art und Weise mit Hilfe konditionaler Logit-Modelle testen. Anhand der Gleichung

$$p_i / p_j = \exp [\sum (x_{ki} - x_{kj}) \beta_k]$$

läßt sich nachvollziehen, daß im konditionalen Logitmodell das Wahrscheinlichkeitsverhältnis zweier Verhaltensalternativen i und j durch die Differenz der Prädiktorvariablen X_{ki} und X_{kj} bestimmt wird, deren Einfluß auf p_i/p_j durch den Regressionskoeffizienten β_k abgebildet wird. Die unabhängigen Variablen X_{k1} ; X_{k2} , ... X_{kt} die im konditionalen Logit-Modell über ein gemeinsames Regressionsgewicht β_k die Kategorien der abhängigen Variablen beeinflussen, werden als generische Variablen bezeichnet.

Modell 1 in Tabelle 5 stellt die Ergebnisse der ML-Schätzung eines konditionalen Logitmodells dar, das neben einer Regressionskonstanten die beiden generischen Variablen „schnell“ (mit den Elementen Bus schnell; PKW-schnell) und „preiswert“ (Bus preiswert; PKW preiswert) als Prädiktoren der abhängigen Variablen „Busnutzung“ (Ja=1; Nein=0) enthält. Wie sich Tabelle 5 entnehmen läßt, übt in Modell 1 nur die generische Variable „schnell“ einen statistisch signifikanten Einfluß auf die beobachtete Verkehrsmittelnutzung aus. In Modell 2 wurde der Koeffizient der Variablen „preiswert“ auf Null gesetzt. Da es sich bei den Modellen um hierarchisch geschachtelte Modelle handelt, läßt sich die Gültigkeit dieser Modellrestriktion mittels LR-Test statistisch testen.

In Tabelle 6 sind für das Konstantenmodell (M_K) und die beiden Logit-Modelle jeweils die Anzahl der Regressionskoeffizienten ($K+1$), der Wert der jeweiligen Minimierungs-

funktion (LL) und der um die Anzahl der Regressionskoeffizienten korrigierte Likelihood-Ratio-Index (R') aufgeführt. So beträgt beim Vergleich des Modells 2 mit dem Modell 1 der LR-Wert 0.87 und damit kleiner als der 0.95-Quantilwert der χ^2 -Verteilung mit einem Freiheitsgrad, der 3,84 beträgt. Die Nullhypothese, daß der Koeffizient der Variablen „preiswert“ nur zufällig von Null abweicht, kann damit nicht abgelehnt werden.

Tabelle 5: Schätzung der zwei konditionalen Logit-Modelle

Modell 1					
Prädiktor		Koeffizient	S.E.	t-Wert	p-Wert
SCHNELL	β_1	1.725	0.677	2.547	0.011
PREIS	β_2	-0.510	0.582	-0.877	0.380
Konstante	β_3	-3.199	0.863	-3.708	0.000
Modell 2					
Prädiktor		Koeffizient	S.E.	t-Wert	p-Wert
SCHNELL	β_1	1.789	0.688	2.602	0.009
Konstante	β_2	-2.770	0.590	-4.693	0.000

Tabelle 6: Fitmaß der drei konditionalen Logit-Modelle

Modell	K+1	LL	R'
$M_K: [1]$	1	-17.364	-----
M 1: $[1, \beta_1, \beta_2]$	3	-11.790	32,1 %
M 2: $[1, \beta_1]$	2	-12.226	29,6 %

4. Diskussion

Wie theoretisch erwartet, übt in der untersuchten Stichprobe Wettenberger Bürger und Bürgerinnen, die ständig über einen PKW verfügen können, die Differenz der mit der Nutzung der Verkehrsmittel Bus und PKW verbundenen Zeitkosten einen statistisch signifikanten Einfluß auf die aktuelle Bus- bzw. PKW-Nutzung aus. Die subjektiv wahrgenommene Geldkostendifferenz hingegen übt bei Kontrolle der Zeitkostendifferenz keinen eigenständigen Effekt auf die Verkehrsmittelnutzung aus. Nach den Ergebnissen des konditionalen Logitmodells beträgt der Effektkoeffizient der Differenzwertvariablen „schnell“ 5.98 ($= e^{1.789}$), d.h. wenn die Differenz zwischen "Bus schnell" und "PKW

schnell" um eine Einheit größer wird, verschiebt sich die Wahrscheinlichkeitsrelation zwischen den Verkehrsmittelalternativen Bus und PKW um das 5.98-fache.

Ferner konnte die Hypothese bestätigt werden, daß die globale Zutreffenswahrscheinlichkeit der Eigenschaften „schnell“ und „preiswert“ auf das Verkehrsmittel Bus substantiell von der Beurteilung mit der Busnutzung verbundener spezifischer Zeit- und Geldkostenaspekte beeinflusst wird. So erklären die beurteilte Fahrtdauer und das aktuelle Buswissen zusammen 28 Prozent der Varianz der Differenzvariablen „schnell“, die Beurteilung des Preises eines Einzelfahrscheins 22 Prozent der Varianz der Differenzvariablen „preiswert“.

Ebenfalls bestätigt wurde die Hypothese, wonach Rentner die mit der Busnutzung verbundenen Zeitkosten geringer einschätzen als Erwerbstätige, die Geldkosten hingegen höher. Auch für die Variable „formaler Bildungsabschluß“ konnte der erwartete Einfluß bestätigt werden: Personen mit Abitur beurteilen die mit der Busnutzung verbundenen Geldkosten deutlich geringer als Personen ohne Abitur, hingegen die Zeitkosten noch höher als Personen ohne Abitur.

Becker erhebt den Anspruch, eine allgemein gültige Verhaltenstheorie zu formulieren, d.h. eine Theorie, die prinzipiell jedes Verhalten erklären kann, egal ob es sich um den Kauf eines Stücks Seife, um Heirat, Scheidung, ethnische Diskriminierung oder Verbrechen handelt. Hier wurde der Versuch unternommen, diese Theorie zur Erklärung der individuellen Entscheidung, den Bus oder den PKW zu nutzen, heranzuziehen. Zusammenfassend kann gesagt werden, daß sich aus Beckers theoretischen Überlegungen auch für diese Verhaltensweise eine Reihe von direkt empirisch testbaren Annahmen ableiten lassen, die sicherlich nicht auf der Hand liegen. Dabei überzeugt besonders Beckers (1965) sehr sparsame, theoretisch geschlossene und stark formalisierte Modellierung. So lassen sich in Anlehnung an Becker Unterschiede in der individuellen Verkehrsmittelnutzung ausschließlich auf Unterschiede in den Restriktionen Zeit und Geld zurückführen. Der empirisch beobachtbare Effekt soziodemografischer Variablen läßt sich theoretisch stringent durch Unterschiede in den jeweiligen Zeit- und Geldkosten erklären und nicht, wie sonst üblich, als Wirkung von „Präferenzen“ oder „Geschmacksunterschieden“.

Aus den vorliegenden Ergebnissen lassen sich ferner direkte Empfehlungen ableiten, welche Maßnahmen von Verkehrsunternehmen durchgeführt werden sollten, um mehr PKW-Nutzer dazu zu bringen, auf den Bus umzusteigen: Neben Maßnahmen, die die Fahrzeit des Busses verringern, sind danach besonders Maßnahmen als wirksam anzuse-

hen, die auf eine Erhöhung des aktuellen Wissens der PKW-Nutzer über das Busangebot abzielen.

Abbildung 4: Entwicklung der Linie 800 (ohne innerstädtischen Verkehre Gießen)

Quelle: Verkehrsverbund Gießen März 96

Die Ergebnisse der vorliegenden Studie wurden von dem regionalen Busunternehmen als Grundlage bei der Konzeption einer ganzen Reihe von angebotsverbessernden Maßnahmen auf der Linie Wettenberg-Gießen benutzt. Im Mittelpunkt standen dabei Maßnahmen, die auf eine Erhöhung des Busnutzungsattributs „schnell“ abzielten. So wurde das Fahrtangebot um 25 Prozent gesteigert, um die Wartezeit bis zur Abfahrt des nächsten Busses spürbar zu verkürzen. Ferner wurden die Informationskosten über das Busangebot dadurch minimiert, daß ein leicht zu merkender Halb-Studentakt

eingeführt wurde. Weiter wurde der mit der Busnutzung bei Fahrten nach Gießen verbundene Zeitaufwand durch Einrichtung neuer, zentral gelegener Haltestellen, durch Bereinigung des Linienvverlaufs sowie die Meidung staugefährdeter Hauptverkehrsstraßen im Gießener Stadtgebiet reduziert. Aus der Erkenntnis heraus, daß es nicht ausreicht, nur die tatsächliche Qualität des Busangebots zu verbessern, sondern daß diese Verbesserungen auch die potentiellen und aktuellen Kunden erreichen müssen, hat das Busunternehmen vor und bei Einführung der Maßnahmen eine intensive Öffentlichkeitsarbeit betrieben. Dazu gehörten eine umfassende Medienarbeit, Großplakate an wichtigen Ausfallstraßen, Informationsplakate an allen Haltestellen, öffentlichen Einrichtungen, Geschäften sowie in sämtlichen Bussen, Szeneplakate in Gaststätten und kulturellen Einrichtungen mit besonderen Hinweisen auf das nächtliche Fahrangebot, ein Infoblatt mit Taschenfahrplan für alle Wettenberger Haushalte, ein Eröffnungsfest, ein Eröffnungspreis Ausschreiben in den drei örtlichen Tageszeitungen sowie kostenlose Schnupperfahrtscheine, die in den Tageszeitungen und im örtlichen Mitteilungsblatt abgedruckt wurden. Ausführlicher wird die Maßnahmenkonzeption in Bamberg (1995) dargestellt.

Abbildung 4 stellt die von dem Verkehrsunternehmen für die Linie Gießen-Wettenberg pro Monat ermittelten Fahrgastzahlen sieben Monate vor Einführung der Maßnahmen (August 1994) und 16 Monate nach Einführung der Maßnahmen dar. Danach stiegen auf dieser Linie die beförderten Personen von durchschnittlich 27.000 Personen pro Monat auf durchschnittlich 46.000 Personen pro Monat nach Einführung der Maßnahmen an. Das entspricht eine durchschnittlichen Steigerung der Beförderungsfälle nach Einführung der Maßnahmen um ca. 70 Prozent.

Anmerkungen

- 1) Der Beitrag entstand im Rahmen des DFG-Schwerpunktprogramms „Globale Umweltveränderung - sozial- und verhaltenswissenschaftliche Dimension“, Teilprojekt Prof. Dr. Schmidt „Verbot, Anreiz oder Bewußtsein? Verkehrsmittelwahl in einer Region“
- 2) In der multivariaten Analyse werden zur empirischen Testung des in Abbildung 2 dargestellten Kausalmodells zwei statistische Verfahren verwendet: Die theoretisch postulierte Beziehung zwischen den soziodemografischen Variablen, den spezifischen Kostenaspekten und den globalen Zeit- und Geldkosten wird mittels eines linearen Strukturgleichungsmodells an den Daten getestet, der Einfluß der globalen Zeit- und Geldkosten auf die berichtete Busnutzung mittels eines konditionalen Logit-Modells. Der Grund für dieses Vorgehen besteht darin, daß es derzeit im Kontext linearer Strukturgleichungsmodelle noch nicht möglich ist, Modelle mit dichotomen bzw. polytomen abhängigen Variablen zu schätzen.

3) In der vorliegenden Untersuchung ist es davon auszugehen, daß die Daten auf ordinalem Skalenniveau gemessen wurden. Streng genommen setzt jedoch die Parameterschätzung mittels ML in LISREL Intervallskalenniveau voraus. Jöreskog/Sörbom (1989) schlagen bei ordinalen Daten vor, die Parameter mittels WLS zu schätzen, wobei die asymptotische Kovarianzmatrix als Input zu verwenden ist. Die korrekte Schätzung der asymptotischen Kovarianzmatrix setzt jedoch eine Stichprobe größer als $N=200$ voraus. Angesichts dieser Restriktionen bin ich bei der ML-Schätzung geblieben.

Literatur

- Ajzen, I./Fishbein, M., 1980: Understanding Attitudes and Predicting Social Behavior. New Jersey: Englewood Cliffs.
- Bamberg, S., 1995: Marketingmaßnahmen in Verkehrsunternehmen - ein Fallbeispiel. Der Nahverkehr, 9/95: 33-40.
- Bamberg, S./Schmidt, P., 1994: Auto oder Fahrrad? Empirischer Test einer Handlungstheorie zur Erklärung der Verkehrsmittelwahl. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 46: 80-102.
- Bamberg, S./Bien, W., 1995: Angebot (des ÖV) nach Wunsch (des MIV-Nutzers) Handlungstheoretische Erklärungsmodelle der individuellen Verkehrsmittelwahl als Basis für nachfrageorientiertes ÖV-Marketing. Internationales Verkehrswesen, 47: 108-116.
- Becker, G. S., 1965: A Theory of the Allocation of Time. The Economic Journal, 75: 493-517.
- Ben-Akiva, M./Lerman, S., 1985: Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge: MIT Press.
- Brüderl, J./Preisendörfer, P., 1994: Der Weg zum Arbeitsplatz. S. 69-88 in: A. Diekmann/A. Franzen (Hrsg.), Kooperatives Umwelthandeln. Bern: Ruegger.
- Deaton, A./Muellbauer, J., 1980: Economics and Consumer Behavior. Cambridge: Cambridge University Press.
- Diekmann, A., 1994: Umweltbewußtsein oder Anreizstrukturen? Empirische Befunde zum Energiesparen, der Verkehrsmittelwahl und zum Konsumverhalten. S. 39-68 in: A. Diekmann/A. Franzen (Hrsg.), Kooperatives Umwelthandeln. Bern: Ruegger.
- Domencich, T./McFadden, D., 1975: Urban Travel Demand. Amsterdam: North-Holland.
- Erke, H., 1990: Straße und Verkehr. S. 548-554 in: L. Kruse/C. Graumann/E. Lantermann (Hrsg.), Ökologische Psychologie. München: Psychologie Verlags Union.
- Frey, B., 1990: Ökonomie ist Sozialwissenschaft. München: Vahlen.

- Held, M., 1982: Verkehrsmittelwahl der Verbraucher. Berlin: Duncker & Humblot.
- Hensher, D./Dalvi, Q., 1978: Determinants of Travel Choice. Westmead: Saxon House.
- Jöreskog, K.G./Sörbom, D., 1989: LISREL 7: A Guide to the Program and Applications. Chicago: SPSS Inc.
- Jöreskog, K.G./Sörbom, D., 1993: Lisrel 8: Structural Equation Modeling with the SIMPLIS Command Language. Hillsdale, NJ: Erlbaum.
- Labonte, B., 1996: Lebensstile und Alltagsmobilität. Unveröffentlichte Diplomarbeit am Fachbereich 03, Universität Gießen.
- Maier, G./Weiss, P., 1990: Modelle diskreter Entscheidungen. Wien/New York: Springer.
- Molt, W., 1990: Verkehrsmittelnutzung. S. 555-559 in: L. Kruse/C. Graumann/E. Lantermann (Hrsg.), Ökologische Psychologie. München: Psychologie Verlags Union.
- Stigler, G.J./Becker, G.S., 1977: „De Gustibus Non Est Disputandum“. American Economic Review, 67 : 76-90.
- Varian, H.R., 1984: Mikroökonomie. München: Oldenbourg.

Korrespondenzadresse

*Sebastian Bamberg
Fachbereich 03 - Gesellschaftswissenschaften
Justus-Liebig-Universität Gießen
Karl-Glückner-Straße 21E
D-35394 Gießen*

UNIMODALITÄT UND UNIMODALITÄTSTESTS

SIEGFRIED GABLER UND INGWER BORG

Kennzahlen spielen in der Beschreibung von Daten eine wichtige Rolle, wobei der Mittelwert als Lageparameter die beliebteste Kennzahl ist. Die Aussagekraft des arithmetischen Mittels sinkt allerdings gravierend, wenn die Verteilung etwa U-förmig oder allgemeiner, mehrgipfelig ist. Klassierte Daten lassen sich auch in Histogrammen darstellen. Ob aus ihrer Darstellung aber auf die der Population zugrunde liegende Verteilung geschlossen werden kann, hängt häufig von der gewählten Intervalleinteilung ab. Keines der gängigen Statistik-Lehrbücher beschreibt einen für die Frage nach der Unimodalität angemessenen Test. Überraschend ist auch, daß keines der großen Statistikpakete einen solchen Test anbietet. Alles, was man findet, sind Tests, bei denen die Anpassung an eine fest vorgegebene Verteilung vorgenommen wird. Um die grundsätzliche Wichtigkeit der Thematik in das Blickfeld der empirischen Sozialwissenschaften zu rücken, beschreiben wir einen wenig bekannten Test auf Unimodalität, den sogenannten DIP-Test. Anhand von Skalenwerten erläutern wir das Vorgehen.

Index statistics play an important role in describing data. The most important such statistic in practice is the mean. It is often overlooked, however, that the interpretability of the mean depends strongly on the shape of the distribution. If the distribution is multi-peaked or U-shaped, the mean is misleading. To check the shape of the distribution by using histogram plots is tricky, because histograms can change substantially if the number of bars is slightly modified. None of the major statistics textbooks describes an inferential test for unimodality, and none of the major statistics program packages provides such a test either. What they offer are tests for particular distributions such as the normal. To direct the attention of social scientists to the importance of the unimodality issue, we here describe a little known unimodality test, the DIP test. An application is given, and practical issues are discussed.

1. Zur Bedeutsamkeit von Maßen zentraler Tendenz

Die vermutlich meistverwendeten Statistiken in der sozialwissenschaftlichen Berichterstattung sind Mittelwerte wie etwa die bekannten durchschnittlichen Popularitätswerte

für Politiker. Eine sinnvolle Interpretation des Mittelwertes unterstellt, daß dieser tatsächlich eine zentrale Tendenz beschreibt in dem Sinn, daß die Beobachtungen um diesen Wert herum klumpen. Wenn sich dagegen z.B. die Präferenzen für politische Entscheidungen auf einer Links-Rechts-Dimension polarisieren, d.h. die Verteilung U-förmig ist, dann ist der Mittelwert irreführend, weil nur ganz wenige oder überhaupt keine der Befragten die Parteien der Mitte bevorzugen, wie der Mittelwert nahelegen würde. Es ist in diesem Fall offensichtlich, daß aus der naiven Interpretation des Mittelwertes als dem mehr oder weniger "typischen" Wert grobe Fehler in der Vorhersage des Wahlscheids entstehen können.

Vielverwendet werden in den Sozialwissenschaften auch kumulierte Prozentangaben, um die Verteilung auf mehrstufigen Skalen zu charakterisieren. Berichtet wird dann etwa, daß sich 60% "zustimmend" und 40% "ablehnend" zu einem bipolaren Item geäußert haben. Auch bei diesen Statistiken tritt das Problem der zentralen Tendenz auf, weil Zustimmungs- und Ablehnungsprozente natürlicherweise als Schnittstellen einer kontinuierlichen, eingipfeligen Verteilung gedeutet werden (Thurstone 1927). Das ist im allgemeinen auch eine empirisch recht gut begründete Annahme, weil sich für Einstellungs- und Meinungsitems zeigt, daß Zustimmungs- und Ablehnungsprozente hoch mit Mittelwerten korrelieren. Wenn also z.B. 60% einem Item zustimmen und 40% dieses ablehnen, dann kann man hieraus schließen, daß die Schnittstelle zwischen Zustimmung und Ablehnung bei etwa $z=+0.26$ liegt (für eine z-standardisierte Variable). Insofern kann man derartige Prozente auch an Stelle von Mittelwerten in der Sozialberichterstattung verwenden (Borg 1989).

Die Voraussetzung für eine sinnvolle Anwendung solcher Maße ist nicht unbedingt eine Normalverteilung, sondern vor allem die Frage, ob man davon ausgehen kann, daß die beobachtete Verteilung aus einer eingipfeligen Population stammt. Ist das nicht der Fall, ist es in jedem Fall überlegenswert, ob es nicht theoretisch fruchtbarer ist, die Population in verschiedene Klassen zu partitionieren. Coombs/Avrunin (1977) haben beispielsweise eine Theorie für Präferenzverhalten entwickelt, die die Eingipfeligkeit von Präferenzfunktionen über einem Merkmalskontinuum aus zwei zugrundeliegenden Funktionen über diesem Merkmalskontinuum ableitet. Beobachtet man mehrgipfelige Präferenzverteilungen, dann sind nach dieser Theorie die Objekte des Auswahlraums grundsätzlich nicht vergleichbar, wie z.B. die Objekte "Eistee" und "heißer Tee", die eine zweigipfelige Verteilung über der Merkmalsdimension "Temperatur" erzeugen. In diesem Fall sind die Daten auch nicht sinnvoll in einem Unfolding-Modell darstellbar (Coombs 1964).

Keines der gängigen Statistik-Lehrbücher beschreibt einen für die Frage nach der Unimodalität der einer beobachteten Verteilung zugrundeliegenden Verteilung angemessenen Test. Noch wichtiger ist, daß auch keines der großen Statistikpakete (z.B. SPSS, SYSTAT, STATISTICA, SAS) einen solchen Test anbietet. Beschrieben bzw. angeboten werden nur besondere Tests, wie z.B. solche zur Prüfung auf Normalität der Verteilung. In der Praxis werden sie aber selten angewendet, wohl deshalb, weil eine zumindest approximative Normalität der Verteilung in den meisten Fällen einfach unterstellt wird. Diese Unterstellung ist sicher nicht unberechtigt. Sie kann aber im Einzelfall falsch sein und führt dann zu schwerwiegenden Fehlinterpretationen der Statistiken. Die folgende Darstellung soll deshalb auch dazu dienen, die grundsätzliche Wichtigkeit der Thematik in das Blickfeld der empirischen Sozialwissenschaften zu rücken.

2. Unimodalität bei Histogrammen

Eine beliebte grafische Darstellung von Daten sind die Histogramme. Sie sind als universelle grafische Darstellungsform für Verteilungen in der Statistik in praktisch allen Statistikprogrammen zu finden. In der Regel werden auf der horizontalen Achse Intervalle gleicher Breite verwendet und die Höhe der Balken durch die absolute oder relative Zahl der Beobachtungen im entsprechenden Intervall ermittelt. Gibt es nur wenige Ausprägungen, wie bei den Likertskalen, die in den Sozialwissenschaften weit verbreitet sind, geben die Histogramme ein genaues Abbild der Häufigkeiten wider, da eine Gruppierung der Daten nicht notwendig ist. Streuen die Daten aber innerhalb eines Intervalls, gehen wertvolle Informationen bei der Darstellung der Daten durch Histogramme verloren.

Unimodalität eines Histogramms ist dann gegeben, wenn die Höhen der Balken von links nach rechts erst bis zu einem bestimmten Punkt wachsen und dann wieder sinken. In diesem Sinne liefern die Daten des ALLBUS für viele Einstellungstems unimodale Histogramme. Da viele der bekannten diskreten Verteilungen (Binomial, Poisson) bei dieser Definition unimodal sind, kann man natürlich auch durch einen Anpassungstest überprüfen, ob die Daten aus einer solchen Verteilung stammen. Vielfach hat man aber die Vorstellung, daß in der Population eine stetige Verteilung vorliegt und die beobachteten Daten Realisierungen aus dieser Verteilung sind. Bekannte Anpassungstests prüfen dann etwa auf Normalität. Stellt man die Daten in Form von Histogrammen dar, geben sie uns nur noch eine mehr oder weniger gute Annäherung für die in der Population zugrunde liegende Verteilung. Ob Unimodalität des Histogramms zum Vorschein kommt oder nicht, ist dabei häufig nur eine Frage der Anzahl der gewählten Intervalle.

3. Eine illustrative Anwendung

Betrachten wir eine Anwendung. Borg/Braun (1992) berichten die Skalenwerte von 125 Mitarbeitern eines Gießener Textilbetriebes auf einer Skala zur Messung der subjektiven Unsicherheit der Arbeitsstelle (SUSA) (Borg 1992). Die Skala unterscheidet zwei Faktoren: 'Bedenken' und 'Sorgen'. Bedenken werden erfaßt mit sechs Items über den wahrscheinlichen Erhalt der Arbeitsstelle, der Gewißheit über die eigene Karriere, der Zuversicht, neue Technologien meistern zu können, usw. Sorgen werden gemessen mit drei Items, bei denen es um die psychische Belastung geht, die der mögliche Verlust der Arbeitsstelle bewirkt. Empirisch zeigt sich bei den 103 non-missing data für 'Sorgen' und 105 non-missing für 'Bedenken', daß Sorgen nur dann auftreten, wenn zumindest gewisse Bedenken gegeben sind. Es gibt aber Personen, die große Bedenken haben, sich aber trotzdem keine Sorgen machen.

Für die Stichprobe ergeben sich die in Abbildung 1 gezeigten Verteilungen der Skalenwerte. Die Abbildungen zeigen Histogramme mit angepaßten Normalverteilungen (fette Glockenkurven) und jeweils einer zweiten gefitteten Funktion (gestrichelte Linien), die eine bessere Anpassung erlaubt (Kernel-Funktionen). Beide Abbildungen wurden mit SYSTAT erzeugt, wobei der Streubereich der X-Achse und der maximale Y-Wert vorgegeben wurden, um bessere Vergleichbarkeit zu erreichen. Die Intervallbildung auf der X-Achse wurde dagegen dem Programm überlassen. Wie man sieht, sind beide Verteilungen nicht sehr normal, vor allem nicht die von Sorgen. Bei Sorgen sieht die Verteilung eher tendenziell zwei-modal aus, wie auch die Kernel-Funktion unterstreicht. Der Kolmogorov-Smirnov-Anpassungstest, kurz K-S Test, lehnt die Normalverteilungshypothese bei Sorgen mit $p < .05$ ab, bei Bedenken aber nicht ($p < .15$). Die letzte Aussage ist allerdings etwas irreführend, weil beim K-S Test der Erwartungswert und die Varianz der Normalverteilung festgelegt werden müssen. Daß bei Bedenken Unimodalität nicht abgelehnt wird, liegt nur daran, daß die Parameter der Verteilung durch die entsprechenden Stichprobengrößen vom Programm automatisch festgelegt wurden. Werden die Daten erst standardisiert und auf Standardnormalverteilung getestet, erhält man den entsprechenden Test mit Lilliefors-Schranken. Für ihn gilt in beiden Fällen $p < .01$. In SYSTAT werden der K-S Test und der Lilliefors Test angeboten und es liegt am Nutzer, den richtigen zu wählen.

Wählt man statt der von SYSTAT intern bestimmten 15 Intervalle für die X-Achse nur 14 solcher Intervalle, also ein geringfügig gröberes Raster, dann ergibt sich Abbildung 2.

Abbildung 1: Häufigkeitsverteilungen (Histogramme über 15 Intervallen der X-Achse) der Skalenwerte 'Bedenken' und 'Sorgen', zusammen mit angepassten Normalverteilungen und Kernel-Funktionen

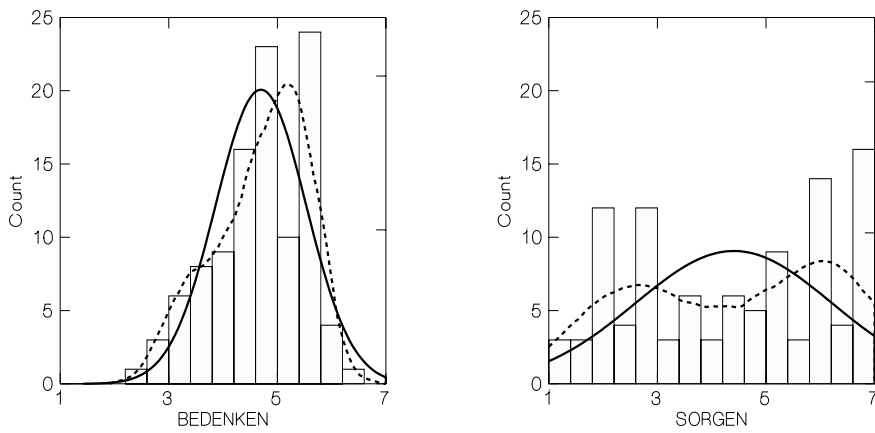
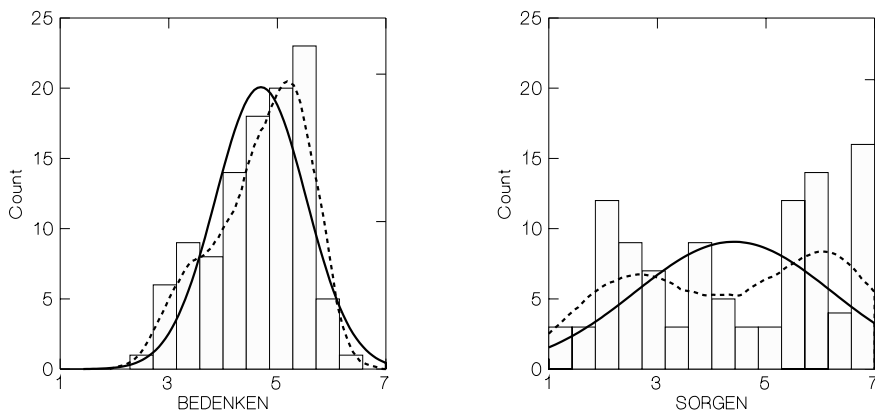


Abbildung 2: Häufigkeitsverteilungen wie in Abbildung 1, mit 14 anstatt 15 Intervallen auf der X-Achse



Man sieht, daß diese kleine Änderung zu einem deutlich anderen Bild der Verteilung der Skalenwerte für Bedenken führt: Der tiefe Einschnitt zwischen den beiden höchsten Säulen in Abbildung 1 ist nun völlig verschwunden ist. (Die gefitteten Funktionen sind dagegen exakt gleich, weil sie nicht dem Histogramm, sondern den Rohdaten angepaßt werden.) Die Programmpakete SPSS und STATISTICA zeigen (per Voreinstellung) ein noch gröberes Raster als 14 Intervalle. Dadurch wird der Eindruck noch weiter verstärkt, daß Bedenken weitgehend normal verteilt sind, während man in Abbildung 1 doch immerhin sieht, daß man auch hier Zweifel haben könnte, ob die Population normal verteilt ist. Insgesamt erkennt man durch dieses Beispiel das grundsätzliche Problem, daß eine Überprüfung der Normalität oder der Eingipfeligkeit mit Histogrammen bei Verwendung verschiedener Auswertungsprogramme zu verschiedenen Entscheidungen führen kann. Das Problem der Rasterung tritt in jedem Fall auf. Im folgenden wird ein Test zur Prüfung der Unimodalität von Verteilungen beschrieben und illustriert.

4. Unimodalität von Verteilungen

Es gibt nur wenige Tests, die die Existenz von mehreren Gipfeln in einer Verteilung überprüfen und nicht jeder dieser Tests ist hinreichend verläßlich unter Bedingungen, die man in der Praxis nicht von vornherein ausschließen kann (Hartigan/Hartigan 1985a). Robust und gleichzeitig transparent ist dagegen der DIP-Test von Hartigan/Hartigan (1985a) für eindimensionale Verteilungen¹⁾. Er basiert auf der DIP-Statistik als Prüfgröße, die die Multimodalität einer Stichprobe mißt als maximale Differenz aller Punkte der empirischen Verteilungsfunktion von der unimodalen Verteilungsfunktion, die diese maximale Differenz minimiert. Die Autoren vermuten, daß die Gleichverteilung die unimodale Verteilung ist, die am schlechtesten von den multimodalen Verteilungen unterschieden werden kann. Sie wird daher als Nullverteilung in der Nullhypothese verwendet.

Es sei F eine Verteilungsfunktion. Dann heißt F *unimodal* mit Modus m , wenn eine Zahl m existiert derart, daß F im Intervall $(-\infty, m]$ konvex und im Intervall $[m, \infty)$ konkav ist. Der Wert m ist nicht notwendig eindeutig. Eine unimodale stetige Verteilung hat also eine Dichte, die rechts vom Modalwert monoton fallend und links vom Modalwert monoton steigend ist. Stetige Verteilungen, deren Dichten logarithmisch konkav sind, sind stets unimodal. Die Cauchy Verteilung ist aber ein Beispiel für eine unimodale Verteilung, deren Dichte nicht logarithmisch konkav ist. Diskrete Verteilungen erzeugen Treppenfunktionen als Verteilungsfunktionen und sind aus diesem Grunde mit Ausnahme der Einpunktverteilung alle multimodal. Für eine ausführliche Darstellung unimodaler Verteilungen verweisen wir auf das Buch von Dharmadhikari/Joag-dev (1988).

Beispiel: Die Abbildungen 3 und 4 zeigen die Verteilungsfunktionen bzw. Dichtefunktionen einer im Intervall $(0,1)$ gleichverteilten, einer standard-normalverteilten und einer U-förmig verteilten Zufallsvariablen.

Abbildung 3: Drei Verteilungsfunktionen (kumulativ normal, rechteckig, U-förmig)

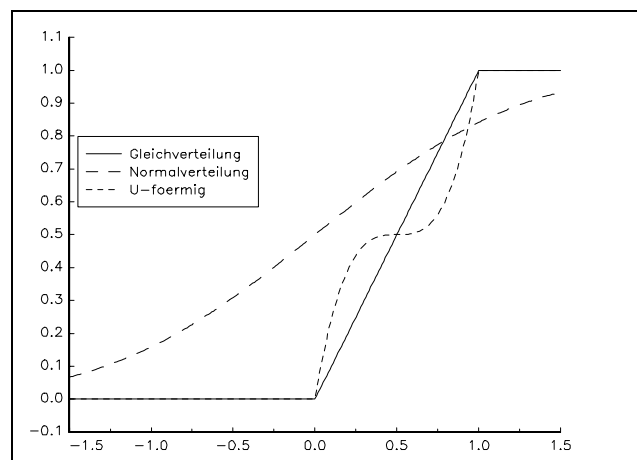
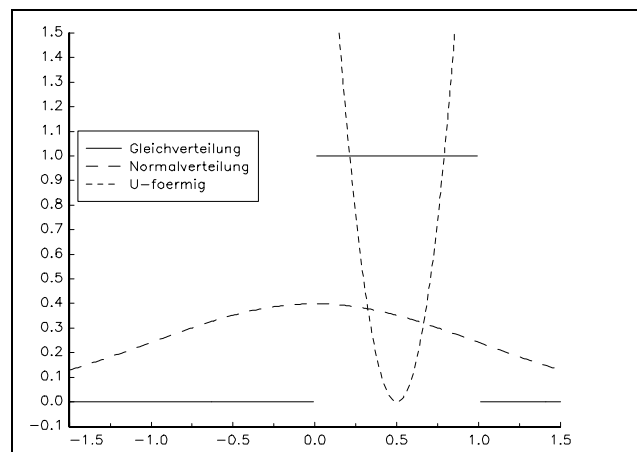


Abbildung 4: Dichtefunktionen für die Verteilungsfunktionen in Abbildung 3



Offensichtlich ist die Standard-Normalverteilung unimodal mit eindeutigem Modus 0, die Gleichverteilung unimodal, wobei für den Modus jeder Wert zwischen Null und Eins genommen werden kann, während die U-förmige Verteilung nicht unimodal, sondern multimodal ist. Beta- und Gamma-Verteilungen sind ebenfalls unimodal, wenn die Parameter größer gleich eins sind. Mischverteilungen sind weitere Beispiele, wie Multimodalität in der Praxis entstehen kann. Hat eine unimodale Verteilungsfunktion die Eigenschaft, daß die Faltung mit irgendeiner anderen unimodalen Verteilungsfunktion wieder zu einer unimodalen Verteilungsfunktion führt, so nennt man sie streng unimodal. Es läßt sich zeigen, daß streng unimodale nichtdegenerierte Verteilungsfunktionen genau solche sind, die stetig sind mit logarithmisch konkaven Dichten.

5. Der DIP-Test als Unimodalitätstest

Um den DIP-Test beschreiben zu können, brauchen wir ein Maß für den Abstand zweier Funktionen. Für zwei beschränkte Funktionen F und G sei ihr Abstand als $\rho(F,G)=\sup_x |F(x)-G(x)|$ definiert. Für eine Klasse \mathcal{A} von beschränkten Funktionen sei $\rho(F,\mathcal{A})=\inf_{G \in \mathcal{A}} \rho(F,G)$. Weiter sei \mathcal{U} die Klasse der unimodalen Verteilungsfunktionen. Der DIP einer Verteilungsfunktion F ist definiert durch $D(F)=\rho(F,\mathcal{U})$. Offensichtlich gilt $D(F)=0$ für $F \in \mathcal{U}$, $D(F)>0$ für $F \notin \mathcal{U}$ und aus der Dreiecksungleichung folgt $D(F_1) \leq D(F_2) + \rho(F_1, F_2)$. Daher mißt der DIP die Abweichung von der Unimodalität.

In der Praxis ist die theoretische Verteilung F in der Regel unbekannt. Was wir beobachten, ist eine Stichprobe X_1, X_2, \dots, X_n aus der Verteilung F . Bezeichnet F_n die empirische Verteilungsfunktion definiert durch $F_n(x) = (1/n) \sum \{X_i \leq x\}$, so folgt aus dem Glivenko-Cantelli'schen Theorem $\rho(F_n, F) \rightarrow 0$ fast überall und daher $D(F_n) \rightarrow D(F)$ fast überall. Ein Test, der auf dem DIP basiert, wird daher asymptotisch irgendeine unimodale Verteilungsfunktion von einer multimodalen unterscheiden.

- Zu testen ist: " H_0 : Verteilung ist unimodal" versus " H_1 : Verteilung ist multimodal".

Da die Nullhypothese zusammengesetzt ist, ist es bei der Entwicklung eines Tests notwendig, eine unimodale Verteilungsfunktion als Nullverteilung zu wählen. Hartigan/Hartigan (1985a) wählen die Gleichverteilung, weil sie vermuten, daß asymptotisch die DIP-Statistik stochastisch am größten für die Gleichverteilung ist. Ein DIP-Wert, der signifikant für diese Nullverteilung ist, wäre auch signifikant bezüglich jeder anderen unimodalen Verteilung. In diesem Sinn wäre die Gleichverteilung als schlechteste unimodale Verteilungsfunktion zu betrachten.

Die Berechnung des Dip-Wertes d erfolgt über einen Algorithmus, der eine unimodale Verteilungsfunktion konstruiert, die von F_n den Abstand d hat. Eine FORTRAN Subroutine wurde von Hartigan (1985b) veröffentlicht, die einen Fehler enthielt und von Sommer/McNamara (1987) korrigiert wurde.

Führen wir einen DIP-Test in den Beispielen durch, ergeben sich die in Abbildung 5 und 6 gezeigten Grafiken. Die Abbildungen zeigen die empirischen Verteilungsfunktionen nach unten und oben durch den Dip-Wert d verschoben. Die gestrichelte Linie gibt die gefittete unimodale Verteilungsfunktion an, die von der empirischen Verteilungsfunktion F_n gerade den Dip-Wert als Abstand hat. Sie besteht im Intervall $[x_1, x_L]$ aus der größten konvexen Minorante von $F_n + d$, im Intervall $[x_U, x_N]$ aus der kleinsten konkaven Majorante von $F_n - d$ und einem Liniensegment im Intervall (x_L, x_U) . Dabei ist x_1 der kleinste und x_N der größte x -Wert im Datensatz. x_L und x_U ergeben sich aus dem Algorithmus von Hartigan (1985b). In unserem Beispiel gilt jeweils $x_L = x_U$, was aber nicht immer so sein muß. Dieser Fall tritt aber insbesondere dann auf, wenn einige Ausprägungen stark dominieren, was bei Variablen mit wenigen Ausprägungen praktisch immer der Fall ist.

Abbildung 5: Verteilung 'Bedenken', nach oben und unten um den Dip-Wert d verschoben

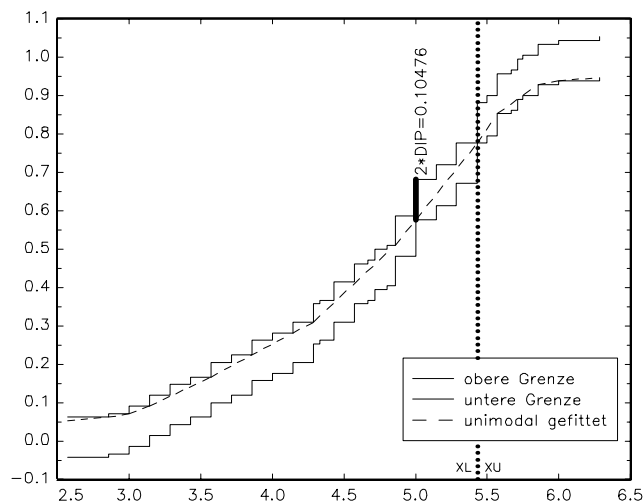
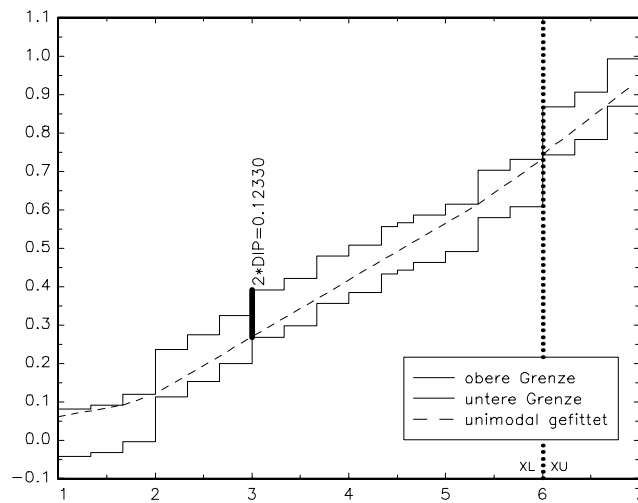


Abbildung 6: Verteilung 'Sorgen', nach oben und unten um den Dip-Wert d verschoben



Für die DIP-Statistik ergibt sich nach dem Algorithmus von Hartigan (1985b) für die Skalenwerte 'Bedenken' der Wert $d_B=0.05238$, für 'Sorgen' der Wert $d_S=0.06165$. Beide liegen nach der in Hartigan/Hartigan (1985a) und Hartigan (1985b) publizierte Tafel im Ablehnungsbereich eines 95% Konfidenzintervalls. Der Test besagt also, daß keiner der Datensätze aus eingipfligen Verteilungen stammt.

Bemerkenswert ist, daß im Falle 'Sorgen' $x_L=x_U$ ist und der berechnete Dip-Wert $d_S=0.06165$ kleiner ist als $\delta=7/103=0.067961$, wobei δ die relative Häufigkeit des Wertes x_L im Datensatz ist. Die gefittete unimodale Verteilungsfunktion ist daher an der Stelle x_L unstetig.

6. Diskussion

Zum DIP-Test ist zunächst anzumerken, daß dann, wenn die wahre unimodale Referenz-Verteilung stark von der Gleichverteilung abweicht, es andere Tests mit größerer Macht gibt (Hartigan/Hartigan1985a). Das sollte aber für den Anwender nicht von großer Bedeutung sein, weil diese wahre Verteilung im allgemeinen unbekannt sein dürfte. Man kann statt der Gleichverteilung auch die gefittete unimodale Verteilungsfunktion als

Nullverteilung verwenden. Simulationsstudien, die die Verteilung der DIP-Statistik bei Gültigkeit der neuen Nullverteilung sichtbar machen, ergaben aber in unseren Beispielen keine Änderung in der Entscheidung.

Lohnt sich der DIP-Test oder kommt man nicht in der Praxis mit dem in einigen Statistik-Paketen angebotenen Kolmogorov-Smirnov-Anpassungstest auf Normalität aus? Eine systematische Antwort läßt sich auf diese Frage zur Zeit nicht geben. Es ist aber bemerkenswert, wie das obige Anwendungsbeispiel zeigt, daß für die Variable 'Bedenken' der K-S Test die Normalverteilungshypothese beibehält, während der DIP-Test die schwächere Unimodalitätshypothese ablehnt. Die Entscheidungen fallen zwar knapp aus, aber dennoch zeigt sich der K-S Test zumindest in diesem Beispiel als der konservativere. Da beide Tests im Fall der Variablen 'Sorgen' jeweils zu einer Ablehnung der Nullhypothese führen, kann man sicher sagen, daß eine routinemäßige Anwendung des K-S Tests besser ist als die völlige Ignorierung der Unimodalitätsfrage. Dabei braucht man sich in den Programmpaketen nicht allein auf die Normalverteilung zu beschränken. Stand-alone Programme für den DIP-Test stehen aber ebenfalls zur Verfügung.

Wie häufig oder wie selten ist der Befund einer nicht-unimodalen Verteilung in der sozialwissenschaftlichen Praxis? Wir haben, um für diesen Artikel ein Anwendungsbeispiel zu finden, verschiedene Datensätze betrachtet. Dabei fanden wir z.B. für verschiedenste Meinungs- und Einstellungsitems des ALLBUS Unimodalität von Histogrammen. Eingipfeligkeit bei den Histogrammen für einzelne Items scheint also der dominierende Normalfall zu sein. Ob aber die den Stichproben zugrunde liegende Verteilung unimodal ist, läßt sich aus diesen Befunden nicht beantworten. Diese Frage wird vom DIP-Test - überraschend, wie es scheint - stets negativ beantwortet. Der Grund hierfür ist, daß im Falle der ALLBUS-Items stets nur eine Verteilung mit wenigen Skalenpunkten vorliegt. Das Kontinuum der X-Achse ist damit auf wenige Punkte "diskretisiert", d.h. die Werte klumpen, und der DIP-Test findet, ganz zu Recht, daß jeweils eine mehrgipfelige (in der Tat: eine viel-gipfelige) Verteilung angenommen werden muß. Im Fall der obigen Testvariablen 'Sorgen' und 'Bedenken' hatten wir dagegen einen anderen Fall, weil diese Variablen summative Werte aus mehreren Items darstellen, die viel feiner abgestuft sind. Die Frage, wie viele Abstufungen einer Variablen ausreichend sind, damit eine Stichprobenverteilung ausreichend viele Abstufungen enthält für eine sinnvolle Anwendung des DIP-Tests, ist zur Zeit noch offen. Eine denkbare praktische Lösung dieses Problems wäre vielleicht, die Verteilung zwischen den Skalenwerten einer grob abgestuften Skala zu interpolieren und dann hieraus eine Zufallsverteilung zu ziehen. In der Praxis sollte man aber, bevor solche Fragen mit einiger Sicherheit beantwortet werden können, bei

Variablen mit nur wenigen Abstufungen eher auf den χ^2 -Anpassungstest, den K-S oder den Lilliefors Test zurückgreifen.

Anmerkung

- 1) Eine Erweiterung auf mehr Dimensionen ist der SPAN-Test, beschrieben in Hartigan (1988).

Literatur

Borg, I., 1989: Zur Präsentation von Umfrageergebnissen. Zeitschrift für Arbeits- und Organisationspsychologie 33: 90-95.

Borg, I., 1992: Überlegungen und Untersuchungen zur Messung der subjektiven Unsicherheit der Arbeitsstelle. Zeitschrift für Arbeits- und Organisationspsychologie 36: 107-116.

Borg, I./Braun, M., 1992: Arbeitsethik und Arbeitsinvolvement als Moderatoren der psychologischen Auswirkungen von Arbeitsplatzunsicherheit. Zeitschrift für Arbeits- und Organisationspsychologie 36: 167-176.

Coombs, C. H., 1964: A theory of data. New York: Wiley.

Coombs, C. H./Avrunin, G., 1977: Single-peaked functions and the theory of preference. Psychological Review 84: 216-230.

Dharmadhikari, S./Joag-dev, K., 1988: Unimodality, Convexity, and Applications. Boston: Academic Press.

Hartigan, J.A./Hartigan, P.M., 1985a: The DIP-Test of unimodality. The Annals of Statistics 13: 70-84.

Hartigan, P.M., 1985b: Computation of the DIP-Statistic to test for unimodality. Applied Statistics 34: 320-325.

Hartigan, J.A., 1988: The SPAN test for unimodality. In H. H. Bock (Hrsg.) Classification and Related Methods of Data Analysis. Elsevier Science Publishers B.V. (North Holland).

Sommer, C.J./McNamara, J.N., 1987: Power considerations for the DIP-Test of unimodality using mixtures of normal and uniform distributions. American Statistical Association; Proceedings of the Statistical Computing Section: 186-191.

Thurstone, L.L., 1927: A law of comparative judgment. Psychological Review 34: 273-286.

GASTARBEITER ODER AUSLÄNDER? ERGEBNISSE DES SPLITS MIT DEN REFORMULIERTEN GASTARBEITERFRAGEN IM ALLBUS 1994¹⁾

THOMAS BLANK UND MARTINA WASMER

Die von Blank/Schwarzer (1994) vorgeschlagene Reformulierung der Gastarbeiter-Skala aus dem ALLBUS sowie die klassischen Gastarbeiter-Items wurden in einem ALLBUS-Split 1994 erhoben. Der Aufsatz beschäftigt sich mit der Analyse der Äquivalenz beider Formulierungen in bezug auf interne und externe Gültigkeit. Diese wird durch konfirmatorische Faktorenanalysen und multiple Gruppenvergleiche nachgewiesen. Die Befunde von Blank/Schwarzer werden somit an einer repräsentativen Studie für Gesamtdeutschland bestätigt.

The revision of the foreign-worker-scale of the ALLBUS suggested by Blank/Schwarzer (1994) and the origin foreign-worker-scale were collected in an ALLBUS-split 1994. This paper is engaged with the analysis of the equivalence of the question wording effects with regard to internal and external validity. These will be demonstrated by confirmatory factor analyses and the multiple group comparison-option of LISREL. The results of Blank/Schwarzer are confirmed in a representative study for whole Germany.

1. Einleitung und Übersicht

Die sogenannten Gastarbeiterfragen der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) - eine Skala mit vier Items, in denen die Diskriminierungsbereitschaft gegenüber Gastarbeitern erhoben wird, und vier Fragen nach Kontakten mit Gastarbeitern in unterschiedlichen Lebensbereichen (vgl. Übersicht 1) - werden regelmäßig repliziert und von den Nutzern des ALLBUS häufig ausgewertet. Allerdings wurden in den letzten Jahren mitunter Bedenken gegenüber dem in den Fragen verwendeten Gastarbeiterbegriff geäußert (vgl. z.B. Blank/Schwarzer 1994; Koch et al. 1994; Kühnel/

Terwey 1994). Diese Kritik, die hier noch einmal kurz zusammengefaßt werden soll, bezieht sich *auf sprachlicher Ebene* zum einen auf die Begriffskomponenten 'Gast' und 'Arbeiter'. Diese Begriffsverengung auf bestimmte Statuspositionen und Personen mit nur vorübergehender Aufenthaltsdauer impliziert bereits eine gewisse Diskriminierung. Zum anderen wird darauf hingewiesen, daß der Begriff 'Gastarbeiter' insbesondere in den Medien, aber auch in der Alltagssprache zunehmend ungebräuchlicher wird. Dies hängt sicherlich mit den Veränderungen *auf gesellschaftlicher Ebene* zusammen, die, so die zweite Hauptargumentationsrichtung der Kritik, dazu geführt haben, daß mit dem Stimulus 'Gastarbeiter' nicht mehr, wie zum Zeitpunkt der Entwicklung der ALLBUS-Skala, nach der prototypischen Zielgruppe fremdenfeindlicher Einstellungen gefragt wird. Andere Fremdgruppen, wie z.B. Asylbewerber und Kriegsflüchtlinge, sind hinzugekommen und heute eher Zielobjekt der Fremdenfeindlichkeit, so daß die Stärke genereller diskriminierender Einstellungen gegenüber Ausländern, die mit der ALLBUS-Skala gemessen werden sollte, unter Umständen nicht mehr angemessen mit Items, die den Gastarbeiterbegriff verwenden, abgebildet werden kann.

In drei regionalen Studien ersetzten Blank und Schwarzer (1994) in den ALLBUS-Einstellungstests den Begriff 'Gastarbeiter' deshalb durch den Begriff 'in Deutschland lebende Ausländer' und fanden, daß die so reformulierte Skala ein sehr reliables und valides Instrument zur Messung einer allgemeinen Diskriminierungstendenz gegenüber Fremdgruppen darstellt. Dieses Instrument soll in Zukunft auch im ALLBUS eingesetzt werden. Um einen die Zeitreihe nicht gefährdenden Umstieg auf die reformulierte Fassung zu erleichtern, wurden die Fragen im ALLBUS 1994 in zwei Splitversionen erhoben. Der einen Hälfte der Befragten wurden die Einstellungs- und Kontaktfragen in der alten Form gestellt, der anderen Hälfte mit der Formulierung 'in Deutschland lebende Ausländer'. Durch den Vergleich dieser Splits wollen wir hier versuchen, die Effekte dieser Formulierungsänderung zu identifizieren und zu einem Urteil in Hinblick auf die Äquivalenz der beiden Messungen zu gelangen.

Im nächsten Abschnitt beschreiben wir zunächst die Stichprobe und die Meßinstrumente. Dabei soll untersucht werden, ob die Variation des Fragetextes in den beiden Splitversionen zu unterschiedlichen deskriptiven Ergebnissen auf der Ebene der Einzelitems geführt hat. Die Analyse der internen und externen Validität der vier Einstellungstests mittels exploratorischer bzw. konfirmatorischer Faktorenanalysen ist Gegenstand der folgenden zwei Abschnitte. In Abschnitt 5 fassen wir unsere Ergebnisse zusammen und diskutieren sie in ihrer Bedeutung für die Dauerbeobachtung der Fremdgruppendifferenzierung durch den ALLBUS.

2. Stichprobe und Meßinstrumente

2.1 Stichproben

Die 3450 Befragten des ALLBUS 1994 (2342 in Westdeutschland und 1108 in Ostdeutschland) stellen eine Zufallsstichprobe aus der erwachsenen Bevölkerung in Privathaushalten in West- und Ostdeutschland dar. Detaillierte Informationen zur Stichprobenbildung - 1994 erstmals auf der Grundlage von Einwohnermeldeamtsdaten - enthält der Methodenbericht zum ALLBUS 1994 (Koch et al. 1994). 156 Befragte besaßen nicht die deutsche Staatsbürgerschaft und bekamen die hier zur Diskussion stehenden Fragen nicht gestellt.

Im folgenden werden wir bei unseren Analysen vier Substichproben unterscheiden: Westdeutsche Befragte des Splits 1 (Gastarbeiterformulierung), Westdeutsche Befragte des Splits 2 (Ausländerformulierung), Ostdeutsche Befragte des Splits 1 (Gastarbeiterformulierung) und Ostdeutsche Befragte des Splits 2 (Ausländerformulierung).

Die Differenzierung nach dem Erhebungsgebiet ist dabei aus zwei Gründen notwendig. Zum einen sind aufgrund der disproportionalen Schichtung der Gesamtstichprobe die Ostdeutschen im ALLBUS überrepräsentiert. Zum anderen bestehen deutliche Ost-West-Unterschiede bezüglich der ideologisch-historischen und aktuellen Bedingungen. So war der Ausländeranteil an der Gesamtbevölkerung in der ehemaligen DDR wesentlich geringer als in der alten BRD und ist es auch heute noch (lt. Statistischem Bundesamt Ende 1993 1,5 Prozent in den neuen Ländern ohne Berlin gegenüber 9,8 Prozent in den alten Ländern). Ein weiteres Argument für nach dem Erhebungsgebiet differenzierte Analysen gründet sich auf die spezifischen Lebenslagen in den beiden Teilen Deutschlands. Es ist zu erwarten, daß die höhere Arbeitslosigkeit im Osten und andere statusrelevante Unterschiede die Wahrnehmung von Fremdgruppen, etwa als Konkurrenten um knappe Ressourcen, beeinflussen.

Unterschiede *innerhalb* der beiden Erhebungsgebiete zwischen den Befragten der beiden Splits sollten angesichts der zufälligen Zuordnung der Fragebogenversionen nicht auftreten. Geringfügige - Abweichungen sind zwar beim (Nicht-)Erwerbsstatus der Befragten zu beobachten. Hinsichtlich der anderen untersuchten zentralen demographischen Variablen 'Geschlecht', 'Alter', 'Allgemeinbildender Schulabschluß' und 'Familienstand' unterscheiden sich die Befragten der beiden Splits jedoch weder im Westen noch im Osten voneinander, so daß man im großen und ganzen von in der demographischen Struktur vergleichbaren Stichproben ausgehen kann.

Relativ deutliche Unterschiede zwischen den Splits treten bei den Interviewermerkmalen auf.²⁾ In den alten Bundesländern wurden die Interviews mit den reformulierten Fragen signifikant häufiger als die des Splits 1 von männlichen, jüngeren und gebildeteren Interviewern durchgeführt. In den neuen Bundesländern sieht es hinsichtlich des Geschlechts und der Bildung der Interviewer genauso aus, allerdings waren hier ältere Interviewer im Split 2 stärker vertreten.

2.2 Meßinstrumente

Die ALLBUS-Gastarbeiterfragen bzw. -Ausländerfragen umfassen zum einen vier Items, mit denen diskriminierende *Einstellungen* gemessen werden sollen. Ob es angemessener ist, diese Items als Operationalisierungen einzelner spezifischer Meinungen oder als multiple Indikatoren einer generellen Einstellung aufzufassen, ist eine weder theoretisch noch empirisch eindeutig zu klärende Frage, wie die am Beispiel dieser Fragen grundsätzlich geführte Diskussion zwischen Jagodzinski et al. und Saris et al. über *true-score-models* vs. *factor-models* zeigt (Jagodzinski et al. 1987; Saris/van den Putte 1988; Jagodzinski et al. 1988; Jagodzinski et al. 1990; Saris/Hartman 1990). Wir werden im folgenden dieser Frage nicht näher nachgehen, jedoch insofern versuchen, beiden Sichtweisen gerecht zu werden, als wir sowohl die Auswirkungen der Formulierungsänderung auf die Antwortverteilungen der einzelnen Items als auch die Auswirkungen auf die Skalenqualität bei Unterstellung eines Faktorenmodells untersuchen werden.

Das erste Item beinhaltet eine eher allgemeine und schwache („ein bißchen“) Forderung nach Anpassung des Lebensstils, die anderen drei beziehen sich auf Diskriminierungen in verschiedenen sozialen Handlungsfeldern, im Arbeitsbereich, im Bereich der politischen Betätigung und im privaten Bereich. Das Ausmaß der Zustimmung zu den vier Aussagen wird mit Hilfe einer siebenstufigen Skala mit benannten Endpunkten erhoben. Die Skala ist nicht balanciert, bei jedem Item indiziert ein höheres Maß an Zustimmung stärkere Diskriminierung. Zum anderen wird im ALLBUS nach dem Bestehen von *Kontakten* mit Gastarbeitern und deren Familien gefragt, wobei zwischen verschiedenen Kontexten - Familie/Verwandschaft, Arbeitsplatz, Nachbarschaft, Freunde/Bekannte - differenziert wird. Da lediglich erhoben wird, ob in diesen Bereichen Kontakte mit Gastarbeitern bzw. Ausländern bestehen oder nicht, sind Rückschlüsse auf die Intensität der Kontakte und das Ausmaß ihrer Freiwilligkeit nur begrenzt möglich, indem man etwa im Freundeskreis eher Freiwilligkeit der Kontakte, in der Nachbarschaft eher geringere Intensität usw. unterstellt.

In Übersicht 1 sind die Fragen zusammengestellt. Um auf der rein deskriptiven Ebene einen Eindruck von der Äquivalenz der Messungen zu gewinnen, werden im folgenden

zunächst die Antwortverteilungen dieser acht Fragen in den verschiedenen Substichproben dargestellt.

Übersicht 1: Die Meßinstrumente in ihrer ursprünglichen und reformulierten Fassung

Einstellungsitens

LEBEN	<i>Gastarbeiter/Die in Deutschland lebenden Ausländer</i> sollten ihren Lebensstil ein bißchen besser an den der Deutschen anpassen
ARBEIT	Wenn Arbeitsplätze knapp werden, sollte man die <i>Gastarbeiter/ in Deutschland lebenden Ausländer</i> wieder in ihre Heimat zurückschicken.
POLITIK	Man sollte <i>Gastarbeitern/den in Deutschland lebenden Ausländern</i> jede politische Betätigung in Deutschland untersagen.
HEIRAT	<i>Gastarbeiter/Die in Deutschland lebenden Ausländer</i> sollten sich ihre Ehepartner unter ihren eigenen Landsleuten auswählen.
Antwortkategorien: 7-stufige Skala mit benannten Endpunkten (1=„stimme überhaupt nicht zu“, 7=„stimme voll und ganz zu“)	

Kontaktitens

Haben Sie persönlich Kontakte zu <i>Gastarbeitern oder zu deren Familien/in Deutschland lebenden Ausländern</i> , und zwar ...	
FAMILIE	In Ihrer eigenen Familie oder näheren Verwandtschaft?
ARBEITSPLATZ	An Ihrem Arbeitsplatz?
NACHBARSCHAFT	In Ihrer Nachbarschaft?
FREUNDESKREIS	In Ihrem sonstigen Freundes- und Bekanntenkreis?
Antwortkategorien: ja/nein	

2.2.1 Die Einstellungsitens der ALLBUS-Skala

In Tabelle 1 sind die Mittelwerte der vier Einstellungsitens für alle vier Substichproben aufgeführt. Die Rangfolge der Items hinsichtlich der Höhe der durchschnittlichen Zustimmung der Befragten unterscheidet sich zwar zwischen West und Ost - im Gegensatz zum Westen wird der Forderung nach Remigration bei knapper Arbeit im Osten stärker zugestimmt als der Forderung nach einem Verbot politischer Betätigungen, was mit der dort besonders angespannten Arbeitsmarktlage zusammenhängen mag. Unterschiede in der Rangfolge der Items zwischen den Splitversionen treten jedoch nicht auf. Im Westen sind die Unterschiede zwischen den Mittelwerten der beiden Splits bei allen Items nur

gering und in keinem Fall signifikant. Im Osten treten etwas deutlichere Unterschiede auf, die jedoch auch nur bei einem Item statistisch signifikant sind. Der Forderung nach dem Wiederheimschicken bei knapper Arbeit wird stärker zugestimmt, wenn sie sich in der Frageformulierung auf Gastarbeiter bezieht, was angesichts der begrifflichen Nähe von Item und Stimulus nicht verwundert.

Tabelle 1: Arithmetische Mittelwerte der Fremdgruppendifferenzierung und Mittelwertdifferenzen zwischen den Splitversionen nach Erhebungsgebiet

	Westen		Osten	
	Gastarbeiter (n=1058)	Ausländer (n=1095)	Gastarbeiter (n=543)	Ausländer (n=537)
Leben	4,43*	4,34	4,40	4,61
	+0,09		-0,21	
Arbeit	3,06*	2,95	3,73	3,46
	+0,11		+0,27 [#]	
Politik	3,50	3,57	3,26	3,37*
	-0,07		-0,11	
Heirat	2,66*	2,55	3,00	3,21
	+0,11		-0,21	

= auf 5 %-Niveau signifikante Mittelwertdifferenz (t-Test).

* = auf 5 %-Niveau signifikant größere Varianz (Levene's Test).

Mit einem Sternchen sind in Tabelle 1 die Mittelwerte derjenigen Splitversion versehen, bei der das jeweilige Item eine signifikant größere Varianz aufwies. Im Westen ist dies im allgemeinen bei der Gastarbeiterformulierung der Fall. Betrachtet man die Antwortverteilungen, so sieht man, daß bei allen drei Items mit signifikanten Varianzunterschieden - Leben, Arbeit, Heirat - die Extremkategorien im Gastarbeitersplit etwas stärker besetzt sind, vor allem die Kategorie '7' der stärksten Zustimmung. Der Stimulus 'Gastarbeiter' ist also im Westen eher als der allgemeinere Ausländerbegriff in der Lage, extreme Urteile hervorzurufen.

Für den Osten, wo die direkte Konfrontation mit einer nennenswerten Gastarbeiterpopulation im Alltag fehlt, gilt dies nicht. Hier zeigt sich lediglich beim Politikitem ein signifikanter Varianzunterschied, im Gegensatz zum Westen mit der größeren Streuung im Ausländersplit. Dieses Item fällt auch im Westen insofern auf, als hier *nicht* der oben für die anderen Items beschriebene Varianzunterschied auftritt. Hier scheint also ein item-spezifischer Effekt entgegenzuwirken.

Nach den Ergebnissen von Tabelle 1 scheint die Verwendung der reformulierten Einstellungsfragen kaum zu anderen Ergebnissen als die alte Fassung zu führen. Die unterschiedlichen Ergebnisse in West und Ost machen dabei deutlich, daß die Begriffsvariation nicht für alle Bevölkerungsgruppen dieselbe Bedeutung hat. Da wir auch bei Personen mit unterschiedlicher formaler Bildung und unterschiedlichem Alter annehmen, daß sie sich in den Einstellungen zu den jeweils subsumierten Gruppen und eventuell auch im Verständnis der beiden verwendeten Begriffe unterscheiden, wollen wir noch einen kurzen Blick auf die Mittelwertdifferenzen in entsprechenden Subgruppen werfen.

Tabelle 2 zeigt die Mittelwertunterschiede der vier Einstellungsitems für Personen mit niedriger formaler Bildung, das heißt mit höchstens Hauptschulabschluß, und Befragte mit höherer formaler Bildung. An dieser Stelle interessieren uns dabei lediglich die auftretenden Unterschiede zwischen den Splits. Die Frage, inwiefern durch die Begriffsvariation in den Splits die deutlich erkennbare Bildungabhängigkeit der Fremdenfeindlichkeit beeinflusst wird, wird später Gegenstand des Abschnitts zur externen Validierung sein.

Im Westen treten bei den höheren Bildungsgruppen keine signifikanten Mittelwertdifferenzen zwischen beiden Skalen auf, während die Personen mit geringer formaler Bildung dazu zu tendieren scheinen, den diskriminierenden Aussagen in der Gastarbeiterformulierung stärker zuzustimmen. Signifikant ist diese Tendenz allerdings nur beim Heiratsitem.

Auch im Osten sind im allgemeinen vor allem in der Subgruppe der Personen mit niedriger Bildung zum Teil recht deutliche Unterschiede zwischen den Splits zu verzeichnen, allerdings in entgegengesetzter Richtung. Hier wird dem Heirats- und dem Lebensstilitem in der Ausländerformulierung signifikant stärker zugestimmt. Diese bei Niedriggebildeten im Osten zu beobachtende Tendenz, eher in Bezug auf Ausländer diskriminierende Einstellungen zu äußern, ist bei den höhergebildeten Ostdeutschen nicht zu finden. Sie weisen vielmehr beim Arbeitsitem in Split 1 einen signifikant höheren Mittelwert auf.

Das generelle Ergebnis, daß die Begriffsvariation in den Splits sich vor allem bei Personen mit niedrigerer formaler Bildung auswirkt und dies in Ost und West unterschiedlich, ist schwer interpretierbar. Es könnte mit einem eventuell weniger stark abstrahierenden Antwortverhalten dieser Gruppe und der im Osten verbreiteten Verengung des Gastarbeiterbegriffs auf die westdeutsche Ausländerpopulation zusammenhängen.

Tabelle 2: Arithmetische Mittelwerte der Fremdgruppendifferenzierung und Mittelwertdifferenzen zwischen den Splitversionen nach Erhebungsgebiet und Bildung

	Westen				Osten			
	niedrigere Bildung		höhere Bildung		niedrigere Bildung		höhere Bildung	
	Gastarb. (n=554)	Ausländ. (n=592)	Gastarb. (n=490)	Ausländ. (n=487)	Gastarb. (n=235)	Ausländ. (n=219)	Gastarb. (n=299)	Ausländ. (n=311)
Leben	4,89	4,72	3,92*	3,91	4,66	5,05	4,20	4,30
	+0,17		+0,01		-0,39 [#]		-0,10	
Arbeit	3,54	3,37	2,50	2,46	3,95	3,92	3,55	3,15
	+0,17		+0,04		+0,03		+0,40 [#]	
Politik	4,01	3,98	2,94	3,08	3,50	3,79*	3,08	3,07
	+0,03		-0,14		-0,29		+0,01	
Heirat	3,27	3,02	1,98	2,00	3,56	4,06	2,55	2,60
	+0,25 [#]		-0,02		-0,50 [#]		-0,05	

= auf 5 %-Niveau signifikante Mittelwertdifferenz (t-Test).

* = auf 5 %-Niveau signifikante größere Varianz (Levene's Test).

Auch wenn man verschiedene Altersgruppen betrachtet, findet man in einigen Fällen signifikante Mittelwertunterschiede, allerdings so vereinzelt, daß sie hier nicht in einer Tabelle dokumentiert sind. Wiederum fallen die Effekte in Ost und West entgegengesetzt aus und treten zudem auch in verschiedenen Altersgruppen auf (im Westen bei den Personen ab 45, im Osten bei den Jüngeren). Während im Westen die älteren Befragten den Forderungen nach Lebensstilanpassung und Endogamie stärker zustimmen, wenn in den Aussagen der Begriff 'Gastarbeiter' verwendet wird, ist es im Osten die Altersgruppe der unter 30-jährigen, die der Forderung nach Anpassung des Lebensstils deutlich stärker zustimmt, wenn nach 'in Deutschland lebenden Ausländern' gefragt wird. Im Osten ebenfalls signifikant ist die Mittelwertdifferenz beim Arbeitsitem in der Altersgruppe der 30- bis 44jährigen. Diese voll im Erwerbsleben stehende und damit gerade im Osten stark mit dem Problem der Arbeitslosigkeit konfrontierte Altersgruppe stimmt der Forderung nach dem Heimschicken bei knapper Arbeit in der Gastarbeiterformulierung signifikant stärker zu.

Wir wollen hier nicht im einzelnen über die jeweiligen Ursachen der Mittelwertdifferenzen zwischen den Splits spekulieren. Als inhaltliche Gründe kommen zunächst Unterschiede in der Bedeutung der verwendeten Begriffe in Betracht, z.B. was die unter den jeweiligen Begriff subsumierten Fremdgruppen angeht. Dabei können diese Unterschiede durchaus zwischen verschiedenen Befragten variieren. So nehmen viel-

leicht die Ostdeutschen den Begriff 'Gastarbeiter' aus Mangel an konkreter Erfahrung wörtlicher als die Westdeutschen und verbinden damit auch stärker die Vorstellung eines nur vorübergehenden Aufenthalts. Und auch wenn das Begriffsverständnis weitgehend das gleiche sein sollte, kann die Reaktion verschiedener Bevölkerungsgruppen auf die Formulierungsänderung dennoch unterschiedlich ausfallen, wenn eben die Einstellungen gegenüber den je subsumierten Fremdgruppen subgruppenspezifisch sind. So ist es zum Beispiel denkbar, daß zwar alle Befragten unter 'Gastarbeitern' im Gegensatz zu 'Ausländern' eher statusniedrige, erwerbstätige Personen verstehen, daß dieser Statusaspekt des Begriffs aber nur die Einstellungen der besonders stark in Konkurrenz zu den ausländischen Arbeitnehmern stehenden Befragtengruppen beeinflußt. Daneben könnten auch Interviewereffekte aufgrund der oben erwähnten Verzerrungen der Stichproben bezüglich der Interviewermerkmale eine Rolle spielen.

Auf jeden Fall muß festgehalten werden, daß die Reformulierung der Skala in den einzelnen Subgruppen zu in Richtung und Intensität unterschiedlichen Effekten führt. Besonders sensibel scheinen die jüngeren und weniger gebildeten Ostdeutschen auf die Variation der Begriffe zu reagieren. Man gewinnt den Eindruck, daß zumindest für diese Gruppen der Stimulus 'Gastarbeiter', der in der ehemaligen DDR nur als Bezeichnung für die ausländischen Arbeitnehmer im Westen gebräuchlich war, nicht geeignet ist, um fremdenfeindliche Tendenzen im allgemeinen zu messen.

2.2.2 Die Kontaktfragen des ALLBUS

In Tabelle 3 sind für die beiden Splitversionen die Anteile derjenigen, die berichten, im jeweiligen Kontext Kontakt zu Gastarbeitern bzw. in Deutschland lebenden Ausländern zu haben, aufgeführt sowie die jeweiligen Prozentsatzdifferenzen zwischen den Splits. Im Gegensatz zu den Einstellungssitem fallen hier die Unterschiede zwischen den Splitversionen recht deutlich aus. Im Westen ist die Prozentsatzdifferenz zwischen den beiden Splitversionen am höchsten bei der Frage nach Kontakten in der Nachbarschaft, am niedrigsten bei der Frage nach Kontakten am Arbeitsplatz - eventuell ein Hinweis darauf, daß bei der alten Gastarbeiterfrage mitunter der Zusatz „und deren Familien“ nicht beachtet wurde. Im Osten sind die Unterschiede im Anteil derjenigen, die von Kontakten berichten, in Anbetracht der generell aufgrund der geringen Ausländerdichte sehr niedrigen Werte ebenfalls bemerkenswert. Im Gegensatz zum Westen fallen sie hier allerdings beim Arbeitsitem am stärksten und beim Nachbarschaftsitem am geringsten aus. Die Absolutzahlen sind jedoch so gering, daß wir auf eine Interpretation dieser Ergebnisse verzichten.

Tabelle 3: Prozentsatz der Befragten mit Kontakten zu Fremdgruppen und Prozentsatzdifferenzen zwischen den Splitversionen nach Erhebungsgebiet und Bildung

Westen	insgesamt		niedrigere Bildung		höhere Bildung	
	Gastarb. (n=1050)	Ausländ. (n=1067)	Gastarb. (n=551)	Ausländ. (n=576)	Gastarb. (n=483)	Ausländ. (n=476)
Familie	12,4%	20,3%	11,0%	15,9%	14,1%	25,4%
	+7,9 [#]		+4,9 [#]		+11,3 [#]	
Arbeitsplatz	39,0%	44,6%	36,8%	37,8%	41,7%	52,9%
	+5,6 [#]		+1,0		+11,2 [#]	
Nachbarschaft	25,8%	37,6%	25,3%	35,5%	26,1%	39,9%
	+11,8 [#]		+10,2 [#]		+13,8 [#]	
Freundeskreis	37,9%	47,9%	29,3%	39,5%	47,3%	57,5%
	+10,0 [#]		+10,2 [#]		+10,2 [#]	

Osten	insgesamt		niedrigere Bildung		höhere Bildung	
	Gastarb. (n=538)	Ausländ. (n=521)	Gastarb. (n=232)	Ausländ. (n=211)	Gastarb. (n=297)	Ausländ. (n=301)
Familie	2,7%	6,5%	1,7%	3,2%	3,7%	8,9%
	+3,8 [#]		+1,5		+5,2 [#]	
Arbeitsplatz	10,0%	15,4%	6,9%	4,7%	12,5%	22,4%
	+5,4 [#]		-2,2		+9,9 [#]	
Nachbarschaft	3,1%	5,7%	2,1%	3,2%	4,0%	7,6%
	+2,6 [#]		+1,1		+3,6	
Freundeskreis	10,5%	15,4%	4,2%	4,6%	15,3%	22,9%
	+4,9 [#]		+0,4		+7,6 [#]	

= auf 5 %-Niveau signifikante Mittelwertdifferenz (Chi-Quadrat).

Neben den Ergebnissen für die Gesamtstichprobe enthält Tabelle 3 auch die Ergebnisse für die bereits bei den Einstellungstems unterschiedenen Bildungsgruppen. Wir interessieren uns auch hier nur für die Prozentsatzdifferenzen zwischen den Splits, nicht für die generelle Bildungsabhängigkeit der Kontakthäufigkeiten. Die einzige nicht-signifikante Prozentsatzdifferenz im Westen ist die zwischen den berichteten Kontakten am Arbeitsplatz bei Personen mit niedriger Bildung. Bei diesen Personen bewirkt die Formulierungsvariation auch beim Familienitem relativ wenig. Im Osten berichten lediglich die Höhergebildeten eher im Ausländersplit Kontakte.

Was die Interpretation dieser Ergebnisse angeht, so deuten die recht deutlichen Differenzen zwischen den Splitversionen darauf hin, daß der Ausländerbegriff der inhaltlich weitere ist und zusätzliche Personengruppen umfaßt. Wenn man die Bildungsvariable hier als Statusindikator auffaßt und davon ausgeht, daß im allgemeinen - je nach Kontext unterschiedlich stark - statushomogene Beziehungen dominieren, ergeben sich außerdem bei den Subgruppenergebnissen wiederum einige Hinweise auf den bereits erwähnten Statusaspekt des Gastarbeiterbegriffs.

Bei der Konstruktion der ALLBUS-Gastarbeiterfragen waren die Kontaktfragen nicht zuletzt als mögliche Determinanten der Einstellungen gegenüber Fremdgruppen konzipiert worden (vgl. Krauth/Porst 1984). Das deskriptive Ergebnis, daß die Ersetzung des Gastarbeiterbegriffs durch die Formulierung 'in Deutschland lebende Ausländer' zwar bei den Kontakten deutliche Effekte hat, sich aber auf die berichteten Einstellungen kaum auswirkt, wirft die Frage auf, wie sich die Formulierungsänderung auf die Stärke des Zusammenhangs zwischen Kontakten und Einstellungen auswirkt. Wir wollen hier keine bestimmte Richtung des Zusammenhangs postulieren, sondern die Möglichkeit des Vorliegens von Wechselwirkungen zwischen Kontakten und Einstellungen (die z.B. auch Reuband 1989 betont) miteinbeziehen. Ein entsprechendes 'Feedback'-Modell, in dem davon ausgegangen wird, daß die Einstellung zu Gastarbeitern und der Kontakt mit Gastarbeitern im Freundeskreis interdependent verknüpft sind, haben Pfeifer und Schmidt mit ALLBUS-Daten quantitativ geschätzt und erhielten eine gute Modellanpassung (vgl. Pfeifer/Schmidt 1987: 82ff.).

Betrachtet man die in Tabelle 4 zusammengestellten Korrelationskoeffizienten zwischen den Kontakt- und den Einstellungsvariablen in den vier Substichproben, so zeigt sich, daß der theoretisch postulierte Zusammenhang in der reformulierten Fassung der Skala deutlicher ausfällt. Ausnahmen bilden im Westen einige Zusammenhänge der Kontaktfragen mit dem Lebensstilitem sowie die Korrelation zwischen Kontakten im Freundeskreis und dem Arbeitsitem, die im Gastarbeitersplit minimal stärker sind. Im Osten ist dies nur bei den Korrelationen zwischen Kontakten im Freundeskreis und den Lebensstil- und Politikitemen der Fall, während ansonsten die meisten Korrelationen überhaupt nur in der Ausländerformulierung die Signifikanzgrenze überschreiten, was teilweise sicherlich eine Folge der extrem schiefen Verteilung der Kontaktfragen im Osten - insbesondere im Gastarbeitersplit - ist. Insgesamt spricht dieses Ergebnis für eine eher höhere externe Validität der ALLBUS-Skala in der reformulierten Fassung. Im übernächsten Abschnitt wird die Frage der externen Validität der beiden Skalen mit Hilfe von weiteren externen Kriterien wiederaufgegriffen werden.

Tabelle 4: Signifikante Korrelationen (Kendall's tau-c) zwischen Einstellungs- und Kontaktitems in den vier Substichproben

Einstellungen		Kontakte			
		Familie	Arbeitsplatz	Nachbarschaft	Freundeskreis
Leben	West (Gastarbeiter)	.08	.15	.10	.23
	West (Ausländer)	.05	.18	.09	.19
	Ost (Gastarbeiter)				.14
	Ost (Ausländer)	.06			.13
Arbeit	West (Gastarbeiter)	.06	.13	.08	.28
	West (Ausländer)	.11	.15	.11	.27
	Ost (Gastarbeiter)				.12
	Ost (Ausländer)	.07	.10		.19
Politik	West (Gastarbeiter)	.05	.12	.06	.26
	West (Ausländer)	.14	.21	.13	.28
	Ost (Gastarbeiter)				.17
	Ost (Ausländer)	.06	.09		.16
Heirat	West (Gastarbeiter)	.05	.13		.24
	West (Ausländer)	.15	.21	.13	.30
	Ost (Gastarbeiter)				.15
	Ost (Ausländer)	.07	.14	.07	.22

Die Ergebnisse der deskriptiven Analysen deuten in ihrer Gesamtheit darauf hin, daß durch die Formulierungsänderung zwar die Bezugsgruppe vergrößert wurde - in den Kontaktfragen werden in der Ausländerformulierung deutlich mehr Kontakte berichtet. Daß aber gleichzeitig nur geringfügige Unterschiede zwischen den Splits in den Antworten auf die Einstellungsfragen zu verzeichnen sind, spricht für die Äquivalenz der beiden Messungen des Konstruktes Fremdenfeindlichkeit.

3. Reliabilität und interne Validität

Ausgangspunkt unserer Überlegungen zur Reliabilität und internen Validität der beiden Einstellungsskalen ist die Annahme, daß sich das Antwortverhalten auf die manifesten Items bei hoher inhaltlicher Äquivalenz der Gastarbeiter- und Ausländerskala faktorenanalytisch jeweils auf den latenten Faktor "Fremdenfeindlichkeit" zurückführen läßt. Darüber hinaus sollten die Faktorenladungen, die Faktorenvarianzen und die Meßfehler nicht signifikant zwischen beiden Skalen variieren. Signifikante Variationen zwischen den Skalen widersprechen der Annahme inhaltlicher Äquivalenz. Unsere erste Hypothese lautet deshalb:

H 1: Die unstandardisierten Faktorenladungen, Faktorenvarianzen und Meßfehler sind bei beiden Skalen numerisch gleich und weisen identische Vorzeichen auf. Dies gilt in der Weststichprobe ebenso wie in der Oststichprobe.

$$\Lambda_x^{\text{Gast}} = \Lambda_x^{\text{Ausl}}; \Phi^{\text{Gast}} = \Phi^{\text{Ausl}}; \theta_\delta^{\text{Gast}} = \theta_\delta^{\text{Ausl}}$$

Weiterhin dürfen bei hoher Äquivalenz der beiden Skalen weder für die Gastarbeiter- noch die Ausländerskala signifikante Unterschiede in den jeweiligen Faktorenladungen beim Vergleich zwischen unabhängigen Stichproben auftreten, wenn beide Skalen eine generelle Einstellungstendenz gegenüber Fremdgruppen messen. Als Hypothese formuliert bedeutet dies:

H 2a: Die unstandardisierten Faktorenladungen, Faktorenvarianzen und Meßfehler der Gastarbeiterskala variieren numerisch nicht zwischen Ost und West und haben identische Vorzeichen.

$$\Lambda_x^{\text{Gast-West}} = \Lambda_x^{\text{Gast-Ost}}; \Phi^{\text{Gast-West}} = \Phi^{\text{Gast-Ost}}; \theta_\delta^{\text{Gast-West}} = \theta_\delta^{\text{Gast-Ost}}$$

Ost

H 2b: Die unstandardisierten Faktorenladungen, Faktorenvarianzen und Meßfehler der Ausländerskala variieren numerisch nicht zwischen Ost und West und haben identische Vorzeichen.

$$\Lambda_x^{\text{Ausl-West}} = \Lambda_x^{\text{Ausl-Ost}}; \Phi^{\text{Ausl-West}} = \Phi^{\text{Ausl-Ost}}; \theta_\delta^{\text{Ausl-West}} = \theta_\delta^{\text{Ausl-Ost}}$$

Eine für alle vier Substichproben getrennt berechnete explorative Faktorenanalyse (PAF) (vgl. Tabelle 5) zeigt sehr homogene Strukturen. Im Osten ist die Reliabilität und interne Gültigkeit für die Ausländerformulierung tendenziell höher als für die Gastarbeiterformulierung. Im Westen zeigen sich diese Tendenzen nicht.

Explorative Faktorenanalysen erlauben jedoch keine Signifikanzprüfungen der Intergruppenunterschiede und setzen Normalverteilung auf Itemebene voraus. Insofern liefert dieses Ergebnis keinen harten Test unserer Hypothesen und berücksichtigt nicht die signifikante Verletzung der Normalverteilungsannahme auf der Einzelitem-Ebene. Wir testen deshalb im folgenden unsere Hypothesen simultan in einem multiplen Gruppenvergleich mit LISREL 8.12 (Jöreskog/Sörbom 1993a). Dazu verwenden wir Kovarianzen und, aufgrund der signifikant verletzten Normalverteilungsannahmen bei den manifesten Variablen, die asymptotischen Varianz-Covarianzen als Input-Matrizen. Diese wurden se-

parat für jede der vier Stichproben mit PRELIS 2 (Jöreskog/Sörbom 1993b) berechnet. Zur Modelltestung wurde der *Weighted-Least-Square*-Schätzer (WLS) herangezogen. Entsprechend unserer Hypothesen haben wir ein Modell spezifiziert, in dem sämtliche Parameter (Faktorenladungen, Faktorvarianzen, Meßfehler) für alle vier Stichproben gleichgesetzt sind. Die globalen Kennwerte dieser konfirmatorischen Faktorenanalyse sind nicht ideal, aber in Anbetracht der Problemstellung - in ihrer semantischen Bandbreite relativ heterogene und zudem siebenstufig gefaßte Items - durchaus zufriedenstellend [Chi-Quadrat=81.52, df=32, p=.0000034, GFI=.99, RMSEA=.022, *p-value of close fit*=1.00, AIC(sat.)=97,52(80.00)]. Die Nullhypothese muß - rein formalstatistisch gesehen - auf der Grundlage des Chi-Quadrat-Tests zurückgewiesen werden (= die auf der Basis des postulierten Modells "zurückgerechnete" Korrelationsmatrix weicht überzufällig von der Ausgangsmatrix ab). Dies ist aber in Anbetracht der Gesamtstichprobengröße (3130 Personen) nicht weiter verwunderlich.³⁾ Deshalb verwenden wir zur Modellbeurteilung den *Goodness-of-Fit-Index* (GFI) aufgrund seiner relativen Unabhängigkeit von Stichprobengrößen und den *Root Mean Square Error of Approximation* (RMSEA) (Steiger 1990; Browne/Cudeck 1992).

Tabelle 5: Faktorenladungen, erklärte Varianzanteile und Reliabilitäten der Fremdgruppendifferenzierung für alle vier Stichproben

	West		Ost	
	Gastarb. (N=1026)	Ausländ. (N=1062)	Gastarb. (N=519)	Ausländ. (N=523)
Leben	.48	.44	.42	.48
Arbeit	.72	.66	.69	.71
Politik	.67	.69	.62	.71
Heirat	.60	.63	.55	.70
erklärte Varianz	39.1%	37.7%	33.7%	42.9%
Cronbach's Alpha	.71	.70	.66	.74

Die unstandardisierten Faktorenladungen variieren nicht signifikant zwischen den Stichproben. Die mit den stichprobenspezifischen Varianzen vollstandardisierten Faktorenladungen (vgl. Tabelle 6) können als Validitätskoeffizienten interpretiert werden. Nach Bollen (1989) ist die interne Validität eines Items durch die Korrelation zwischen beobachteter und latenter Variable bestimmt. Für drei der vier Items ergeben sich somit befriedigende Validitätskoeffizienten. Lediglich die Frage der Lebensstilanpassung fällt deutlich dahinter zurück. Dieses Item weist jedoch bereits seit der ersten Verwendung

der Gastarbeiterskala im ALLBUS 1980 die geringste Validität auf (Kühnel/Terwey 1990).

Tabelle 6: Meßmodell Fremdenfeindlichkeit: Stichprobenspezifisch vollstandardisierte Faktorenloadungen (Validitätskoeffizienten) und Residuen

	West		Ost	
	Gastarbeiter (N=1026)	Ausländer (N=1062)	Gastarbeiter (N=519)	Ausländer (N=523)
	Faktorenloadungen (Validitätskoeffizienten)			
Leben	.46	.46	.46	.46
Arbeit	.70	.70	.70	.70
Politik	.68	.68	.68	.68
Heirat	.63	.63	.63	.63
	Residuen			
Leben	.79	.79	.79	.79
Arbeit	.52	.52	.52	.52
Politik	.54	.54	.54	.54
Heirat	.60	.60	.60	.60

Im Gegensatz zur explorativen Faktorenanalyse stellt LISREL keine direkte Schätzung der durch den Faktor erklärten Varianz zur Verfügung. Deshalb ist hier nur eine Schätzung auf der Ebene der Einzelitems möglich. In unserem Modell ist die Reliabilität der Items als die durch den Faktor erklärte Varianz je Item definiert (Bollen 1989). Sie berechnet sich für jeden Indikator über 1-standardisiertem Residuum. Diese sind zwischen den Stichproben stabil (LEBEN: .21; ARBEIT: .48; POLITIK: .46; HEIRAT: .40) und entsprechen in etwa anderen Befunden (Kühnel/Terwey 1990). Eine solche Interpretation der Meßfehler ist insofern problematisch, weil eine Trennung zwischen systematischen und zufälligen Meßfehlern hier nicht vorgenommen wurde (Saris 1982; Alwin 1989). Deshalb sind unsere Reliabilitäten vermutlich leicht unterschätzt. Neben der itemspezifischen Reliabilitätsschätzung kann zur Gesamtbeurteilung des Modells der GFI herangezogen werden. Er beschreibt, inwieweit das Modell die empirischen Kovarianzen zwischen den Items erklärt. Im vorliegenden Fall erklärt das Modell bei einem GFI von .99 zu 99 Prozent die empirischen Kovarianzen.

Aufgrund dieser Ergebnisse können die Ausländer- und die Gastarbeiterskala als äquivalente Messungen des gleichen Konzepts gelten. Dies gilt sowohl in West- wie auch in Ostdeutschland.⁴⁾

4. Externe Validierung

Als externe Kriterien (Faktoren) haben wir die Variablen Alter, Bildung und politische Links-Rechts-Orientierung⁵⁾ ausgewählt. In Tabelle 7 sind die Mittelwerte und Standardabweichungen dieser Indikatoren für alle vier Stichproben dargestellt.

Tabelle 7: Alter, Bildung und politische Orientierung: Deskriptive Parameter

	West				Ost			
	Gastarbeiter (N=1026)		Ausländer (N=1062)		Gastarbeiter (N=519)		Ausländer (N=523)	
	\bar{x}	s	\bar{x}	s	\bar{x}	s	\bar{x}	s
Bildung	2,83	1,13	2,77	1,09	2,83	1,00	2,85	1,00
Links-Rechts	5,23	1,72	5,24	1,66	4,65	1,54	4,65	1,56
Geburtsjahr 19..	47,6	16,97	46,9	16,89	45,7	16,76	46,5	15,93

Weder zwischen beiden West- noch zwischen beiden Oststichproben bestehen signifikante Verteilungsunterschiede hinsichtlich der externen Kriterien. Beide West- und beide Oststichproben stammen aus den jeweiligen Grundgesamtheiten West bzw. Ost.

Einige systematisch signifikante Unterschiede bestehen dagegen zwischen Ost und West. Beim Bildungsniveau ist die Varianz im Westen signifikant größer als im Osten. Die Ostdeutschen sind signifikant stärker links-orientiert als die Westdeutschen. Darüber hinaus ist in den Oststichproben die Varianz der politischen Orientierung signifikant geringer als in den Weststichproben. Die Ostdeutschen weisen somit eine homogenere politische Orientierung auf. Bis auf eine Ausnahme (West-1 und Ost-1) bestehen zwischen Ost und West keine signifikanten Unterschiede beim Alter der Befragten.

Die Korrelationsstruktur zwischen den externen Kriterien ist in den beiden Weststichproben identisch (vgl. Tabelle 8). Dies bestätigt die Annahme, daß beide Weststichproben aus der gleichen Grundgesamtheit stammen. Dagegen variieren insbesondere die Korrelation zwischen formalem Bildungsniveau und politischer Orientierung erheblich zwischen den beiden Oststichproben. Dies widerspricht der Annahme, daß beide Stichproben aus der gleichen Grundgesamtheit stammen. Nach unseren vorläufigen Befunden sind diese korrelativen Unterschiede auf Interviewereffekte zurückzuführen (vgl. Abschnitt 2.1 und Reinecke 1991).

Tabelle 8: Produkt-Moment-Korrelationen der externen Kriterien für alle vier Stichproben

		Bildung	Links-Rechts
Links-Rechts	West-1 (Gastarbeiter)	-.08**	
	West-2 (Ausländer)	-.10**	
	Ost-1 (Gastarbeiter)	-.20**	
	Ost-2 (Ausländer)	-.05	
Geburtsjahr	West-1 (Gastarbeiter)	.26**	-.13**
	West-2 (Ausländer)	.26**	-.13**
	Ost-1 (Gastarbeiter)	.38**	.05
	Ost-2 (Ausländer)	.36**	.07

** = signifikant auf dem 1%-Niveau.

Rein deskriptiv ist in Ost und West ein hohes Bildungsniveau mit einer eher politischen Links-Orientierung verknüpft. Dieser Zusammenhang ist allerdings in der Oststichprobe mit dem Ausländerstimulus nicht signifikant. Alter und Bildung sind dahingehend miteinander verknüpft, daß Ältere ein niedrigeres Bildungsniveau besitzen als Jüngere. Dieser Zusammenhang ist im Osten etwas stärker als im Westen. Ein ost-west-spezifischer Unterschied scheint bei der Beziehung zwischen Alter und politischer Orientierung zu bestehen. Während im Westen Ältere eher über eine politische Rechtsorientierung verfügen als Jüngere, besteht im Osten kein signifikanter Zusammenhang zwischen Alter und politischer Orientierung. Für die externe Validierung der beiden hier diskutierten Skalen postulieren wir in Übereinstimmung mit der Literatur folgende Korrelationshypothesen:

H 3: Mit zunehmendem Lebensalter steigt die Bereitschaft, Fremdgruppen abzuwerten (Jagodzinski et al. 1990; Kühnel/Terwey 1994; Blank/Schwarzer 1994).

H 4: Mit zunehmendem formalen Bildungsniveau nimmt die Bereitschaft, Fremdgruppen abzuwerten, ab (Jagodzinski et al. 1990; Kühnel/Terwey 1994; Blank/Schwarzer 1994).

H 5: Mit zunehmender politischer Rechtsorientierung steigt die Bereitschaft, Fremdgruppen abzuwerten (Kühnel/Terwey 1994; Blank/Schwarzer 1994).

Diese Hypothesen sollen im Westen wie im Osten gelten. Weiterhin nehmen wir an, daß sich die unstandardisierten Kovarianzen der externen Kriterien mit den beiden Skalen sowohl in den beiden Weststichproben wie auch in den beiden Oststichproben nicht signifikant unterscheiden. Unklar ist, inwieweit Ost-West-Unterschiede hinsichtlich der Kovarianz zwischen Alter, Bildung, politischer Orientierung und Fremdgruppenabwertung bestehen. Einerseits ist zu erwarten, daß sich die strukturellen, ideologischen und politischen Unterschiede zwischen BRD und DDR in der Zeit von 1948 bis 1989/90,

aber auch die vereinigungsbedingten strukturellen Unterschiede zwischen den alten und den neuen Bundesländern in systematischen Korrelationsunterschieden zwischen Ost und West widerspiegeln könnten. Andererseits sind unsere Daten, wie oben angedeutet, möglicherweise durch Interviewereffekte verzerrt. Zur Hypothesenprüfung haben wir deshalb zwei geschachtelte konfirmatorische Faktorenmodelle (Bollen 1989) mit den Konstrukten Alter, Bildung, politische Orientierung und Fremdgruppenabwertung spezifiziert (Jagodzinski et al. 1990):

Modell 1: Alle Parameter (Faktorenladungen, Meßfehler, Faktorenkorrelationen) für alle vier Stichproben sind gleichgesetzt. Dies bedeutet, es bestehen keine Ost-West-Unterschiede hinsichtlich der externen Gültigkeit.

$$\Lambda_X^{\text{Gast-West}} = \Lambda_X^{\text{Gast-Ost}} = \Lambda_X^{\text{Ausl-West}} = \Lambda_X^{\text{Ausl-Ost}}$$

$$\theta_o^{\text{Gast-West}} = \theta_o^{\text{Gast-Ost}} = \theta_o^{\text{Ausl-West}} = \theta_o^{\text{Ausl-Ost}}$$

$$\Phi^{\text{Gast-West}} = \Phi^{\text{Gast-Ost}} = \Phi^{\text{Ausl-West}} = \Phi^{\text{Ausl-Ost}}$$

Modell 2: Die Korrelationen der vier Faktoren werden für Ost und West separat geschätzt. Die Faktorenladungen und die Meßfehler sind in allen vier Stichproben gleichgesetzt.

$$\Lambda_X^{\text{Gast-West}} = \Lambda_X^{\text{Gast-Ost}} = \Lambda_X^{\text{Ausl-West}} = \Lambda_X^{\text{Ausl-Ost}}$$

$$\theta_o^{\text{Gast-West}} = \theta_o^{\text{Gast-Ost}} = \theta_o^{\text{Ausl-West}} = \theta_o^{\text{Ausl-Ost}}$$

$$\text{diag}\Phi^{\text{Gast-West}} = \text{diag}\Phi^{\text{Gast-Ost}} = \text{diag}\Phi^{\text{Ausl-West}} = \text{diag}\Phi^{\text{Ausl-Ost}}$$

$$(\Phi_s^{\text{Gast-West}} = \Phi_s^{\text{Ausl-West}}) \neq (\Phi_s^{\text{Gast-Ost}} = \Phi_s^{\text{Ausl-Ost}}),$$

wobei s die Elemente unterhalb der Hauptdiagonalen indiziert. Dies spiegelt die oben angesprochenen Ost-West-Unterschiede hinsichtlich der externen Kriterien wider.

Beide Modelle wurden jeweils in einem multiplen Gruppenvergleich zwischen den vier Stichproben unter Verwendung von Kovarianzen und der asymptotischen Varianz-Kovarianz-Matrix mit LISREL 8.12 (Jöreskog/Sörbom 1993a) geschätzt. In Tabelle 9 sind die Anpassungsparameter beider Modelle dargestellt.

Tabelle 9: Anpassungsparameter der beiden Modelle

	Modell 1 invariante Faktorenkorrelationen	Modell 2 ost-west-spezifische Faktorenkorrelationen	Differenz
df	95	89	$\text{diff}_{df} = 6$
Chi-Quadrat	391,0	298,36	$\text{diff}_{\chi^2} = 92,64$
p	.000	.000	
N	3110	3110	
RMSEA	.032	.027	
p-value of close fit	1.0	1.0	
GFI	.96	.97	
AIC (sat.)	425,01 (224,0)	344,36 (224,0)	
df / Chi-Quadrat	4,12	3,35	

Beide Modelle erreichen einen befriedigenden Modellfit im Hinblick auf den GFI, den RMSEA und den *p-value of close fit* und sind damit empirisch nicht widerlegt. Das Modell 2 stellt gegenüber dem ersten Modell bei Verwendung des Chi-Quadrat-Differenzentests (Bollen 1989; Jöreskog/Sörbom 1993a) eine signifikante Verbesserung dar. Durch die Aufgabe von sechs Freiheitsgraden reduziert sich der Chi-Quadrat-Wert um fast 100. Den größten Anteil an dieser Verbesserung liefert dabei die ost-west-spezifische Korrelationsschätzung der externen Kriterien untereinander.⁶⁾

Im folgenden diskutieren wir deshalb nur das zweite Modell mit ost-west-spezifischen Faktorenkorrelationen. Wir wollen hier nicht die Ursachen der unterschiedlichen Korrelationsstrukturen zwischen Alter, Bildung und politischer Orientierung in Ost und West analysieren, sondern haben in diesem Modell lediglich die empirischen Fakten berücksichtigt. Es muß weiteren Analysen überlassen werden, ob die korrelativen Ost-West-Unterschiede hinsichtlich der externen Kriterien inhaltlich und/oder methodisch zu erklären sind.

Die mit den stichprobenspezifischen Varianzen vollstandardisierten Faktorenladungen der Gastarbeiter- bzw. Ausländeritems sind über alle vier Stichproben gleich (vgl. Tabelle 10). Alle Faktorenladungen sind hoch signifikant. Die in Abschnitt 3 bereits nachgewiesene strukturelle Äquivalenz bleibt somit auch bei der Einführung externer Kriterien erhalten und robust.

Tabelle 10: Meßmodell Fremdenfeindlichkeit: Stichprobenspezifisch vollstandardisierte Faktorenladungen (Validitätskoeffizienten) und Residuen

	West		Ost	
	Gastarbeiter (N=1026)	Ausländer (N=1062)	Gastarbeiter (N=519)	Ausländer (N=523)
	Faktorenladungen			
Leben	.50	.50	.50	.50
Arbeit	.67	.67	.67	.67
Politik	.66	.66	.66	.66
Heirat	.66	.70	.66	.66
	Residuen			
Leben	.77	.77	.77	.77
Arbeit	.56	.56	.56	.56
Politik	.55	.55	.55	.55
Heirat	.54	.54	.54	.54

Die durch den Faktor erklärte Varianz pro Item (Reliabilität) ist ebenfalls bei allen vier Stichproben gleich und beträgt für das Item LEBEN .23, für das Item ARBEIT .44, für das Item Politik .45 und für das Item HEIRAT .46. Gegenüber dem internen Validitätsmodell (vgl. Tabelle 6) bestehen nur geringfügige Unterschiede in den Faktorenladungen.

Die oben als Kriterien externer Validität formulierten Korrelationshypothesen H 3 - H 5 bestätigen sich für beide West- und beide Oststichproben (vgl. Tabelle 11). Je älter die Befragten sind, je niedriger das formale Bildungsniveau und je stärker die politische Rechtsorientierung, desto stärker ist die Diskriminierungstendenz gegenüber Ausländern. Weder im Westen noch im Osten bestehen dabei Unterschiede zwischen den beiden Skalenformulierungen.

Unterschiede bestehen zwischen Ost und West. Alter und Fremdgruppenabwertung sind im Westen doppelt so stark korreliert wie im Osten. Im Gegensatz dazu ist die korrelative Beziehung zwischen politischer Orientierung und Fremdgruppenabwertung im Osten deutlich stärker als im Westen. Aus den oben genannten Gründen (Interviewereffekte) verbietet sich jedoch eine voreilige inhaltliche Interpretation dieser Ost-West-Unterschiede.

Tabelle 11: Fremdenfeindlichkeit und externe Kriterien: Stichprobenspezifisch vollstandardisierte signifikante Faktorenkorrelationen

		Geburtsjahr	Bildung	Links-Rechts
Fremdenfeindlichkeit	West (Gastarbeiter)	-.39	-.39	.20
	West (Ausländer)	-.39	-.39	.20
	Ost (Gastarbeiter)	-.19	-.41	.31
	Ost (Ausländer)	-.19	-.41	.31

Tabelle 12: Externe Kriterien: Stichprobenspezifisch vollstandardisierte signifikante Faktorenkorrelationen

		Geburtsjahr	Bildung	Links-Rechts
Geburtsjahr	West 1	1.00		
	West 2	1.00		
	Ost 1	1.00		
	Ost 2	1.00		
Bildung	West 1	.25	1.00	
	West 2	.25	1.00	
	Ost 1	.38	1.00	
	Ost 2	.38	1.00	
Links-Rechts	West 1	-.12	-.08	1.00
	West 2	-.12	-.08	1.00
	Ost 1	.08	-.12	1.00
	Ost 2	.08	-.12	1.00

Da wir oben bereits auch auf die signifikanten Unterschiede in den beiden Oststichproben hinsichtlich der Korrelationen der externen Kriterien untereinander hingewiesen haben, sind der Vollständigkeit halber in Tabelle 12 auch die entsprechenden Korrelationsschätzungen des zweiten Modells dargestellt. Entsprechend unserer Modellspezifikation spiegeln sich hier nicht die Korrelationsunterschiede zwischen den beiden Oststichproben wider (vgl. Tabelle 8). Es zeigt sich jedoch, daß im Osten Alter und Bildung stärker korreliert sind als im Westen. Geringe Ost-West-Unterschiede bestehen auch hinsichtlich der korrelativen Beziehung zwischen Alter, Bildung und politischer Orientierung. Darüber hinaus weist die Korrelation zwischen Alter und politischer Orientierung unterschiedliche Vorzeichen in Ost und West auf. Ungeachtet dieser Problematik ist als zentrales Ergebnis festzustellen, daß auch durch die Einführung externer Kriterien die Äquivalenz der beiden Skalen nachgewiesen ist. Die Ergebnisse von Blank/Schwarzer (1994) werden bestätigt.

5. Zusammenfassung und Ausblick

Seit 1996 werden die Gastarbeiterfragen im ALLBUS ausschließlich in der reformulierten Fassung "die in Deutschland lebenden Ausländer" erhoben. Wir haben deshalb die reformulierten Gastarbeiteritems mit den klassischen Gastarbeiteritems als Indikatoren einer allgemeinen Diskriminierungsbereitschaft auf der Grundlage repräsentativer Daten für Gesamtdeutschland verglichen. Damit wurde der Reformulierungsvorschlag von Blank/Schwarzer (1994), der lediglich an regionalen Stichproben entwickelt wurde, einer erneuten Validierung unterzogen.

Die deskriptiven Vergleiche der Splitversionen hinsichtlich der Mittelwerte bei den Einstellungsitens und der Kontaktraten ergaben, daß die Formulierungsänderung zwar bei den Kontaktfragen zu einem recht deutlichen Anstieg der berichteten Kontakte führt, also anscheinend wirklich eine Begriffserweiterung darstellt. Gleichzeitig waren jedoch im allgemeinen nur geringfügige Unterschiede in den Antworten auf die Einstellungsfragen zu berichten. Die Stärke der in beiden Splits gemessenen Fremdenfeindlichkeit war weitgehend ähnlich.

Durch konfirmatorische Faktorenanalysen in einem multiplen Gruppenvergleich konnten wir die strukturelle Äquivalenz beider Skalen sowohl intern wie auch durch die Einführung externer Kriterien zeigen. Die Ergebnisse belegen, daß mit der neuen Kurzskaala ein Instrument zur Messung einer generellen Diskriminierungsbereitschaft zur Verfügung steht, daß dem theoretischen Konzept semantisch besser Rechnung trägt und unempfindlicher ist gegenüber proportionalen Veränderungen in den gesellschaftlich relevanten Fremdgruppen.

Auch wenn die Ergebnisse unserer Analysen bestätigen, daß so das latente Konstrukt "Fremdenfeindlichkeit" genauso reliabel und valide gemessen wird, bleibt die Frage, ob die Äquivalenz der alten und neuen Messung auch auf der rein deskriptiven Ebene gegeben und damit die Zeitreihenfähigkeit sichergestellt ist. Zu bezweifeln wäre dies nach unseren deskriptiven Analysen am ehesten bei den Ostdeutschen. Hier wurde die Zeitreihe jedoch 1994 erst begonnen, und da der Begriff 'Gastarbeiter' in der ehemaligen DDR nur für die Ausländer im Westen gebraucht wurde, sind die in Ostdeutschland zum Teil zu beobachtenden recht deutlichen Mittelwertdifferenzen wohl eher ein Argument für die reformulierte Fassung. Im Westen sind die Mittelwertdifferenzen bei den Einstellungsitens mit der alten und der neuen Formulierung in der Gesamtstichprobe so gering, daß man sie praktisch vernachlässigen kann. Bei subgruppenspezifischen Analysen ist allerdings eine gewisse Vorsicht geboten. Bei Längsschnittanalysen in bezug auf einzelne Subgruppen sowie vor allem auch generell bei den Kontaktfragen sollte man

daher in Zukunft bei der Interpretation von Zeitreihen den Split des ALLBUS 1994 nutzen, um den Anteil der auf die Formulierungsänderung zurückgehenden Effekte abzuschätzen.

In Zukunft wird zu prüfen sein, inwieweit die Skala auch in der Dauerbeobachtung eine robuste Faktorenstruktur aufweist. Dies wäre von einer validen Skala zu fordern und ist bei den klassischen Gastarbeiteritems gegeben (Kühnel/Terwey 1990). Ein Instrument für die Dauerbeobachtung sollte nicht nur spezifische gesellschaftliche Veränderungen messen, die sich bei einzelnen Items der Skala zeigen, sondern insbesondere sensibel gegenüber grundsätzlichen Einstellungsveränderungen sein. Dies würde sich in Mittelwertsveränderungen auf der Faktorebene widerspiegeln. Inwieweit die reformulierten Fragen diesem Anspruch genügen, wird noch zu untersuchen sein.

Anmerkungen

- 1) Wir danken Horst-Alfred Heinrich, Achim Koch, Peter Schmidt und Stefan Schwarzer für die kritische Durchsicht des Manuskripts.
- 2) Dies hängt sicherlich zum Teil mit der beim ALLBUS 1994 praktizierten Form der Splitzuordnung zusammen, bei der jeder Interviewer aus erhebungstechnischen Gründen nur mit einer Fragebogenversion arbeitete.
- 3) Vergleiche zu dieser Problematik Bollen (1989), Browne/Cudeck (1992) und Marsh (1994).
- 4) Wie aus Tabelle 1 zu entnehmen ist, weisen bei beiden Skalen die Ostdeutschen auf Itemebene höhere Mittelwerte auf als die Westdeutschen. Sie sind bei t-Test-Vergleichen zwischen Ost und West signifikant. In einem hier nicht näher dargestellten, um die empirischen Item-Mittelwerte erweiterten Modell ergeben sich unter Beibehaltung der Faktorenladungsinvarianz signifikante Mittelwertunterschiede auf der latenten Ebene zwischen Ost und West. Bei beiden Skalen äußern die Ostdeutschen eine gegenüber den Westdeutschen signifikant höhere latente Fremdgruppenabwertung. Dies steht in Übereinstimmung mit anderen Befunden (z.B. Runge 1990; Herrmann/Schmidt 1995; Blank/Schmidt 1996).
- 5) Diese Variablen gehen mit folgenden Skalierungen in die Analysen ein. *Alter*: Geburtsjahr des Befragten; *Bildung* [höchster erreichter allgemeinbildender Schulabschluß: kein Abschluß; Volks-/Hauptschulabschluß bzw. Polytechnische Oberschule mit Abschluß 8. oder 9. Klasse; Mittlere Reife, Realschulabschluß bzw. Polytechnische Oberschule mit Abschluß 10.Klasse; Fachhochschulreife (Abschluß einer Fachoberschule etc.); Abitur bzw. Erweiterte Oberschule mit Abschluß 12.Klasse (Hochschulreife)]; *Links-Rechts-Orientierung* (Selbsteinstufung der eigenen politischen Ansichten): Skalenwert zwischen 1 (links) und 10 (rechts).

6) Ein ebenfalls guter Modellfit ist zu erreichen, wenn man eine Schätzung der Korrelation unter den externen Kriterien erlaubt, die dem Muster der Produkt-Moment-Korrelationen (vgl. Tabelle 8) entspricht. Solch ein Modell berücksichtigt die vorhandenen Interviewereffekte expliziter ($df=93$, $\text{Chi-Quadrat}=316.94$, $p=.000$, $\text{RMSEA}=.028$, $p\text{-value of close fit}=1.0$, $\text{GFI}=.96$, $\text{AIC (sat.)}=354.94$ (224.00)). Da wir jedoch Interviewereffekte in unseren Analysen nicht systematisch berücksichtigt haben, verzichten wir auf die Darstellung dieses Modells.

Literatur

Alwin, D. F., 1989: Problems in the estimation and interpretation of the reliability of survey data. *Quality and Quantity* 23: 277-331.

Blank, T./Schwarzer, S., 1994: Ist die Gastarbeiterskala noch zeitgemäß? Die Reformulierung einer ALLBUS-Skala. *ZUMA-Nachrichten* 34: 97-115.

Blank, T./Schmidt, P., 1996: Konstruktiver Patriotismus im vereinigten Deutschland? Ergebnisse einer repräsentativen Studie. In: A. Mummendey/B. Simon (Hrsg.), *Identität und Verschiedenheit*. Bern: Huber (im Druck).

Bollen, K. A., 1989: *Structural equations with latent variables*. New York: John Wiley & Son.

Browne, M.W./Cudeck, R., 1993: Alternative ways of assessing model fit. *Sociological Methods & Research* 21: 230-258.

Herrmann, A./Schmidt, P., 1995: Autoritarismus, Anomie und Ethnozentrismus. S. 287-319 in: G. Lederer/P. Schmidt (Hrsg.), *Autoritarismus und Gesellschaft. Trendanalysen und vergleichende Jugenduntersuchungen 1945-1993*. Opladen: Leske + Budrich.

Jagodzinski, W./Kühnel, S./Schmidt, P., 1987: Is there a „Socratic Effect“ in nonexperimental panel studies? Consistency of an attitude toward Guestworkers. *Sociological Methods and Research* 15: 259-302.

Jagodzinski, W./Kühnel, S./Schmidt, P., 1988: Is the true score model or the factor model more appropriate? Response to Saris and Putte. *Sociological Methods and Research* 17: 158-164.

Jagodzinski, W./Kühnel, S./Schmidt, P., 1990: Searching for parsimony: are true-score models or factor models more appropriate? *Quality and Quantity* 24: 447-470.

Jöreskog, K. G./Sörbom, D., 1993a: *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.

Jöreskog, K. G./Sörbom, D., 1993b: *PRELIS 2 User's reference guide*. Chicago: Scientific Software International.

Koch, A./Gabler, S./Braun, M., 1994: Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 1994. ZUMA-Arbeitsbericht 94/11. Mannheim.

Krauth, C./Porst, R., 1984: Sozioökonomische Determinanten von Einstellungen zu Gastarbeitern. S. 233-266 in: K.U. Mayer/P. Schmidt (Hrsg.), Allgemeine Bevölkerungsumfrage der Sozialwissenschaften: Beiträge zu methodischen Problemen des ALLBUS 1980. Frankfurt: Campus.

Kühnel, S./Terwey, M., 1990: Exploring attitudes towards foreign "Guestworkers" in Western Germany by different approaches. Paper presented at the XIIth World Congress of Sociology in Madrid, 9.-13. Juli 1990.

Kühnel, S./Terwey, M., 1994: Gestörtes Verhältnis? Die Einstellungen der Deutschen zu Ausländern in der Bundesrepublik. S. 71-105 in: M. Braun/P. Ph. Mohler (Hrsg.), Blickpunkt Gesellschaft 3. Einstellungen und Verhalten der Bundesbürger. Opladen: Westdeutscher Verlag.

Marsh, H. W., 1994: Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling* 1: 5-34.

Pfeifer, A./Schmidt, P., 1987: LISREL. Die Analyse komplexer Strukturgleichungsmodelle. Stuttgart: Gustav Fischer Verlag.

Reinecke, Jost, 1991: Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte. Opladen: Westdeutscher Verlag.

Reuband, K.-H., 1989: Wächst in der Bundesrepublik der Ausländerfeindlichkeit? *Neue Praxis* 89/3: 270-274.

Runge, I., 1990: *Ausland DDR - Fremdenhaß*. Berlin.

Saris, W.E., 1982: Different questions, different variables? S. 78-95 in: C. Fornell (Hrsg.), *The second generation of multivariate statistics*. Band 2: Measurement and evaluation. New York: Praeger.

Saris, W.E./van den Putte, B., 1988: True scores or factor models: A secondary analysis of the ALLBUS-test-retest data. *Sociological Methods and Research* 17: 123-157.

Saris, W.E./Hartman, H., 1990: Common factors can always be found but can they also be rejected? *Quality and Quantity* 24: 471-490.

Steiger, J. H., 1990: Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research* 25: 173-180.

VERTRAUEN IN INSTITUTIONEN IN DEUTSCHLAND ZWISCHEN 1991 UND 1995

DIETER WALZ

Vereint - und doch gespalten“ - so wurde vielfach der Zustand der „inneren Einheit“ Deutschlands nach Vollzug der Wiedervereinigung beschrieben. Tatsächlich hängt das Gelingen der politischen Integration in Deutschland auch davon ab, daß sich im Verlaufe des Wiedervereinigungsprozesses weitgehend übereinstimmende Einstellungen und Verhaltensmuster im vereinten Deutschland etablieren. Wenn auch die Ansichten über das unverzichtbare Ausmaß an politischer Unterstützung für das politische System weit auseinandergehen, sind dauerhafte Zweifel an der Vertrauenswürdigkeit der politischen Ordnung mit der Idee der Demokratie nicht vereinbar. Nachdem zu Beginn der Wiedervereinigung vor allem das West-Ost Gefälle im Institutionenvertrauen als problematisch perzipiert worden war, haben sich die Einstellungen der Bevölkerung in Ost und West im Grunde bei allen untersuchten politischen Institutionen in den letzten Jahren sehr stark angenähert. Damit gibt es heute mehr Gemeinsamkeiten als Unterschiede in beiden Teilen Deutschlands. Ob man deshalb allerdings schon von einer gemeinsamen politischen Kultur ausgehen kann, ist damit nicht gesagt.

Unified - but still divided“ - so people described the situation about the „internal Unification“ in Germany after the formal Unification process. But it is indeed important for the integration in Germany, that there will be the same political attitudes toward the (newly) political system. One major political attitude toward the political system is the confidence people have in their major political institutions. After there were more differences than common ground in 1990, the paper investigate whether these differences are still present, whether they are even greater or whether there took a real political acculturation place. The empirical analyse showed that the political attitudes in East and West toward almost all political institutions under investigation are converged. Even if there is a townturn in political trust in the West, there are more common ground than differences today in Germany. However, it is not easy to say whether there is

already a common political culture in Germany - but further research will show, whether not only the formal but also the political integration was successful.

1. Einleitung

„Vereint - und doch gespalten“ - so wurde vielfach der Zustand der „inneren Einheit“ Deutschlands nach Vollzug der formalen Integration beider deutschen Teilstaaten im Jahre 1990 beschrieben. Zu Beginn des Wiedervereinigungsprozesses war es sicherlich auch gerechtfertigt zu vermuten, daß die „Trennung über die Dauer von zwei Generationen, politische Sozialisation in konträren politischen Systemen, berufliche Anpassung und Eingliederung in unterschiedliche Wirtschaftssysteme, Teilhabe an politischen Entscheidungen auf der einen, politische Bevormundung, ja Unterdrückung auf der anderen Seite ... zu unterschiedlichen Mentalitäten der Deutschen geführt“ hat (Feist 1991: 21). So war es sicherlich ebenfalls gerechtfertigt, 1990 (noch) von zwei getrennten politischen Kulturen, d.h. von unterschiedlichen Einstellungen der Bevölkerung in Ost- und Westdeutschland gegenüber dem (neuen) politischen System und seinen Teilbereichen auszugehen (vgl. Kaase 1993). Allerdings wird vielfach konzediert, daß das Gelingen der politischen Integration Deutschlands sowie die Stabilität und Persistenz des politischen Systems auch maßgeblich davon abhängt, daß sich in beiden Teilen Deutschlands nach einer gewissen Übergangsphase weitgehend übereinstimmende politische Einstellungen und Verhaltensmuster herausbilden. Almond und Verba (1965) sprechen in diesem Zusammenhang von einer notwendigen Kongruenz von politischer Kultur und politischer Struktur als Faktor der Stabilität und Durabilität des politischen Systems, denn „kein demokratisches System kann ohne die Akzeptanz und Unterstützung durch die Bevölkerung existieren“ (Gluchowski/Zelle 1992: 251; vgl. Miller 1974a; Almond/Powell 1978; Gabriel 1986; Westle 1989a/b, 1990). Eine besondere Relevanz erhält die Frage nach der Kongruenz von politischer Kultur und politischer Struktur also insbesondere dann, wenn - wie im Falle der deutschen Wiedervereinigung - die Bevölkerung bzw. Teile der Bevölkerung in Folge eines Regimewechsels mit neuen politischen Strukturen konfrontiert wird. In dieser Situation befinden sich nun die Bürgerinnen und Bürger der ehemaligen DDR seit dem Jahre 1990. Die Wiedervereinigung ist praktisch mit einem kompletten Institutionentransfer von West nach Ost einhergegangen, wobei die Bevölkerung der fünf neuen Bundesländer formal nach Artikel 23 Grundgesetz in das politische System der (alten) Bundesrepublik integriert wurde. Allerdings zeigen historische Erfahrungen, daß einer formalen Übernahme demokratischer Strukturen nicht automatisch deren Akzeptanz bzw. Unterstützung folgen muß, d.h. daß damit „noch lange keine ‚Akkulturation‘ erreicht ist“

(Schacht 1991: 14; vgl. Feist 1991; Gabriel 1993). Insofern ist es „in dieser kritischen Übergangsphase“ (Gabriel 1993: 3) auch unter Stabilitätsgesichtspunkten von großer Bedeutung, daß sich im Verlaufe des Wiedervereinigungsprozesses weitgehend übereinstimmende Einstellungen der Bevölkerung gegenüber dem - teils neuen - politischen System etablieren.

Bezugnehmend auf Eastons Modell politischer Unterstützung (1975) kann die Struktur des politischen Systems in drei zentrale Elemente zerlegt werden: a) die politische Gemeinschaft, b) die politische Ordnung (Normen, Werte und Institutionen), sowie c) die politischen Herrschaftsträger. Die Systemkultur differenziert sich bei Easton in spezifische und diffuse (Vertrauen und Legitimität) Unterstützung. Während spezifische Unterstützung als instrumentelle Outputbewertung aufgefaßt werden kann, die sich zudem primär auf die politischen Herrschaftsträger bezieht, stellt diffuse Unterstützung eine auf alle drei Objekte gerichtete affektive, moralische und wertbezogene Einstellungsart dar (vgl. Westle 1989, 1992: 461-462). In Erweiterung der Ausführungen von Almond und Verba werden also „die Objekte und Arten der Unterstützung nicht als gleich bedeutsam angesehen, sondern aus ihren unterschiedlichen Eigenschaften wird auf eine systemfunktionale Hierarchie der politischen Unterstützung geschlossen“ (Westle 1992: 462). Für die Stabilität und Funktionsfähigkeit politischer Systeme ist eine diffuse politische Unterstützung der politischen Ordnung und der politischen Gemeinschaft wesentlich bedeutsamer, als die spezifische politische Unterstützung für die politischen Herrschaftsträger. „Dieser Vorstellung entspricht der in Demokratien institutionell vorgesehene Mechanismus des Regierungswechsels durch Wahlen als möglicher Ausdruck politischer Unzufriedenheit mit den jeweils amtierenden Herrschaftsträgern, der ein Übergreifen negativer Bewertungen auf das System vermeiden soll“ (Westle 1989b: 224; vgl. Fuchs 1981). Über das unverzichtbare Ausmaß an diffuser Unterstützung für die politische Ordnung gehen die Auffassungen allerdings weit auseinander (vgl. zusammenfassend Gabriel 1993). Während eine in der Tradition des Konzeptes demokratischer Elitenherrschaft stehende Sicht des Institutionenvertrauens eine breite Vertrauensbasis fordert, und dies vor allem mit dem Erfordernis einer effektiven Erfüllung staatlicher Aufgaben begründet (vgl. Gamson 1968), wird jedoch andererseits auch gefordert, daß eine Demokratie keine vertrauensvolle, sondern eher eine mißtrauische Öffentlichkeit benötigt, die - um Mißbrauch vorzubeugen - das Handeln der Führung kritisch überwacht und einem ständigen Rechtfertigungsdruck aussetzt (vgl. Parry 1976; Döring 1990).¹⁾ Einige Arbeiten verweisen schließlich auf die Notwendigkeit einer gesunden Mischung von Vertrauen und Mißtrauen in der Gesellschaft (vgl. Sniderman 1981). Diese Mischung ist gegeben, wenn Teile der Bevölkerung der Führung vertrauen, andere ihr mißtrauen, wenn das Vertrauensniveau

schwankt und wenn es sowohl Dissens als auch Konsens gibt. Gabriel (1993) verweist darauf, daß man diesen Interpretationen allerdings nicht den Status ausformulierter, empirisch bewährter oder auch nur prüfbarer Theorien zubilligen kann. Auseinandersetzungen über das systemfunktionale Niveau an politischer Unterstützung lassen sich ebenfalls mit den Mitteln der empirischen Forschung nicht beilegen, insofern dienen die referierten Ansätze bestenfalls als Hintergrund für die Interpretation empirischer Befunde in diesem Beitrag. Ungeachtet der unterschiedlichen Auffassungen über das in einer Demokratie erforderliche Mindestmaß an politischer Unterstützung „sind weitverbreitete und dauerhafte Zweifel an der Vertrauenswürdigkeit der politischen Ordnung und der sie tragenden Institutionen mit der Idee der Demokratie unvereinbar“ (Gabriel 1993: 3).

Als eine maßgebliche Orientierungsart gegenüber dem politischen System wird das Vertrauen - als Teildimension diffuser politischer Unterstützung - der Bevölkerung in öffentliche Einrichtungen und Institutionen - als Teil der politischen Ordnung - angesehen (vgl. Miller 1974a/b; Citrin 1974; Easton 1975; Barber 1983; Gabriel 1993). Die Entwicklung des Institutionenvertrauens seit 1984 im Westen bzw. seit 1991 in Gesamtdeutschland wollen wir in diesem Beitrag dokumentieren. Bevor allerdings empirische Befunde zum Ausmaß und zur Entwicklung des Vertrauens für Institutionen in der Bundesrepublik seit der Wiedervereinigung präsentiert werden, erfolgt zunächst eine kurze Einführung in die „Vertrauensforschung“, ein Überblick über den Forschungsstand sowie einige Anmerkungen zur Datensituation.

2. Theoretische Basis, Forschungsstand, Daten und Variablen

Der Begriff des Vertrauens wird trotz seiner offensichtlichen Bedeutung - nicht nur für den politischen Bereich - in der Forschung bisher alles andere als eindeutig bzw. einheitlich verwendet (vgl. Preisendörfer 1987; Bentele 1992; zusammenfassend Walz 1995b). Wir setzen bei der Bestimmung des Vertrauensbegriffes bei Stockes an, der unter Vertrauen zunächst nur „... a basic evaluative orientation toward government“ versteht, ohne allerdings zu spezifizieren, auf welche Dimensionen des Vertrauensbegriffes er sich explizit bezieht (Stockes 1962: 67). Gamson übernimmt diesen Definitionsvorschlag von Stockes, klassifiziert das Vertrauen jedoch ferner als „... more generalized, cumulative, and more stable“ als „... attitudes on specific issues“. Vertrauen sei zudem „... the general expectation people have about the quality of the products (and outputs) that the political system produces“ (Gamson 1968: 40, 1971: 71; Hervorhebungen DW). Mögliche inhaltliche Dimensionen dieser Erwartungen im Rahmen des Vertrauens beschreibt Abramson, wenn er darauf hinweist, Vertrauen sei „... the belief that they

(leaders of the political system, DW) will usually be honest ... and ... they will usually act in the interest of the people. Feelings of trust may also involve the belief that leaders are competent“ (Abramson 1972: 1245; vgl. auch Miller/Listhaug 1990; Gabriel 1992b; Hervorhebungen DW). Auch für Miller (1974b) spielen Erwartungen bei der Vergabe von Vertrauen eine wichtige Rolle: „... (trust is) the belief that government is operating according to one's normative expectations of how government (authorities and institutions DW) should function“ (Miller 1974b: 989; vgl. Miller/Listhaug 1990). Während auch bei Barber (1983; vgl. Jäckel 1990) das Vertrauen stark mit Erwartungen zusammenhängt, definiert Coleman (1982) Vertrauen als „Übertragung von Kontrolle über Handlungen, Ressourcen und Ereignisse“, womit „immer Risiko ... und immer ein Zukunftsbezug“ mit eingeschlossen ist (Schneeberger 1982: 302; vgl. Jäckel 1990). Für Luhmann (1968) ist Vertrauen schließlich ein Mechanismus zur „Reduktion sozialer Komplexität“ (vgl. Bentele 1992). Die Annahme einer einheitlichen Verwendung des Vertrauensbegriffes ist also nicht gerechtfertigt, vielmehr finden sich zahlreiche Umschreibungen dafür in der Literatur wieder. Ferner ist darauf hinzuweisen, daß sich Vertrauen nicht nur auf ein Objekt beziehen kann (bspw. Institutionen), sondern daß „... political trust is best regarded as a differentiated attitude toward different levels of the political system - toward the public philosophy which justifies the regime, its political institutions, or the authorities who hold office at a particular time“ (Gamson 1971: 42; vgl. Jäckel 1990; Gabriel 1992b). Damit gibt es auch nicht mehr das politische Vertrauen, sondern entsprechend objektabhängig beispielsweise das Personen-, Institutionen- oder Systemvertrauen (vgl. Bentele 1993). Wir wollen hier zusammenfassend unter Vertrauen für Institutionen eine Einstellung verstehen, die zunächst einmal im Sinne Colemans als „Übertragung von Kontrolle über Handlungen, Ressourcen und Ereignisse“ verstanden werden kann, jedoch untrennbar verbunden mit der Erwartung, daß öffentliche Einrichtungen und Institutionen die ihnen übertragenen Aufgaben anhand bestimmter Kriterien („Qualities“, wie z.B. gemeinwohlorientiert, ehrlich, kompetent) im Grunde zufriedenstellend erfüllen. In diesem Sinne wird Vertrauen zu einem - wenigstens temporären - „reservoir of good will“ (vgl. Easton 1975), d.h. zu einer wichtigen Teildimension diffuser politischer Unterstützung (vgl. Gabriel 1992b; Walz 1995b).

Akzeptiert man nun das Vertrauen als eine Einstellungsart, mit der sich das Verhältnis zwischen Bevölkerung und Institutionen beschreiben läßt, so bleibt - neben der notwendigen Abgrenzung des Vertrauensbegriffes von anderen Konzepten (vgl. Gabriel 1992b) - noch die Frage nach den zentralen Problemstellungen der „Vertrauensforschung“. Sehr treffend faßt Gabriel mögliche Aspekte folgendermaßen zusammen: „Despite of their prominent position in public opinion research as well as in

democratic theory, the conceptual status of political trust has remained largely unclear. Considerable confusion exists on the definition and operationalization of the concept, their dimensionality, their relationship to other attitudes, the adequacy of measurement instruments, and finally the consequences of varying levels of trust ... for the political system“ (Gabriel 1992b: 2; vgl. ähnlich Preisendörfer 1987; Walz 1995b/c). Folgende Bereiche können m.E. nun als Kern einer „Vertrauensforschung“ angesehen werden:

1. Wie ist es mit Ausmaß und Entwicklung des Institutionenvertrauens bestellt?
2. Welche Struktur des Institutionenvertrauens läßt sich dabei erkennen?
3. Welche Faktoren bestimmen Ausmaß und Veränderung des Vertrauens?
4. Welche Bedeutung hat das Vertrauen auf unterschiedlichen Ebenen?

Während der erste und zweite Aspekt vor allem eine Aufgabe der Deskription darstellt, bezieht sich der dritte Bereich auf mögliche Determinanten des Vertrauens. Der vierte Aspekt thematisiert schließlich Konsequenzen von Vertrauen und Mißtrauen auf unterschiedlichen Ebenen (vgl. Walz 1995a/b). In diesem Beitrag wollen wir jedoch nur auf den ersten Bereich eingehen, d.h. aufzeigen, wie sich das Vertrauen der Bevölkerung in West- und Ostdeutschland gegenüber ausgewählten politischen Institutionen seit der Wiedervereinigung entwickelt (Deskription) hat (vgl. Lipset/Schneider 1983; Listhaug 1984; Feist 1991; Gabriel 1993). Bevor jedoch unter Rückgriff auf repräsentative Umfragestudien empirische Befunde zum Institutionenvertrauen in der Bundesrepublik Deutschland vor und nach der Wiedervereinigung präsentiert werden, ist - neben einem Überblick über den Forschungsstand - auch eine kritische Würdigung der Datensituation und eine Auseinandersetzung mit den entsprechenden Frageformen und Antwortvorgaben zu leisten (vgl. Löwenhaupt 1993; Walz 1995c).

Die Erforschung des Institutionenvertrauens fristet in der deutschen Sozialforschung „ein Schattendasein“, die „Literaturlage gleicht einem Puzzle ... wobei Teile des Puzzles wiederum neue Fragen aufwerfen“ (Bentele 1992: 95; vgl. Gabriel 1992b, 1993; Plasser/Ullram 1994: 365). In der Bundesrepublik Deutschland scheint eine systematische und regelmäßige Untersuchung der Vertrauensbeziehungen der Bevölkerung gegenüber öffentlichen Einrichtungen und Institutionen bisher kaum als relevantes Forschungsfeld erkannt worden zu sein, wenn auch einige Beiträge vorliegen, die sich auch oder explizit mit dem Institutionenvertrauen auseinandersetzen (vgl. Miller/Listhaug 1990; Jäckel 1990; Döring 1990; Feist 1991; Klages 1990; Fuchs u.a. 1991; Gabriel 1992a, 1993; Krüger 1992; Gluchowski/Zelle 1992; Bentele 1992, 1993, 1994; Rattinger 1993; von Beyme 1993; Löwenhaupt 1993; Plasser/Ullram 1994; Koch u.a. 1994; Walz 1994, 1995 b/c). Nach wie vor gilt die Aussage von Backes (1990: 155), daß „obwohl - oder gerade weil - die Kategorie des `Vertrauens` für das Zusammenleben der Menschen eine so her-

ausragende Bedeutung besitzt, (...) die Politikwissenschaft diese Thematik nicht zum Gegenstand monographischer Bearbeitung gemacht“ hat. An dieser Situation hat sich auch seit 1990 nichts grundsätzlich geändert. Trotz umfangreicher Forschungsaktivitäten, welche die Wiedervereinigung und damit zusammenhängende Fragestellungen thematisieren (vgl. etwa Fuchs u.a. 1990; Bauer 1991; Herbert/Wildenmann 1991; Westle 1992; Jung u.a. 1992; Seibel 1992), liegen bislang kaum neuere Analysen zum Institutionenvertrauen vor (vgl. als Ausnahmen Gabriel 1993; Walz 1995a).

Für den Untersuchungszeitraum 1990 bis 1995 liegen etwa 30 Studien vor, die Fragen nach dem Vertrauen für öffentliche Einrichtungen und Institutionen enthalten und zur Analyse herangezogen werden können (vgl. Walz 1995c). Die Art und Anzahl der erfaßten öffentlichen Einrichtungen und Institutionen variiert jedoch von Studie zu Studie teils beträchtlich, so daß bestenfalls in allen Studien nur eine Art „institutioneller Kernbereich“ einbezogen wurde.²⁾ Auch die Verwendung unterschiedlicher Skalen- bzw. Antwortvorgaben erschweren - unnötigerweise - die Analysen ganz erheblich (vgl. Löwenhaupt 1993). Es ist jedoch möglich, hier empirische Befunde zum Ausmaß und zur Entwicklung des Vertrauens für sieben Institutionen (Bundesverfassungsgericht, Gerichte, Polizei, Bundeswehr, Bundestag, Bundesregierung, Parteien) zu präsentieren (vgl. Walz 1995c; Anmerkung 5).³⁾

3. Empirische Befunde

Die vergleichenden Analysen zum Ausmaß und zur Entwicklung des Vertrauens für politische Institutionen beginnt mit dem Jahre 1991. Für 1990 stellte das INFAS-Institut fest, daß die „Vertrauenswerte in Ost und West ... signifikant unterschiedlich“ sind (Feist 1991: 26). Während 1990 im Westen alle Institutionen mehrheitlich mit Vertrauen bedacht wurden, war „das Bild in der DDR von Mißtrauen und Ratlosigkeit geprägt“ (Feist 1991: 26). Nur das Fernsehen verfügte am Beginn des Vereinigungsprozesses auch im Osten mehrheitlich über Vertrauen (vgl. Walz 1995c). Zu Beginn des Wiedervereinigungsprozesses konnten also mehr Unterschiede als Gemeinsamkeiten in Ost und West ermittelt werden (vgl. Feist 1991; Krüger 1992; Gabriel 1993; Kaase 1993). Für die nachfolgende Präsentation empirischer Befunde werden a) Bundesverfassungsgericht und Gerichte, b) Polizei und Bundeswehr, als auch c) Bundestag, Bundesregierung bzw. Parteien jeweils getrennt behandelt. Bereits in einem anderen Zusammenhang konnte gezeigt werden, daß es theoretisch sinnvoll und empirisch begründet ist, die stärker in den Parteienwertbewerb involvierten Institutionen (Bundestag, Bundesregierung, Parteien) von eher dem Parteienwertbewerb

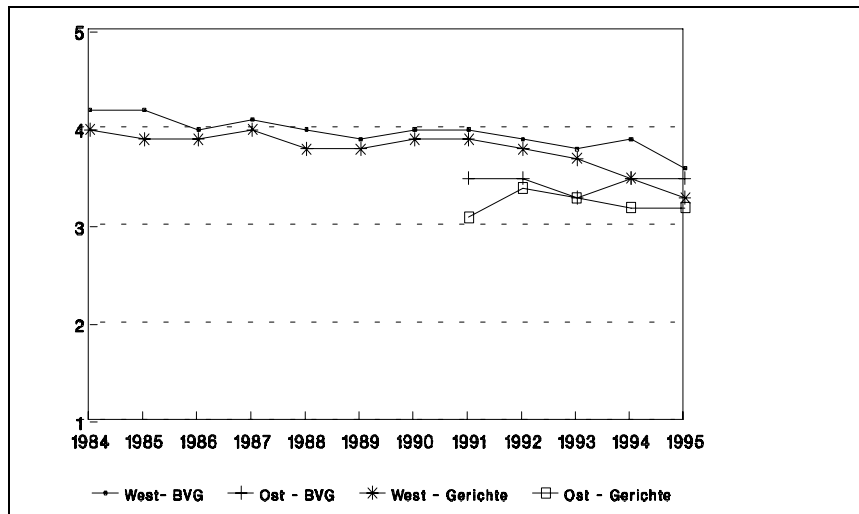
ausgelagerten Institutionen (Bundesverfassungsgericht, Gerichte, Polizei, Bundeswehr) zu trennen (vgl. Pickel/Walz 1995; Gabriel 1993).

3.1 Zum Vertrauen für das Bundesverfassungsgericht und die Gerichte

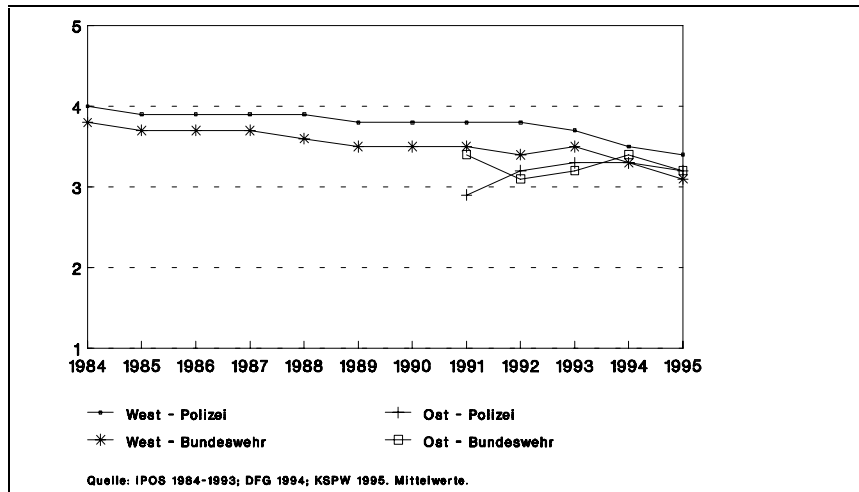
Das größte Ausmaß an Vertrauen der im Westen untersuchten öffentlichen Einrichtungen und Institutionen erreichte das Bundesverfassungsgericht. Zwischen 1984 und 1995 ist allerdings ein Rückgang im Ausmaß an Vertrauen festzustellen, d.h. von einem Mittelwert von 4,2 im Jahre 1984 auf einen Mittelwert von 3,6 im Jahre 1995.⁴⁾ Dieser Rückgang verlief jedoch nicht kontinuierlich, sondern nahm einen leicht zyklischen Verlauf, mit leichtem Anstieg der Vertrauenswerte in den jeweiligen Bundestagswahljahren und einem Vertrauensrückgang in den Jahren dazwischen. Auch in den fünf neuen Ländern genießt das Bundesverfassungsgericht im Vergleich zu den anderen Institutionen das größte Vertrauen, wenn auch das ostdeutsche Vertrauensniveau teils deutlich unter dem des Westens liegt. Die Entwicklung der Vertrauenswerte verlief im Osten, mit Ausnahme eines kleinen Vertrauenseinbruchs 1993, ausgesprochen stabil. Vergleicht man nun die Entwicklung in Ost- und Westdeutschland, so zeigen sich durchaus ähnliche Tendenzen, d.h. leichter Vertrauensrückgang zwischen 1992 und 1993 bzw. ein Anstieg des Vertrauens zwischen 1993 und 1994. Insbesondere seit 1993 nähern sich die Vertrauenswerte für das Bundesverfassungsgericht an (Differenz: 1993: 0,5; 1994: 0,4; 1995: 0,1). Die Entwicklung deutet also auf eine schrittweise Annäherung der Einstellungen der Bevölkerung in den alten und neuen Bundesländern gegenüber dem obersten deutschen Gericht hin.

Über ein relativ hohes Vertrauensniveau verfügen ebenfalls die Gerichte (Justiz) in Westdeutschland. Allerdings gehen auch hier die Vertrauenswerte - nicht kontinuierlich, sondern wiederum zyklisch - zwischen 1984 und 1995 um 0,7 Skalenpunkte zurück. Auch bei den Gerichten zeigt sich ein leichter Anstieg der Vertrauenswerte 1987 und 1990, allerdings nicht mehr im Jahre 1994 (was aber an den unterschiedlichen Skalen liegen kann). Im Osten Deutschlands stieg das Vertrauen für die Gerichte zunächst zwischen 1991 und 1992 stark an, um allerdings zwischen 1992 und 1994 wieder abzusinken. Insgesamt bewegt sich damit das Vertrauensniveau im Osten noch unter dem des Westens. Bei einer Analyse der Entwicklung der Vertrauenswerte in Ost und West zeigt sich ebenfalls vor allem seit 1992 eine ausgesprochen gleichförmiger Verlauf, wobei sich die Vertrauenswerte insbesondere seit 1993 deutlich nähern (1993: 0,4; 1994: 0,3; 1995: 0,1).

Grafik 1: Vertrauen für das Bundesverfassungsgericht und die Gerichte, 1984-1995



Grafik 2: Vertrauen für die Polizei und Bundeswehr, 1984-1995



3.2 Zum Vertrauen für die Polizei und Bundeswehr

Für die Polizei läßt sich konstatieren, daß das Ausmaß an Vertrauen zwischen 1984 und 1995 geringfügig, aber kontinuierlich von einem Mittelwert von 4,0 in 1984 auf einen Mittelwert von 3,7 in 1993 sinkt. Für 1994 und 1995 wurde ebenfalls geringere Vertrauenswerte ermittelt, sind aber aufgrund der unterschiedlichen Skalen nicht unmittelbar mit den Werten bis 1993 vergleichbar. Während in Westdeutschland das Vertrauen vor allem seit 1992 deutlich zurückgegangen ist, läßt sich für den Osten ein Anstieg der Vertrauenswerte zwischen 1991 und 1994 konstatieren. Nur zwischen 1994 und 1995 geht das Vertrauen wieder leicht zurück.

Die Entwicklung zwischen Ost und West zeigt für den gesamten Untersuchungszeitraum eine schrittweise und kontinuierliche Annäherung der Vertrauenswerte von einer Differenz von 0,9 Skalenpunkte in 1991 auf eine Differenz von nur noch 0,2 Skalenpunkte in 1995. Hier haben sich ebenfalls seit 1991 annähernd gleich Einstellungen herausgebildet.

Im Westen der Bundesrepublik zeigt sich zwischen 1984 und 1995 ein fast kontinuierlicher Vertrauensrückgang für die Bundeswehr, der vor allem 1994 und 1995 besonders deutlich wird. In Ostdeutschland geht zunächst zwischen 1991 und 1992 das Vertrauen zurück, um zwischen 1992 und 1994 jedoch deutlich anzusteigen. Zwischen 1994 und 1995 ist dann wieder ein leichter Vertrauensrückgang zu konstatieren. Auch hier ist die Entwicklung in beiden Landesteilen durchaus ähnlich: Vertrauensrückgang zwischen 1991 und 1992 und Vertrauensanstieg zwischen 1992 und 1993 (im Westen) bzw. zwischen 1992 und 1994 (im Osten). Waren 1991 die Vertrauenswerte in Ost und West noch sehr ähnlich, so vergrößerte sich in den Folgejahren die Differenz, um 1994 und 1995 allerdings wieder fast identische Werte anzunehmen. Anzumerken ist noch, daß die Bundeswehr seit 1994 im Osten über mehr Vertrauen verfügt, als im Westen.

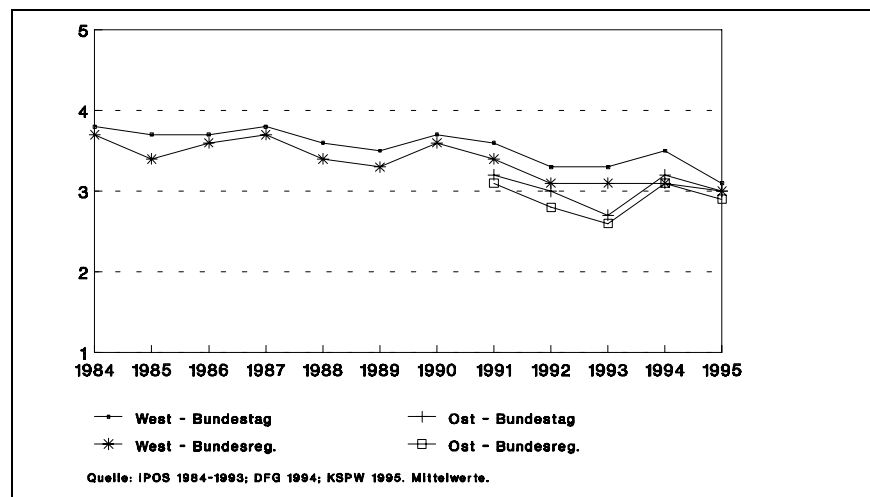
3.3 Zum Vertrauen für den Bundestag, Bundesregierung und Parteien

Das Vertrauen der westdeutschen Bevölkerung für den Bundestag hat ebenfalls seit 1984 schrittweise abgenommen, allerdings wiederum nicht kontinuierlich, sondern zyklisch, wobei im Vergleich zu anderen Institutionen diese Schwankungen wesentlich ausgeprägter sind. Jeweils in den Wahljahren 1987, 1990 und 1994 konnte ein deutlicher Aufwärtstrend für den Bundestag ermittelt werden, wobei jedoch 1990 und 1994 das Niveau aus 1987 nicht mehr erreicht werden konnte (vgl. Gabriel 1993). Tendenziell ging damit das Vertrauen der westdeutschen Bevölkerung in den Bundestag zurück. Wenn auch auf niedrigerem Niveau, so ist die ostdeutsche Entwicklung der im Westen nicht unähnlich, auch hier geht das Vertrauen zwischen 1991 und 1993 deutlich zurück,

steigt zwischen 1993 und 1994 jedoch wieder an. Zwischen 1994 und 1995 gehen die Vertrauenswerte wieder leicht zurück. Zusammenfassend ist festzuhalten, daß tendenziell das Vertrauen für den Bundestag sukzessive immer geringer wird. Ferner ist anzumerken, daß die Entwicklung in Ost und West sehr deutliche Übereinstimmungen aufweist und sich die Vertrauenswerte ebenfalls wieder seit 1993 stark annähern.

Eine sehr ähnliche Entwicklung wie beim Bundestag zeigt sich auch in den alten Bundesländern für die Bundesregierung. Auch hier zeichnet sich nicht etwa ein kontinuierlicher Vertrauensrückgang ab, sondern ein deutlich zyklisch ausgeprägter Kurvenverlauf, d.h. Vertrauensanstieg in den Wahljahren (nicht 1994) und Vertrauensverlust in den Jahren dazwischen.

Grafik 3: Vertrauen für Bundestag und Bundesregierung, 1984-1995



Insgesamt ist das Vertrauen um 0,7 Skalenpunkte zwischen 1984 und 1995 zurückgegangen. Während das Vertrauensniveau im Osten sich ebenfalls wieder unter dem westdeutschen Niveau bewegt, zeigt der Kurvenverlauf jedoch gewisse Ähnlichkeiten: Vertrauensrückgang zwischen 1991 und 1993, Vertrauensanstieg zwischen 1993 und 1994 sowie einen nochmals leichten Rückgang zwischen 1994 und 1995. Während sich dabei die Differenzen im Zeitraum von 1991 bis 1993 zwischen Ost- und Westdeutschland sogar vergrößerten, konnten für 1994 identische Vertrauenswerte

ermittelt werden, und auch 1995 variieren die Vertrauenswerte nur geringfügig. Alles in allem ist für die Bundesregierung ein vor allem seit 1990 deutlich geringeres Vertrauensniveau zu konstatieren, während allerdings die Befunde in Ost und West - nach durchaus ähnlichem Verlauf - konvergierten.

Bei der Präsentation von empirischen Befunden zum Ausmaß und zur Entwicklung des Vertrauens für die politischen Parteien muß für die achtziger und frühen neunziger Jahre auf Ergebnisse des EMNID-Instituts zurückgegriffen werden, da in den IPOS-Studien das Vertrauen für politische Parteien erst ab 1992 erhoben wurde. Für den Zeitraum 1979 bis 1993 kann ebenfalls nicht von einem kontinuierlichen Vertrauensrückgang gesprochen werden, vielmehr verlief auch hier die Entwicklung diskontinuierlich. Allerdings geht das Vertrauen seit 1983 zunächst langsam, seit etwa 1986 deutlich zurück. 1993 hatten nur noch etwa ein Viertel der Befragten im Westen und nur noch 13 Prozent der Befragten im Osten Vertrauen zu den politischen Parteien.

Tabelle 1: Vertrauen für die politischen Parteien, 1979-1993

	West									Ost
Jahr	1979	1981	1982	1983	1984	1986	1988	1990	1993	1993
Parteien	43	40	39	50	47	45	38	26	26	13

Quelle: EMNID Umfrage+Analyse Nr.11-12/1993. Prozentangaben.

Auch eine differenziertere Analyse der Entwicklung zwischen 1990 und 1994 zeigt für Westdeutschland einem Vertrauensrückgang (vgl. Walz 1995c), während das Vertrauensniveau im Osten nur zwischen 1992 und 1993 leicht zurückging, um zwischen 1993 und 1994 sogar wieder anzusteigen. Die Entwicklung in Ost- und Westdeutschland zeigt also folgendes Bild: Während 1992 und 1993 das ostdeutsche Vertrauensniveau noch unter dem des Westens lag, wurden für 1994 identische Vertrauenswerte ermittelt, allerdings im Vergleich zu anderen Institutionen auf einem sehr niedrigen Niveau. Die politischen Parteien stoßen damit sowohl im Westen als auch im Osten überwiegend auf Mißtrauen.

3.4 Zusammenfassung und Diskussion der empirischen Befunde

Hinsichtlich des Ausmaßes und der Entwicklung der Vertrauensbeziehungen der Bevölkerung in Ost und West gegenüber politischen Institutionen lassen sich mehrere Muster identifizieren: Zwischen 1984 und 1995 verloren in den alten Bundesländern alle in die Untersuchung einbezogenen Institutionen an Vertrauen. Am stärksten war der Vertrauensrückgang für den Bundestag und die Bundesregierung, gefolgt von den Parteien, der Polizei und der Bundeswehr. Dieser Vertrauensrückgang verlief ferner bei

einigen Institutionen kontinuierlich (z.B. Polizei und Parteien), bei den anderen Institutionen eher zyklisch (z.B. Bundesverfassungsgericht, Gerichte, Bundestag, Bundesregierung), wobei vor allem in den jeweiligen Bundestagswahljahren ein teils deutlicher Vertrauensanstieg konstatiert werden konnte.

Für Ostdeutschland ist festzustellen, daß zu Beginn des Wiedervereinigungsprozesses für alle Institutionen geringere Vertrauenswerte ermittelt werden konnten als im Westen. Allerdings ist ein genereller Vertrauensverlust für die jeweiligen Institutionen - wie im Westen - nicht festzustellen, vielmehr zeichnen sich unterschiedliche Entwicklungsmuster ab: Eine eher konstante Entwicklung - wenn auch auf unterschiedlichem Niveau - zeigt sich beim Bundesverfassungsgericht und den Parteien, eher zyklische Bewegungen für die anderen Institutionen, wobei die Vertrauenswerte nach diskontinuierlichem Verlauf für die Gerichte und die Polizei leicht angestiegen bzw. für den Bundestag, die Bundesregierung und die Bundeswehr eher gefallen sind.

Da aufgrund unterschiedlicher Skalenvorgaben die Vergleichbarkeit der Befunde (seit 1994) nicht unmittelbar gegeben ist, interessiert vor allem noch die Frage, ob sich im Verlauf des Wiedervereinigungsprozesses sukzessive annähernd übereinstimmende Einstellungen in beiden Landesteilen herausgebildet haben, oder ob die Unterschiede nach wie vor bestehen bzw. ob sie sich gar vergrößert haben (vgl. Tabelle 2).

Tabelle 2: Institutionenvertrauen in Ost- und Westdeutschland, 1991-1995

Institution	1991	1992	1993	1994	1995
BVerfG	0,5	0,4	0,5	0,4	0,1
Justiz/Gerichte	0,8	0,4	0,4	0,3	0,1
Polizei	0,9	0,6	0,4	0,2	0,2
Bundeswehr	0,1	0,3	0,3	-0,1	-0,1
Bundestag	0,4	0,3	0,6	0,3	0,1
Bundesregierung	0,3	0,3	0,5	0,0	0,1
Parteien	--	0,3	0,3	0,0	--

Quelle: IPOS 1991, 1992, 1993; DFG 1994; ALLBUS 1994; KSPW 1995 (vgl. Walz 1995c).
Ein Minuszeichen bedeutet, daß im Osten mehr Vertrauen vorhanden ist, als im Westen.

Dabei zeigen sich interessante Befunde: Während sich die Mittelwertdifferenzen für die Institution Bundeswehr schon zu Beginn des Wiedervereinigungsprozesses kaum unterschieden und auch im Verlaufe des Untersuchungszeitraumes kaum differierten, nähern

sich die Mittelwertdifferenzen für alle anderen Institutionen seit 1991 schrittweise an, um 1995 nur noch um 0,1 bzw. 0,2 Skalenpunkte zu differieren. Tabelle 2 gibt zusammenfassend das Ausmaß und die Entwicklung der Differenzen der Vertrauenswerte zwischen Ost- und Westdeutschland im Zeitraum zwischen 1991 bis 1995 wieder. Während sich also vielfach die Vertrauenswerte deutlich nähern und auch die Entwicklung der Vertrauenswerte durchaus Ähnlichkeiten offenbart, stimmt allenfalls der Trend im Westen zu immer weniger Vertrauen bedenklich.

4. Abschließende Bemerkungen

Gegenstand dieser Arbeit war es, empirische Befunde zum Ausmaß und zur Entwicklung des Vertrauens der Bevölkerung in beiden Teilen Deutschlands für politische Institutionen zu präsentieren. Diese Analysen sind sowohl unter politischen wie auch wissenschaftlichen Gesichtspunkten von Bedeutung, da das Gelingen der politischen Integration Deutschlands auch davon abhängt, daß sich in Ost- und Westdeutschland - nach einer gewissen Übergangsphase - annähernd übereinstimmende Einstellungen gegenüber öffentlichen Einrichtungen und Institutionen herausbilden. Als eine wichtige Einstellungsart wurde in diesem Beitrag das Vertrauen der Bevölkerung in einige zentrale Institutionen des politischen Systems der Bundesrepublik Deutschlands näher untersucht. Während es zu Beginn des Untersuchungszeitraumes hinsichtlich der Einstellungen der ost- und westdeutschen Bevölkerung zu den politischen Institutionen mehr Unterschiede als Gemeinsamkeiten gab (vgl. Feist 1991; Krüger 1992; Gabriel 1993), hat sich das Bild im Verlaufe der letzten fünf Jahre deutlich hin zu einer Annäherung der Einstellungen in beiden Landesteilen verändert. Bei der überwiegenden Zahl der untersuchten Institutionen sind bereits heute praktisch gleiche Vertrauenswerte in Ost und West zu konstatieren. Damit konnten m.E. nennenswerte Fortschritte beim Abbau der kulturellen Spaltung Deutschlands erzielt werden, jedenfalls soweit das Vertrauen der Bevölkerung in einige öffentliche Einrichtungen und Institutionen zur Debatte steht.

Eine demokratietheoretische Interpretation der Befunde hängt schließlich nicht zuletzt auch von der Wahl des entsprechenden Bezugsrahmens ab (vgl. Gamson 1968; Parry 1976; Sniderman 1981; zusammenfassend Gabriel 1993). Ein vorbehaltloses Vertrauen für politische Institutionen ist dabei weder für die Befragten in West- noch in Ostdeutschland typisch (vgl. Gamson 1968). Ebenso wenig sind alle Institutionen mit einem generellen Vertrauensschwund bzw. überwiegend mit Mißtrauen konfrontiert (vgl. Parry 1976; Döring). In Übereinstimmung mit dem von Sniderman (1981) entwickeltem Konzept eines „Balanced Judgment“ variiert vielmehr das Vertrauen zu einzelnen

Institutionen in beiden Teilen Deutschlands. Ferner gibt es sowohl vertrauensvolle als auch mißtrauische Befragte, und es lassen sich im Zeitverlauf teils deutliche Schwankungen konstatieren. Zudem ist festzustellen, daß die dem Parteienwettbewerb eher ausgelagerten Institutionen wesentlich weniger von einem Vertrauensrückgang bzw. von Schwankungen betroffen sind, als die eher in den Parteienwettbewerb involvierten Institutionen des demokratischen Regimes, was aufgrund ihrer systematischen Stellung nicht weiter überrascht.⁵⁾ Während zu Beginn des Wiedervereinigungsprozesses vor allem das Gefälle zwischen West und Ost (vgl. Feist 1991; Gabriel 1993) als problematisch perzipiert worden war, zeigen sich heute deutliche Übereinstimmungen in der Entwicklung der Vertrauenswerte, sowie vielfach auch eine sukzessive Annäherung der Einstellungen der Bevölkerung in beiden Teilen Deutschlands. Dabei handelt es sich m.E. zweifelsohne um einen wichtigen Baustein für eine gemeinsame politische Kultur in Deutschland.⁶⁾ Bedenklich stimmt aber, daß die Vertrauenswerte - vor allem im Westen - langsam immer geringer werden. Ob es sich dabei nur um einen Skalierungseffekt handelt, läßt sich jedoch mit den Mitteln der empirischen Politikforschung ebensowenig bestimmen wie die Frage, wann möglicherweise ein inakzeptables Niveau an Systemunterstützung erreicht sein könnte. Jedoch deuten die Befunde darauf hin, daß nicht nur mit kontinuierlichem Vertrauensverlust, sondern vielfach auch wieder mit einem Anstieg der Vertrauenswerte gerechnet werden kann, und daß dieses Auf und Ab der Demokratie in Deutschland bis dato gut bekommen ist und auch weiterhin als funktional für einen angemessenen politischen Prozeß im vereinten Deutschland interpretiert werden kann.

Korrespondenzadresse

*Dieter Walz
Regnitzstr. 37
91301 Forchheim
E-Mail: d.walz@msh.sub.org*

Anmerkungen

1) Ein Problem der Ausführungen von Döring (1990) u.a. liegt m.E. in der mangelnden Differenzierung zwischen „people running these institutions“ und den Institutionen selbst (vgl. Lipset/Schneider 1983). Die skizzierten Überlegungen beziehen sich ja primär auf das Vertrauen für die Regierenden, nicht jedoch auf die Institutionen bzw. auf die institutionelle Ordnung insgesamt. Auch wenn man konzidiert, daß eine entsprechend Einstellung zu den „authorities“ funktional sein kann - wie verhält es sich mit dem Vertrauen zu den Gerichten

oder zur Polizei? Indiziert auch hier überwiegend Mißtrauen eine „gesunden Zustand“ der politischen Kultur? Vgl. dazu auch die Hinweise in Anmerkung 5.

2) Insgesamt wurden zwischen 1990 und 1995 die Vertrauenswerte für knapp 40 verschiedene Institutionen ermittelt. Allerdings wurden zahlreiche Institutionen nur einmal bzw. sehr unregelmäßig untersucht, so daß eine kontinuierliche Untersuchung zum Ausmaß und zur Entwicklung des Vertrauens nur für einige wenige Institutionen möglich ist. Folgende Studien stehen für die Analysen grundsätzlich zur Verfügung, wobei kein Anspruch auf Vollständigkeit erhoben werden soll: IPOS-Studien 1984-1993; ALLBUS 1994; ISSP 1991; Rattinger-Falter Panel 1990-1992, DFG-Studie 1994; WVS 1990; KSPW-Studie 1995; EMNID 1979-1993; INFAS 1990; DJI 1990; vgl. Walz 1995c.

3) Die Auswahl der untersuchten Institutionen ergibt sich einerseits aus der Verfügbarkeit entsprechender Untersuchungen (vgl. die Hinweise in Anmerkung 2), andererseits handelt es sich aber auch um zentrale Einrichtungen des demokratischen Regimes, wobei vor allem die Trennung von Institutionen, die dem Parteienwettbewerb ausgelagert bzw. in den Parteienwettbewerb eher involviert sind, von Bedeutung ist.

4) Bei den folgenden Analysen wurde eine 5er Skala (1=kein Vertrauen; 5=großes Vertrauen) verwendet. Allerdings ist zu beachten, daß zwischen 1990 und 1993 ursprünglich eine 11er Skala (IPOS) und - teils 1994 (ALLBUS) als auch 1995 (KSPW) eine 7er Skala verwendet wurde. Diese Skalen wurden jeweils zu einer 5er Skala recodiert. Damit ist es allerdings nur unter Vorbehalt möglich, die Mittelwerte vor und nach 1993 zu vergleichen. Möglich ist jedoch ein Vergleich der Mittelwertdifferenzen (vgl. Tabelle 2). Die Skalen wurden folgendermaßen recodiert:

Tabelle 3: Zur Recodierung unterschiedlicher Skalenvorgaben

Skala	Extrem		Mitte		Extrem
5er	1	2	3	4	5
7er	1	2-3	4	5-6	7
11er	1-2	3-4	5-6-7	8-9	10-11

Von der Recodierung sind also die IPOS-Studien 1991, 1992 und 1993, als auch der ALLBUS 1994 und die KSPW-Studie 1995 betroffen.

5) Im Rahmen dieser Arbeit kann zudem nicht darauf eingegangen werden, ob mit den vorliegenden Fragen nach dem *Vertrauen für Institutionen* eine - wie im theoretischen Konzept formuliert - eher output-unabhängige Bewertung der jeweiligen Institutionen, oder aber eine durchaus auch output-abhängige, d.h. instrumentelle Bewertung der jeweiligen Institutionen sowie bspw. die Perzeption und Beurteilung der Performanz der in diesen Institutionen tätigen Akteure einfließt. Vorstellbar ist in diesem Zusammenhang, daß beim Vertrauen für die Bundesregierung bspw. auch die Zufriedenheit mit der wahrgenommenen Problemlösungskompetenz oder die Sympathien für einzelne Regierungsmitglieder in die Vergabe unterschiedlicher Vertrauenswerte einfließen, obwohl nicht die „people running

these institutions“, sondern die Institution selbst bewertet werden soll. Aufgrund möglicher Konfundierungseffekte ergeben sich u.U. natürlich erhebliche Auswirkungen für die Interpretation unterschiedlicher Vertrauenswerte. Solange keine Untersuchungen vorliegen, die derartige Operationalisierungsprobleme lösen, bleibt eine Interpretation der Befunde immer problematisch (vgl. Westle 1989; Löwenhaupt 1993). Insofern steht hier auch eher die Frage nach der Angleichung der Vertrauenswerte in Ost und West im Mittelpunkt des Interesses, und nicht so sehr die demokratietheoretische Bedeutung unterschiedlicher Niveaus.

6) Hinsichtlich der Entwicklung der Struktur des Institutionenvertrauens konnten ebenfalls im Verlaufe des Wiedervereinigungsprozesses große Übereinstimmungen erzielt werden. Auch bei Determinanten und Konsequenzen des Vertrauens sind sich Ost- und Westdeutsche nicht mehr allzu fremd (vgl. dazu ausführlich Walz 1995c).

Literatur

Abramson, P. A., 1972: Political Efficacy and Political Trust among Black Schoolchildren. Two Explanations. *The Journal of Politics* Vol. 34: 1243-1275.

Almond, G. A./Verba, S., 1965: *The Civic Culture. Political Attitudes and Democracy in Five Nations*. Princeton: Princeton University Press.

Almond, G. A./Powell, G. B., 1978: *Comparative Politics*. Boston: Little, Brown and Company.

Backes, U., 1990: Diskussionsbericht. S. 155-161 in: Haungs, Peter (Hrsg.), *Politik ohne Vertrauen?* Baden-Baden: Nomos Verlagsgesellschaft.

Barber, B., 1983: *The Logic and Limits of Political Trust*. New Brunswick: Rutgers University Press.

Bauer, P., 1991: Politische Orientierungen im Übergang. Eine Analyse politischer Einstellungen der Bürger in West- und Ostdeutschland 1990/91. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* Jg. 43: 433-453.

Bentele, G., 1992: *Öffentliches Vertrauen. Eine Literaturlauswertung*. Erstellt im Auftrag des Presse- und Informationsamtes der Bundesregierung. Bamberg.

Bentele, G., 1993: Immer weniger öffentliches Vertrauen. *Bertelsmann Briefe* Mai 1993: 39-43.

Bentele, G., 1994: Öffentliches Vertrauen - normative und soziale Grundlagen für Public Relations. S. 131-158 in: Armbricht, W./Zabel, Ulf J. (Hrsg.), *Normative Grundlagen der Public Relations*. Opladen: Westdeutscher Verlag.

Beyme, K. v., 1993: Der Parteienstaat und die Vertrauenskrise in der Politik. S. 23-42 in: Unseld, S. (Hrsg.), *Politik ohne Projekt? Nachdenken über Deutschland*. Frankfurt/M: Suhrkamp.

- Citrin, J., 1974: Comment: The Political Relevance of Trust in Government. *The American Political Science Review* Vol. 68/3: 973-988.
- Coleman, J. S., 1982: Systems of Trust. A rough theoretical framework. *Angewandte Sozialforschung* Jg. 10/1-2: 277-299.
- Döring, H., 1990: Aspekte des Vertrauens in Institutionen. Westeuropa im Querschnitt der Internationalen Wertestudie. *Zeitschrift für Soziologie* 2: 73-89.
- Easton, D., 1975: A Re-Assessment of the Concept of Political Support. *British Journal of Political Science* 5: 435-457.
- EMNID-Institut, 1993: Umfrage + Analyse Nr. 11-12: 82-91.
- Feist, U., 1991: Zur politischen Akkulturation der vereinten Deutschen. Eine Analyse aus Anlaß der ersten gesamtdeutschen Bundestagswahl. *Aus Politik und Zeitgeschichte* B 11-12: 21-32.
- Fuchs, D., 1981: Dimensionen politischer Unterstützung. S. 208-225 in: Klingemann, H.-D./Kaase, M. (Hrsg.): *Politische Psychologie*. PVS-Sonderheft 12. Opladen: Westdeutscher Verlag.
- Fuchs, D./Klingemann, H.-D./Schöbel, C., 1991: Perspektiven der politischen Kultur im vereinigten Deutschland. *Aus Politik und Zeitgeschichte* B32: 35-46.
- Gabriel, O. W., 1986: Politische Kultur, Postmaterialismus und Materialismus in der Bundesrepublik Deutschland. Opladen: Westdeutscher Verlag.
- Gabriel, O. W., 1992a: Politische Einstellungen und politische Kultur. S. 95-131 in: Gabriel, O. W. (Hrsg.), *Die EG-Staaten im Vergleich. Strukturen - Prozesse - Politikinhalt*. Opladen: Westdeutscher Verlag.
- Gabriel, O. W., 1992b: Values, Trust and Efficacy. Paper for the Fifth Meeting of the Subgroup 'Impact of Values' European Science Foundation Scientific Program on Beliefs in Government. Colchester, Essex, April 22-26, 1992.
- Gabriel, O. W., 1993: Institutionenvertrauen im vereinigten Deutschland. *Aus Politik und Zeitgeschichte* B43: 3-12.
- Gamson, W. A., 1968: *Power and Discontent*. Homewood, Ill.: Dorsey Press.
- Gamson, W. A., 1971: Political Trust and its Ramification. S. 41-55 in: Abcarian, G./Soule, J. (Hrsg.), *Social Psychology and Political Behaviour*. Columbus, Ohio: Merrill Press.
- Gluchowski, P./Zelle, C., 1992: Demokratisierung in Ostdeutschland. Aspekte der politischen Kultur in der Periode des Systemwechsels. S.231-274 in: Gerlich, P./Plasser, F./Ulram, P. A. (Hrsg.), *Regimewechsel. Demokratisierung und politische Kultur in Ost-Mitteleuropa*. Wien: Böhlau Verlag.
- Herbert, W./Wildenmann, R., 1991: Deutsche Identität. Die subjektive Verfassung der Deutschen vor der Vereinigung. S.71-98 in: Wildenmann, R. (Hrsg.), *Nation und Demokratie. Politisch-strukturelle Gestaltungsprobleme im neuen Deutschland*. Baden-Baden: Nomos.

- Jäckel, H., 1990: Über das Vertrauen in der Politik. S. 31-42 in: Haungs, P. (Hrsg.), Politik ohne Vertrauen. Baden-Baden: Nomos Verlagsgesellschaft.
- Jung, M./Roth, D., 1992: Politische Einstellungen in Ost- und Westdeutschland seit der Bundestagswahl 1990. Aus Politik- und Zeitgeschichte B19: 3-16.
- Kaase, M., 1993: Innere Einheit. S. 372-383 in: Weidenfeld, W./Korte, K.-R. (Hrsg.), Handbuch zur deutschen Einheit. Frankfurt am Main: Campus Verlag.
- Klages, H., 1990: Vertrauen und Vertrauensverlust. S. 43-60 in: Haungs, P. (Hrsg.), Politik ohne Vertrauen? Baden-Baden: Nomos Verlagsgesellschaft.
- Koch, A./Gabler, S./Braun, M., 1994: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1994. ZUMA-Arbeitsbericht 94/11. Mannheim.
- Krüger, W., 1992: Vertrauen in Institutionen. S. 100-113 in: Deutsches Jugendinstitut (Hrsg.), Schüler an der Schwelle zur Deutschen Einheit. Politische und persönliche Orientierungen in Ost und West. Opladen: Leske + Budrich.
- Lipset, S. M./Schneider, W., 1983: The Decline of Confidence in American Institutions. Political Science Quarterly Vol. 90/3: 379-402.
- Listhaug, O., 1984: Confidence in Institutions: Findings from the Norwegian Values Study. Acta Sociologica Vol. 27/2: 111-122.
- Löwenhaupt, S., 1993: Vertrauen in Institutionen und Einstellungen der Bürger zur Verwaltung. Bericht zur Datenlage. in: Verwaltungswissenschaftliche Beiträge der Universität Bamberg, Nr. 29. Hrsg. von Prof. Dr. H.-U. Derlin.
- Luhmann, N., 1968: Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität. Stuttgart: Enke.
- Miller, A. H., 1974a: Political Issues and Trust in Government: 1964-1970. American Political Science Review Vol. 68/3: 951-972.
- Miller, A. H., 1974b: Rejoinder to 'Comment' by Jack Citrin: Political Discontent or Ritualism. American Political Science Review Vol. 68/3: 989-1001.
- Miller, A. H./Listhaug, O., 1990: Political Parties and Confidence in Government: A Comparison of Norway, Sweden and the United States. British Journal of Political Science 20: 357-386.
- Parry, G., 1976: Trust, Distrust and Consensus. British Journal of Political Science Vol. 6/1: 129-142.
- Pickel, G./Walz, D., 1995: Politikverdrossenheit. Dimensionen und Ausprägungen politischer Verdrossenheit in Ost- und Westdeutschland. Unveröffentl. Arbeitspapier. Bamberg.
- Plasser, F./Ulram, P., 1994: Politische Systemunterstützung und Institutionenvertrauen in den OZE-Staaten. Österreichische Zeitschrift für Politikwissenschaft Jg. 23/4: 365-379.

- Preisendörfer, P., 1987: Vertrauenskrise in der modernen Gesellschaft? *Angewandte Sozialforschung* Jg. 14/1: 31-39.
- Rattinger, H., 1993: Abkehr von den Parteien? Dimensionen der Parteienverdrossenheit. *Aus Politik und Zeitgeschichte* B 11: 24-35.
- Schacht, K., 1991: Droht eine Polarisierung der deutschen Gesellschaft? S.14 in: *Die tageszeitung*, 7.1.1991.
- Schneeberger, A., 1982: Vertrauen als Eigenschaft von Interaktionssystemen. Entwurf eines theoretischen Bezugsrahmens. *Angewandte Sozialforschung* Jg. 10/3: 301-307.
- Seibel, W., 1992: Notwendige Illusionen: Der Wandel des Regierungssystems im vereinten Deutschland. *Journal für Sozialforschung* Jg. 32/3-4: 337-361.
- Sniderman, P. M., 1981: *A Question of Loyalty*. Berkley: University of California Press.
- Stockes, D. E., 1962: Popular Evaluations of Government: An Empirical Assessment. S. 61-73 in: Cleveland, H./Laswell, H. D. (Hrsg.), *Ethics and Bigness*. New York: Harper.
- Walz, D., 1994: Parteiidentifikation und Institutionenvertrauen in der Bundesrepublik Deutschland. Paper zum DVPW-Kongreß "Einigung und Zerfall. Deutschland und Europa nach dem Ende des Ost-West Konflikts" vom 25.-28.8.1994 in Potsdam.
- Walz, D., 1995a: Parteiidentifikation, Institutionenvertrauen und Systemunterstützung in Ost- und Westdeutschland (i.E.).
- Walz, D., 1995b: Vertrauen für politische Institutionen. Unveröff. Arbeitspapier zur Dissertation.
- Walz, D., 1995c: Einstellungen zu den politischen Institutionen. Paper für die Arbeitsgruppe III/3 - „Mikropolitische Aspekte des Transformationsprozesses“ des KSPW-Projekts „Wandel politischer Orientierungen und Verhaltensmuster in Deutschland seit dem Beitritt der fünf neuen Länder“.
- Westle, B., 1989a: *Politische Legitimität - Theorien, Konzepte, empirische Befunde*. Baden-Baden: Nomos Verlagsgesellschaft.
- Westle, B., 1989b: Einstellungen zu den politischen Parteien und der Demokratie in der Bundesrepublik Deutschland. S. 223-239 in: Wildenmann, R. (Hrsg.), *Volksparteien. Ratlose Riesen?* Baden-Baden: Nomos Verlagsanstalt.
- Westle, B., 1990: Zur Akzeptanz der politischen Parteien und der Demokratie in der Bundesrepublik Deutschland. S. 253-298 in: Kaase, M./Klingemann, H.-D. (Hrsg.), *Wahlen und Wähler*. Opladen: Westdeutscher Verlag.
- Westle, B., 1992: Strukturen nationaler Identität in Ost- und Westdeutschland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* Jg. 44/3: 461-488.

FRAGEN ZU FRAGEN: INSTRUMENTBEDINGTE ANTWORTVERZERRUNGEN?

WILLY H. EIRMBTER UND RÜDIGER JACOB

Neben personenbezogenen Artefakten bei Befragungen werden in der Literatur stets auch instrumentbedingte Verzerrungen des Antwortverhaltens diskutiert. Zur Klärung einiger dieser Fragen (Reihenfolge der Antwortvorgaben, unterschiedliche Benennung der Skalenpunkte bei Likert-Skalen) wurde im Rahmen des ZUMA-Sozialwissenschaften-BUS II/1995 bei einer Befragung über AIDS ein Methodensplit durchgeführt. Bivariate Analysen und Drittfaktor-Kontrollen mit den Merkmalen Alter, Geschlecht, Bildungsstatus und Wohnortgröße haben gezeigt, daß eine unterschiedliche Reihung der Antwortvorgaben wie auch die Verwendung unterschiedlicher Benennungen bei Likert-Skalen keinen systematischen Einfluß auf das Antwortverhalten haben.

In the literature there are two general sources of response effects discussed: features of persons (interviewers and respondents) and features of the questionnaire. This article deals with possible response effects due to the questionnaire. In a survey concerning Aids which was part of the ZUMA-Sozialwissenschaften-BUS II/1995, two versions of a questionnaire were developed changing the order of answers in closed questions and using two different wordings in Likert-scales. Bivariate analysis as well as multivariate analysis (age, gender, education, size of place of residence as controlling factors) have shown that there is no systematic effect of a different wording or the order of answers.

1. Artefakte bei Interviews

Bekanntlich ist das mündliche Interview keine neutrale Erhebungsmethode, sondern grundsätzlich einer Reihe von Verzerrungen oder Reaktivitäten ausgesetzt. Üblicherweise werden hierbei in den Lehrbüchern unterschieden: Interviewer-, Befragten- und Instrumenteffekte (vgl. dazu exemplarisch Schnell/Hill/Esser 1995; Groves 1989; Bradburn 1983). Ergänzen ließen sich hier außerdem Situationseffekte. Der vorliegende Beitrag befaßt sich in diesem Zusammenhang mit einigen möglichen Instrumenteffekten. In-

strument-induzierte Verzerrungen, die in der Literatur immer wieder genannt werden, sind: Verwendung bestimmter Formulierungen, Länge der Fragen, Stellung der Fragen im Fragebogen, Zahl und Reihenfolge der vorgegebenen Antwortkategorien (vgl. Groves 1989: 450-484; Bradburn 1983: 298-308). Untersucht wurden mögliche Effekte, die auf differentielle Formulierungen bei Likert-Skalen und die Reihung der Antwortvorgaben zurückzuführen sind.

Bei solchen instrumentbedingten Antwortverzerrungen spielen aber typischerweise auch inhaltsunabhängige Reaktionseinstellungen, sog. Response-Sets, eine wichtige Rolle, also die Tendenz von Befragten, Fragen nicht aufgrund deren jeweiliger Inhalte, sondern infolge vorgefaßter Dispositionen zu beantworten. Entsprechende Instrumenteffekte lassen sich mithin genau genommen nicht isoliert von Befragteneffekten analysieren.

Der wichtigste Befragteneffekt in diesem Zusammenhang ist die Neigung zur inhaltsunabhängigen Zustimmung ("Akquieszenz", vgl. dazu Esser 1977). Die Wahl der jeweiligen erstgenannten Antwortkategorie ist ein Spezialfall dieses Response-Sets. Dahinter steht die Überlegung, daß die Reihenfolge der Antwortvorgaben als Hinweis auf die gesellschaftlich gängige Position oder die "richtige" Antwort interpretiert werden kann und dadurch die zuerst genannten Kategorien höhere Zustimmungsquoten produzieren.

Ähnliches ist zu vermuten hinsichtlich der Antwortreaktionen auf vorgegebene verbale Kategorien, wie sie vielfach in 5-poligen Likert-Skalen verwendet werden ("stimme sehr zu", "stimme eher zu", "teils-teils", "lehne eher ab", "lehne sehr ab"). Hierzu wurde in der Literatur u.a. auch die Frage aufgeworfen, ob diese Form der Skalierung eindimensional ist (vgl. dazu Duncan/Stenbeck 1987). In diesem Zusammenhang wurde bereits bei der ersten Fassung des Fragebogens 1990 das Problem diskutiert, daß Zustimmung etwas qualitativ anderes ist als Ablehnung und sich nicht auf dem gleichen semantischen Kontinuum bewegt. Die explizite Ablehnung einer Aussage kann mit höheren psychischen Kosten verbunden sein als deren bloße "Nicht-Zustimmung" und damit in besonderer Weise Bejahungstendenzen provozieren. Um einem solchen möglichen Effekt zu begegnen, müßte eine eindimensionale Likert-Skala folgerichtig aus den verbalen Kategorien "stimme sehr zu", "stimme eher zu", "teils-teils", "stimme eher nicht zu" und "stimme überhaupt nicht zu" konstruiert werden.

Dem Argument, daß Zustimmung und Ablehnung semantisch differentielle Konzepte sind, ist nicht zu widersprechen. Allerdings hätte eine entsprechende Neu-Konstruktion der Likert-Skalierung Probleme hinsichtlich der Vergleichbarkeit der so produzierten Verteilungen mit Vorläuferbefragungen geschaffen, weshalb die ursprüngliche Benen-

nung der Kategorien beibehalten wurde. Der Methodensplit nun bot die Möglichkeit, empirisch zu prüfen, ob diese eben skizzierten Vermutungen zutreffen, ob also etwa die sprachlogischen Unterschiede zwischen Zustimmung und Ablehnung von den Befragten auch in dieser Art konnotiert werden und dann auch ein unterschiedliches Antwortverhalten zur Folge haben.

Zur Erklärung von Akquieszenz wird primär auf Persönlichkeitsmerkmale von Befragten verwiesen (vgl. Couch/Kensiston 1960; Hare 1960). Die Tendenz zur Bejahung von Fragen wird als Behauptungsstrategie von unterprivilegierten Befragten mit geringer Ich-Stärke und wenig ausgeprägter Gestaltungskompetenz in unüberschaubaren Situationen aufgefaßt. Dabei wird in der Literatur insbesondere der sozioökonomische Status bzw. der Bildungsstatus von Personen als zentraler Indikator für Akquieszenz diskutiert (Schnell/Hill/Esser 1995: 329; Groves 1989: 441 f). "Unterschichtpersonen neigen zum ACQ-RS [Akquieszenz-Response-Set] - und dies teilweise auch völlig situationsunabhängig" (Esser 1977: 259).

2. Methodik

Um die These der generellen Existenz inhaltsunabhängiger Reaktionen auf Fragen und der erhöhten Verbreitung von solchen Response-Sets in statusniedrigeren Gruppen empirisch prüfen zu können, wurde bei einer Einschaltung im ZUMA-Sozialwissenschaften-BUS II/1995 unter anderem ein Methodensplit durchgeführt. Die Forschungsfrage lautete: Lassen sich unterschiedliche Antwortverteilungen bei Fragen mit unterschiedlicher Reihung der Antwortvorgaben bzw. bei Likert-Skalen mit semantisch unterschiedlichen Skalenbenennungen feststellen. Solche möglichen Artefakte können hier anhand einer sehr großen und für die deutsche Wohnbevölkerung repräsentativen Stichprobe untersucht werden.

Inhaltlich handelt es sich bei der Befragung um die Replikation eines Surveys zu Krankheitsvorstellungen über AIDS im Rahmen des Forschungsprojektes "AIDS und die gesellschaftlichen Folgen",¹⁾ welches seit 1990 an der Universität Trier durchgeführt wird und das vom Bundesgesundheitsamt (jetzt Robert-Koch-Institut) aus Mitteln des Bundesministeriums für Forschung und Technologie gefördert wird (vgl. dazu Eirnbter/Hahn/Jacob 1993). Die Befragung fand in der Zeit vom 30.5. bis zum 15.7.1995 statt. Es wurden 2.077 auswertbare Interviews in den alten und 1.079 Interviews in den neuen Bundesländern realisiert. Die Gesamtfallzahl beträgt somit 3.156 Befragte. Die Ausschöpfungsquote West beläuft sich auf 68,2 Prozent, die Ausschöpfungsquote Ost

auf 70,1 Prozent. Die Auswahl der Zielpersonen erfolgte - wie beim Sozialwissenschaften-BUS üblich - mit Hilfe des ADM-Mastersample.

Für den Methodensplit wurden zwei unterschiedliche Fragebogenversionen erstellt (im folgenden Split A mit Population A und Split B mit Population B genannt). Jeweils die Hälfte der Befragungsgesamtheit sollte mit der einen oder der anderen Version befragt werden; die Auswahl der Zielpersonen erfolgte zufällig. Splitversion A wurde von 1.043 Probanden in den alten und 557 Probanden in den neuen Ländern beantwortet, mit Version B wurden 1.034 Bundesbürger aus den alten und 522 aus den neuen Ländern befragt.

Überprüft wurden - wie schon erwähnt - zum einen mögliche Effekte, die aus der Reihenfolge der Antwortvorgaben resultieren können. Bei Fragen mit ordinal-skalierten Antwortkategorien wurden die Kategorien gedreht (vgl. dazu die im Anhang dokumentierten Fragen). In Split B wurden die Antwortvorgaben der Fragen 3, 4, 6, 7, 10, 20 und 26 gedreht.²⁾

Hinsichtlich der verbalen Kategorien bei Likert-Skalen sollten in der Variante A die vorgegebenen Items aus den Statementbatterien 5, 8, 9, 11, 15 und 19 mittels der herkömmlichen Benennung ("stimme sehr zu" bis "lehne sehr ab") und in der anderen Fassung mit einer modifizierten Form ("stimme sehr zu" bis "stimme überhaupt nicht zu") eingeschätzt werden (vgl. dazu den Anhang). Außerdem wurde bei Frage 21, wo es um die Akzeptanz der Einführung verschiedener staatlicher Maßnahmen zur Bekämpfung von AIDS ging, in Variante A mit den dichotomen Vorgaben "stimme zu" und "lehne ab", in Variante B mit "stimme zu" und "stimme nicht zu" gearbeitet.

3. Ergebnisse

3.1 Soziodemographische Struktur der Testpopulationen

Zunächst wurden die beiden Testpopulationen daraufhin untersucht, ob es in bezug auf die sozialstrukturelle Zusammensetzung (Merkmale Alter, Geschlecht, Bildungsstatus und Wohnortgröße) signifikante Unterschiede gibt. Entsprechende Analysen zeigen, daß die Altersstruktur, die Geschlechterproportionen und die Verteilung auf Dörfer, Klein-, Mittel- und Großstädte sich nicht signifikant voneinander unterscheiden, wohl aber die Bildungsstruktur (Sig. = .032). In Split A sind Personen mit niedrigerem Bildungsstatus leicht überrepräsentiert. Allerdings ist diese Abweichung sehr gering und bezieht sich insbesondere auf Personen ohne Schulabschluß.³⁾ Es kann also davon ausgegangen werden, daß die Teilstichproben sozialstrukturell weitestgehend identisch sind und mithin

mögliche signifikante Unterschiede in der Beantwortung der Versionen A und B nicht auf sozialstrukturelle Unterschiede zurückzuführen sind.

3.2 Unterschiede zwischen Version A und B: Signifikanztests

Die Frage nach differentiellen Antwortmustern in den Versionen A und B wurde im nächsten Schritt analysiert. Dazu wurden alle Fragen (auch diejenigen, die in beiden Varianten in identischer Form gestellt worden sind) auf signifikante Unterschiede untersucht. Das Ergebnis dieses Signifikanztestes findet sich für die in den Methodensplit aufgenommenen Fragen im Anhang.

Wesentliches Ergebnis dieser bivariaten Analyse ist, daß es nur wenige signifikante Unterschiede zwischen den Antwortverteilungen von Split A und Split B gibt. Betrachtet man zunächst die Fragen, die in beiden Versionen in identischer Form gestellt wurden, so zeigt sich, daß hier die Antwortmuster nicht differieren. Dies war bei zufällig ausgewählten Testpopulationen auch nicht anders zu erwarten und hätte im gegenteiligen Fall ernsthafte Zweifel an der Qualität des Auswahlverfahrens aufgeworfen.

Wenn man sich nun den Fragen zuwendet, die in unterschiedlicher Form gestellt wurden, so mehren sich zwar signifikante Unterschiede zwischen Population A und Population B, von einem systematischen Effekt kann aber auch hier nicht gesprochen werden. Von insgesamt 52 Fragen und Items, die in unterschiedlicher Form vorgegeben worden sind, weichen nur 24, also weniger als die Hälfte, signifikant voneinander ab. Es gibt - mit Ausnahme der Fragen 6 und 7 - auch keine durchgängigen signifikanten Unterschiede bei einzelnen Fragen. Weder die Drehung von Antwortvorgaben noch semantische Änderungen der Likert-Skalierung haben systematische Änderungen der Antwortmuster der Testpopulationen zur Konsequenz.

Insbesondere die Differenzierung von Ablehnung und Nicht-Zustimmung scheint für die Befragten bedeutungslos zu sein und hat keinen durchgängigen Effekt auf deren Antwortverhalten. Von 32 Items aus Likert-Skalen und aus Frage 21 weisen lediglich 7, also weniger als ein Viertel, signifikante Abweichungen auf. Auch bei Frage 21, wo in Version A "stimme zu" und "lehne ab" als dichotome Antwortvorgaben unmittelbar nebeneinander stehen, bestehen bis auf 19K überhaupt keine signifikanten Unterschiede zu Version B (Kategorien "stimme zu" vs. "stimme nicht zu"). Hier ließe sich zwar einwenden, daß bei dieser dichotomen Antwortvorgabe kein Ausweichen auf die neutrale Mittelkategorie möglich ist. Dennoch müßten sich, wenn die Benennung der Antwortvorgaben einen inhaltsunabhängigen Effekt auf das Antwortverhalten hätte, hier erkennbare Unterschiede zwischen den Versionen A und B zeigen.

Bei den signifikanten Unterschieden zwischen den Versionen A und B bewegen sich die Differenzen in der Regel in einem Bereich zwischen 2 und 6 Prozent. Zudem weisen die als signifikant ausgewiesenen Prozentsatzdifferenzen zwischen den beiden Befragungspopulationen keine einheitliche Tendenz auf. Bei Frage 4 etwa sind die Anteilswerte derer, die die genannten Krankheiten als ansteckend einschätzen, in Population B höher, bei Frage 6 dagegen sind mehr Befragte aus Population A der Ansicht, daß die genannten Krankheiten sehr bedrohlich sind. Bei Frage 7 wiederum ist zwar in Population A die Besetzung der Kategorie "sehr bedrohlich" höher als in B, dort aber ist im Gegenzug die Kategorie "bedrohlich" stärker besetzt. Falls also in Anbetracht der ohnehin nicht systematisch auftretenden Unterschiede zwischen Split A und B von Effekten einer unterschiedlichen Reihung von Antwortvorgaben gesprochen werden kann, so sind diese nicht einheitlich. Denn es sind nicht *durchgängig* die zuerst genannten Antwortalternativen, die die höheren Besetzungen aufweisen.

3.3 Einfluß von Drittfaktoren

Auch wenn auf der bivariaten Ebene keine systematischen Effekte differentieller Fragekonstruktionen feststellbar sind, so ist es natürlich denkbar, daß das Antwortverhalten *bestimmter* Gruppen überzufällig häufig auch von Formulierungen oder Reihungen der Antwortvorgaben beeinflusst wird. Die entsprechende These lautete ja gerade, daß insbesondere Personen mit niedrigerem sozialen Status und geringer Ich-Stärke zu Akquieszenz neigen.

Um zu überprüfen, ob und inwieweit sozialstrukturelle Merkmale als Indikatoren für differentielle Sozialisationschicksale und Lebenserfahrungen und damit auch für soziale Ungleichheit Einflußfaktoren bei einer inhaltsunabhängigen Reaktion auf Interviewfragen darstellen, wurden mit den Merkmalen Alter, Geschlecht, Bildung und Wohnort⁴⁾ sowie der regionalen Herkunft (alte und neue Länder) und dem Merkmal "Methodensplit" bei den ordinalskalierten Fragen mehrfaktorielle Varianzanalysen gerechnet.⁵⁾ Wenn diese Merkmale einen entscheidenden Einfluß auf das Antwortverhalten haben, dann müssen sich signifikante Interaktionseffekte feststellen lassen. Die Ergebnisse dieser Varianzanalysen finden sich in den Tabellen 1 und 2, wo außerdem auch die Haupteffekte des Merkmals "Methodensplit" ausgewiesen werden. Dabei zeigt sich, daß im Vergleich zum Test der Prozentsatzdifferenzen noch weniger signifikante Unterschiede ausgewiesen werden, die zentrale Tendenz in den Populationen A und B ist häufig identisch.

3.3.1 Reihenfolge der Antwortvorgaben

Zunächst wurden mögliche Effekte der unterschiedlichen Reihung von Antwortvorgaben untersucht (vgl. Tabelle 1).

Tabelle 1: Varianzanalysen, Haupteffekte des Merkmals Methodensplit (SPSS-Prozedur "Oneway") und Interaktionseffekte (SPSS-Prozedur "Anova") mit den Merkmalen Alter, Geschlecht, Bildung, Wohnortgröße und Herkunftsregion (alte/neue Länder), Signifikanz von F

Frage Nr.	Split	Alter	Gesch.	Bildung	Wohnort	Region (Ost/West)
3	.634	.084	.890	.769	.075	.525
4A	.838	.991	.909	.314	.628	.247
4B	.083	.820	.385	.046	.659	.015
4C	.754	.700	.224	.879	.698	.334
4D	.393	.216	.203	.225	.033	.229
4E	.782	.990	.056	.309	.688	.018
6A	.000	.644	.980	.785	.007	.166
6B	.000	.653	.845	.984	.166	.127
6C	.004	.912	.556	.202	.268	.109
6D	.008	.955	.992	.927	.168	.014
6E	.000	.188	.543	.914	.619	.005
6F	.000	.138	.842	.515	.391	.001
6G	.001	.532	.317	.650	.083	.004
7A	.000	.230	.978	.540	.819	.565
7B	.018	.230	.177	.897	.052	.964
7C	.027	.230	.557	.501	.418	.713
7D	.004	.713	.768	.424	.437	.940
10	.184	.404	.387	.787	.751	.635
26	.000	.312	.395	.592	.003	.079

Generell läßt sich feststellen, daß keiner der verwendeten Drittfaktoren einen systematischen Einfluß auf das Antwortverhalten hat. Signifikante Interaktionseffekte treten nur selten auf (von 95 Interaktionen sind 10 signifikant) und folgen keinem erkennbaren Muster. Die Merkmale Alter und Geschlecht produzieren überhaupt keinen Effekt, von einem Einfluß des Status-Indikators Bildung kann bei nur einem signifikanten Interaktionseffekt ebenfalls nicht ernsthaft gesprochen werden. Auch das Merkmal Wohnortgröße ist mit drei signifikanten Interaktionseffekten hier wohl zu vernachlässigen.

Etwas anders sieht die Sache dagegen bei der regionalen Herkunft aus. Sechs von insgesamt 19 Interaktionen, also rund 30 Prozent sind signifikant. Auch hier ist aber kein einheitliches Muster erkennbar. Sieht man sich diese Mittelwertsunterschiede im Detail an, so ergibt sich vielmehr ein höchst heterogenes Bild. Mal unterscheiden sich nur die Mittelwerte der Befragten West, mal nur die der Befragten Ost. Dabei sind bei einigen Fragen die Mittelwerte von Population A niedriger, bei anderen die von Population B. Systematische Zusammenhänge sind mithin nicht zu erkennen.

Allerdings zeigt sich auch bei Mittelwertvergleichen, daß bei den Fragen 6 und 7 durchgängig signifikante Unterschiede zwischen Version A und B bestehen. Hier ist es in der Tat so, daß in beiden Populationen die jeweils zuerst genannte Kategorie (sehr bedrohlich bzw. überhaupt nicht bedrohlich) häufiger genannt wird als in der Vergleichsgruppe. Diese inhaltsunabhängige Tendenz zur Bevorzugung der ersten Antwortalternative läßt sich aber nicht auf Effekte soziodemographischer Variablen zurückführen, wie auch dreidimensionale Tabellenanalysen gezeigt haben. Gleichwohl müßten hier, wenn man theoretische Überlegungen zu spezifischen Response-Sets wie Akquieszenz berücksichtigt, spezifische Persönlichkeitsmerkmale der Befragten ursächlich sein. Dabei wird, wie oben schon erwähnt, typischerweise auf den sozialen Status der Befragten verwiesen, wonach insbesondere unterprivilegierte Personen solche Reaktionseinstellungen aufweisen. Zwar produziert in unserer Analyse das Statusmerkmal "Bildung" keine dementsprechenden signifikanten Effekte, dafür lassen sich solche theoriekonformen Effekte aber beobachten, wenn als Drittfaktoren Skalen zur Kontrollüberzeugung der Befragten als Indikatoren für Selbstbewußtsein und Ich-Stärke verwendet werden. Sechs Items zu Kontrollüberzeugungen sind in einem Selbstausfüllbogen erhoben und nach faktorenanalytischer Prüfung zu einer Skala interner und einer Skala externer Kontrollüberzeugungen zusammengefaßt worden.⁶⁾

Zumindest in der Tendenz weisen Personen, die internen Kontrollüberzeugungen eher fernstehen und externen Kontrollüberzeugungen eher zuneigen, eher spezifische Reaktionseinstellungen auf, die zur Bevorzugung der jeweils erstgenannten Kategorie führen. Da dieser Effekt aber, wie schon erwähnt, nicht bei allen Fragen zu beobachten ist, scheint auch der Inhalt der Fragen eine Rolle zu spielen. Allerdings läßt sich aus dem vorliegenden Material keine Prognose ableiten, wie sich dieser Frageninhalt denn auswirken könnte. So ist es nicht zu erwarten und auch nicht verständlich, daß bei Fragen nach der subjektiv empfundenen Bedrohlichkeit von Krankheiten oder Nebenwirkungen technischer Entwicklungen (Fragen 6 und 7) zwar eine akquieszenzähnliche Reaktionseinstellung zu beobachten ist, bei der Einschätzung der Infektiosität

verschiedener Krankheiten (Frage 4) aber gerade nicht. Hier werden im Gegenteil die letztgenannten Kategorien bei einigen Krankheiten sogar häufiger genannt.

3.3.2 Unterschiedliche Benennung der Likert-Skalenpunkte

Auch bei der Verwendung unterschiedlicher Benennungen bei Likert-Skalen sind systematische Effekte der verwendeten Dritt-Faktoren nicht zu erkennen (vgl. Tabelle 2). Von 110 Interaktionen sind lediglich 20 signifikant. Auch hier produzieren das Alter und das Geschlecht überhaupt keinen signifikanten Effekt, der Einfluß des Merkmals Bildung ist ebenfalls vernachlässigbar gering. Wieder sind es nur Indikatoren des sozialräumlichen Kontextes - Herkunftsregion West/Ost (8 signifikante Interaktionen = 36 Prozent) und diesmal auch Wohnortgröße (10 signifikante Interaktionen = 45 Prozent) - die einen gewissen Einfluß haben.

Ein Blick auf die Details vermittelt bei den Befragten aus Ost- und Westdeutschland wiederum das oben schon skizzierte völlig heterogene Bild, ein systematischer und gleichsinniger Einfluß dieses Merkmals ist nicht erkennbar. Etwas anders sieht dagegen der Effekt der Variablen "Wohnortgröße" aus. Signifikante Effekte sind hier auf das unterschiedliche Antwortverhalten der Bewohner kleinerer Gemeinden zurückzuführen, während zwischen Großstädtern kaum differentielle Antwortmuster zu beobachten sind. Befragte kleiner Gemeinden weisen bei den Fragen, wo signifikante Unterschiede ausgewiesen sind, in Split A niedrigere Mittelwerte auf, wählen also häufiger die erstgenannten Kategorien bzw. vermeiden es häufiger, Aussagen explizit abzulehnen. Selbst von diesem ja nicht systematisch auftretenden "Muster" gibt es aber in zwei Fällen Abweichungen (Fragen 9B und 11C).

Auch wenn sich für die Gesamtskala keine systematischen signifikanten Differenzen zwischen den beiden Varianten feststellen lassen, so könnte es ja immerhin sein, daß im Fall der Verwendung der Kategorien "lehne eher ab" und "lehne sehr ab" (Split A) in Anbetracht der schon angeführten Argumente die Mittelkategorie "teils-teils" vergleichsweise stärker besetzt ist, weil Personen, auch wenn ein Item ihrer Meinung widerspricht, dies ungern als Ablehnung artikulieren und sich in die neutrale Mittelkategorie flüchten. Die Differenzen zwischen den Anteilswerten der Kategorien "teils-teils" und "lehne eher ab" müßten mithin größer sein als die Differenzen zwischen "teils-teils" und "stimme eher nicht zu" (Split B). Denn wenn explizite Ablehnung tatsächlich mit größeren psychischen Kosten verbunden ist als bloße "Nicht-Zustimmung", dann würde die "teils-teils" Kategorie häufiger, die Folgekategorie "lehne eher ab" dagegen seltener gewählt werden als in der Vergleichsgruppe.

Tabelle 2: Varianzanalysen, Haupteffekte des Merkmals Methodensplit (SPSS-Prozedur "Oneway") und Interaktionseffekte (SPSS-Prozedur "Anova") mit den Merkmalen Alter, Geschlecht, Bildung, Wohnortgröße und Herkunftsregion (alte/neue Länder), Signifikanz von F

Item Nr.	Split	Alter	Gesch.	Bildung	Wohnort	Region (Ost/West)
5A	.157	.060	.497	.312	.001	.269
5B	.974	.801	.438	.743	.222	.693
5C	.202	.134	.857	.600	.803	.495
5D	.012	.282	.792	.882	.030	.744
5E	.204	.255	.542	.040	.011	.051
5F	.000	.353	.462	.285	.644	.778
8A	.115	.468	.698	.672	.003	.609
8B	.013	.460	.829	.074	.103	.005
8C	.175	.247	.728	.146	.518	.003
8D	.247	.622	.620	.234	.050	.005
9A	.011	.787	.227	.122	.017	.040
9B	.770	.638	.071	.053	.010	.000
9C	.162	.718	.624	.018	.514	.000
11A	.984	.746	.639	.122	.401	.030
11B	.000	.843	.986	.092	.427	.000
11C	.673	.724	.682	.523	.005	.002
15A	.081	.895	.663	.438	.226	.294
15B	.153	.643	.140	.135	.038	.093
15C	.380	.094	.205	.416	.001	.180
19A	.325	.956	.503	.795	.058	.522
19B	.738	.792	.884	.310	.232	.069
19C	.072	.327	.283	.702	.010	.605

Tatsächlich ist dies aber nicht der Fall, wie Tabelle 3 zeigt, wo die Differenzen zwischen den Anteilswerten dieser Kategorien aus den im Fragebogen dokumentierten Verteilungen berechnet wurden. Die ausgewiesenen Signifikanzniveaus beziehen sich auf einen unmittelbaren Vergleich nur dieser Ausprägungen unter Ausschluß der übrigen Skalen- ausprägungen. Systematische Zusammenhänge sind wiederum nicht erkennbar. Von 22 getesteten Differenzen sind nur 4 signifikant. Zudem ist die Richtung des Zusammenhangs nicht eindeutig: bei 13 Differenzpaaren ist tatsächlich die Differenz in Split A größer als in B, bei 7 dagegen ist es umgekehrt, 2 sind identisch.

Tabelle 3: Differenzen zwischen den Anteilswerten der Kategorie "teils-teils" und den jeweiligen Folgekategorien in Split A ("lehne eher ab") und Split B ("stimme eher nicht zu"), Signifikanzen dieser Differenzen

Item	Differenz zwischen Kat. "teils-teils" und Folgekat. Split A (Split B)	A>B	B>A	Sig.
5A	11,3 (6,7)	*		.082
5B	-6 (-10,4)	*		.044
5C	10,7 (13,9)		*	.312
5D	-4,7 (-4,6)			.996
5E	-13,7 (-12,4)		*	.772
5F	14,4 (12,9)	*		.015
8A	28,8 (25,0)	*		.002
8B	6,4 (3,9)	*		.270
8C	10,9 (10,4)	*		.643
8D	16,0 (13,5)	*		.125
9A	20,0 (21,3)		*	.955
9B	23,9 (22,4)	*		.374
9C	18,4 (17,7)	*		.392
11A	19,5 (20,2)		*	.799
11B	19,1 (11,5)	*		.000
11C	10,1 (10,1)			1.000
15A	3,5 (4,3)		*	.708
15B	14,5 (13,6)	*		.892
15C	18,2 (19,3)		*	.960
19A	7,7 (6,4)	*		.569
19B	4,5 (0,8)	*		.142
19C	13,8 (16,8)		*	.368

Auch hier wurde der mögliche Einfluß von Dritt-Faktoren auf Verschiebungen der Prozentsatzdifferenzen zwischen den Kategorien "teils-teils" und "lehne eher ab" bzw. "stimme eher nicht zu" untersucht. Dabei treten ebenfalls keine durchgängigen systematischen Effekte auf, die große Mehrzahl der dreidimensionalen Tabellen wurde als nicht signifikant ausgewiesen. Es zeigt sich aber bei den wenigen signifikanten Zusammenhängen, daß die Kategorie "stimme eher nicht zu" häufiger gewählt wird als die Kategorie "lehne eher ab" und zwar insbesondere bei Befragten mit niedrigerem formalen Bildungsstatus oder Wohnsitz in kleineren Gemeinden. Dabei ist ein Trend zu sozial wünschenswertem Antwortverhalten erkennbar: Aussagen wie "Weil die Ansteckungswege bekannt sind, kann man sich vor AIDS schützen" (5F), "Personen, die

sich mit dem AIDS-Virus angesteckt haben, verdienen unser Mitgefühl" (8A), "Mit dem deutschen Gesundheitssystem können wir sehr zufrieden sein" (9A) oder "Die nun strengeren Kontrollen schließen weitere Infektionen durch verseuchte Blutpräparate aus" (11B) werden in der Tendenz häufiger mit "stimme eher nicht zu" beantwortet als mit der Kategorie "lehne eher ab". Bei Vorgabe der Antwortkategorien "lehne eher ab" und "lehne sehr ab" ist die "teils-teils"-Kategorie signifikant stärker besetzt. Allerdings - um es nochmals zu betonen - mehr als nur vage Hinweise für einen solchen Trend lassen sich dem vorliegenden Material nicht entnehmen.

4. Schlußbemerkungen

Systematisch auftretende Änderungen des Antwortverhaltens der Befragten durch die Änderung der Reihenfolge von Antwortvorgaben konnten nicht festgestellt werden. Insbesondere zeigt sich keine eindeutige inhaltsunabhängige Tendenz zur Bevorzugung der ersten Antwortalternative und zwar auch dann nicht, wenn soziodemographische Merkmale als kontrollierende Drittfaktoren eingesetzt werden.

Auch die Verwendung der alternativen Kategorien "stimme nicht zu" und "lehne ab" bei Likert-Skalen führte nicht zu systematischen signifikanten Unterschieden zwischen den beiden Testpopulationen. Auch hier ließen Drittfaktor-Kontrollen mit soziodemographischen Merkmalen keine systematischen Zusammenhänge oder gruppenspezifische "Anfälligkeiten" für Instrumenteffekte erkennen.

Am ehesten noch scheinen zwar in der Tat Befragte mit niedrigerem Bildungsstatus, Befragte mit geringerer Ich-Stärke und Einwohner kleinerer Gemeinden sich in ihrem Antwortverhalten auch von der inhaltsunabhängigen Art und Weise von Fragen beeinflussen zu lassen, indem sie spezifische Reaktionseinstellungen aufweisen oder im Sinne sozialer Wünschbarkeit antworten. Aber auch hier sind die Ergebnisse nicht eindeutig gerichtet und liefern nicht mehr als vage Hinweise.

Unter pragmatischen und forschungsökonomischen Gesichtspunkten sind diese Ergebnisse natürlich sehr begrüßenswert. Offenbar läßt sich die herkömmliche Likert-Skalierung weiter verwenden, ohne daß man sich mit der Frage der Reliabilität der Skala aufgrund fehlender Eindimensionalität auseinandersetzen müßte, da dies empirisch kein Problem zu sein scheint. Darüber hinaus scheint es aber auch unproblematisch zu sein, die Verteilungen inhaltlich gleicher Likert-Items zu vergleichen, die sich lediglich durch die Benennung der Skalenpunkte unterscheiden. Ob also die Einschätzung eines Sachverhaltes mit einer Skala von "stimme sehr zu" bis "lehne sehr ab" oder von "stimme sehr zu" bis "stimme überhaupt nicht zu" erhoben wird, ist entgegen meßtheoretischer

Annahmen für das Antwortverhalten nicht von Bedeutung. Replikationen entsprechender Tests wären natürlich sehr wünschenswert - vielleicht ließe sich ein entsprechender Methodensplit zur weiteren Klärung dieser Probleme für den ALLBUS realisieren.

Korrespondenzadresse

*Prof. Dr. Willy H. Eirmbter, Dr. Rüdiger Jacob
Universität Trier
FB IV Soziologie
54286 Trier
Tel.: 0651-201-2653*

Anmerkungen

- 1) Beteiligte Personen: Prof. Dr. Willy H. Eirmbter, Prof. Dr. Alois Hahn, Dr. Rüdiger Jacob; seit 1995: Dr. Frank Lettke, Diplom-Kauffrau Claudia Hennes.
- 2) Aus Platzgründen wird hier nicht der gesamte Fragebogen dokumentiert, sondern nur die Fragen, die in unterschiedlichen Versionen erhoben worden sind. Der Gesamtfragebogen kann bei den Autoren angefordert werden.
- 3) Aufgrund der großen Fallzahl der Stichprobe werden auch schon vergleichsweise kleine Differenzen als signifikant ausgewiesen und sollten, wenn sie weniger als 5 Prozentpunkte betragen nicht überinterpretiert werden. Umso stärker ist hervorzuheben, daß sich kaum signifikante und insbesondere keine systematischen Unterschiede zwischen den Versionen A und B nachweisen lassen.
- 4) Zur Verwendung von Alter, Bildung und Wohnortgrößenklassen als Indikatoren für eine differentiell ausgeprägte generelle Ressource "Handlungs- und Gestaltungskompetenz" vgl. Eirmbter/Hahn/Jacob 1993.
- 5) Meßtheoretisch erfüllen rangskalierte Merkmale zwar nicht die für das metrische Meßniveau notwendige Voraussetzung der Äquidistanz der einzelnen Meßpunkte. Verschiedene Untersuchungen haben aber gezeigt, daß die Analyse solcher Variablen auch mit Verfahren, die ein metrisches Meßniveau erfordern (wie etwa die Varianzanalyse), nicht zu nennenswerten Verzerrungen der Analyseergebnisse führt (vgl. dazu Allerbeck 1978 und Labovitz 1970. Siehe auch Diehl/Kohr 1987: 374).
- 6) Die Items im einzelnen: 1. Bei Problemen und Widerständen finde ich in der Regel Mittel und Wege, um mich durchzusetzen. 2. Erfolg ist oft weniger von Leistung, sondern vielmehr von Glück abhängig. 3. Es hat sich für mich als gut erwiesen, selbst Entscheidungen zu treffen, anstatt mich auf das Schicksal zu verlassen. 4. Ich übernehme gern Verantwortung. 5. Bei wichtigen Entscheidungen orientiere ich mich oft an dem

Verhalten von anderen. 6. Ich habe häufig das Gefühl, daß ich wenig Einfluß darauf habe, was mit mir geschieht.

Literatur

- Allerbeck, K., 1978: Meßniveau und Analyseverfahren - Das Problem "strittiger Intervallskalen". Zeitschrift für Soziologie, 7: 199-214.
- Bradburn, N. M., 1983: Response Effects. S. 289-328 in: Rossi, P.H./J. D. Wright/A. B. Anderson (Hrsg.): Handbook of Survey Research. Orlando, Florida: Academic Press.
- Couch, A./Kensiston, K., 1960: Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable. Journal of Abnormal and Social Psychology, 60: 150-174.
- Diehl, J. M./Kohr, H. U., 1987: Deskriptive Statistik, Frankfurt, 7. Auflage: Fachbuchhandlung für Psychologie.
- Duncan, O. D./Stenbeck, M., 1987: Are Likert Scales Unidimensional?. Social Science Research 16: 245-259.
- Eirmbter, W. H./Hahn, A./Jacob, R., 1993: AIDS und die gesellschaftlichen Folgen. Frankfurt: Campus.
- Esser, H., 1986: Response Set - Methodische Problematik und soziologische Interpretation. Zeitschrift für Soziologie 6: 253-263.
- Groves, R. M., 1989: Survey Errors and Survey Costs. New York: Wiley.
- Hare, A. P., 1960: Interview Responses: Personality or Conformity? Public Opinion Quarterly, 24: 679-685
- Labovitz, S., 1970: The Assignment of Numbers to Rank Categories. American Sociological Review 35: 515-524.
- Schnell, R./Hill, P. B./Esser, E., 1995: Methoden der empirischen Sozialforschung. München, 5. Auflage: Oldenbourg.

Anhang: Fragen und Items des Methodensplits

3	Für wie wahrscheinlich halten Sie es, daß es einmal dazu kommen könnte, daß man sich wegen AIDS kaum noch auf die Straße wagen kann. Halten Sie das für	wahrscheinlich	Split A 7,0	Split B 5,2
		weniger wahrscheinlich	34,7	39,4
		unwahrscheinlich	58,3	55,4
		N: Sig.=.0006	1578	1545
Split B: Reihenfolge der Kategorien gedreht.				

4 Im folgenden nenne ich Ihnen einige Krankheiten. Bitte sagen Sie mir, für wie ansteckend Sie diese Krankheiten halten.

Split A	sehr ansteckend	weniger ansteckend	überh. nicht ansteckend	weiß nicht	
Split B	überh. nicht ansteckend	weniger ansteckend	sehr ansteckend	weiß nicht	Sig.:
A Grippe	63.9 2.5	32.4 30.7	3.0 65.3	0.7 1.4	.134
B Krebs	4.1 78.1	9.4 13.4	79.9 3.2	6.6 5.3	.001
C Tuberkulose	59.5 5.9	26.1 25.1	7.4 61.0	7.0 8.0	.197
D Masern	64.4 2.4	27.6 27.4	4.6 66.0	3.4 4.1	.009
E AIDS	51.6 9.1	33.8 31.4	10.6 54.0	4.0 5.5	.048

5 Im folgenden werden Ihnen einige Aussagen vorgestellt. Sagen Sie mir zu jeder Aussage auf der Liste, inwieweit Sie dieser zustimmen bzw. diese ablehnen. Nennen Sie mir hierzu jeweils den Kennbuchstaben der Aussage und die dazugehörige Antwortkennziffer aus der Skala oben auf der Liste.

Split A	stimme sehr zu	stimme eher zu	teils/ teils	lehne eher ab	lehne sehr ab	
Split B	stimme sehr zu	stimme eher zu	teils/ teils	stimme eher zu	stimme überhaupt nicht zu	Sig.:
A Im Kreis von Kranken und Behinderten fühle ich mich eher unwohl	8.7 7.7	20.3 20.2	33.4 31.3	22.1 24.6	15.5 16.2	.364

B	Wenn man es genau bedenkt, gibt es gar keinen wirksamen Schutz vor AIDS	7.6 6.3	13.4 16.6	24.9 21.7	30.9 32.1	24.1 23.3	.018
C	Wenn man sich vor allen Gefahren schützen wollte, wäre das Leben zwar sicherer, aber nicht mehr lebenswert	17.2 16.3	29.6 27.2	27.5 30.4	16.8 16.5	8.9 9.7	.318
D	AIDS-Gefahren lauern überall	9.1 7.2	15.9 14.2	25.4 26.1	30.1 30.7	19.5 21.8	.117
E	Die Gefahren v. AIDS sind mir bewußt, aber weil ich auch Spaß im Leben haben will, nehme ich best. Risiken in Kauf	2.6 1.6	7.9 8.4	13.7 13.4	27.4 25.8	48.3 50.8	.202
F	Weil die Ansteckungswege bekannt sind, kann man sich vor AIDS schützen	49.2 41.4	29.8 35.1	17.0 17.5	2.6 4.6	1.5 1.5	.000

6 Für wie bedrohlich halten Sie die folgenden Krankheiten, wenn Sie jetzt einmal an sich, Ihre Familie und Ihren engeren Freundeskreis denken. Bitte antworten Sie anhand dieser Liste.

Split A	sehr bedrohlich	bedrohlich	weniger bedrohl.	überh. nicht bedrohl.	
Split B	überh. nicht bedrohl.	weniger bedrohlich	bedrohlich	sehr bedrohlich	Sig.:
A Tuberkulose	9.5 36.1	22.4 41.9	40.5 16.2	27.6 5.9	.000
B Herz- und Kreislauferkrankungen	21.5 11.6	44.2 30.7	25.6 40.9	8.7 16.9	.000
C Rinderwahnsinn (BSE)	11.0 33.6	21.9 36.4	37.1 22.2	29.9 7.9	.010
D AIDS	16.2 29.2	21.0 37.0	35.7 21.8	27.1 12.1	.011

E	Gelbsucht (Hepatitis)	8.3 25.7	27.8 44.6	42.9 24.1	20.9 5.5	.000
F	Krebs	21.4 13.8	40.3 31.0	27.1 38.5	11.2 16.8	.000
G	Alte und neue Tro- penkrankheiten	5.6 46.8	17.5 35.1	35.4 13.9	41.4 4.2	.002

7 Die moderne technische Entwicklung kann neben allen positiven Wirkungen auch negative Folgen haben, die vielleicht zu schweren Katastrophen führen. Für wie bedrohlich halten Sie:

Split A	sehr bedrohlich	bedrohlich	weniger bedrohlich	überh. nicht bedrohl.	Sig.:
Split B	überh. nicht bedrohl.	weniger bedrohlich	bedrohlich	sehr bedrohlich	
A	33.3 5.3	39.9 26.8	21.9 41.6	4.9 26.4	.000
B	29.2 7.0	38.2 27.3	26.7 40.9	5.8 24.8	.030
C	49.3 2.3	37.2 10.5	11.3 44.4	2.2 42.8	.000
D	50.8 1.3	39.2 10.2	8.81 43.2	1.2 45.3	.019

8 Wir haben hier einige Aussagen zusammengestellt, die man häufig im Zusammenhang mit AIDS hört. Sagen Sie mir bitte zu jeder Aussage auf der Liste, inwieweit Sie dieser zustimmen bzw. diese ablehnen. Nennen Sie hierzu jeweils wieder den Kennbuchstaben und die dazugehörige Antwortkennziffer.

Split A	stimme sehr zu	stimme eher zu	teils/ teils	lehne eher ab	lehne sehr ab	Sig.:
Split B	stimme sehr zu	stimme eher zu	teils/ teils	stimme eher nicht zu	stimme überh. nicht zu	
A	23.4 21.7	34.3 34.5	34.1 33.2	5.3 8.2	2.8 2.4	.021
B	9.0 6.0	17.9 16.1	28.5 29.9	22.1 26.0	22.5 22.1	.003

C	Unsere heutige Zeit mit ihrer Vergnügungssucht und all den Ausschweifungen hat AIDS erst möglich gemacht	15.0 11.7	25.9 26.8	28.0 29.4	17.1 19.0	13.9 13.1	
							.055
D	Schuld an AIDS sind die Hauptrisikogruppen	15.7 15.1	26.2 23.8	30.2 30.9	14.2 17.4	13.6 12.8	
							.111

9 Unser Gesundheitssystem wird unterschiedlich beurteilt. Wir haben im folgenden einige Aussagen zusammengestellt, die man in diesem Zusammenhang häufig hört. Sagen Sie mir bitte zu jeder Aussage auf der Liste, inwieweit Sie dieser zustimmen bzw. diese ablehnen. Nennen Sie hierzu jeweils wieder den Kennbuchstaben und die dazugehörige Antwortkennziffer.

Split A	stimme sehr zu	stimme eher zu	teils/teils	lehne eher ab	lehne sehr ab		
Split B	stimme sehr zu	stimme eher zu	teils/teils	stimme eher nicht zu	stimme überh. nicht zu	Sig.:	
A	Mit dem dt. Gesundheitssystem können wir sehr zufrieden sein.	17.8 15.8	36.8 35.0	30.9 32.6	10.9 11.3	3.6 5.2	
							.096
B	Beim heutigen Gesundheitssystem steht nicht mehr das Helfen im Vordergrund, sondern das Geldverdienen	18.4 19.7	30.2 28.5	35.9 35.8	12.0 13.4	3.4 2.6	
							.359
C	Sogar im Krankenhaus kann man sich heute mit dem AIDS-Virus infizieren	22.1 20.7	35.2 34.0	29.1 30.2	10.7 12.5	2.8 2.5	
							.438

10 Es gibt unterschiedliche Ansichten darüber, ob man erkennen kann, daß jemand mit dem AIDS-Virus infiziert ist. Manche Personen sagen, daß sie mit Sicherheit erkennen könnten, ob jemand mit dem AIDS-Virus infiziert ist. Andere Personen sind sich dabei nicht so sicher, und wieder andere sagen, daß sie eine Infektion nicht erkennen könnten. Wie ist das bei Ihnen? Welcher Meinung neigen Sie eher zu?

	Split A	Split B
Ich würde mit Sicherheit erkennen, daß jemand mit dem AIDS-Virus infiziert ist	3.4	2.3
Ich würde wahrscheinlich erkennen, daß jemand mit dem AIDS-Virus infiziert ist	8.0	8.4
Ich würde nicht erkennen, daß jemand mit dem AIDS-Virus infiziert ist	64.6	70.4

INT.: nicht vorlesen Man kann nicht erkennen, daß jemand mit dem AIDS-Virus infiziert ist N: Sig.: = .001	23.9 1596	19,0 1551
Split B: Reihenfolge der Kategorien gedreht.		

11 Anfang letzten Jahres wurde in der Öffentlichkeit viel über den Skandal um AIDS-infizierte Blutpräparate gesprochen. Wir haben hierzu einige Aussagen zusammengestellt. Sagen Sie mir bitte zu jeder Aussage auf der Liste, inwieweit Sie dieser zustimmen bzw. diese ablehnen. Nennen Sie hierzu jeweils wieder den Kennbuchstaben und die dazugehörige Antwortkennziffer.

Split A	stimme sehr zu	stimme eher zu	teils/ teils	lehne eher ab	lehne sehr ab	
Split B	stimme sehr zu	stimme eher zu	teils/ teils	stimme eher nicht zu	stimme überh. nicht zu	Sig.:
A Der Blut-Skandal hat gezeigt, daß auf unser Gesundheitssystem kein Verlaß mehr ist	19.1 18.0	26.4 27.9	35.2 35.5	15.7 15.3	3.6 3.3	.839
B Die nun strengeren Kontrollen schliessen weit. Infektionen durch verseuchte Blutpräparate aus	12.6 10.9	35.6 31.7	33.1 31.6	14.0 20.1	4.8 5.7	.000
C Wenn man wegen einer geplanten Operation eine Bluttransfusion erhalten muß, sollte man vorher Eigenblut spenden	47.8 49.3	29.9 28.5	15.5 15.4	5.4 5.3	1.4 1.5	.907

15 Sagen Sie mir bitte zu jeder Aussage auf der Liste, inwieweit Sie dieser zustimmen bzw. diese ablehnen. Nennen Sie jeweils wieder den Kennbuchstaben und die dazugehörige Antwortkennziffer.

Split A	stimme sehr zu	stimme eher zu	teils/ teils	lehne eher ab	lehne sehr ab	
Split B	stimme sehr zu	stimme eher zu	teils/ teils	stimme eher nicht zu	stimme überh. nicht zu	Sig.:
A Wenn ich es mir recht überlege, möchte ich mit AIDS-Infizierten - doch lieber gar keinen Kontakt haben	15.8 10.6	21.8 28.1	28.5 26.6	25.0 22.3	12.9 12.4	.002
B Ein AIDS-Kranker in meiner Nach- barschaft würde mich stören	5.1 3.9	12.4 12.8	19.3 18.8	33.8 32.4	29.5 32.2	.290
C Die AIDS-Be- drohung führt noch dazu, daß man sich außer- halb der eig. vier Wände nirgends mehr sicher fühlen kann	2.3 2.0	7.2 5.7	13.6 14.1	31.8 33.4	45.1 44.8	.436

19 Im folgenden werden Ihnen nochmals einige Aussagen vorgestellt. Sagen Sie mir bitte wieder, inwieweit Sie diesen Aussagen zustimmen bzw. diese ablehnen. Nennen Sie jeweils wieder den Kennbuchstaben und die dazugehörige Antwortkennziffer.

Split A	stimme sehr zu	stimme eher zu	teils/ teils	lehne eher ab	lehne sehr ab	
Split B	stimme sehr zu	stimme eher zu	teils/ teils	stimme eher nicht zu	stimme überh. nicht zu	Sig.:
A Heute muß man eigentlich mißtrau- en, weil man ja nicht weiß, wer AIDS-infiziert ist	5.1 5.1	13.7 13.5	23.7 25.6	31.4 32.0	26.2 23.8	.562
B Zumindest in der Freizeit sollten sich die Gesunden Bereiche schaffen dürfen, wo man vor AIDS sicher ist	5.5 4.3	15.5 15.4	22.3 25.7	26.8 26.5	29.9 28.1	.137

C	An meinem Arbeitsplatz würde ich mit jemandem der mit dem AIDS-Virus infiziert ist, ohne weiteres zusammenarbeiten	27.7 25.2	31.2 30.4	25.1 27.9	11.3 11.1	4.6 5.4	
							.253

20	Zur Eindämmung von AIDS gibt es in der Öffentlichkeit gegensätzliche Standpunkte: Die einen halten Aufklärung und freiwillige Maßnahmen für das Richtige, die anderen fordern Zwangsmaßnahmen und Verbote. Was glauben Sie <u>persönlich</u> , ist in <u>erster Linie</u> erfolgversprechend:	Aufklärung und freiwillige Maßnahmen	Split A 86,9	Split B 85,4
		Zwangsmaßnahmen und Verbote	13,1	14,6
		N: Sig.= .217	1560	1498
Split B: Reihenfolge der Kategorien gedreht.				

21 Zur Bekämpfung von AIDS werden verschiedene Maßnahmen vorgeschlagen. Bitte sagen Sie mir, welchen möglichen Maßnahmen zur AIDS-Bekämpfung Sie persönlich zustimmen bzw. welche Sie ablehnen.

Split A	stimme zu	lehne ab	
Split B	stimme zu	stimme nicht zu	Sig.:
A Kondomverteilung an Jugendliche	82.4 81.3	17.6 18.7	.418
B Verstärkte Sexualerziehung in Schulen	94.9 95.9	5.1 4.9	.181
C Verteilen von Einwegspritzen an Drogenabhängige	72.2 71.2	27.8 28.2	.529
D Generelle namentliche Meldepflicht aller Infizierten	62.3 62.7	37.7 37.3	.828
E Zwangstest nur für Hauptrisikogruppen	50.6 53.9	49.4 46.1	.064
F Zwangstest für alle	17.1 18.0	82.9 82.0	.504
G Trennung der AIDS-Kranken von den Gesunden	17.9 19.1	82.1 80.9	.379
H Kennzeichnung aller AIDS-Infizierten	13.0 11.6	87.0 88.4	.210
J Schließung von Bordellen und Verbot der Prostitution	27.5 26.1	72.5 73.9	.371
K Grenzkontrollen und Einreiseverbot für AIDS-Infizierte	41.2 45.3	58.8 54.7	.019

		<i>Split A</i>	<i>Split B</i>
26 Alles in allem: Ist das AIDS-Problem <u>für Sie</u> :	sehr wichtig	17.2	16,3
	wichtig	34.8	38,0
	weniger wichtig	34.1	34,2
	überhaupt nicht wichtig	14.0	11,6
	N:	1597	1549
Sig. = .000			
Split B: Reihenfolge der Kategorien gedreht.			

DER ALTENHEIMSURVEY: DURCHFÜHRUNG UND REPRÄSENTATIVITÄT EINER BEFRAGUNG IN DEN EINRICHTUNGEN DER STATIONÄREN ALTENHILFE¹⁾

THOMAS KLEIN UND SIEGFRIED GABLER

Der Altenheimsurvey (AHS) ist eine Repräsentativerhebung unter Altenheimbewohnern. Er erfasst insbesondere auch den Lebensverlauf und die Gründe des Heimeintritts. Darüber hinaus wird mit dem Altenheimsurvey eine wichtige Lücke im Datenbestand der empirischen Sozialforschung geschlossen: Allgemeine Bevölkerungsumfragen wie der ALLBUS, das Sozio-Ökonomische Panel, der Familiensurvey und andere Umfragen beziehen sich ausschließlich²⁾ auf die Bevölkerung in Privathaushalten. Mit dem Altenheimsurvey werden große Teile des Fragenprogramms allgemeiner Bevölkerungsumfragen in vergleichbarer Weise auch für Altenheimbewohner erhoben.

The survey of homes for the aged is a representative survey among the population in homes for the aged. In particular, it collects the life history and the reasons why respondents move to these homes. In addition, the survey fills an important gap in the stock of data files for empirical social research. General social surveys such as the ALLBUS, the Socioeconomic Panel, the German Family Survey refer exclusively to the population in private households. In the survey of homes for the aged questions usually collected in general surveys for private households are gathered in a comparable manner for the people living in homes for the aged.

1. Ziele und Gegenstand des Altenheimsurvey

Der Altenheimsurvey (AHS) ist eine Repräsentativerhebung unter Bewohnern stationärer Alteinrichtungen. Die Studie zielt zum einen auf die Gründe des Heimeintritts, d.h. auf die Motive und auf die Lebensumstände vor Heimeintritt. Dabei werden die früheren Wohnverhältnisse, der frühere Wohnort, der (subjektive) Grund des Heimeintritts, die Inanspruchnahme ambulanter Dienste, Netzwerkbeziehungen zu den Kindern und vieles mehr erfragt. Hervorzuheben ist in diesem Zusammenhang auch die retrospektive Erfassung biographischer Angaben insbesondere zur Gesundheitsbiographie

und zur Familienbiographie. Anders als in früheren Studien³⁾ wird damit der Versuch unternommen, die Heimeintrittsgründe retrospektiv zu erfassen und den Heimeintritt aus der Perspektive des Lebensverlaufs zu erklären. Demgegenüber steht die aktuelle Lebens- und Pflegesituation nicht im Mittelpunkt der Erhebung.

Ein zweites Ziel des Altenheimsurvey ist zum anderen, eine wichtige Lücke im Datenbestand der empirischen Sozialforschung zu schließen (vgl. auch Schnell 1991). Allgemeine Bevölkerungsumfragen wie der ALLBUS, das Sozio-Ökonomische Panel, der Familiensurvey und andere Umfragen erfassen nur die Bevölkerung in Privathaushalten, während mit dem Altenheimsurvey große Teile des Fragenprogramms allgemeiner Bevölkerungsumfragen in vergleichbarer Weise auch für Altenheimbewohner erhoben werden. Teilweise waren bei der Konstruktion des AHS-Fragebogens forschungsspezifische Erweiterungen oder Reduzierungen des Fragebogens notwendig. Diese Veränderungen sind jedoch aus Vergleichbarkeitsgründen so gering wie möglich gehalten. Der Altenheimsurvey kommt damit einer wiederholten Forderung nach, in verschiedenen Untersuchungen einheitliche Frageformulierungen zu verwenden (vgl. z.B. Friedrichs 1980; Hoffmeyer-Zlotnik/Ehling 1991).

Eine Besonderheit des Altenheimsurvey (gegenüber den im Anhang zusammengestellten Studien) ist schließlich eine Mischung aus Eigeninterviews und Proxi-Interviews für dieselbe Person. Wegen der vor allem aus gesundheitlichen Gründen erwarteten Ausfälle wurden einige Fragen vor der Durchführung des eigentlichen Interviews als Proxi-Interview mit der Einrichtungsleitung geklärt. Eine aktuelle Infratest-Studie (Schneekloth/Müller 1995), die sich insbesondere auf die gegenwärtige Lebens- und Pflegesituation der Bewohner bezieht, beruht sogar ausschließlich auf Proxi-Interviews und ist dadurch sehr gut in der Lage, ein repräsentatives Bild der Lebensbedingungen von Heimbewohnern zu geben. Die betreffenden Angaben bleiben deshalb unabhängig von einer eventuellen Selektivität der (Bewohner-) Ausfälle. Bei der retrospektiven Erfassung biographischer Zusammenhänge sind jedoch Proxi-Interviews nur sehr begrenzt anwendbar.

2. Stichprobenplan

2.1 Grundgesamtheit

Die Grundgesamtheit besteht aus den deutschen und ausländischen Bewohnern, die mindestens 60 Jahre alt sind und in Alten-, Pflege- oder Altenwohnheimen in Deutschland wohnen. Derzeit leben in 8.634 solchen Einrichtungen⁴⁾ rund 680.000 Menschen⁵⁾.

Solche Spezialpopulationen sind in der Regel nur dann durch eine Stichprobe gut abbildbar, wenn eine entsprechende Auswahlgrundlage vorhanden ist. In der vorliegenden Erhebung ist dies das Altenheim-Adreßbuch 94, in dem sämtliche Einrichtungen der stationären Altenhilfe in Deutschland verzeichnet sind, die im Frühjahr 1994 bestanden und postalisch auszumachen waren. Die dort gemachten Angaben beinhalten neben der Adresse auch die Trägerschaft und in den meisten Fällen auch die vorhandene Platzzahl. Das Altenheim-Adreßbuch erfaßt alle Typen stationärer Alteneinrichtungen: Altenwohnheime, Altenheime und Pflegeheime. Im Altenwohnheim leben die Bewohner in Apartments, wo sie ihren Haushalt weitgehend selber führen, aber Gemeinschaftseinrichtungen zur gemeinsamen Nutzung zur Verfügung stehen. Im Altenheim sind die alten Menschen in der Regel nicht mehr in der Lage, einen Haushalt selbständig zu führen. Sie sind aber nicht oder noch nicht pflegebedürftig. In den Pflegeheimen leben Menschen, die auf fremde Hilfe angewiesen sind. Bei den Trägerschaften sind es im folgenden öffentliche (kommunale), kirchliche, freigemeinnützige und privat-gewerbliche Träger. Kirchliche und freigemeinnützige Träger sind im folgenden zu einer Kategorie zusammengefaßt.

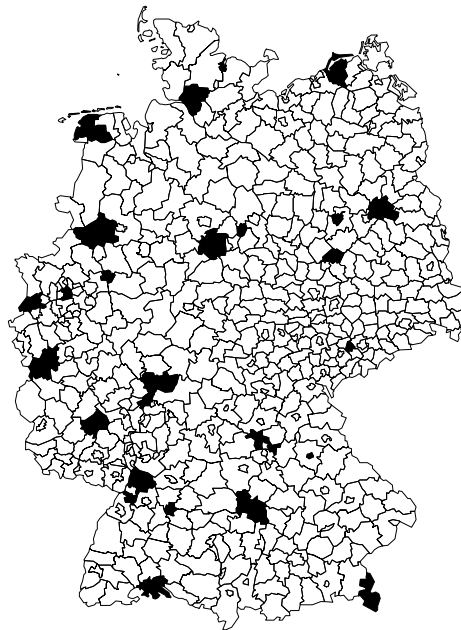
2.2 Stichprobe

Das Auswahlverfahren ist so gestaltet, daß zumindest theoretisch jede in einer Alteneinrichtung lebende Person dieselbe Wahrscheinlichkeit hatte, in die Stichprobe zu gelangen. Darüber hinaus sollte durch regionale Schichtung eine geographische Streuung gewährleistet sein. Da eine Liste der Bewohner von solchen Einrichtungen nicht existiert, wurden die Angaben des Altenheim-Adreßbuch 94 verwendet, um zunächst die Heime auszuwählen. Aus Kostengründen mußte die Zahl der Kreise, in denen befragt werden sollte, auf 25 beschränkt werden, aus denen jeweils fünf Alteneinrichtungen ausgewählt wurden. Auf der ersten Stufe erfolgte die Auswahl der Kreise nach regionaler Anordnung durch größenproportionale, systematische Zufallsauswahl. Dazu wurde aus dem Altenheim-Adreßbuch, soweit dort angegeben, die Zahl der Plätze aller Einrichtungen in einem Kreis ermittelt. Es zeigte sich, daß von den 8.634 Heimen⁶⁾ bei über 2.000 keine Platzzahlangabe vorlag. Um dennoch eine proportionale Auswahl vornehmen zu können, wurde auf Bundeslandebene die Differenz zwischen der Platzzahl der amtlichen Heimplatzstatistik und der aggregierten Platzzahl des Altenheim-Adreßbuches ermittelt, und diese Zahl gleichmäßig auf die Heime aufgeteilt, bei denen keine Platzangabe gemacht wurde. Dabei wurde von der Vorstellung ausgegangen, daß es in der Regel die kleineren Heime sein würden, bei denen eine Platzzahlangabe fehlt.⁷⁾ Die für die Auswahl notwendigen Angaben wurden in eine Datei geschrieben, die neben einer Heimnummer die Kreiskennziffer, die Einwohner pro Quadratkilometer, das Bundesland, den

Verbandsträger und die Platzzahl enthielt. Einen Überblick über die regionale Streuung der ausgewählten Kreise gibt Abbildung 1.

Bei der zweiten Auswahlstufe wurden die Heime innerhalb eines ausgewählten Kreises nach Trägerschaft und Einwohnerdichte sortiert und dann durch größenproportionale, systematische Zufallsauswahl fünf Heime⁸⁾ je Kreis ausgewählt. Auf der dritten und letzten Stufe wurden dann uneingeschränkt zufällig⁹⁾ je 30 Bewohner der entsprechenden Alteneinrichtung ausgewählt und befragt. Für die ausgewählten Personen wurden in aller Regel zunächst Proxi-Interviews und dann, falls möglich, Eigen-Interviews durchgeführt.

Abbildung 1: Überblick über die regionale Streuung der ausgewählten Kreise

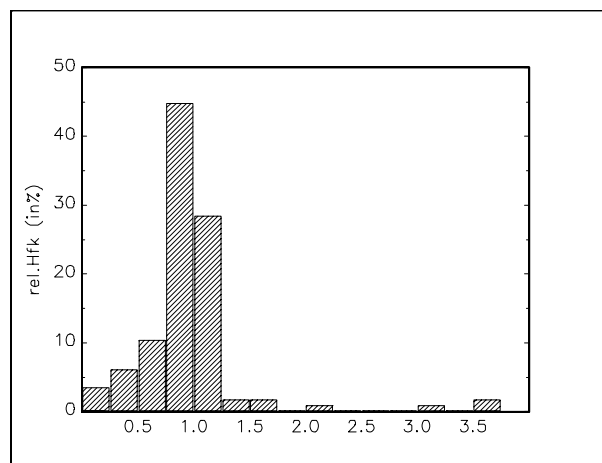


Zu dem gewählten Auswahlvorgehen ist folgendes zu bemerken:

- Nicht alle im Sinne der Grundgesamtheit definierten Einrichtungen stehen im Altenheim-Adreßbuch. Dennoch dürfte es die zur Zeit umfangreichste und aktuellste Datengrundlage für eine Auswahl sein (vgl. Klein/Salaske 1996a).¹⁰⁾

- Die Zahl der ausgewählten Kreise ist relativ gering. Es besteht daher die Gefahr eines Klumpeneffekts. Es wurden allerdings mittels Simulationen Vorstudien über die Güte von Parameterkonstellationen durchgeführt, um zu sehen, welchen Einfluß die Zahl der ausgewählten Kreise und die Zahl der darin ausgewählten Heime auf gewisse bekannte Variablen wie Einwohnerdichte, Bundesland und Verbandsträger hat. Zu beachten ist außerdem, daß die im Mittelpunkt stehende retrospektive Erfassung der Lebensbiographie keinen engen Bezug zur aktuellen Geographie aufweist. Der Klumpeneffekt wäre unter Umständen anders zu beurteilen, wenn gegenwärtige Lebensbedingungen oder die gegenwärtige Pflegesituation im Mittelpunkt der Untersuchung gestanden hätten.

Abbildung 2: Histogramm für das Verhältnis zwischen der Bewohnerzahl bei Befragung und der für die Auswahl zugrundegelegten Platzzahl



- Das durchgeführte Auswahlverfahren führt zu keiner selbstgewichtenden Stichprobe der Bewohner, wenn der Belegungsgrad der Heime unterschiedlich ist. Nach den Erfahrungen der Interviewer dürfte dieser Fehlereinfluß gering sein, da die meisten Heime vollständig belegt sind. Die Platzzahlangaben im Altenheim-Adreßbuch, erweitert durch Ersetzung der fehlenden Werte, stimmen dennoch nicht mit der aktuellen *Bewohnerzahl* überein. In Abbildung 2 ist dargestellt, wie sich die tatsächliche Bewohnerzahl und die für die Auswahl verwendete Platzzahl unterscheiden. Dabei ist festzuhalten, daß es sich bei den Extrema nicht um ersetzte missings handelt,

sondern um Werte, die im Altenheim-Adreßbuch so stehen. Ursachen für größere Abweichungen waren etwa verändertes Platzangebot durch Umbaumaßnahmen vor allem im Osten, oder die Angaben waren Aggregate von mehreren Heimen, die von einer Leitung verwaltet wurden.

3. Durchführung und Felderfahrungen

Die Interviews wurden von August 1995 bis März 1996 durchgeführt. Entsprechend den Auswahlstufen wurde erst mit den gezogenen Einrichtungen und dann mit den Bewohnern Kontakt aufgenommen. Für die Kontaktaufnahme mit den Einrichtungen war zunächst um Unterstützung der Verbandsträger nachgesucht worden. Von neun angesprochenen Verbandsträgern haben sechs die Untersuchung befürwortet und unterstützt.¹¹⁾ Insgesamt lag bei 60,3% der ausgewählten Einrichtungen eine Verbandsunterstützung vor. Den Einrichtungsleitern wurde die Untersuchung mit einem Anschreiben angekündigt, in dem Sinn und Zweck der Erhebung erklärt und gegebenenfalls das Unterstützungsschreiben des Verbandsträgers beigelegt wurde. Mitgeschickt wurde außerdem immer auch ein Aushang, der die Bewohner schon frühzeitig auf die Befragung vorbereitet hat. Der darauffolgende, telefonische Kontakt mit den Einrichtungsleitern war nicht den Interviewern überlassen, sondern zentral organisiert. Die dadurch weitgehend bei einer Person akkumulierten Erfahrungen mit der telefonischen Kontaktaufnahme haben sich als zunehmend sehr hilfreich erwiesen. Die zentrale Organisation der Telefonate mit den Einrichtungen diente außerdem der Terminabsprache und war aus erhebungslogistischen Gründen erforderlich. Bei den schließlich beteiligten Einrichtungen kann nach übereinstimmender Einschätzung aller Interviewer davon ausgegangen werden, daß die volle Bandbreite der denkbaren Pflege- und Betreuungssituationen abgedeckt wurde.

Die Auswahl der Bewohner innerhalb der Einrichtungen beruht auf Zufallszahlen. Hierzu war zuvor für jede mögliche Bewohnerzahl eine Liste mit jeweils 30 Zufallszahlen zusammengestellt worden. Im Anschluß an die Auswahl wurden (meist mit der Einrichtungsleitung) die Proxi-Interviews durchgeführt.¹²⁾ Bei den darauf folgenden Interviews mit den Bewohnern war es eine ausgesprochene Unterstützung, wenn der Interviewer zuerst vorgestellt wurde und ggf. sogar Interviewtermine mit den Bewohnern vereinbaren konnte. Solche Terminabsprachen haben sich insbesondere im Wohnheimbereich als sinnvoll erwiesen. Im Unterschied zu den bei allgemeinen Bevölkerungsumfragen in Privathaushalten gemachten Erfahrungen mit der Befragung älterer Menschen wurde die Befragung im Rahmen des Altenheimsurvey von den Altenheimbewohnern ganz überwiegend positiv aufgenommen. Für viele war die Befragung eine willkommene Abwechslung. Für einige Bewohner war die Befragung darüber hinaus mit einem sehr

starken Gesprächsbedürfnis verbunden. Es kam sogar vor, daß sich Bewohner nachhaltig für die Untersuchung „beworben“ haben, wobei schwer zu erklären war, daß dies mit einer Zufallsauswahl nicht zu vereinbaren ist. Allerdings gab es auch 9,6% Verweigerungen. Diese gingen in wenigen Fällen mitunter auch auf Angehörige zurück. Zum größeren Teil konzentrieren sich diese Verweigerungen allerdings in Altenwohnanlagen, in denen die Bewohner vergleichsweise selbständig leben und wo oft keine Ankündigung der Untersuchung und keine Vorstellung des Interviewers stattfand.¹³⁾

In einigen Fällen war die Befragung mit einem standardisierten Fragebogen nicht unproblematisch. So war es vor allem bei einem starken Redebedürfnis oder bei geringer Konzentrationsfähigkeit des Bewohners mitunter schwierig, die Fragen in der vorgesehenen Reihenfolge durchzugehen. In einzelnen Interviews, bei denen die Befragbarkeit des Bewohners nicht allzu hoch einzuschätzen war, wurden die Themenkomplexe deshalb nicht in der vorgesehenen Reihenfolge abgearbeitet. Ein (Erinnerungs-) Problem war tendenziell mit den Zeitangaben verbunden; in einigen Fällen wurden Hilfsmittel wie der Ehering oder die Heiratsurkunde bemüht. Wenn dem Bewohner in diesem Zusammenhang der Rückgang seiner geistigen Leistungsfähigkeit bewußt wurde und ein Interviewabbruch drohte, wurde die Fragenreihenfolge in wenigen Fällen auch auf Initiative des Interviewers geändert. Als guter Einstieg hat sich schon im Pretest der Fragenkomplex zur früheren Wohnsituation bewährt. Eher schlecht liefen die Fragen zur finanziellen Situation, die deshalb erst zum Ende des Interviews gestellt wurden. Die durchschnittliche Interviewzeit betrug etwa 20 Minuten, sie variierte aber beträchtlich je nach Umfang der biographischen Angaben und je nach der geistigen Leistungsfähigkeit.

Aufgrund der geschilderten (und einer Unzahl weiterer) Schwierigkeiten wurden die Interviews von nur wenigen, hochmotivierten und sozialwissenschaftlich vorgebildeten Interviewern durchgeführt, die nach einer Pretest- und Schulungsphase zunehmende Erfahrung mit der Befragung alter Menschen bekamen. Dieser Prozeß wurde darüber hinaus vor allem in den ersten Erhebungswochen von einer intensiven Rücksprache und Rückkoppelung begleitet und unterstützt, bei der auch eine frühe Sichtung der schon durchgeführten Interviews zu einer weiteren Verbesserung beigetragen hat. Insgesamt waren neun Interviewer eingesetzt, wobei die große Mehrzahl der Interviews von nur sechs Personen durchgeführt wurde.¹⁴⁾ Eine im Vergleich zu allgemeinen Bevölkerungsumfragen in Privathaushalten derart geringe Zahl von Interviewern ist natürlich mit Klumpungseffekten verbunden, die Zahl der eingesetzten Interviewer ist aber andererseits klein (sic!) genug, um Interviewereffekte bei der Auswertung statistisch kontrollieren zu können. Die im folgenden dokumentierte, hohe Ausschöpfung auch innerhalb der Einrichtungen beruht nicht zuletzt auf dem praktizierten Interviewerkonzept.

Tabelle 1: Ausfälle von Einrichtungen nach Trägerschaft, Platzzahl und Bundesland

Einrichtungsmerkmale ¹⁾	Ausfälle	
	absolut	in % der jeweiligen Kategorie in der Bruttostichprobe
insgesamt	15	12,9
<u>Verbandsträgerschaft</u>		
freigemeinnützig	9	12,3
kommunal	1	4,8
privat	4	23,5
ohne Verbandsträger	1	20,0
<u>Platzzahl</u>		
1-50	1	3,7
51-100	4	12,5
101 und mehr	10	17,5
<u>Bundesland²⁾</u>		
Schleswig-Holstein	1	10,0
Niedersachsen	1	7,1
Nordrhein-Westfalen	5	20,0
Baden-Württemberg	4	26,7
Bayern	1	5,6
Sachsen	2	40,0
Berlin	1	20,0
Hessen	0	0,0
Rheinland-Pfalz	0	0,0
Brandenburg	0	0,0
Mecklenburg-Vorpommern	0	0,0
Sachsen-Anhalt	0	0,0

¹⁾ Kategorisierung mit nacherhobenen Informationen.

²⁾ Die Bruttostichprobe enthält keine Heime der Bundesländer Hamburg, Bremen, Thüringen und Saarland.

Quelle: Altenheimsurvey (AHS) 1995/96.

4. Ausfallgründe

Im folgenden sind die Ausfälle und die Ausfallgründe zusammengefaßt und auf wichtige Strukturmerkmale hin analysiert. Ausgehend von 25 Kreisen, fünf Einrichtungen (sample points) in jedem Kreis und 30 Interviews in jeder Einrichtung (bzw. jedem sample point) waren bis zu 3.750 Interviews angestrebt. Da 12 Einrichtungen bei der Befragung weniger als 30 Bewohner hatten (die zumindest 60 Jahre alt waren), beläuft sich die Bruttostichprobe auf 3.623. Davon wurden 3.121 Interviews zumindest mit dem Proxi-Fragenteil realisiert (Nettostichprobe I).¹⁵⁾ Unter diesen sind 1.678 auch mit dem Bewohner durchgeführte Interviews (Nettostichprobe II).

In Tabelle 1 sind zunächst die Ausfälle der Einrichtungen - d.h. die Ausfälle der zweiten Auswahlstufe - dokumentiert. Von den 116 gezogenen Einrichtungen haben sich 101 an der Erhebung beteiligt (87,1%). Ausgefallen sind 15 Einrichtungen (12,9%).¹⁶⁾ Die Beteiligung auf Heimebene liegt damit sogar noch über der von Infratest Burke, die im Altenheimbereich eine Ausschöpfung von 81% erreicht hat (Schneekloth/Müller 1995: 18). Wie aus Tabelle 1 hervorgeht, waren die Ausfälle bei den privaten Einrichtungen mit 23,5% am höchsten, gefolgt von den Einrichtungen ohne Verbandsträger, freigemeinnützigen Einrichtungen und den kommunalen Einrichtungen mit 12,3 bzw. 4,8% Ausfällen. In bezug auf die Platzzahl konzentrieren sich die Ausfälle auf die größeren Einrichtungen. In bezug auf die regionale Verteilung der Einrichtungsausfälle sind überproportionale Ausfallquoten in Nordrhein-Westfalen, Baden-Württemberg, Sachsen und Berlin zu verzeichnen (Tabelle 1). Da in jeder Einrichtung dieselbe Personenzahl erfaßt wurde, entspricht die in Tabelle 1 wiedergegebene Struktur der Einrichtungsausfälle weitgehend¹⁷⁾ der der Bewohnerausfälle bei den Proxi-Interviews.

In Tabelle 2 sind - ausgehend von der Nettostichprobe I (d.h. von den Proxi-Interviews) - die Bewohnerausfälle bei den Eigeninterviews nach dem Alter, der Wohndauer, dem Gesundheitszustand, der Pflegebedürftigkeit und dem Ausfallgrund aufgeschlüsselt. Insgesamt konnten 46% der Proxi-Interviews nicht mit Eigenangaben vervollständigt werden, wofür unterschiedliche Ausfallgründe maßgeblich waren. Wie aus Tabelle 2 ersichtlich, dominiert dabei dauerhafte Nichtbefragbarkeit: Über 20% der Bewohner wurde von vornherein als dauerhaft nicht befragbar eingestuft.¹⁸⁾ Diese teilen sich mit 11% und 10,8% zu fast gleichen Teilen auf Ausfälle, die der Interviewer beim ersten Kontakt als nicht befragbar eingestuft hat, und Ausfälle, die nur von der Heimleitung als nicht befragbar eingestuft wurden, ohne daß ein persönlicher Kontakt mit dem Bewohner stattgefunden hat.¹⁹⁾ Hinzuzurechnen sind 6,8% Interviewabbrüche, die in aller Regel vom Interviewer veranlaßt wurden, weil sich der Befragte trotz intensiver Versuche als nicht

oder kaum befragbar gezeigt hat. Für die zusammengenommen fast 30% der Ausfälle wegen Nichtbefragbarkeit waren fast ausschließlich geistige und psychische Beeinträchtigungen verantwortlich, die sich in der Interviewerbeurteilung unter Demenz subsumieren. Angesichts des von Schneekloth/Müller (1995: 31) festgestellten 28%-Anteils der Bewohner, die von den Pflegern oder Betreuern „häufig ... räumlich, zeitlich oder zur eigenen Person desorientiert“ eingestuft wurden, können diese knapp 30% Ausfälle wegen Nichtbefragbarkeit als ausgesprochen gering eingestuft werden.²⁰⁾ Immerhin haben jedoch auch fast 10% verweigert, und 6% waren innerhalb des Interviewzeitraums von in der Regel zwei Tagen nicht erreichbar.

Aufgeschlüsselt nach dem Alter zeigt Tabelle 2 erwartungsgemäß eine um so geringere Ausfallquote wegen Nichterreichbarkeit, je höher das Alter: Während die 60- bis 70jährigen mit 10% wegen Nichterreichbarkeit ausfallen, sind die 90jährigen und Älteren nur zu 3% nicht erreichbar. Die Verweigerungen sind hingegen nur bei den 60- bis 70jährigen überdurchschnittlich. Die Nichtbefragbarkeitsquote ist hingegen um so höher, je höher das Alter. Eine Ausnahme bildet allerdings die eher altersunabhängige Quote der laut Heimleitung nicht Befragbaren, zu denen kein persönlicher Kontakt stattgefunden hat. Faßt man die verschiedenen Ausfallgründe zusammen, so fallen nur die 90jährigen und Älteren mit einer Ausfallquote von über 50% aus dem Rahmen (vgl. Tabelle 2).

Im Hinblick auf die Wohndauer in der Einrichtung zeigen sich in Tabelle 2 kaum systematisch unterschiedliche Ausfallquoten. Lediglich die Nichterreichbarkeit scheint mit zunehmender Verweildauer abzunehmen, wofür vermutlich der zuvor beschriebene Alterseffekt verantwortlich ist (vgl. Tabelle 2).

Aufgeschlüsselt nach dem Ausmaß gesundheitlicher Beeinträchtigung und nach der Pflegebedürftigkeit (-seinschätzung durch die Heimleitung) ist die Nichterreichbarkeitsquote unter den Nichtbeeinträchtigten und den Nichtpflegebedürftigen deutlich am höchsten. Die Nichtbefragbarkeitsquoten der verschiedenen Art sind um so höher, je schlechter der Gesundheitszustand bzw. je höher der Grad der Pflegebedürftigkeit (vgl. Tabelle 2). Insgesamt entspricht die Struktur der Ausfälle weitgehend den Erwartungen.

5. Analysen zur Repräsentativität

Was resultiert aus dem Zufallsfehler und aus den Ausfällen für die Repräsentativität? Da der Altenheimsurvey sowohl in bezug auf Einrichtungsmerkmale wie in bezug auf Bewohnermerkmale auswertbar ist, ist neben der repräsentativen Abbildung der Bewohnerstruktur auch von Interesse, inwieweit die Einrichtungsstrukturen richtig erfaßt sind.

Tabelle 2

Tabelle 3

5.1 Repräsentativität der Einrichtungen (2. Auswahlstufe)

In Tabelle 3 ist zunächst die Repräsentativität der Einrichtungen analysiert. Als Maßstab bzw. als Grundgesamtheit dient auch hierbei das Altenheim-Adreßbuch. Wie aus Tabelle 3 hervorgeht, sind vor allem die freigemeinnützig getragenen Einrichtungen mit 59,9% in der Bruttostichprobe gegenüber 66,6% in der Grundgesamtheit unterrepräsentiert, während die kommunalen Einrichtungen mit 25,9% gegenüber 17,3% überrepräsentiert sind. Dieser Vergleich beruht allerdings auf der Kategorisierung der Trägerschaft im Altenheim-Adreßbuch, bei dem ein Großteil der Trägerschaftsinformation fehlt. Ein ähnlicher Unterschied besteht im Hinblick auf die letztlich interessierende Nettostichprobe der an der Erhebung beteiligten Einrichtungen. Tabelle 3 informiert weiterhin über die Repräsentativität der ausgewählten Einrichtungen in bezug auf Größe und regionale Verteilung. Erwartungsgemäß recht gut ist die Verteilung der Größe der Einrichtungen und die Verteilung auf die Bundesländer abgebildet, wenn man davon absieht, daß vier kleine Bundesländer - mit zusammen nur 7% der Einrichtungen - nicht in der Stichprobe enthalten sind.

5.2 Repräsentativität der Bewohner (3. Auswahlstufe)

In Tabelle 4 sind entsprechende Analysen zur Repräsentativität der Bewohner nach der Trägerschaft und der Größe ihrer Einrichtung sowie der geographischen Verteilung zusammengestellt. In bezug auf die Trägerschaft ist die Bewohnerstruktur in der Stichprobe gut abgebildet, wobei die Bewohner von gemeinnützigen Einrichtungen nur leicht unter-, die von kommunalen Einrichtungen leicht überrepräsentiert sind. In bezug auf die Einrichtungsgröße stimmt die Stichprobe gut mit der Grundgesamtheit des Altenheim-Adreßbuchs überein, wenn man auch in der Stichprobe von den Platzangaben des Altenheim-Adreßbuchs ausgeht. Unter Zugrundelegung der nacherhobenen, tatsächlichen Bewohnerzahlen bei Befragung bekommen hingegen die Bewohner in kleinen Einrichtungen ein größeres Gewicht. Im weiteren sei die Repräsentativität der Bewohner auch im Hinblick auf Geschlecht, Alter, Wohndauer und Familienstand beleuchtet. Hierzu wird der Altenheimsurvey in Tabelle 5 mit zwei anderen Untersuchungen verglichen: der Studie von Infratest Burke (Schneekloth/Müller 1995) sowie der Untersuchung von Krug/Reh (1992). Die Infratest-Untersuchung enthält nur Proxi-Interviews und ist in bezug auf die Proxi-Interviews mit 3.015 Fällen nur geringfügig kleiner als der Altenheimsurvey. Die Untersuchung von Krug/Reh wendet sich zum Teil an die Verwaltungsbehörden, die für den Heimaufenthalt aufkommen, und zum anderen Teil an die selbstzahlenden Bewohner. Die Untersuchung von Krug/Reh fand postalisch statt und hat trotz der insgesamt über 8.000 zurückerhaltenen Fragebögen bei den Selbstzahlern eine ausgesprochen geringe Rücklaufquote von nur 9% (Krug/Reh 1992: 17).

Tabelle 4

Tabelle 5: Repräsentativität der Bewohner nach Geschlecht, Familienstand, Altersgruppe und Verweildauer (in %)

Merkmale	Altenheimsur.	Infratest Burke (Schneekloth/Müller 1995)	Krug/Reh (1992)
Erhebungsjahr	1995/96	1994	1989/90
insgesamt			
Nettostichprobe I			
absolut (in %)	3.121 (100)	3.015 ¹⁾ (100)	8.478 ²⁾ (100)
Nettostichprobe II			
absolut (in %)	1.678 (100)	-	-
<u>Geschlecht</u>			
Männer	18,1 ³⁾	21	22
Frauen	81,9 ³⁾	79	78
<u>Alter⁴⁾</u>			
60 bis unter 70	8,2 ³⁾	9,5	10
70 bis unter 80	20,9 ³⁾	20,0	22
80 bis unter 90	52,3 ³⁾	52,6	50
90 und älter	18,6 ³⁾	17,9	18
<u>Verweildauer</u>			
unter 1 Jahr	19,0 ³⁾	19	19 ⁶⁾
1 bis unter 3 Jahre	29,5 ³⁾	31	30 ⁶⁾
3 bis unter 7 Jahre	30,2 ³⁾	31	31 ⁶⁾
7 Jahre und mehr	21,3 ³⁾	19	20 ⁶⁾
<u>Familienstand</u>			
ledig	19,3 ⁵⁾	21	27
verheiratet	7,9 ⁵⁾	7	7
verwitwet	67,5 ⁵⁾	64	60
geschieden	5,2 ⁵⁾	8	6

1) Reine Proxi-Interviews.

2) 7.462 reine Proxi-Interviews bei Behörden und 1.016 reine Eigen-Interviews.

3) Nettostichprobe I.

4) Bezogen auf die Bewohner im Alter von 60 und darüber. Die Studien von Schneekloth/Müller und von Krug/Reh erfassen darüber hinaus auch jüngere Bewohner von Alteneinrichtungen.

5) Nettostichprobe II.

6) Angaben für die Eigen-Interviews, die nur mit den Selbstzahlern durchgeführt wurden.

In bezug auf das Geschlecht sind Männer mit 18,1% im Altenheimsurvey etwas geringer repräsentiert als in den beiden anderen Studien (vgl. Tabelle 5). Die Altersverteilung der drei Studien ist hingegen recht ähnlich, wobei im Altenheimsurvey eher die jüngeren Alten zwischen 60 und 70 etwas geringer vertreten sind. In bezug auf die Wohndauer (bis zum Befragungszeitpunkt!) zeigen alle drei Erhebungen sogar nahezu dieselbe Verteilung. Bezüglich des Familienstands entspricht der Altenheimsurvey stärker der aktuelleren Infratest-Erhebung. Da der Familienstand im Altenheimsurvey aus den Bewohnerinterviews hervorgeht, während die Infratest-Erhebung durchgehend auf Proxi-Interviews beruht, ist die hohe Übereinstimmung hierbei besonders erfreulich. Auch die mit den Bewohnern selbst durchgeführten Interviews sind somit (soweit vergleichbar) gut gelaufen.

6. Ausblick

Der Altenheimsurvey läßt erwarten, daß die Gründe und die Motive der Übersiedlung in eine Einrichtung der stationären Altenhilfe besser untersucht werden können als dies mit bisherigen Erhebungen der Fall war (vgl. zum Überblick auch Klein 1994, 1996; Klein/- Salaske 1994a, 1994b, 1994c). Dabei ist allerdings zu beachten, daß die zweite Stufe der Pflegeversicherung eine Reihe von Veränderungen mit sich bringen kann. Dennoch bietet der Altenheimsurvey gegenwärtig auch eine Ergänzung sozialstruktureller Informationen über die Altenbevölkerung in der Bundesrepublik. In diesem Zusammenhang sind auch typische Unterschiede des Lebensverlaufs späterer Heimbewohner von Interesse. Darüber hinaus ist anzunehmen, daß die Auswertungen allgemein einen besseren Einblick in die Bevölkerungsgruppe der Alten geben, die im Zuge der Bevölkerungsalterung zunehmend an Bedeutung gewinnt.

Übersicht 1: Empirische Studien zu Bewohnern von Einrichtungen der stationären Altenhilfe in Deutschland (chronologisch geordnet)

Autor	Titel	Erscheinungsjahr	Zielpopulation	Bruttostichprobe/realisierte Stichprobe	Zeitraum der Studie
Vettiger, G., Jaffe, A., Vogt, A.	Alte Menschen in Altersheimen - Soziologische, psychologische und medizinische Studien im Altersheim der Stadt Basel	1951	Menschen in einem Altersheim der Stadt <i>Basel</i>	203 / 203	Winter 1947 - 27. Februar 1948
Beske, F.	Das Gemeinschaftsleben in Altersheimen - Sozialhygienische Analyse der Existenzbedingungen von Altersheim- und Pflegeheimbewohnern	1960	alle Bewohner von Alters- und Pflegeheimen öffentlicher und privater Trägerschaft in der Stadt <i>Kiel</i> im Jahre 1956	605 (Totalerhebung) / 575	-
Blume, O.	Alte Menschen in einer Großstadt - Ergebnisse einer empirischen Untersuchung in Köln	1962	alle im Jahre 1960 65jährigen und älteren Bürger <i>Köln's</i> , die in „normalen Wohnungen“ (=Privathaushalte) leben + alle Bürger Kölns, die im Jahre 1960 in Alterswohnhäusern, Stiften und Altersheimen (von der Stadt und von privaten Wohlfahrtsverbänden getragen) leben	Privathaushalte: 1.600 / 1.200 + Wohnhäuser und Stifte: 158 / 155 Altersheime: 162 / 160	Privathaushalte: April-Mai 1960
Schmelzer, H., Teibert, W.	Alter und Gesellschaft - Eine soziologische Untersuchung der sozialen Voraussetzungen von Maßnahmen der Altenhilfe	1969	alle 65jährigen und älteren Einwohner (incl. Altersheimbewohner) von <i>Rheinland-Pfalz</i>	1.279 / 1.112 (incl. Altersheimbewohner) + zusätzlich 84 Altersheimbewohner	20. Oktober - Ende November 1968

Autor	Titel	Erscheinungsjahr	Zielpopulation	Bruttostichprobe/realisierte Stichprobe	Zeitraum der Studie
Lohmann, S.	Die Lebenssituation älterer Menschen in der geschlossenen Altersfürsorge	1970	65jährige und Ältere in privaten Haushalten + Bewohner von Altenheimen, von Altenpflegeheimen, von Wohnstiften, von Altenwohnheimen und von Altenwohnhäusern in <i>sechs deutschen Städten</i> und <i>einem deutschen Landkreis</i>	- + <i>Personen in Alten- und Pflegeheimen:</i> -/ 1.752 <i>Personen in Altenwohnheimen:</i> -/ 427	1960-1964
Schmitz-Scherzer, R., Schick, I., Kühn, D.	Altenwohnheime, Personal und Bewohner - Eine empirische Studie in der Stadt Braunschweig	1978	Bewohner von 20 Alten(wohn)- und Pflegeheimen in <i>Braunschweig</i>	-/ 1.002	September / Oktober 1975
Zimmermann, R., Bernhart, S.	Das soziale Umfeld und Betreuungsformen Hochbetagter	1981	über 74jährige in einem großstädtischen, einem mittelstädtischen und einem ländlichen Gebiet	883, darunter 301 Heimbewohner	Befragung von August 1979 - Januar 1980
Saup, W.	Übersiedlung ins Altenheim: Belastende Umwelbedingungen in Altenheimen und Bewältigungsreaktionen von Altenwohnheimbewohnern	1981	alle 1981 in West-Berlin existierenden Seniorenheime	-/ 26 Heime	1981
Bickel, H., Jaeger, J.	Die Inanspruchnahme von Heimen im Alter	1986	alle im Jahre 1982 verstorbenen Mannheimer Bürger im Alter von über 65 Jahren	2.585 (Totalerhebung) / 2.585	1982

Autor	Titel	Erscheinungsjahr	Zielpopulation	Bruttostichprobe/realisierte Stichprobe	Zeitraum der Studie
Lehmkuhl, D., Bosch, G., Steinhart, I.	Alte Menschen in Heimen	1986	alle über 65jährigen Bewohner sämtlicher Einrichtungen der geschlossenen Altenhilfe des Berliner Bezirks <i>Charlottenburg</i>	- / 907	1977
Rückert, W.	Demographische Grundlagen zur Altenhilfe-Planung	1987	Datengrundlage: - Daten der Volkszählung von 1970 - Daten einer Erhebung in Heimen <i>Baden-Württembergs</i> im Jahre 1974	-	1970, Volkszähl. 1974; Erhebung in Heimen in B.-W.
Schmidt, B., Witte, F.-D.	Pflegebedürftige alte Menschen in einem Mecklenburgischen Landkreis	1991	alle Einwohner <i>Parchims</i> im Alter von 70 Jahren und älter (darunter 181 Heimbewohner)	3.239 (Totalerhebung) / 3.236	Oktober 1987 - März 1988
Krug, W., Reh, G.	Pflegebedürftige in Heimen - Statistische Erhebungen und Ergebnisse	1992	alle stationär versorgte Pflegebedürftige in <i>Westdeutschland</i>	Empfänger von „Hilfe zur Pflege“: ca. 12.200 / 7.462 Selbstzahler: 8.387 / 1.016	Stichtags-erhebung 1989/90
Kruse, A.	Konflikt- und Belastungssituationen in stationären Einrichtungen der Altenhilfe und Möglichkeiten ihrer Bewältigung	1992	Bewohner von 26 Alten(wohn)- und Pflegeheimen (privat, gemeinnützig, öffentlich) in <i>Westdeutschland</i>	300 / 241	-
Klein, T., Salaske, I.	Determinanten des Heimeintritts im Alter und Chancen seiner Vermeidung. Eine Längsschnittuntersuchung für die Bundesrepublik Deutschland	1994	Heimeintritte im Zeitverlauf (Westdeutschland, ohne Ausländer) auf Basis des Sozio-ökonomischen Panels	1.230, darunter 55 Heimeintritte	1984-1991

Autor	Titel	Erscheinungsjahr	Zielpopulation	Bruttostichprobe/realisierte Stichprobe	Zeitraum der Studie
Bickel, H.	Zur Inanspruchnahme von stationärer Pflege und Versorgung im Altersverlauf	1994	alle in 3 Viermonatszeiträumen zwischen 1991 und 1993 verstorbenen Mannheimer Bürger im Alter von über 65 Jahren	2.507 / 2.507	1991-1993
Schneekloth, U., Müller, U.	Möglichkeiten und Grenzen selbständiger Lebensführung in Einrichtungen - Hilfe- und Pflegebedürftige in Heimen	1996	Bewohner von Wohn- einrichtungen der Alten- und Behindertenhilfe aller Altersklassen	Befragt wurden die Pfleger/Be- treuer von 3.015 Bewohnern aus 377 Alteinrichtungen und von 1.251 Bewohnern aus 158 Wohn- einrichtungen der Be- hindertenhilfe	Ende 1994

Quelle: eigene Zusammenstellung.

Literatur

- Beske, F. 1960: Das Gemeinschaftsleben in Altersheimen, Sozialhygienische Analyse der Existenzbedingungen von Altersheim- und Pflegeheimbewohnern. Stuttgart.
- Bickel, H. 1994: Zur Inanspruchnahme von stationärer Pflege und Versorgung im Altersverlauf. Das Gesundheitswesen, 56: 363-370.
- Bickel, H./Jäger, J. 1986: Die Inanspruchnahme von Heimen im Alter. Zeitschrift für Gerontologie, 19: 30-39.
- Blume, O. 1962: Alte Menschen in einer Großstadt, Ergebnisse einer empirischen Untersuchung in Köln. (Bd. 6). Göttingen.
- Friedrichs, J. 1980: Methoden der empirischen Sozialforschung. Opladen.
- Hoffmeyer-Zlotnik, J./Ehling, M. 1991: Demographische Standards für Deutschland. Ein Instrumentenentwurf. ZUMA-Nachrichten, 28: 29-40.
- Klein, T. 1994: Der Heimeintritt im Alter. Neue Befunde für die Bundesrepublik Deutschland. Sozialer Fortschritt, 43: 44-50.
- Klein, T. 1996: Determinants of Institutionalization in Old Age. In: Roland Eisen/Frank A. Sloan (Eds.), Alternatives for Ensuring Long-Term Care (in Druck). Boston: Kluwer.
- Klein, T./Salaske, I. 1994a: Die Bedeutung des Gesundheitszustands für den Heimeintritt alter Menschen. Das Gesundheitswesen, 56: 235-240.

- Klein, T./Salaske, I. 1994b: Die Bedeutung sozialer Beziehungen für den Heimeintritt im Alter. Theoretische Überlegungen und empirische Befunde. *Zeitschrift für Sozialreform*, 10: 641-661.
- Klein, T./Salaske, I. 1994c: Determinanten des Heimeintritts im Alter und Chancen seiner Vermeidung. Eine Längsschnittuntersuchung für die Bundesrepublik Deutschland. *Zeitschrift für Gerontologie*, 27: 442-455.
- Klein, T./Salaske, I. 1996a: Regionale Disparitäten im stationären Versorgungsangebot für alte Menschen. *Zeitschrift für Gerontologie und Geriatrie*, 29: 65-75.
- Klein, T./Salaske, I. 1996b: Strukturen der stationären Altenversorgung in der Bundesrepublik Deutschland. *Sozialer Fortschritt*, 45: 41-48.
- Krug, W./Reh, G. 1992: Pflegebedürftige in Heimen: Statistische Erhebungen und Ergebnisse. Schriftenreihe des Bundesministeriums für Familie und Senioren.
- Kruse, A. 1992: Konflikt- und Belastungssituationen stationärer Einrichtungen der Altenhilfe und Möglichkeiten ihrer Bewältigung. Stuttgart.
- Lehmkuhl, D./Bosch, G./Steinhart, I. 1986: Alte Menschen in Heimen. *Zeitschrift für Gerontologie*, 19: 56-64.
- Lohmann, S. 1970: Die Lebenssituation älterer Menschen in der geschlossenen Altersfürsorge. Hannover.
- Rückert, W. 1987: Demographische Grundlagen zur Altenhilfe-Planung. S. 59-97 in: Brandt, H./Dennebaum, E. M./Rückert, W. (Hg.), *Stationäre Altenhilfe. Problemfelder - Rahmenbedingungen - Perspektiven*. Freiburg im Breisgau.
- Saup, W. 1984: Streß und Streßbewältigung bei der Heimübersiedlung älterer Menschen. *Zeitschrift für Gerontologie*, 17: 198-204.
- Schmelzer, H./Tebert, W. 1969: *Alter und Gesellschaft. Eine soziologische Untersuchung der sozialen Voraussetzungen von Maßnahmen der Altenhilfe*. Bonn.
- Schmidt, B./Witte, F. D., 1991: Pflegebedürftige alte Menschen in einem Mecklenburgischen Landkreis. *Archiv für Wissenschaft und Praxis der sozialen Arbeit*, 22: 3-19.
- Schmitz-Scherzer, R./Schick, I./Kühn, D. 1978: *Altenwohnheime, Personal und Bewohner, eine empirische Studie in der Stadt Braunschweig*. (Bd. 57). Stuttgart, Berlin, Köln, Mainz.
- Schnell, R. 1991: Wer ist das Volk? Zur faktischen Grundgesamtheit bei „allgemeinen Bevölkerungsumfragen“: Undercoverage, Schwererreichbare und Nichtbefragbare. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 43: 106-137.
- Schneekloth, U./Müller, U. 1995: Hilfe- und Pflegebedürftigkeit in Heimen. S. 13-52 in: Bundesministerium für Familie, Senioren, Frauen und Jugend (Hg.), *Schnellbericht zur Repräsentativerhebung im Rahmen des Forschungsprojekts "Möglichkeiten und Grenzen selbständiger Lebensführung in Einrichtungen"*. Bonn.

Vettiger, G./Jaffe, A./Vogt, A. 1951: Alte Menschen im Altersheim. Soziologische, psychologische und medizinische Studien im Altersheim der Stadt Basel. Basel.

Zimmermann, R./Bernhart, S. 1981: Das soziale Umfeld und Betreuungsformen Hochbetagter. Aktuelle Gerontologie, 11: 141-143.

Anmerkungen

- 1) Die hier vorgestellte Erhebung ist im Rahmen des von Thomas Klein (Heidelberg) durchgeführten Forschungsprojekts „Der Heimeintritt alter Menschen und Chancen seiner Vermeidung“ entstanden, das von der Deutschen Forschungsgemeinschaft (DFG) gefördert wird. Stichprobenplan und Fragebogen wurden in enger Zusammenarbeit mit ZUMA (Mannheim) entwickelt.
- 2) Im Sozio-Ökonomischen Panel werden immerhin Umzüge in Altenheime weiterverfolgt, dabei ist aber von einer beträchtlichen Untererfassung auszugehen.
- 3) Eine Zusammenstellung findet sich in Übersicht 1 im Anhang.
- 4) Umfang der Heime in der Auswahlgesamtheit (Altenheim-Adreßbuch 1994).
- 5) Laut Heimplatzstatistik für 1994.
- 6) Etwa 200 Einrichtungen aus dem Adreßbuch wurden nicht verwendet, da sie nicht im engeren Sinn zu den Einrichtungen der stationären Altenhilfe zählen (vgl. im einzelnen auch Klein/Salaska 1996a, 1996b).
- 7) Es zeigte sich bei der Befragung, daß diese Annahme nicht in allen Fällen richtig war.
- 8) Die 5 Heime sind nicht notwendig alle verschieden. Es kam aufgrund des hohen Platzangebots einiger Heime vor, daß diese mehrfach ausgewählt wurden. Entsprechend wurde in solchen Heimen ein Vielfaches von 30 an Heimbewohnern befragt. Genauer wurden also 5 "sample points" ausgewählt. In einem Fall waren zwei Heime zusammengelegt worden. 60 Bewohner wurden deshalb befragt. Die 125 sample points haben sich dadurch auf 116 Einrichtungen reduziert.
- 9) Die Interviewer arbeiteten mit zuvor erzeugten Zufallszahlen.
- 10) Das Altenheim-Adreßbuch 94 diente auch bei der Untersuchung des Bundesministeriums für Familie, Senioren, Frauen und Jugend "Hilfe- und Pflegebedürftige in Heimen" als Auswahlgrundlage (siehe Schneekloth/Müller 1995).
- 11) Von den folgenden Verbandsträgern haben wir Unterstützung erhalten: dem Diakonischen Werk, dem Deutschen Roten Kreuz, dem Bundesverband privater Alten- und Pflegeheime, dem Deutschen Caritasverband, dem Deutschen Städtetag und dem Deutschen Landkreistag. All diesen Verbandsträgern sei auch an dieser Stelle nochmals gedankt! Zwei weitere Verbandsträger haben wegen ihrer föderalen Organisation auf die regionalen Verbände verwiesen. Nur ein Verband hat sich gegen die Untersuchung gesperrt.

12) Die betreffenden Fragen sind im Fragebogen so formuliert, daß sie ersatzweise auch dem Bewohner selbst hätten gestellt werden können.

13) Soweit in diesen Anlagen keine zentrale 'Heim'-Leitung existierte, fehlen bei den betreffenden Personen auch die Proxi-Interviews. Es handelt sich somit um Totalausfälle.

14) Innerhalb einer Einrichtung waren bis zu vier, in aller Regel ein bis zwei Interviewer tätig.

15) Zusätzlich sind weitere maximal 30 Interviews noch durchführbar, da eine Einrichtung wegen Umzugs erst für einen späteren Zeitpunkt zugesagt hat. Bei den folgenden Bewohneranalysen sind die Interviews dieser Einrichtung noch nicht enthalten, bei den einrichtungsbezogenen Auszählungen ist jedoch die betreffende Einrichtung bereits mitgezählt.

16) Unter den 15 Heimabsagen wurden sechs Absagen mit Personalüberlastung begründet. Jeweils zwei Einrichtungen hatten mit anderen Studien schlechte Erfahrungen gemacht, hielten die Studie für nicht wünschenswert oder haben keinen Grund angeben können. Jeweils eine Einrichtung hat die Absage mit einem Umbau, mit dem Schutz der Bewohner oder mit einer anderen, gerade laufenden Studie begründet.

17) Geringe Abweichungen ergeben sich lediglich aufgrund einiger Einrichtungen mit weniger als 30 Plätzen und einiger Altenwohnanlagen, soweit in diesen keine Proxi-Interviews durchführbar waren.

18) Dabei hatte Priorität, daß der Interviewer mit der ausgewählten Person Kontakt aufnimmt und eine Nichtbefragbarkeit gegebenenfalls selbst feststellt. Lediglich wenn die Einrichtungsleitung ausdrücklich von einer Befragung abgeraten hat oder wenn ethische Bedenken bestanden (z.B. bei Selbstmordgefahr oder bei Sterbenden), wurde die Nichtbefragbarkeit nicht selbst festgestellt.

19) Auf Nachfrage wurde die Nichtbefragbarkeit in aller Regel dahingehend beantwortet, daß der Betreffende "nicht in der Lage sei, sinnvoll zu antworten".

20) Einige Interviews sind natürlich auch an „gelegentlicher“ Desorientierung, die Schneekloth/Müller (1995: 31) auf 21% der Bewohner schätzt, gescheitert. Die Durchführung des Altenheimsurveys hat allerdings gezeigt, daß die Befragbarkeit tendenziell etwas besser war als vom Personal eingeschätzt.

BUCHBESPRECHUNGEN

DURCHBRÜCHE IN DER MODERNEN STATISTIK

Wie jede andere Wissenschaft ist auch die theoretische Statistik kein in sich geschlossenes, gewissermaßen vom Himmel gefallenes System. Statt dessen ist sie eine historisch gewachsene Disziplin, in der sich nicht nur bestimmte Methoden auf unterschiedlichem Wege entwickelt haben, sondern auch hinsichtlich der eigentlichen Grundlagen des Faches völlig konträre Positionen bestehen. Dennoch werden in den Sozialwissenschaften ganz bestimmte Ansichten als Selbstverständlichkeiten vermittelt. Sichtbarster Ausdruck dieser eingefahrenen Denkweisen ist die fast einer 'Einführungsweihe' gleichkommende, stark verschulte und einseitige Statistik-Grundausbildung. Einen "Durchbruch" zu einer offeneren Sichtweise könnte das vorliegende, der Aufmerksamkeit der Sozialwissenschaften bislang entgangene, zweibändige Sammelwerk darstellen.

SAMUEL KOTZ/NORMAN L. JOHNSON (Eds.)

Breakthroughs in Statistics.

I. Foundations and Basic Theory. Springer: New York 1992 (1. Ndr. 1993).

II. Methodology and Distribution. Springer: New York 1992 (2. Ndr. 1996).

631/600 S., je 88 Mark; ISBN: 0-387-94037-5 / 0-387-94039-1

Die beiden vorliegenden Bände enthalten 39 Originalbeiträge bzw. Auszüge aus Originalbeiträgen, die für die Entwicklung der theoretischen und angewandten Statistik des 20. Jahrhunderts von den Vertretern des Faches als "Durchbruch" angesehen werden. Für die Zeit bis etwa 1900 ist die Fixierung solcher Ereignisse wie überhaupt eine Geschichte der Entwicklung der Statistik recht unproblematisch. Das bezeugen die mittlerweile zahlreich vorhandenen Monographien und Aufsatzsammlungen, die die entscheidenden Entwicklungen sowohl in theoretischer als auch angewandter Sicht aufgearbeitet haben. KOTZ und JOHNSON datieren für die Zeit von 1657 - 1889 elf solcher Durchbrüche, von C. HUYGHENS *De Ratiociniis in Aleae Ludo* bis F. GALTONS *Natural Inheritance* (Bd. 1, S. VIII f).

Je weiter man sich der Gegenwart nähert, desto schwieriger wird es naturgemäß, "historische", d.h. für die weitere Entwicklung eines Faches bedeutende Schriften festzumachen. Die Herausgeber haben zu diesem Zweck eine Expertenbefragung durchgeführt: Sie haben weltweit an rund fünfzig führende Statistiker die Frage gerichtet, welche Schriften sie für den Zeitraum ab 1890 als die wichtigsten ansehen. Dabei sollten mindestens fünf genannt werden; der Modalwert der ersten dreißig Antworten lag schließlich bei acht. Während über die Auswahl früherer Schriften bzw. Autoren weitgehend Einigkeit bestand (vor allem über K. PEARSON, "Student", R. A. FISHER, J. NEYMAN und E. S. PEARSON), waren die Angaben für die Jahre nach 1940, insbesondere aber nach 1960, sehr heterogen. Lediglich B. EFRONS Bootstrapping-Arbeit von 1979 sei wieder von den meisten genannt worden.

KOTZ und JOHNSON entwickelten die Idee zu dieser Sammlung während ihrer Tätigkeit als Herausgeber der *Encyclopedia of Statistical Sciences* (1979-1989) und einer systematischen Untersuchung der "Presidential Adresses" an die Royal Statistical Society, das International Statistical Institute und die American Statistical Association der letzten 50 Jahre. Ihre Intention war eine, wie sie sagen, erschreckende Feststellung: die zunehmende Balkanisierung ihres Faches. Um dennoch die Statistik als eine einheitliche Wissenschaft zu retten, war ihnen daran gelegen, eine Sammlung von Texten zusammenzustellen und kommentieren zu lassen, die für die Entwicklung der statistischen Methodologie allgemein als Durchbruch angesehen werden.

Die Zuordnung zu den zwei Bänden ist nach Aussage der Herausgeber zum Teil willkürlich und durch den Wunsch nach gleichem Umfang beider Bände bestimmt. Bezüglich der Auswahl haben sich KOTZ und JOHNSON bis auf zwei oder drei Ausnahmen, die sie persönlich für bedeutend halten (Bd. 1, S. XI), an die Liste der häufigsten Nennungen gehalten. (Leider haben sie die "Hit"liste der häufigsten Nennungen nicht veröffentlicht.) Eine dieser Ausnahmen ist die 1956 vor der American Statistical Association gehaltene Presidential Address "*Statistical Frontiers*" von Gertrude COX, die 1957 im *Journal of the American Statistical Association* veröffentlicht wurde und die als erster Text abgedruckt ist.

Die meisten hier wiederveröffentlichten Beiträge sind auch heute noch leicht zugänglich (23 der 39 Beiträge sind aus dem *Journal of the Royal Statistical Society* (A oder B), dem *Journal of the American Statistical Association*, den *Annals of Mathematical Statistics* oder *Biometrika*). Mit Ausnahme des bereits erwähnten ersten Textes von G. COX, an den sich Kommentare von G. BARNARD, I. J. GOOD, D. V. LINDLEY, F. MOSTELLER und P. K. SEN anschließen, sind alle Texte mit einem einleitenden Kommentar

versehen, der in der Regel den abgedruckten Beitrag in Leben und Werk des Autors sowie in die allgemeine Entwicklung der Statistik einordnet.

Diese Kommentare machen den eigentlichen Reiz der Bände aus. Sie sind ausnahmslos von renommierten Statistikern verfaßt, zum Teil von solchen, die ebenfalls zum Thema des erläuterten Textes grundlegende Beiträge geleistet haben. Gelegentlich findet sich sogar der Kommentar eines Autors, von dem selbst ein Text wiedergegeben ist, wie z. B. von I. J. GOOD zu H. E. ROBBINS Entwicklung des empirischen Bayes-Ansatzes (*An empirical Bayes approach to statistics*, 1955). GOODS 1952 erschienener Beitrag *Rational decisions* wird von D. V. LINDLEY kommentiert, D. A. S. FRASER, dessen Einführung eines strukturellen Wahrscheinlichkeitskonzeptes abgedruckt ist (*Structural probability and a generalization*, 1966) hat die Einleitung zu M. S. BARTLETTS Untersuchung der Fisherschen Statistik (*Properties of Sufficiency and Statistical Tests*, 1937) verfaßt, B. EFRON schließlich, dessen bereits erwähnter Text zum Bootstrapping-Verfahren enthalten ist, kommentiert den Beitrag von W. JAMES und C. M. STEIN über Schätzungen unter einer quadratischen Verlustfunktion (*Estimation with quadratic loss*, 1961), den EFRON als "the most striking theorem of post-war mathematical statistics" bezeichnet (Bd. 1, S. 437).

Über die grundsätzliche Bedeutung der einzelnen Artikel läßt sich nicht streiten, sie können hier auch nicht im einzelnen besprochen werden. Ich will mich im folgenden auf wenige Bemerkungen zu den Beiträgen beschränken, die erstens für die Sozialwissenschaften relevant und zweitens grundsätzlicher und nicht technischer Natur sind.

Die Grundlage der modernen Statistik wurde ohne Zweifel von R. A. FISHER gelegt. FISHER machte 1922 in einem umfangreichen Beitrag erstmals die klare Trennung zwischen theoretischen Parametern und ihren Statistiken deutlich (*On the Mathematical Foundations of Theoretical Statistics*). Zu dieser Zeit waren Begriffe wie Mittelwert, Standardabweichung etc. sowohl für theoretische Parameter als auch für deren Schätzungen in Gebrauch. FISHER postulierte, daß sich die statistische Inferenz auf theoretische und somit feste Parameter (und nicht auf gegebene Daten) zu beziehen habe und markierte damit die Richtung für Forschungen auf dem Gebiet der theoretischen Statistik für die nächsten 50 Jahre. Eine Reihe von zentralen Konzepten wird in diesem Beitrag zum ersten Mal behandelt oder zumindest erstmals systematisch eingeordnet: so etwa die Forderung nach Konsistenz, Effizienz und Suffizienz eines Schätzers. Weiterhin zeigte er, daß die bis dahin hauptsächlich verwendeten Statistiken, die von K. PEARSON propagierten Momentenschätzer, in der Regel weder effizient noch konsistent noch erwartungstreu sind. Statt dessen schlägt er die (für einen speziellen Fall bereits 1912 von

ihm eingeführte) Maximum-Likelihood Methode als allgemeines Schätzprinzip vor, da diese Methode automatisch Schätzer liefere, die die genannten Kriterien erfüllen.

S. GEISSER bezeichnet in seinem Kommentar diesen Beitrag zu Recht als "landmark event in theoretical statistics". Die Bedeutung FISHERS für die moderne theoretische Statistik ist kaum zu überschätzen. Neben den hier angesprochenen Inferenzkonzepten basieren zwei weitere auf seinen Ideen: zum einen der Ausbau der von ihm entwickelten Likelihood-Methode zu einem allgemeinen Likelihood-Prinzip (siehe dazu unten), zum anderen die bis heute umstrittene und viel diskutierte Fiduzialinferenz. Darüber hinaus verdanken wir ihm zahlreiche technische Detaillösungen und schließlich die bis heute grundlegenden Ausführungen über das Problem des "*Design of Experiments*".

Als weiterer Meilenstein der theoretischen Statistik werden einhellig die Arbeiten von J. NEYMAN und E. S. PEARSON angesehen. Während FISHER in bezug auf das Testen von Hypothesen nur die Alternativen "Ablehnung" und "keine Aussage möglich" gelten lassen wollte, entwickelten NEYMAN und PEARSON eine geschlossene Testtheorie, die Ablehnungs- und Annahmehereiche unterschied und Begriffe wie die "Macht" eines Tests, "Fehler erster und zweiter Art" und "gleichmäßig bester Test" einführte. Bis zum Ende des 19. Jahrhunderts basierte das Testen von Hypothesen auf Verteilungen von Teststatistiken, die (1) bestenfalls für große Stichproben angemessen waren und (2) aus intuitiven Gründen herangezogen wurden. Durch die Einführung der t -Verteilung durch W. S. GOSSET 1908 (der Text ist in der vorliegenden Sammlung enthalten) und die Beiträge FISHERS, der die exakten Verteilungen der Prüfgrößen χ^2 , F und einiger Korrelationskoeffizienten in Normalverteilungen ableitete, konnte zumindest Problem (1) überwunden werden. Mit der Lösung dieses Problems stellte sich dann die Frage nach einer formal befriedigenden Testtheorie. E. S. PEARSON berichtete in einer Rückschau, die Idee zu dieser Theorie durch einen Hinweis von GOSSET bekommen zu haben: "It is the simple suggestion that the only valid reason for rejecting a statistical hypothesis is that some alternative hypothesis explains the observed events with a greater degree of probability." (Bd. 1, S. 68). GOSSET argumentierte in diesem Brief, daß ein noch so kleiner Wahrscheinlichkeitswert wie etwa 0.0001 für eine Zufallsstichprobe per se nicht zu einer Ablehnung einer Hypothese führen könne. Lediglich der Vergleich mit einer *Alternativ*hypothese, "which will explain the occurrence of the sample with a more reasonable probability, say .05 (such as that it belongs to a different population or that the sample wasn't random or whatever will do the trick) you will be very much more inclined to consider that the original hypothesis is not true." (Ebda.).

Diese Idee wurde dann gemeinsam von NEYMAN und PEARSON 1928 in einer umfangreichen zweiteiligen Arbeit in der *Biometrika* zum Konzept des Likelihood-Quotiententests

ausgebaut. Während PEARSON darin bereits die einheitliche Methode sah, nach der sie gesucht hatten, zeigte sich NEYMAN noch nicht zufrieden. Das Ergebnis weiterer Überlegungen war dann der Beitrag von 1933, in dem sich auch das berühmte Neyman-Pearson Lemma findet, das besagt, daß in der Klasse aller Tests mit Sicherheitswahrscheinlichkeit α die Gütefunktion des Likelihood-Quotientests die Gütefunktion jedes anderen Tests dominiert (d.h. jeder andere Test hat eine größere Fehlerwahrscheinlichkeit zweiter Art).

NEYMAN und PEARSON zeigten anhand einer Reihe von Beispielen die Anwendung dieses Prinzips und legten damit den Grundstein für eine weithin anerkannte allgemeine Testtheorie, die heute zusammen mit den ebenfalls von NEYMAN formulierten "Konfidenz"-intervallen als "klassisch" bezeichnet wird.

Das von NEYMAN und PEARSON postulierte Vorgehen war im eigentlichen Sinne nichts anderes als eine Verhaltensrichtlinie. Die beiden Autoren formulierten die hinter ihrer Theorie stehende Überzeugung wie folgt: "Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong." (Bd. 1, S. 74) Inferenzaussagen sind demnach hypothetisch-deduktiv und nur möglich, *bevor* sich Ereignisse realisieren und beziehen sich damit nicht auf spezifische Hypothesen, sondern auf die "lange Sicht". Konsequenterweise wurde dieser Ansatz später von A. WALD zu einer reinen Entscheidungstheorie erweitert (WALDS Beitrag *Statistical decision functions* ist hier ebenfalls wiederabgedruckt), und auch NEYMAN hat in seinen späteren Arbeiten immer wieder diesen verhaltenstheoretischen Aspekt betont.

Vehemente Kritik kam dagegen von keinem geringeren als R. A. FISHER, der die Neyman-Pearson Theorie eventuell für Situationen, in denen permanent Entscheidungen getroffen werden mußten, anerkennen wollte; auf keinen Fall aber für inferenzstatistische Beurteilungen in einem *wissenschaftlichen* Sinne.

Auch von anderer Seite wurden dieser Vorgehensweise Vorbehalte entgegengebracht: Zum einen würden *in der Praxis* Modelle meist im Lichte der Daten ausgewählt, außerdem oftmals nicht nur eine, sondern mehrere Hypothesen anhand der selben Daten untersucht, und schließlich sei in vielen Situationen die Reduktion der Inferenz auf eine Ja/Nein-Entscheidung nicht angemessen.

Weiterhin hat sich auch gezeigt, daß optimale (im Sinne gleichmäßig mächtigster) Tests nur für eingeschränkte Situationen existieren oder aber (im Sinne der Maximierung der minimalen Macht) so komplex sind, daß ihre Anwendung erhebliche Probleme bereitet.

Es ist aber zu betonen, daß diese Vorbehalte die Ausnahme darstellen und die überwiegende Mehrzahl, vor allem im Bereich der angewandten Statistik, den Neyman-Pearson Ansatz vorbehaltlos akzeptiert hat, er sich dort geradezu zu einem Paradigma entwickelt hat, auch wenn sich die Statistiker bis heute darüber streiten, wo denn nun genau die Unterschiede zwischen diesem Ansatz und der Fisherschen Testkonzeption liegen [vgl. etwa LEHMANN, E. L. (1993), The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two?, in: *Journal of the American Statistical Association* 88, S. 1242-1249].

Der Neyman-Pearson Ansatz fand schnell starke Verbreitung und entwickelte sich zur dominierenden Inferenzkonzeption. Eine Alternative ergab sich aus einer Weiterentwicklung der Likelihood-Ideen von R. A. FISHER, die vor allem von G. A. BARNARD geleistet und zu einem Likelihood-Prinzip ausgebaut wurde. BARNARD zeigte anhand eines Beispiels eines Parameters, der nur zwei Ausprägungen annehmen konnte, daß das Likelihood-Prinzip für zwei Daten, aus denen das gleiche Likelihood-Verhältnis resultierte, zur selben Entscheidung führen mußte, im inferenzstatistischen Sinne beide Situationen also "gleich" sind. Diese Folgerung stand in starkem Widerspruch zu der dominierenden Neyman-Pearson Theorie.

Weitreichende Beachtung fand das Likelihood-Prinzip dann durch eine Arbeit von BIRNBAUM (1962). BIRNBAUM wies nach, daß das Likelihood-Prinzip (LP) zum einen aus zwei bis dahin allgemein akzeptierten Prinzipien, der Suffizienz (S) und der Konditionalität (K), folgte, zum anderen diese beiden aber auch impliziert: $(LP) \Leftrightarrow (S) \wedge (K)$. Das Likelihood-Prinzip hatte radikale Konsequenzen. Es konstatierte, daß die gesamte Evidenz eines Datums in der Likelihood-Funktion enthalten sei: somit sei, *nachdem* die Daten erhoben wurden, der Stichprobenraum irrelevant. Das bedeutet, daß Evidenzmaße, die sich auf den Raum aller möglichen Daten beziehen (auf den Wahrscheinlichkeits- oder Parameterraum), wie etwa P -Werte oder Konfidenzniveaus, *nach* Realisation eines konkreten Datums für die Inferenz ohne Belang sind. Dies war eine Absage an die frequentistische Position, ohne auf Bayesianische Argumente zurückgreifen zu müssen.

Noch einen Schritt weiter gingen die Vertreter einer Inferenzsicht, die unter dem Eponym "Bayesianer" zusammengefaßt werden. Zumindest drei Texte dieser Richtung werden hier ebenfalls zu den Durchbrüchen der modernen Statistik gerechnet. Für "Bayesianer" ergeben sich bekanntlich aposteriori-Wahrscheinlichkeiten aus dem Produkt von apriori-Wahrscheinlichkeiten und den Likelihood-Werten. Eine zentrale Frage ist hierbei die Herkunft der apriori-Wahrscheinlichkeiten. Für B. DE FINETTI, Vertreter einer radikalen subjektivistischen Sicht, stand 1937 fest, daß die Grundlage jeder Wahrscheinlichkeit subjektiver Natur ist (*Foresight: Its Logical Laws, Its Subjective Sour-*

ces). Eine zentrale Rolle spielt dabei DE FINETTIS "Vertauschbarkeitstheorem": dies besagt, grob gesprochen, daß Zufallsgrößen x_1, x_2, \dots, x_n vertauschbar sind, genau dann wenn ihr gemeinsames Wahrscheinlichkeitsmaß invariant gegenüber ihrer Permutation ist. Ob dies nun, wie R. E. BARLOW in seinem Kommentar zu DE FINETTIS Arbeit behauptet, als eine Lösung des philosophischen Problems der Induktion angesehen werden kann, mag dahingestellt sein. Jedenfalls gebührt DE FINETTI das Verdienst, den damals (und heute) vorherrschenden Konzeptionen von R. A. FISHER und J. NEYMAN/E. S. PEARSON eine kongeniale Alternative entgegengesetzt zu haben, die an die Tradition von T. BAYES und P. S. LAPLACE anknüpfte und zu einem Grundstein der modernen Bayes-Inferenz wurde.

Einen anderen Weg schlug 1955 H. E. ROBBINS ein (*An Empirical Bayes Approach to Statistics*). Im Gegensatz zu DE FINETTI setzte er keine epistemischen, sondern "physikalische" apriori-Wahrscheinlichkeiten voraus. Er ging von der Frage aus, ob man den Bayes-Ansatz auch verwenden könne, wenn die apriori-Wahrscheinlichkeit eines Parameters zwar unbekannt sei, aber "existiere". Die Annahme einer physikalisch existierenden apriori-Wahrscheinlichkeit wird von den meisten "Bayesianern" nicht geteilt (oder positiv gesprochen: nicht benötigt). I. J. GOOD zählt jedoch in seinem Kommentar in der ihm eigenen Art eine ganze Batterie von unterschiedlichen "hierarchischen", "pseudo-" und "quasi-" Bayes-Ansätzen auf, in denen dieser empirische Bayes-Ansatz mittlerweile anzutreffen sei.

Der Durchbruch in praktischer Hinsicht gelang einem Bayes-Ansatz schließlich 1963 durch die Arbeit von W. EDWARDS, H. LINDMAN und L. J. SAVAGE (*Bayesian Statistical Inference for Psychological Research*), der Bayessche Inferenzstatistik erstmals außerhalb theoretischer Debatten zur Sprache brachte. Ungeachtet der offensichtlichen Anerkennung dieser Ansätze seitens der Vertreter der theoretischen Statistik sind Bayessche Auffassungen und ihre Implementation in der heutigen sozialwissenschaftlichen Statistikausbildung zumeist verpönt. Hier dominiert in der Regel der Neyman-Pearson-Ansatz, auch in bezug auf Situationen, in denen eine solche "repeated sampling" Konzeption zumindest fragwürdig ist und sich alternative Entwürfe wie das Likelihood-Prinzip oder ein Bayes-Ansatz eher rechtfertigen ließen. Wie die hier versammelten Kommentare zeigen, verfügen zumindest die Statistiker bislang nicht über eine einhellige Ansicht oder einen "Königsweg".

Auf Probleme ganz anderer Art hat 1962 in einem ebenso bahnbrechenden Beitrag J. W. TUKEY hingewiesen. Diese Arbeit, in der TUKEY zum Teil bereits an anderer Stelle vorgelegene Aspekte systematisiert, zum Teil neue Fragen aufwirft, ist in mehrfacher Hinsicht bemerkenswert. Das Manuskript ging bei der Redaktion der *Annals of Mathe-*

mathematical Statistics (AMS) am 1. Juli 1961 ein, dem selben Tag, an dem ein neuer Herausgeber (J. L. HODGES) sein Amt übernahm. Es erschien als Leitartikel der ersten Ausgabe von 1962, während eine Reihe von Beiträgen, die vom vorhergehenden Herausgeber (W. KRUSKAL) bereits zur Publikation vorgesehen waren, erst Ende 1962 erschienen. Der Beitrag umfaßt insgesamt 67 Seiten. Von den 147 Artikeln, die 1962 in der *AMS* publiziert wurden, haben 25% bis zu 6 Seiten, 50% bis zu 8 Seiten und 75% bis zu 14,5 Seiten Umfang. Der zweitlängste Beitrag umfaßt 29 Seiten. Im Gegensatz zu anderen Artikeln der *AMS* enthält dieser Beitrag keine formalen Ableitungen und Beweise, propagiert keine Optimalitätskriterien. Statt dessen stellt er in 48 durchnummerierten Abschnitten fest, daß bei der Analyse von "real-world data" andere Probleme relevant sind. TUKEY, der nach einem Chemie-Studium in Mathematik promovierte, legt Wert darauf, "Mathematiker" und "Statistiker" zu trennen, wobei er unter letzteren, sich eingeschlossen, "Datenanalytiker" versteht. Sein Beitrag beginnt nicht mit einem Satz wie "Let X be a random variable..." (Das ist z.B. der Beginn des Beitrags von ROBBINS), sondern mit dem folgenden Statement: "For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. [...] All in all, I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data." Er fordert statt dessen neue Fragen, neue Antworten in "unfamiliar summaries", insgesamt "more realistic frameworks", die die Existenz von "spotty data" berücksichtigen.

Gerade in den Sozialwissenschaften genießt TUKEY heute als Begründer der mittlerweile sehr beliebten sog. "Explorativen Datenanalyse" hohes Ansehen. Eine genaue Lektüre seiner Texte kann dabei einige Fehlinterpretationen beseitigen: TUKEY redet hier nicht etwa nichtparametrischen oder verteilungsfreien Methoden das Wort. Diese sollten sich beschränken "(i) upon situations where relative ranks are really all the data available, and (ii) upon situations where unusually quick or portable procedures are needed. In other situations it will be possible, and often desirable, to analyze the data more thoroughly and effectively by other methods." (S. 411) Die Absage an den allgemeinen Gebrauch der Normalverteilung bedeutet für ihn lediglich, diese durch eine andere, angemessenere Verteilung zu ersetzen. Die Konstruktion von Konfidenzintervallen und eine Forderung wie Effizienz gehören ebenso zu seiner Konzeption von "Data Analysis" wie "Robustness" und Monte Carlo Simulationen (S. 418, 421, 446). Andererseits fordert er aber auch, "to give up the vain hope that data analysis can be founded upon a lo-

gico-deductive system [...] and to face up to the fact that *data analysis is intrinsically an empirical science*. (Bd. 2, S. 448f). Wie sind solche Aussagen einzuordnen? Es drängt sich die Frage auf, weshalb sich TUKEY nicht etwa explizit zu einer bayesianischen Konzeption bekennt. Vielleicht sollte man die einzelnen wörtlichen Ausführungen aber auch nicht überstrapazieren. Wie bei anderen genialen Statistikern (R. A. FISHER, J. NEYMAN, G. E. P. BOX) lassen sich auch bei TUKEY bei genauem Hinsehen widersprüchliche Aussagen finden. Bei den über 800 Veröffentlichungen TUKEYS ist allerdings die Gefahr besonders groß, daß man durch (Fehl-) Auswahl einzelner Aussagen das Bild verzerrt.

Die empirischen Sozialwissenschaften können jedenfalls aus den hier kurz vorgestellten Texten lernen, daß nicht alle Daten, die ihnen vorliegen, stichprobenartiger Natur sind und damit die Neyman-Pearson Maschinerie angeworfen werden muß. Sie kann sich auf das - nicht ungefährliche - Terrain der "explorativen" Datenanalyse im Sinne einer "Maldies-mal-jenes" Vorgehensweise begeben oder sich von vornherein auf alternative Inferenzkonzeptionen beziehen, die von der theoretischen Statistik angeboten werden. Die Anwendungsbezogenheit bzw. "Realitätsnähe", die oftmals als Forderung TUKEYS postuliert wird, ist dabei nicht das erste Kriterium. Auch dies kann man aus den hier versammelten Texten lernen. Interessant erscheint mir ein Vergleich von P. J. HUBER und J. W. TUKEY durch F. R. HAMPEL. HUBER sei "one of those newcomers to statistics who, equipped with a broad and deep general mathematical knowledge and unspoiled by unreflected statistical traditions, could take a fresh look at the field." (Bd. 2, S. 484). Stark beeinflusst durch eine Veröffentlichung von TUKEY habe er sich mit der Möglichkeit robuster Schätzungen beschäftigt, aber im Gegensatz zu TUKEYS Arbeiten den Rahmen der mathematischen Statistik niemals verlassen. Dennoch haben seine rein mathematischen Arbeiten eine direkte praktische Relevanz. Nicht zuletzt deswegen gilt die Arbeit *Robust Estimation of a Location Parameter* eben, wie alle anderen hier versammelten Texte, als "Durchbruch".

Wie wird nun die zukünftige Entwicklung aussehen? Ein Aspekt wird in seiner Bedeutung überproportional zunehmen: die Unterstützung durch Computer. Die Bootstrapping-Verfahren wären ohne Computer nicht möglich gewesen und ohne Frage wird die zukünftige Entwicklung der Statistik massiv von diesem technischen Hilfsmittel beeinflusst werden. Bereits die Anwendung verallgemeinerter linearer Modelle durch J. A. NELDER und R. W. M. WEDDERBURN (1972) konnte sich erst allgemein durchsetzen, nachdem mit der Software GLIM ein (unter der Mitarbeit von NELDER entwickeltes) relativ leicht handhabbares Computerprogramm Verbreitung fand. Schon jetzt sind nicht nur in der angewandten Statistik computerintensive Verfahren wie die Monte Carlo Simulation oder das Gibbs Sampling auf dem Vormarsch. Der vor einigen Jahren noch oft zu ver-

nehmende Vorwurf, solche Vorgehensweisen würden das analytische Denken verdrängen (billige CPU-Rechenzeit im Vergleich zu teurer "Denk"zeit) sind mittlerweile einer Einsicht in die Notwendigkeit eines Miteinanders gewichen.

Eine Reihe weiterer interessanter, vor allem wissenschaftshistorischer Aspekte kann hier nicht angesprochen werden, etwa die Frage nach der fachspezifischen Zuordnung, der Diffusionsprozesse oder etwa der Karriereverläufe der Wissenschaftler (die jeweils in Kurzbiographien dargestellt werden). Auch für die Beantwortung solcher Fragen bietet sich die vorliegende Sammlung in hervorragender Weise an. Ein Zitat aus S. C. PEARCES Kommentar zu FISHERS berühmtem Lehrbuch *Statistical Methods for Research Workers* soll stellvertretend für die meisten hier versammelten Texte stehen und damit allgemein die Lektüre der Bände empfehlen (Bd. 2, S. 64): "Fisher was a genius and in these passages he was pointing the way as no one had pointed it before. Like many geniuses, he has suffered from uncritical disciples, but reading the words he actually wrote shows that his claims were modest and completely justified."

THOMAS RAHLF
UNIVERSITÄT HALLE-WITTENBERG
KRÖLLWITZER STRASSE 44
06120 HALLE
E-Mail: Rahlf@t-online.de

* * *

Gerhard Brosius ist vielen SPSS-Anwendern durch seine beiden früheren Bücher „SPSS/PC+ Basics und Graphics“ und „SPSS/PC+ Advanced Statistics und Tables“ bekannt. Die Autoren bieten in dem vorliegenden Band aber ein völlig neues Buch, das keine Überarbeitung der früheren Publikationen darstellt. Erläutert wird der Umgang mit der SPSS-Windows-Version 6.01, wobei das Buch auch für die Version 6.1 zu verwenden ist. Die Autoren beschränken sich auf die Module „Base System“ und „Professional Statistics“. Eine Begleitdiskette mit verschiedenen Datensätzen, anhand derer der Leser die statistischen Prozeduren nachrechnen kann, rundet das Buch ab.

GERHARD BROSIUS/FELIX BROSIUS
SPSS. Base System und Professional Statistics.
1. Auflage. International Thomson Publishing: Bonn 1995
1002 Seiten; 89 Mark. ISBN 3-929821-62-1

Das Buch beginnt mit einführenden Kapiteln zum Arbeiten mit SPSS unter Windows. Die dann folgenden Abschnitte beschäftigen sich mit Datenmanagement-Funktionen (z.B. Transformation, Gewichtung, Fallauswahl, Aggregation). Bei den statistischen Verfahren werden z.B. die Häufigkeitsauszählung, Kreuztabellen, t-Tests, ein- und mehrfaktorielle Varianzanalysen, Korrelationen, einfache und multiple Regressionsanalysen sowie nichtparametrische Testverfahren behandelt. Hier findet sich auch ein Abschnitt zur explorativen Datenanalyse, der dem neuen Trend des grafischen Zugangs zu den Daten Rechnung trägt. Bis dahin sind die gängigsten statistischen Verfahren abgedeckt, und es folgt eine ausführliche Darstellung der Grafikoptionen. Im Anschluß finden sich fortgeschrittenere statistische Verfahren, wie z.B. die Diskriminanz-, Faktoren- und Clusteranalyse. Das Buch schließt mit Aufgaben und Lösungen zu den gebräuchlichsten Auswertungsverfahren.

Einführung und Grundlagen. Zur Einführung zählt der Umgang mit elementaren Windows-Techniken (z.B. Fenstergröße anpassen, Hilfemenü) sowie die möglichen individuellen Einstellungen für das Arbeiten mit SPSS. Anhand eines kurzen Beispiels wird der Weg vom Fragebogen bis zu einer ersten statistischen Auswertung gezeigt. Zu den Grundlagen zählen auch die Dateneingabe und Variablendefinitionen. Nach dem Kapitel zum Aufbau der Datendatei wird das Arbeiten mit den Ausgabe- und Syntaxdateien beschrieben. Auch hier finden sich immer noch Hinweise zu grundlegenden Windows-Techniken, wie z.B. dem Bewegen des Cursors in der Ausgabeliste. Die Darstellung allgemeiner Windows-Grundlagen und programmspezifischen Bedienungshinweisen hätte man besser in getrennten Kapiteln abhandeln sollen, um dann diese Kenntnisse als selbstverständlich für die weitere Behandlung im Buch vorauszusetzen.

Datenmanagement. Beschrieben werden in mehreren Kapiteln die Möglichkeiten der Variablentransformation einschließlich der verschiedenen mathematischen Funktionen, Datenselektionsverfahren und Techniken zur Modifikation der Daten (wie z.B. Aggregieren und Zusammenführen). Es fehlen Hinweise zu portablen SPSS-Dateien sowie zum Datenaustausch mit Datenbank- und Tabellenkalkulationsprogrammen. Auch das Einlesen von Daten im ASCII-Format wird nicht erwähnt.

Statistik. In den Kapiteln zu den statistischen Prozeduren werden zunächst die statistischen Grundlagen für die jeweils besprochene Analyse ausführlich behandelt. Erst danach finden sich Hinweise, mit welchen Menüpunkten diese Analyse in SPSS für Windows anzufordern ist. Dabei werden Datensätze verwendet, die bereits für die Erklärung der Statistiken herangezogen wurden, ohne allerdings die Analyse vollständig zu berechnen. Man sollte die jeweils interessierenden Kapitel vollständig lesen, da sonst die Beispiele nicht nachvollzogen werden können. Die Erläuterungen zu den statistischen Ver-

fahren richten sich an den Statistik-Anfänger und sind überaus verständlich. Am Ende des Buches finden sich zur praktischen Umsetzung des Erlernten Aufgaben, die man selbständig anhand der mitgelieferten Daten bearbeiten und mit den beigelegten Lösungen vergleichen kann.

Grafik. Auf mehr als 200 Seiten wird auf die Grafikoptionen in SPSS eingegangen. Ein Überblicksartikel beschreibt die generelle Erstellung von Grafiken und das Arbeiten mit dem Grafik-Karussell. Im Anschluß daran werden die Möglichkeiten der Modifizierung von Grafiken, d.h. z.B. der Änderung von Farben in Muster oder dem Einfügen von Legenden und Beschriftungen, beschrieben. Die verschiedenen Grafiktypen werden in separaten Kapiteln ausführlich behandelt. Auf die Übergabe von Grafiken in andere Programme (z.B. die Einbettung in eine Textverarbeitung) wird leider nicht näher eingegangen.

Wie schon in früheren SPSS-Büchern liegt eine der Stärken von Gerhard Brosius in der verständlichen Erklärung statistischer Verfahren. Dies ist ihm auch im vorliegenden Buch gelungen. Daher ist das Buch für Statistik-Anfänger zu empfehlen. Für diese Zielgruppe wäre es allerdings ausreichend gewesen, sich auf die wichtigsten statistischen Verfahren zu beschränken, so daß das Buch insgesamt handlicher geworden wäre. Für den erfahrenen Statistik- und SPSS-Anwender ist es dagegen zu ausführlich angelegt und für einen kurzen Überblick weniger geeignet. Etwas verwundert haben uns die vielen Tippfehler bei einem Preis von 89 Mark, zumal Textverarbeitungsprogramme heute Rechtschreibprüfungen anbieten. Hier sollten die Autoren und der Verlag bei einer Neuauflage des Buches mehr Sorgfalt darauf verwenden.

MICHAELA THOMA/CORNELIA ZÜLL

* * *

Am Anfang stand eine Tagung im Jahr 1992 zum Thema "Qualitative Research Process and Computing", bei der das Angebot einschlägiger Software und deren Anwendung im Vordergrund standen. Das vorliegende Buch ist allerdings kein Tagungsband, obwohl ein großer Teil der Autoren mit den bei der Tagung 1992 anwesenden Softwareentwicklern identisch ist. Udo Kelle ist es gelungen, statt eines Tagungsbandes einen Band vorzulegen, der sich den unterschiedlichen Aspekten der computerunterstützten Analyse qualitativer Daten zuwendet und sich vor allem mit Theorie und Methodologie auseinandersetzt. Daß dieses anschaulich nicht ohne Software geschehen kann und soll, ist selbstverständlich.

UDO KELLE (HG.)

Computer-Aided Qualitative Data Analysis. Theory, Methods and Practice.

London: Sage 1995

224 Seiten, ISBN 0-8039-7761-1, £ 13,95 (Paper), £ 37,50 (Cased)

Durch die Breite des Spektrums der Auseinandersetzung mit dem Thema der computerunterstützten Analyse qualitativer Daten gliedert sich das Buch in vier Teile: Nach einer Einführung durch den Herausgeber werden im ersten Teil sehr generelle methodologische Fragen angesprochen. Zunächst diskutieren Kelle und Laurie die Frage der Validität der Ergebnisse qualitativer Sozialforschung bei Einsatz des Computers mit Blick auf Textauswahl und Kodierung. Sodann zeichnen Lee und Fielding des Nutzers Erfahrungen und Nöte im Kennenlernen und im Umgang mit Software zur Analyse qualitativer Daten nach und machen potentiellen Nutzern Mut. Nachfolgend diskutiert Lonkula die "Grounded Theory" vor dem Hintergrund einer computerunterstützten Analyse und hebt deren Einfluß auf die Entwicklung einer spezifischen Software hervor. Diesen Teil abschließend diskutieren Seidel und Kelle unterschiedliche Funktionen des Kodierens für die Analyse von Textdaten.

Im zweiten Teil wird der Zusammenhang zwischen Computereinsatz und Theoriebildung betrachtet. Hierbei wird auf die unterschiedlichen Zusammenhänge zwischen Textsegmenten, Codes und Memos und deren spezifischer Signifikanz für die qualitative Theoriebildung eingegangen und die Möglichkeit diskutiert, per Computer aus diesen Textelementen "Netzwerke" zu konstruieren. Dieser Teil beginnt mit Dey's Diskussion der Probleme einer Zerlegung von Textdaten und einer anschließenden computerunterstützten analytischen, kontextorientierten Verbindung zwischen Textsegmenten über "Hypertext"-Verknüpfungen. Richards und Richards folgen mit einer Betrachtung der hierarchischen Struktur zwischen Codes erster und zweiter Ordnung und diskutieren unterschiedliche methodologische Aspekte der Konstruktion und Analyse hierarchischer Kategorien. Den Teil zwei abschließend demonstriert Araujo seine Erfahrungen der Entwicklung vom Codeschema zum hierarchischen Netzwerk von Kategorien.

Der dritte Teil beschäftigt sich mit dem Einsatz des Computers bei der Hypothesenprüfung. Hierbei werden rechnerische und methodologische Aspekte des Einsatzes des Computers bei der Organisation der Textdaten für die qualitative Hypothesenprüfung diskutiert. Dieser Teil beginnt mit einer Betrachtung der logischen und rechnerischen Grundlagen computerunterstützter Hypothesenprüfung von Sibert und Shelly, um dann durch eine zweite Strategie der Hypothesenprüfung von Hesse-Biber und Dupuis ergänzt zu werden. Die unterschiedlichen Strategien sind bedingt durch eine unterschiedliche

Definition von "Hypothese": Sibert und Shelly sehen hierin eine vage Idee über die Beziehung zwischen zwei oder mehr Konzepten, Hesse-Biber und Dupuis betrachten eine Hypothese als eine logische Feststellung über An- und Abwesenheit oder Beziehungen zwischen gewissen Tatsachen (vgl. Seite 111). Teil drei abschließend berichtet Huber über spezifische Techniken der Ableitung für die qualitative Theoriebildung eines deduktiven Systems.

Der vierte Teil beschäftigt sich mit dem Computereinsatz für die Triangulation, verknüpft also hermeneutische Methoden des Verstehens mit Techniken der Reduktion und Standardisierung von Informationen. Hier geht es um Verfahren, die in einer Kombination zwischen qualitativen und quantitativen Operationen quantitative Daten aus Textmaterial generieren. Dieser Teil beginnt mit Kuckartz's Demonstration einer typologisierenden fallorientierten Klassifikation, aufbauend auf den theoretischen und methodologischen Überlegungen von Weber und Schütz. Es folgen Überlegungen von Roller, Mathes und Eckert zu einer hermeneutisch-klassifikatorischen Inhaltsanalyse, deren Ausgangsmaterial im Gegensatz zum üblichen qualitativen Datenmaterial halbstrukturierte Interviews aus großen Stichproben sind. In einem dritten Beitrag beschäftigt sich Ragin mit einer Methode der Formalisierung des intuitiven Prozesses multipler Vergleiche.

Im letzten Kapitel des Buches, dieses wäre besser als Teil fünf zu bezeichnen, geben Prein, Kelle und Bird einen Überblick über die von den Autoren benutzte und oft auch entwickelte Software. Hierbei wird kein Anspruch auf Vollständigkeit erhoben. Dieses leisten andere Publikationen, wie z.B. WEIZMAN & MILES, 1995: Computer Programs for Qualitative Data Analysis. London: Sage.

Alles in allem gibt das vorliegende Buch einen guten Einblick in die Möglichkeiten des Computereinsatzes bei der qualitativen Datenanalyse. Erfreulich ist zu vermerken, daß Kelle eine große Anzahl namhafter Kollegen für dieses Buchprojekt begeistern konnte. Die Breite des angesprochenen Themenspektrums und die Nähe zur konkreten Analyse, die nicht zuletzt auch aus der intimen Kenntnis der Software durch die Autoren herrührt, bringen dem interessierten Forscher die Notwendigkeit eines Computereinsatzes bei der qualitativen Analyse nahe. Natürlich kommt es hierbei dem Buche zu Gute, daß die meisten Autoren gleichzeitig auch Softwareentwickler sind. Dennoch darf man das Buch nicht als ein Anpreisen von Softwareprodukten mißverstehen, sondern es handelt sich um eine ehrliche Auseinandersetzung mit Theorie und Methodologie sowie mit den Daten qualitativer Analyse.

JÜRGEN H.P. HOFFMEYER-ZLOTNIK

SUDMAN, BRADBURN und SCHWARZ bieten eine Zusammenstellung kognitionspsychologischer Forschungen und Befunde zum Thema Umfragemethodologie. Besonderes Augenmerk wird in fast allen Kapiteln auf die praktische Bedeutung der Ergebnisse für die Konstruktion von Fragebögen gelegt. Das Buch enthält 11 Kapitel, beginnend mit einer allgemeinen Einführung in die kognitionspsychologische Perspektive (Kapitel 1), einer Übersicht über Methoden zur Aufdeckung von kognitiven Prozessen und Problemen in Fragebögen (Kapitel 2) und einer Beschreibung der kognitiven und kommunikativen Prozesse im Interview (Kapitel 3). Kapitel 4 bis 6 beschäftigen sich mit Kontexteffekten, im wesentlichen bei der Messung von Einstellungen. In Kapitel 7 bis 10 wird die Messung objektiver Sachverhalte abgehandelt. Im letzten Kapitel tragen die Autoren nochmals die Ergebnisse der vorangegangenen Kapitel zusammen und folgern daraus einige Grundprinzipien für die Konstruktion von Fragebögen.

S. SUDMAN/N. BRADBURN/N. SCHWARZ
**Thinking about Answers. The Application of Cognitive
Processes to Survey Methodology**
1. Auflage. San Francisco: Jossey-Bass Publishers. 1996
304 Seiten, ISBN 0-7879-0120-2, \$ 33.95

Das zweite Kapitel enthält eine Darstellung der gegenwärtig verwendeten Methoden, um die bei der Beantwortung von Fragen stattfindenden kognitiven Prozesse aufzudecken. Die vorgestellten Methoden, wie z.B. das Erstellen verbaler Protokolle („Thinkalouds“), Kognitive Labors oder die Codierung der Befragten-Interviewer-Interaktion (Behavioral Coding), werden jeweils dahingehend analysiert, inwieweit sie sich zur Aufdeckung von Verständnisproblemen im Fragebogen eignen.

Das dritte Kapitel führt in die kognitionspsychologischen Grundlagen ein, die für die empirischen Studien der weiteren Kapitel maßgebend sind. Um eine Frage in einem Interview zu beantworten, muß der Befragte mehrere Aufgaben bewältigen: Er muß die Frage verstehen, entsprechende Informationen zur Beantwortung heranziehen und verarbeiten, eine Antwort bilden und sie in die zur Verfügung stehenden Antwortalternativen einpassen. Diese Aufgaben des Befragten werden im Buch sequentiell dargestellt. Die Autoren weisen allerdings darauf hin, daß die einzelnen Aufgaben teilweise gleichzeitig abgearbeitet werden. Der durch frühere Fragestellungen konstruierte Kontext ist in der Interviewsituation häufig noch aktiviert, wenn spätere Fragen beantwortet werden. Diesem Umstand schreiben die Autoren viele der in dem Buch dargestellten Kontexteffekte zu.

In den folgenden drei Kapiteln finden sich ausführliche Darstellungen von Kontexteffekten. Der Kontext, in dem Fragen gestellt werden, kann die kognitiven Prozesse des Befragten auf jeder Stufe des Frage-Antwort-Prozesses unterschiedlich beeinflussen. Dabei werden wichtige Unterscheidungen eingeführt, die zur Vorhersage von Art und Stärke von Kontexteffekten wichtig sind: zwischen mentalen Repräsentationen des *Frageinhalts* oder des *Standards*, an dem der Frageinhalt zu messen ist, einerseits, und zwischen permanent oder nur temporär (z.B. durch den Einfluß vorheriger Fragen) zugänglicher Information andererseits. Eingehend werden Kontexteffekte diskutiert, die dadurch entstehen, daß bei vorhergehenden Fragen Informationen aktiviert wurden. Diese Informationen beeinflussen den Frage-Antwortprozeß und führen entweder zu Assimilations- oder Kontrasteffekten. Schließlich wird der Einfluß der Reihenfolge von Antwortalternativen (response order effects) behandelt. Für Primacy-Effekte (häufigere Nennungen des Items, das am Anfang der Liste steht) und Recency-Effekte (d.h. Items werden häufiger genannt, wenn sie am Ende einer Liste stehen), werden empirische Befunde und eine theoretische Konzeptionalisierung vorgestellt.

In weiteren vier Kapiteln beschäftigen sich die Autoren mit Verhaltensberichten. Wenn Befragte über zurückliegendes Verhalten Auskunft geben sollen, müssen sie einen bestimmten Teil ihres Gedächtnisses, das sogenannte autobiographische Gedächtnis, nach relevanten Informationen durchsuchen. Die Autoren beschreiben, wie das autobiographische Gedächtnis organisiert ist und diskutieren die Implikationen für die Umfrageforschung.

Manchmal will der Forscher nicht nur wissen, ob ein bestimmtes Ereignis in einer spezifischen Zeitperiode eingetreten ist, sondern auch, wann das war. Die genaue Zeitangabe (event dating) ist eine schwierige kognitive Aufgabe. Befragte werden über genaue Daten zunehmend unsicherer, je länger das Ereignis zurückliegt. Aus verschiedenen Studien weiß man, daß Befragte eher mehr Ereignisse in einer Referenzperiode angeben als sich tatsächlich ereignet haben („Telescoping“). Wenn es zu schwierig ist, eine Frage bezüglich der Häufigkeit eines bestimmten Ereignisses aus dem Gedächtnis zu beantworten, verwenden Befragte Schätzstrategien. Die Autoren diskutieren, wann und wie Schätzstrategien herangezogen werden.

Verhaltensberichte werden nicht nur über das eigene Verhalten der Befragten erfaßt, sondern der Befragte soll häufig auch Auskünfte über andere Personen geben, die mit im gleichen Haushalt leben oder die er gut kennt („Proxy Reporting“). Aus den vorgestellten Experimenten kann geschlossen werden, daß diese Berichte über andere bei vielen Verhaltensweisen nicht ungenauer sind als Berichte über eigenes Verhalten.

Alle Fragen werden in einem Kontext gestellt. Von daher ist es nicht möglich Kontexteffekte auszuschließen. Was dem Forscher von den Autoren nahegelegt wird, ist dann auch eher zu versuchen, Kontrolle über die Kontexteffekte zu erlangen, d.h. über die ablaufenden Prozesse Klarheit zu erlangen. Allenfalls wird vorgeschlagen, Fragen nicht in einem Kontext zu präsentieren, der stark von dem Kontext abweicht, in dem normalerweise über diese Fragen nachgedacht wird. Andererseits kann es aber auch sinnvoll sein, zur Testung bestimmter Hypothesen Kontexteffekte bewußt zu produzieren.

Die Autoren weisen darauf hin, daß es sehr schwierig sein kann, bei einer gegebenen Frage zu prognostizieren, welche Effekte welche Wirkung haben. Letztlich können nur methodische Studien vor der Haupterhebung Klarheit schaffen. Die zu ziehenden Schlußfolgerungen für die Gestaltung von Fragebögen können folglich nicht in Form von Rezepten präsentiert werden, weil dies der Komplexität eines Fragebogens nicht gerecht wird. Stattdessen wird für die Verwendung der 'Think Aloud'-Technik und anderer Methoden, gegebenenfalls aber auch multipler Vorstudien bei der Konstruktion von Fragebögen argumentiert.

Es liegt noch keine umfassende Theorie des Umfrageprozesses vor, die alle Aspekte bei der Beantwortung von Fragen mit einer begrenzten Anzahl theoretischer Prinzipien erklären kann. Darin bildet auch das vorliegende Werk keine Ausnahme. Es verwendet Theoriestücke sowohl aus psycholinguistischer, sozialer und kognitiver Psychologie. Aber dennoch ist es den Autoren in dem anschaulich geschriebenen und leicht lesbaren Buch gelungen, die relevanten Theoriestücke und empirischen Ergebnisse differenziert und klar verständlich zu präsentieren. Das Buch bietet einen Überblick über die kognitionspsychologische Forschung der letzten Jahrzehnte zum Thema Umfragemethodologie. Es schließt an frühere Werke von Sudman, Bradburn und Schwarz zum gleichen Thema an. Das Besondere am vorliegenden Band ist, daß sehr deutlich die Implikationen der Befunde für die Fragebogenkonstruktion herausgearbeitet werden. Das Buch enthält eine Vielzahl nützlicher Hinweise, wie Fragebögen besser zu schreiben sind. Ratschläge in dieser Form sind von Umfragepraktikern in früheren Publikationen zum Thema Kognition und Umfrageforschung vermißt worden. Man kann es sowohl den Praktikern und Theoretikern der Umfrageforschung als auch Kognitionspsychologen empfehlen.

MICHAELA THOMA/MICHAEL BRAUN

Das Buch von WALTER ASSENMACHER „Deskriptive Statistik“ ist in der Reihe Springer Lehrbuch erschienen und soll einen umfassenden Überblick über Methoden der deskriptiven Statistik geben, die durch einige Verfahren der explorativen Datenanalyse ergänzt wurden. Es sei daher ein Vergleich mit dem 1994 in derselben Reihe erschienenen Buch von POLASEK: *EDA Explorative Datenanalyse. Einführung in die deskriptive Statistik* erlaubt, das in den ZUMA-Nachrichten 36 besprochen wurde. Insgesamt fällt das Buch von Assenmacher hinter das Buch von Polasek zurück, obwohl ersteres eine saubere und fehlerfreie Darstellung der deskriptiven Statistik liefert. Bei einem Preisunterschied von 9 Mark erhält man bei Polasek ein um 100 Seiten umfangreicheres Buch, das wegen einzeiligen Abstands zudem mindestens die doppelte Information liefert.

WALTER ASSENMACHER

Deskriptive Statistik

Springer Verlag 1996: Heidelberg u.a.
252 Seiten, 36 Mark, ISBN 3-540-60715-3

Die äußere Form ist bei Assenmacher eher langweilig und zum Lesen ermüdend. Man vergleiche nur etwa die Einteilung der Verhältniszahlen bei Polasek auf S. 261 mit der von Assenmacher auf Seite 220. Sein Text wird nur unterbrochen durch ein paar Formeln, Tabellen und Abbildungen. Zum Vergleich: Im Buch von Assenmacher sind 44 Abbildungen und 46 Tabellen, bei Polasek sind es 135 Abbildungen und 77 Tabellen.

Eine kleine Auflockerung bei einem Buch sind Unterschiede in der Auszeichnung der Schrift bei Definitionen. Dies ist nicht nur eine Geschmacksfrage, sondern zudem noch sehr hilfreich. Sucht man im Sachverzeichnis einen Begriff, etwa *Funktionalparameter*, wird bei Assenmacher auf Seite 63 verwiesen. Da Begriffe beim ersten Erscheinen weder unterstrichen, kursiv noch im Fettdruck stehen, muß man praktisch die ganze Seite lesen, bis man die entsprechende Stelle findet.

Die Beispiele sind bei Polasek praxisorientierter als bei Assenmacher. Mit keinem Wort werden bei Assenmacher Softwareprogramme erwähnt, die heute die ganze Rechenarbeit erledigen und eine Fülle von Möglichkeiten der deskriptiven Statistik bieten. Das Buch berechnet nach herkömmlicher Art und Weise unter Verwendung von Umformungsregeln die Kenngrößen, wie das bei Studenten in der Prüfung eben der Fall ist. Interpretationen oder Sensitivitätshinweise sind eher selten. Welches Kriterium der Autor hat, um bei einem korrigierten Kontingenzkoeffizient von 0,3768 zu schreiben, daß zwischen den betrachteten Variablen ein Zusammenhang besteht, der jedoch nicht sehr

ausgeprägt ist, wohingegen beim gleichen Problem ein Assoziationskoeffizient von $-0,6757$ den Schluß auf hohe Korrelation zuläßt, bleibt sein Geheimnis. $x_{(1)}$ bezeichnet auf Seite 57 die kleinste Beobachtung, auf Seite 123 die größte Beobachtung. Ein Literaturverzeichnis gibt es in Form einer Literaturliste, die nur Bücher, aber keine Softwarehinweise enthält, und Aufgabensammlungen. Merkwürdigerweise wird aber das in derselben Reihe erschienene Buch von Polasek nicht aufgeführt.

Vielleicht kommt das Buch von Assenmacher etwas zu schlecht weg, was jedoch nicht daran liegt, daß der Autor in seinen Ausführungen über Statistische Institutionen ZUMA als repräsentatives Glied der Markt- und Meinungsforschungsinstitute einordnet. Zu guter Letzt seien daher auch ein paar positive Dinge gesagt. Stoffmäßig enthält das Buch alle klassischen Bereiche der deskriptiven Statistik wie Häufigkeitsverteilungen, Parameter ein- und zweidimensionaler Datensätze, elementare Regressions- und Zeitreihenanalyse, sowie Verhältnis- und Indexzahlen. Den einzelnen Abschnitten folgen auch Übungsaufgaben, deren Lösungen am Schluß des Buches angegeben sind. Gut gefallen hat mir die klare Ausdrucksweise und die axiomatische Grundlagendarstellung bei der Konstruktion der Parameter. Lobenswert - weil mittlerweile nicht immer üblich - ist schließlich auch, daß das Buch ziemlich frei von lästigen Druckfehlern ist.

SIEGFRIED GABLER

BUCHHINWEISE

SOFTSTAT '95

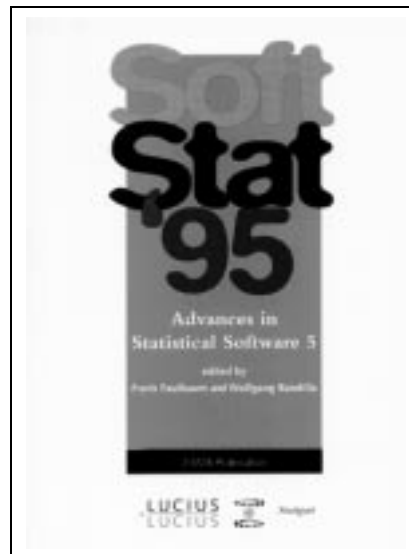
X ADVANCES IN STATISTICAL SOFTWARE 5

Frank Faulbaum/Wolfgang Bandilla

Lucius & Lucius: Stuttgart 1996

650 Seiten, 124 Mark, ISBN 3-8282-0001-X

This volume reports on the 8th Conference on the Scientific Use of Statistical Software (SoftStat '95). The purpose of the conference series is to provide an overview of the different ways in which methods and tools of computer science can be made useful for statistics. In particular the conference deals with new developments, applications, comparisons and evaluations of statistical analysis systems including the various mathematical procedures and algorithms which underly the different software realizations. In addition it is concerned with the role of statistical analysis systems in the research process, in teaching and education as well as in the solution of concrete problems in special fields of application. The topics dealt with at this conference were: Algorithmic Aspects of Statistical Data Analysis, Computer Programs for Statistical Data Analysis, Experimental Design, Exploratory Data Analysis, Geographical Information Systems, Interactive Statistical Graphics, Metadata and Statistical Information Systems, Scaling and Classification, Simulation, Software Tools in Qualitative Social Research, Computer Programs for Statistical Modeling, Statistics and Neural Network Computing, The Role of Internet Services, Training in Statistics and Statistical Software.



BELIEFS IN GOVERNMENT SERIES

Die fünf Bücher der Serie „Beliefs in Government“, herausgegeben von *MAX KAASE*, *KENNETH NEWTON* und *ELINOR SCARBROUGH*, sind das Ergebnis eines sechsjährigen Forschungsprogramms der European Science Foundation. Alle fünf Bücher sind 1995 im Verlag *Oxford University Press* erschienen.

CITIZENS AND THE STATE

HANS-DIETER KLINGEMANN, DIETER FUCHS (EDS.)

Fears that representative democracy in western Europe is in crisis are examined on the basis of trends in mass attitudes over the past two or three decades. The evidence suggests not crisis but a changing relationship between governments and citizens.

ISBN 0-19-827955-8. Vol. 1. 35 £.

THE SCOPE OF GOVERNMENT

OLE BORRE, ELINOR SCARBROUGH (EDS.)

Long-term survey data reveal persistently high levels of support across western Europe for the public provision of welfare and social security services. This book reaches surprising conclusions about theories of ungovernability, overload, welfare state backlash, and tax revolt.

ISBN 0-19-827954-X.
Vol. 3. 35 £.

PUBLIC OPINION & INTERNATIONALIZED GOVERNANCE

OSKAR NIEDERMAYER, RICHARD SINNOTT (EDS.)

New agencies of international government are among the most innovative and experimental aspects of late twentieth century politics. This book examines mass opinion about the European Union, its structures, operations, enlargement, and legitimacy, but it also deals with the UN and NATO.

ISBN 0-19-827958-2. Vol. 2. 35 £.

THE IMPACT OF VALUES

JAN W. VAN DETH, ELINOR SCARBROUGH (EDS.)

Declining religiosity, waning class values, rising postmaterialism, along with Green values, postmodernism, feminism, are indicative of profound and widespread change in the values of citizens. This book tracks these changes and analyses their impact on political efficacy, interest, activity, trust, voting and involvement in new social movements.

ISBN 0-19-827957-4. Vol. 4.

BELIEFS IN GOVERNMENT

MAX KAASE, KENNETH NEWTON

This overview book pulls together the empirical findings of the other four volumes and sets them in the broad context of mass politics in modern Europe. Its conclusions about political participation, mass political opinion, postmaterialism and post-modernism, and democratic change and stability break new ground in political science.

ISBN 0-19-827956-6. Vol. 5. 25 £.

**ALLE, NICHT JEDER.
EINFÜHRUNG IN DIE METHODEN DER DEMOSKOPIE**

Elisabeth Noelle-Neumann/Thomas Petersen

München: dtv 1996

656 Seiten, 38 Mark, ISBN 3-423-04688-0



Dieses Buch ist zuerst 1963 unter dem Titel „Umfragen in der Massengesellschaft“ in der Reihe „rowohlts deutsche enzyklopädie“ erschienen. Jetzt ist das Buch überarbeitet und mit doppeltem Umfang bei dtv neu erschienen, mit Thomas Petersen als Koautor und neuem Haupttitel: „Alle, nicht jeder“. Um den Titel zu verstehen, braucht man sich nur an Conan Doyle zu halten, der seinen Meisterdetektiv Sherlock Holmes, so die Autoren, sagen läßt: „Der einzelne Mensch ist ein unlösbares Rätsel, aber zusammengenommen werden die Menschen zu einer statistischen Gewißheit. Man kann, zum Beispiel, niemals voraussagen, was ein bestimmter Mensch tun wird, aber mit großer Präzision kann man sagen, was der Durchschnitt einer großen Zahl von Menschen im Schilde führt. Die Individuen sind äußerst verschieden und wechselhaft, aber Prozentzahlen sind konstant. Das lehren uns die Statistiker.“ In acht Hauptkapiteln wird die Methode der Demoskopie vorgeführt, mit theoretischen Prinzipien, zahlreichen Beispielen aus der Praxis des Allensbacher Instituts und anderen Forschungsinstituten, unter anderem auch mit über hundert Beispielen zu Frageformulierungen, 35 Analyse-Beispielen und den Ergebnissen von 30 Feldexperimenten.

AUF DER SUCHE NACH GIEßEN. ERGEBNISSE VON BÜRGERBEFRAGUNGEN UND STADTFORSCHUNG

Eckhard Dommer/Peter Schmidt

Gießen: Ferber 1995

199 Seiten, 28 Mark, ISBN 3-927835-68-4



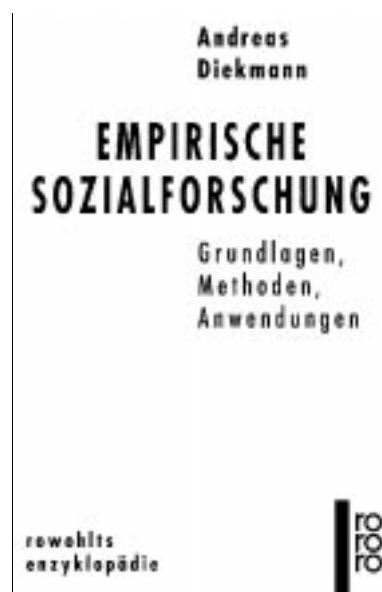
Sowohl die Repräsentanten der Stadt Gießen als auch die Justus-Liebig-Universität Gießen stellten in den letzten Jahren fest, daß Image und Realität der Stadt auseinanderfallen. Daher entstand ein starkes Interesse, sich mit der Wahrnehmung der "objektiven" Gegebenheiten (z.B. Zahl der Arbeitsplätze, Straßennetz, ÖPNV) und den subjektiv wahrgenommenen Merkmalen, (d.h. dem Image) der Stadt auseinanderzusetzen. Es interessierte insbesondere, ob sich die Meinung der Gesamtbevölkerung und die Einschätzungen der Angehörigen der Universität, besonders der Studierenden unterscheiden, da die Studenten einen großen Anteil der Stadtbevölkerung bilden. Inhalt: Gießen - auf der Suche nach der verlorenen Identität und der aktiven Gestaltung einer modernen Stadt; Das Image der Stadt Gießen - aktuelle und historische Dimensionen; Leitfadeninterviews mit Interessenvertretern und Vertretern aus verschiedenen Wirtschaftssektoren; Die Bürgerbefragung: Das Image der Stadt in der Meinung der Bürger; Umzug und Abwanderung; Gießen, kommunale Probleme und Einstellungswandel; Verkehr in Gießen - der unvermeidbare Preis der Zentralität?

**EMPIRISCHE SOZIALFORSCHUNG
GRUNDLAGEN, METHODEN, ANWENDUNGEN**

Andreas Diekmann

Hamburg: Rowohlt 1995

639 Seiten, 32,90 Mark, ISBN 3-499-55551-4



Behandelt werden grundlegende Methoden der modernen empirischen Sozialforschung. Die Vor- und Nachteile der einzelnen Untersuchungstechniken werden kritisch herausgearbeitet und an zahlreichen Beispielen aus der Forschungspraxis illustriert. Im Mittelpunkt dieses Lehrbuchs stehen: Untersuchungsplanung, Stichproben, Messung und Skalierung von Einstellungen, Querschnitts-, Panel- und Kohortenstudien, experimentelle und quasi-experimentelle Designs, Erhebungsverfahren, nicht-reaktive Methoden, Datenanalyse. Die Kenntnis dieser Methoden, die praktisch in allen sozialwissenschaftlichen Disziplinen verwendet werden, aber auch in der Markt-, Meinungs-, Wahl- und Medienforschung sowie in den statistischen Ämtern, ist unerlässlich für jeden, der sich auseinandersetzen will mit gesellschaftlichen Entwicklungen und Zusammenhängen, die auf Daten und Zahlen beruhen.

ZUMA-NACHRICHTEN SPEZIAL
TEXT ANALYSIS AND COMPUTERS

Cornelia Züll/Janet Harkness/Jürgen Hoffmeyer-Zlotnik (Hrsg.)

Mannheim: ZUMA 1996

132 Seiten, 25 Mark, ISBN 3-924220-11-5

The ZUMA-Nachrichten Spezial "Text Analysis and Computers" documents a cornerstone in an initiative to bring together scholars from different disciplines engaged in the computer-assisted analysis of texts. It presents reprints of the talks given by four keynote speakers and the abstract of all the papers presented at the Text Analysis and Computers Conference held in Mannheim from September 18-21, 1995. The papers address four broad areas - content analysis, qualitative approaches in the social sciences, information processing and corpus linguistics; each of the keynote speakers' paper reviews one of these areas:

Computer-Aided Qualitative Data Analysis: An Overview (*Udo Kelle*); Principles of Content Analysis for Information Retrieval Systems (*Jürgen Krause*); Machine-Readable Text Corpora and the Linguistic Description of Languages (*Christian Mair*); Computer Assisted Content Analysis (*Erhard Mergenthaler*).

One of the main aims of the conference was to provide a forum for an exchange on text analysis procedures and potentials across disciplines, an ambitious undertaking in view of the diversity of perspectives and interests involved. The conference undoubtedly accomplished some of the ground work necessary for an interdisciplinary discourse to begin; a number of cooperative projects have been started - on new tools for text analysis, on establishing an internet discussion forum and on planning more intensive research cooperation between the humanities and the social sciences.

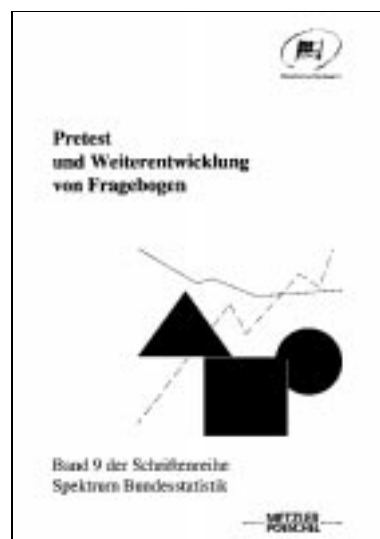


PRETEST UND WEITERENTWICKLUNG VON FRAGEBOGEN

Statistisches Bundesamt

Stuttgart: Metzler-Poeschel 1996

119 Seiten, 17,80 Mark, ISBN 3-8246-0472-8



Pretestverfahren gewinnen als Instrument der Qualitätsverbesserung sowohl in der Umfrageforschung als auch in der Umfragepraxis zunehmend an Bedeutung. Sie werden zur Vorbereitung von Erhebungen nicht nur von Marktforschungs- und sozialwissenschaftlichen Instituten, sondern auch in der amtlichen Statistik eingesetzt. Schwerpunkt dieses Bandes sind die einzelnen Beiträge der wissenschaftlichen Tagung „Pretest und Weiterentwicklung von Fragebogen“, die im Juni 1995 im Statistischen Bundesamt stattfand. Die Beiträge befassen sich einerseits mit Methoden zum Test von Erhebungsinstrumenten und andererseits mit der praktischen Umsetzung von Pretests. Ergänzt werden die Tagungsbeiträge um die Erfahrungen aus zwei Pretests, die das Statistische Bundesamt in Zusammenarbeit mit dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) durchgeführt hat.

ZUMA-ARBEITSBERICHTE

Nachfolgend sind die ZUMA-Arbeitsberichte, die seit Dezember 1995 publiziert worden sind, in Form von Abstracts kurz dargestellt. ZUMA-Arbeitsberichte werden Interessenten auf Anfrage zugesandt. Bestellungen sind zu richten an:

Zentrum für Umfragen, Methoden und Analysen
ZUMA-Publikationen
Postfach 12 21 55
68072 Mannheim

* * *

Chr. Michael Dunke: Regionale Unterschiede von Medienverbreitung und Mediennutzung. Ausgewählte Ergebnisse der MEDIA-Analyse '94 im Regierungsbezirk Stuttgart. ZUMA-Arbeitsbericht 95/07.

„Region als Bedarfskategorie“? Diese Frage hat nichts an Aktualität eingebüßt. Seit Ende der siebziger, verstärkt in den frühen achtziger Jahren stehen öffentlich-rechtlicher und privater Rundfunk in der Diskussion. Die erst jüngst aufgekommene Diskussion um die ARD als föderal strukturierte öffentlich-rechtliche Einrichtung zeigt die medienpolitische Brisanz des Themas 'lokale/regionale Kommunikationsräume'. Darüber hinaus ist die Funktion lokalen/ regionalen Rundfunks als „Vermittler der Nahwelt“ aufgrund seiner schwierigen Erfassung immer kontroversen Auffassungen und Diskussionen ausgesetzt. Am Beispiel des Senders Radio T.O.N. Gold ist ersichtlich, daß lokaler/ regionaler Hörfunk lokale Identität vermitteln kann und die Unterschiedlichkeit von Lebensräumen widerspiegelt. Trotz eingeschränkter Nachweisbarkeit in seiner Funktion als lokaler Integrationsfaktor kann lokaler Hörfunk bei mangelnder Alternative (z.B. der Tagespresse) als Vermittler von örtlichen Lebenswelten eine tragende Rolle spielen.

* * *

Peter R. Schrott/Michael F. Meffert: How to Test 'Real' Campaign Effects: Linking Survey Data to Content Analytical Data. ZUMA-Arbeitsbericht 96/01.

In electoral campaign research we usually find two different approaches depending on what discipline the researcher is coming from. Political scientists usually study campaign effects by looking at variables such as candidate evaluations or issues. If these attitudes change over the period of a campaign by holding other intervening factors constant the effects are attributed to the actual campaign messages, coverage, advertisements etc. Communication researchers, on the other hand, are often looking at the actual coverage of campaigns and try to interpret the election outcome in view of the campaign battle as

it was fought in the media. Both approaches suffer from the deficit of not looking at both, the content of the campaign coverage and the survey answers of the respondents. In this paper we try to bring both approaches together. First, a detailed content analysis of the German election campaign of 1990 will provide the actual data of what was covered, and how was it covered in the campaign. These data are then linked to a survey of that time period. The data come from the German part of an international project (which involves researchers from the US, Germany, Great Britain, Japan, and Spain), and provides us with unique data on media usage, political attitudes, and the probably most extensive content analysis up-to-date. The important point, however, is that through the linkage of content analytical and survey data a direct effect of campaign coverage can be measured and analyzed. This German data set thus provides us with methodological opportunities which can not be followed up with most data sets on electoral campaigns. Substantially, the paper is also an attempt to show the potential effect of actual media coverage on the voters' attitudes. Special attention will be given on issues (such as the unification issue), and on the candidate evaluations of the chancellor candidates. Preliminary results suggest that Chancellor Kohl could gain more support through most of the media (in this case news on private as well as public television channels) than his challenger Lafontaine.

* * *

Michael Häder: Zur Evaluation der Delphi-Technik. Eine Ergebnisübersicht. ZUMA-Arbeitsbericht 96/02.

Die Delphi-Methode verspricht ihrem Anwender, über Sachverhalte, zu denen ihm zunächst nur unsicheres und unvollständiges Wissen vorliegt, zuverlässigere Informationen zu liefern. Einer der Hauptansatzpunkte von Kritik an der Nutzung von Delphi-Befragungen besteht gegenwärtig in der noch mangelhaften bzw. unvollständigen Beurteilung der Qualität der mit diesem Verfahren gewonnenen Ergebnisse. Im vorliegenden Bericht wird zunächst auf einige typische Ansätze bei den bisherigen Evaluationsbemühungen hingewiesen. Danach wird die Methodik solcher Ansätze kritisch betrachtet. Schließlich werden die Ergebnisse der Evaluation von zwei konkreten Delphi-Ansätzen vorgestellt. Abschließend wird dann nochmals die Frage aufgegriffen, wie eine geeignete Methodik zur Beurteilung von Delphi aussehen kann. Das Ergebnis erbrachte unterschiedliche, mitunter widersprüchliche Aussagen. Kein Kriterium und damit kein einzelnes Ergebnis konnte die Hauptaussage zum Funktionieren von Delphi liefern. Alle Kriterien enthalten indes spezifische Anhaltspunkte zur Evaluation von Delphi-Befragungen.