Zu Stichprobenfehlerberechnungen im Rahmen des ADM-Stichprobenplans

Kirschner, Hans-Peter

Zu Stichprobenfehlerberechnungen im Rahmen des ADM-Stichprobenplans

1. Einleitung

Es gehört gemeinhin nicht zu den Hauptsorgen eines empirisch arbeitenden Sozialwissenschaftlers, der Analysen auf der Basis einer Zufallsstichprobe von Personen durchführt, zu erfahren, mit welchen Stichprobenfehlern eigentlich die Populationsparameter, die er aus eben dieser Stichprobe schätzt, behaftet sind. Werden komplexere Analyseverfahren gerechnet, steht in aller Regel die Lösung nicht-trivialer Modell-Anpassungsprobleme im Vordergrund, und dafür werden die Daten so genommen wie sie sind, also, ohne ein eventuell disproportionierendes Stichprobendesign zu berücksichtigen oder mögliche Stichprobenfehler in die Modellansätze einzubringen.

Aus diesem Ist-Zustand einen Vorwurf abzuleiten, wäre allerdings sehr ungerechtfertigt. Zum einen gehen fast sämtliche statistische Theorien für multivariate Analyseverfahren von dem Ideal aus, daß die verwendeten Daten einem wohlausbalancierten Erhebungsprozeß entstammen (simple random sampling); folglich besteht auch kein Grund, in der zugehörigen Software explizit die Einbeziehung von möglichen Stichprobenfehlern zu unterstützen. Selbst wenn der Forscher diese also berücksichtigen wollte, hätte er mit Schwierigkeiten zu rechnen. Zum anderen gehört es nicht zum Standardangebot auf seiten derjenigen Institutionen, die mit der Realisierung von Zufallsstichproben befaßt sind, selbst für so grundlegende Parameter wie Mittelwerte wichtiger Variablen auf den Stichprobenplan bezogene Stichprobenfehler anzugeben. Der Forscher erhält also in der Regel noch nicht einmal im Vorfeld von Analysen einen Überblick über die Größenordnungen der Stichprobenvarianzen z.B. von Anteilswerten.

Insgesamt kann also schwerlich ein Vorwurf erhoben werden, wenn es in der Forschungspraxis der Sozialwissenschaften wenig verbreitet ist, den Datenerhebunsprozeß selber in seinen möglichen Einflüssen auf Analyseergebnisse zu berücksichtigen.

Allerdings wäre es auch unangemessen, zu laut seitens der Sozialwissenschaften darüber zu klagen, daß es so sehr unpopulär ist - im übrigen national wie international - Stichprobenfehler auszuweisen für Daten, die mit Hilfe komplexer Stichprobenpläne gewonnen wurden. Dies durchzuführen kann nämlich sowohl von der Stichprobentheorie her als auch in bezug auf die praktische Berechnung überaus schwierig sein.

Ein Anliegen dieser Arbeit ist es daher zu verdeutlichen, von welcher Art die Schwierigkeiten sind, die beim Versuch der Bestimmung von Stichprobenfehlern auftreten; zugleich soll aber demonstriert werden - und dies ist das vorrangige Ziel - daß es durchaus möglich ist, mit einfachen Mitteln in erster Näherung sich einen Eindruck von der Größenordnung der potentiellen Variabilität von Schätzungen von Mittelwerten zu verschaffen.

Alle diesbezüglichen Überlegungen sind zugeschnitten auf das in der Bundesrepublik am meisten angewendete komplexe Stichprobendesign, den sogenannten ADM-Stichprobenplan (ADM = Arbeitskreis Deutscher Marktforschungsinstitute). Es wird dabei durchgehend unterstellt, daß dem Leser dieses Ziehungsverfahren in seinen Grundzügen bekannt ist. Beispielrechnungen werden auf der Basis der Daten des ALLBUS 1982 durchgeführt.[1)]

Besonderes Gewicht wird in der Darstellung darauf gelegt, exemplarisch deutlich zu machen, wie für eine 'typische' Zufallsstichprobe, die mit Hilfe des ADM-Stichprobenplans gewonnen wurde - nämlich des ALLBUS 1982 - mit den Hilfsmitteln des sehr verbreiteten Programmpakets SPSS (Version 9) Stichprobenfehler berechnet werden können. Dies erscheint umso wichtiger, als so viele methodisch interessierte Forscher selber in die Lage versetzt werden, für Variablen, die in ihrem Forschungskontext von besonderer Bedeutung sind, die Größenordnung der Stichprobenfehler von ungewichtet oder gewichtet berechneten Mittelwerten zu ermitteln. In einem Anhang ist daher auch ein entsprechendes SPSS-Setup angegeben.

Abschließend sei darauf hingewiesen, daß im Programmpaket OSIRIS IV Spezial-Software zur Berechnung von Stichprobenvarianzen enthalten ist (vgl. VINTER, 1980). Leider ist im universitären Bereich in der Bundesrepublik dieses im übrigen wenig portable Programmpaket in seiner neuesten Version nicht verfügbar, so daß sich eine Diskussion seiner Möglichkeiten an dieser Stelle erübrigt.

2. Mittlere quadratische Abweichung und Design-Effekt

Hat man mit Hilfe eines Zufallsmechanismus eine Stichprobe ζ aus der interessierenden Grundgesamtheit gezogen, wird man ihre Qualität in bezug auf die Schätzung des Mittelwerts $\bar{Y}$ einer Variablen Y daran messen, wie Schätzung $\hat{\bar{Y}}(\zeta)$ und wahrer Wert $\bar{Y}$ im Schnitt, d.h. im Mittel über alle möglichen Stichproben ζ, voneinander abweichen. Ist die Wahrscheinlichkeit dafür, daß man gerade die Stichprobe ζ zieht $P(\zeta)$, wird man also als Maßzahl für die Qualität der Stichprobe, oder genauer für die Qualität des Stichprobenplans, die mittlere quadratische Abweichung von $\hat{\bar{Y}}$ heranziehen, d.h.

$$\text{(1)} \qquad MSE(\hat{\bar{Y}}) = \sum_{\zeta} P(\zeta)\, (\hat{\bar{Y}}(\zeta) - \bar{Y})^2,$$

wobei sich die Summation über alle möglichen Stichproben ζ erstreckt (MSE = mean squared error). Man sieht unmittelbar, daß $MSE(\hat{\bar{Y}})$ die mit den Wahrscheinlichkeiten für die einzelnen Stichproben gewichtete Summe der quadrierten Abweichungen des Schätzwerts vom wahren Wert ist. $MSE(\hat{\bar{Y}})$ ist offensichtlich ein intuitiv sehr naheliegendes Maß.

Die Formel (1) ist für sich genommen selbstverständlich lediglich eine zunächst inhaltsleere formale Beschreibung einer Idee, die Qualität eines Stichprobenplans in bezug auf die Schätzung des Mittelwerts einer Variable zu messen. Von großer praktischer Bedeutung ist dieser Ansatz jedoch, falls $MSE(\hat{\bar{Y}})$ aus dem Datenmaterial einer Stichprobe ζ geschätzt werden kann. Es leuchtet unmittelbar ein, daß dazu genauere Kenntnisse der Werte $P(\zeta)$ unabdingbar sind. Als Beispiel hierzu sei für einen sehr wichtigen Spezialfall eines Stichprobendesigns die Schätzbarkeit von $MSE(\hat{\bar{Y}})$ demonstriert.

Angenommen, die Stichprobe ζ ist durch n-maliges Ziehen ohne Zurücklegen zustande gekommen. Die Größe der Grundgesamtheit sei N. Dann ist bekanntlich $P(\zeta)$ stets $\binom{N}{n}^{-1}$, also konstant.[2] Da unter diesem Stichprobenplan das ungewichtete Mittel $\hat{\bar{Y}}(\zeta) = \sum_{\zeta} y_i/n$ unverzerrt ist, wird $MSE(\hat{\bar{Y}})$ zur Varianz von $\hat{\bar{Y}}$ und nach einer leichten Umformung erhält man

$$\text{(2)} \qquad Var(\hat{\bar{Y}}) = (1 - \frac{n}{N}) \frac{1}{n} S_Y^2 ,$$

wobei $S_Y^2 = \sum_{i=1}^{N} (y_i-\bar{Y})^2/N-1$ (vgl. z.B. DES RAJ, 1968:35).

Zu schätzen bleibt in (2) S_Y^2, was in naheliegender Weise durch die Stichprobenvarianz $s^2 = \sum_{\zeta} (y_i-\bar{y})^2/n-1$ möglich ist - hier und im folgenden bedeutet $\sum_{\zeta}$ stets Summation über die in der Stichprobe enthaltenen Einheiten.

Formel (2) ist insofern von großer Bedeutung, als man die mittlere quadratische Abweichung, die man unter komplexeren Stichprobenplänen u.U. schätzen kann, gerne vergleicht mit derjenigen quadratischen Abweichung, die man unter dem Stichprobenplan "einfaches Ziehen ohne Zurücklegen" erhalten hätte. Formal bildet man den Quotienten der Schätzung von $MSE(\hat{\bar{Y}})$ bzgl. des komplexen Stichprobenplans geteilt durch den (2) entsprechenden Schätzwert und nennt diese Zahl den Design-Effekt (vgl. SUDMAN, 1976:184f.). Wir werden später (Abschnitt 6) im Zusammenhang mit Schätzungen aus dem ALLBUS 1982 auf diesen Begriff zurückkommen.

Der Design-Effekt ist eine unter praktischen Gesichtspunkten bedeutende Maßzahl, deren Wert noch dadurch besonders betont wird, daß $MSE(\hat{\bar{Y}})$ aus (2) für sich alleine genommen im Bereich der Sozialwissenschaften von beschränkter Aussagekraft ist. Ein Beispiel möge dies verdeutlichen:

Die Variable Y sei durch eine 7-Punkte-Skala beschrieben mit den Werten 1 bis 7.[3] Die Größe der Grundgesamtheit sei rund N=44 000 000 (z.B. mindestens 18jährige deutsche Bevölkerung in Privathaushalten in der Bundesrepublik) und die Stichprobengröße sei n=3000. Man kann dazu zeigen, daß völlig unabhängig von der Verteilung der Variablen Y in der Grundgesamtheit die Quadratwurzel aus der Varianz von $\hat{\bar{Y}}$ niemals größer werden kann als $3/\sqrt{2999} = 0.055$.[4] D.h. unabhängig vom Wert des zu schätzenden Mittelwerts $\bar{Y}$, der irgendwo zwischen 1 und 7 liegt, ist bei diesen Stichprobengrößen der "Standardfehler", der vermöge des Zufallsexperiments "Stichprobenziehen" erzeugt wird, selbst im schlechtesten Fall von einer so geringen Größenordnung, daß sich der Wert dieser Zahl eher im Vergleich zu Stichprobenfehlern aus komplexen Stichprobenplänen begründet - diese sind in aller Regel größer - als in den Möglichkeiten einer direkten Interpretation.

Der Vollständigkeit halber sei angemerkt, daß für eine dichotome Variable Y in 0-1-Codierung der "Standardfehler" unter dem Stichprobenplan "einfaches

Ziehen ohne Zurücklegen" äußerstenfalls $0.5/\sqrt{2999} = 0.0091$ beträgt - also für eine Fallzahl von 2999 -, sich demnach ungünstigstenfalls im Bereich von 1% bewegt, gänzlich unabhängig davon, wie die Verteilung von Y in der Population beschaffen ist.

Der Stichprobenansatz "einfaches Ziehen ohne Zurücklegen" ist selbstverständlich praktisch nicht umsetzbar. Die vielen Millionen von bei üblichen Querschnitten potentiell befragbaren Personen befinden sich eben nicht als Kugeln in einer Urne, aus der man dann ohne Zurücklegen zufällig 3000 Kugeln entnimmt. Stichproben wie die für den ALLBUS 1982 entstehen vielmehr durch das Zusammenwirken mehrerer komplexer Zufallsmechanismen, die keineswegs alle so beschrieben werden können, daß stets die Wahrscheinlichkeiten $P(\zeta)$ aus Formel (1) befriedigend beschreibbar sind. M.a.W. Schätzungen von $MSE(\hat{\bar{Y}})$ sind dann - wenn überhaupt - lediglich approximativ möglich. Wie diese Problematik für das ADM-Design angegangen werden kann, ist Gegenstand des folgenden Abschnitts.

3. Die Randomisierungsprozesse beim ADM-Design

Fast alle größeren Querschnittsbefragungen in der Bundesrepublik und West Berlin werden zur Zeit mit Hilfe des dreistufigen ADM-Stichprobenplans durchgeführt. Dessen grundlegende Eigenschaften finden sich z.B. beschrieben in KIRSCHNER (1984:118ff.) und sollen hier nicht im einzelnen erläutert werden. Leider gehört es nicht zum Standard der solche Befragungen durchführenden und mehrheitlich privatwirtschaftlich verfaßten Institute, Aussagen über die Stichprobenfehler von Mittelwertschätzungen zu wichtigen Variablen zu machen, oder aber gar Design-Effekte anzugeben. In aller Regel ist der 'Abnehmer' solcher Umfragen darauf angewiesen, selber zu versuchen, Zahlen dieser Art zu berechnen. Da die gängige Literatur dazu praktisch keine Hilfestellung gibt, sei in diesem Abschnitt zunächst erläutert, welche Zufallsprozesse bei der Exekutierung des ADM-Designs überhaupt anfallen. Nach zwei weiteren Abschnitten, die eher theoretischen Grundlagen gewidmet sind, wird im Abschnitt 6 dann beschrieben, wie mit Hilfe von SPSS 9 Schätzungen von $MSE(\hat{\bar{Y}})$ und damit auch vom Design-Effekt konkret berechnet werden können.

Bekanntlich werden in der ersten Stufe des ADM-Stichprobenplans in systematischer Weise Stimmbezirke bzw. synthetische Stimmbezirke gezogen mit einer

Ziehungswahrscheinlichkeit, die ihrem Bedeutungsgewicht (= geschätzte Anzahl der Privathaushalte) proportional ist. Die Zufälligkeit dieses Schrittes besteht ausschließlich darin, daß eine einzige Zufallszahl (maschinell) erzeugt wird, die den Startpunkt für das systematische Ziehen festlegt. Da alle Algorithmen, die diesen Vorgang steuern, bekannt sind, kann man diesen Randomisierungsprozeß als vollständig beschreibbar betrachten, und in der Tat lassen sich seine wichtigen Eigenschaften stichprobentheoretisch gut darstellen.[5)]

Komplexer stellt sich die Situation auf den folgenden Stufen des Designs dar. Die letztendliche Auswahl einer Befragungsperson in einer Primäreinheit läßt sich nämlich auffassen als das Resultat von vier unterschiedlichen Zufallsexperimenten.

Das erste besteht darin, daß die random route des Interviewers (vgl. KIRSCHNER, 1984:128ff., dort random route = standard random) im Zuge der Feldsteuerung seitens des durchführenden Instituts festgelegt wird. Man kann diesen Schritt bei der Stichprobenziehung im Prinzip als zufällige Entnahme eines Klumpens von Klingelschild-Adressen aus dem Adressen-Pool des Stimmbezirks beschreiben. Grundsätzlich gilt hier wie beim ersten Randomisierungsprozeß zur Ziehung der Stichprobe von Primäreinheiten (Primäreinheit = Stimmbezirk), daß das Zufallsexperiment vollständig dargestellt werden kann, insbesondere im Hinblick auf Auswahlwahrscheinlichkeiten. Gerade dies gilt nicht für das sich überlagernde Zufallsexperiment, dessen Ausprägungen in der Umsetzung der Begehungsvorschriften durch die Interviewer gesehen werden können.

Beide zufälligen Prozesse zusammen resultieren in einer Stichprobe zufälliger Größe von Haushalten aus der jeweiligen Primäreinheit. Ob dabei alle Haushalte enes Stimmbezirks gleiche Auswahlwahrscheinlichkeit haben, muß grundsätzlich für jede Studie einzeln diskutiert werden. Allein schon die Tatsache, daß bei verschiedenen kommerziellen Markt- und Meinungsforschungsinstituten die Interviewerstäbe unterschiedlich strukturiert sind, kann einen Einfluß darauf haben, ob z.B. dem äußeren Anschein nach 'schwierigere' Haushalte mit verminderter Auswahlwahrscheinlichkeit in die Stichprobe gelangen.

Die beiden Randomisierungsprozesse der Vorgabe der random route sowie deren Implementierung durch den Interviewer, wirken vor einer Interaktion des Interviewers mit Personen des zu kontaktierenden Haushalts. Gelingt es dem Interviewer, in den Haushalt zu gelangen, also eine Interaktion mit irgendwelchen Haushaltsmitgliedern zuwege zu bringen, sind wiederum zwei zufällige Prozesse für die erfolgreiche Auswahl und Befragung einer Person des Haushalts verantwortlich. Der erste, vollständig theoretisch als einfache und uneingeschränkte Zufallsauswahl beschreibbar, besteht in der Anwendung der dem Interviewer im vorhinein von seinem Institut übergebenen Zufallszahlenreihe zur Ermittlung der Zielperson im Haushalt (vgl. KIRSCHNER, 1984:132f.). Diese Prozeduren, die allen Haushaltsmitgliedern, die der Grundgesamtheit angehören, gleiche Auswahlwahrscheinlichkeiten garantieren, werden Schwedenschlüssel - vorwiegend in der Bundesrepublik - oder Kish table bzw. Kish grid genannt (vgl. KISH, 1949). Der weitaus weniger beschreibbare zufällige Prozeß, der mit darüber bestimmt, ob ein Interview erfolgreich abgeschlossen werden kann, liegt in dem überaus komplexen Wechselspiel zwischen dem Interviewer und den Mitgliedern des kontaktierten Haushalts bzw. der Zielperson selbst. Ein Versuch, den Effekt auf die Auswahlwahrscheinlichkeiten für die einzelnen Personen in der Stichprobe zu ermitteln, ist bisher für das ADM-Design nicht unternommen worden. Dies ist allerdings kaum überraschend, da eine Untersuchung des Phänomens bedeuten würde, das sehr vielschichtige non response Problem bei persönlichen Befragungen im Rahmen des ADM-Stichprobenplans in den Griff zu bekommen.

Die vier letztgenannten Randomisierungsprozesse sind natürlich (stochastisch) unabhängig von der Zufallsauswahl der Stichprobe von Stimmbezirken. Sie sind vielmehr bezogen auf den einzelnen Stimmbezirk, auf die dortigen soziodemographischen Verhältnisse, auf die für diesen Stimmbezirk verfügbaren Interviewer usw. M.a.W. diese vier Randomisierungsprozesse beschreiben für jede Primäreinheit - unabhängig davon, ob diese in die Stichprobe gelangt oder nicht - im Resultat ein Zufallsexperiment, dessen Ausgang in einer Stichprobe von Personen aus eben diesem Stimmbezirk besteht. Dabei kann diese Stichprobe durchaus leer sein - beim ALLBUS 1982 waren in 23 der Primäreinheiten keine Interviews zustandegekommen. Der Ziehungsalgorithmus für die Stichprobe der Primäreinheiten bestimmt also lediglich, in welchen dieser Primäreinheiten das 'zugehörige' Zufallsexperiment zur Bestimmung der Befragungspersonen exekutiert werden soll.

Ganz allgemein läßt sich das 'Zufallsexperiment' auf Stimmbezirksebene so charakterisieren, daß ein 'objektivierbarer' Ziehungsalgorithmus überlagert wird von schwer beschreibbaren und als zufällig anzusehenden Implementierungsprozessen, die vorwiegend im Verhalten der Interviewer ihren Ursprung haben. Unter dem Gesichtspunkt, daß man die Qualität von Parameterschätzungen mit Hilfe der Schätzung von mittleren quadratischen Abweichungen ermitteln will, bedeutet dies, daß die potentielle Variabilität der Schätzungen nicht nur darin begründet liegt, daß allein schon über den Stichprobenplan Variation erzeugt wird, sondern daß auch das Feldgeschehen im allgemeinen mit zu den Wahrscheinlichkeiten beiträgt, mit denen die verschiedenen möglichen Stichproben auftreten.

Es sei darauf hingewiesen, daß diese Argumentation bewußt unterdrückt, daß zusätzliche Variation in bezug auf die erfragten Daten über das Befragtenverhalten selber erzeugt wird. Bei bestimmten Fragen wird man z.B. unterstellen können, daß einige der Zielpersonen schlicht - bewußt oder unbewußt - die Unwahrheit sagen. Da es jedoch den Rahmen dieses Beitrags sprengen würde, auch diese Problemstellung in die Diskussion einzubeziehen, sei nochmals betont, daß sich die Überlegungen der folgenden Abschnitte auf ein 'Modell' beschränken, in dem die potentielle Variabilität einer Schätzung ausschließlich über die in diesem Abschnitt beschriebenen fünf Randomisierungsprozesse begründet wird.

4. Erwartungstreue von gewichtetem und ungewichtetem Mittel

In der Forschungspraxis der Sozialwissenschaften wird bei Auswertungen von Daten aus großen Querschnitten in aller Regel zur Schätzung von Anteilen oder, allgemeiner, von Mittelwerten das ungewichtete arithmetische Mittel herangezogen. Insbesondere wenn komplexe statistische Analysen gerechnet werden, kommt eine gewichtete Berechnung von Verteilungsparametern selten in Betracht, nicht zuletzt deshalb, weil eine Gewichtung keineswegs immer zu einer Verbesserung der Schätzwerte führt (vgl. KIRSCHNER, 1984:166ff.). Selbstverstandlich entlastet dieses Argument nicht davon, darüber nachzudenken, was das einfache arithmetische Mittel zum einen denn eigentlich schätzt, und zum anderen, wenn es das schätzt, was es schätzen soll, wie genau es dann schätzt. Der Vollständigkeit halber sollen im folgenden diesbezügliche Überlegungen auch angestellt werden für mit der Anzahl der Personen der Grundgesamtheit im Haushalt (= reduzierte Haushaltsgröße) gewichtete Schätzungen.

Sowohl arithmetisches Mittel als auch gewichtetes Mittel sind sogenannte Verhältnisschätzer, d.h. Zähler <u>und</u> Nenner sind zufällige Größen. Etwas verallgemeinernd kann man dies wie folgt darstellen:

Faßt man die vier im vorigen Abschnitt letztgenannten Randomisierungsprozesse in der Weise zusammen, daß man nur am gemeinsamen Resultat, nämlich einer Zufallsstichprobe S(i) von Personen aus dem i-ten Stimmbezirk interessiert ist, läßt sich der Verhältnisschätzer, der sich auf eine Stichprobe von <u>k Primäreinheiten</u> stützt, für zwei beliebige Variablen X und Y schreiben als

$$(3) \qquad \hat{T}(X,Y) = \sum_{i=1}^{k} \sum_{S(i)} y_j \Big/ \sum_{i=1}^{k} \sum_{S(i)} x_j = \sum_{i=1}^{k} \hat{T}_i(S(i),Y) \Big/ \sum_{i=1}^{k} \hat{V}_i(S(i),X),$$

mit $\hat{T}_i(S(i),Y) = \sum_{S(i)} y_j$, $\hat{V}_i(S(i),X) = \sum_{S(i)} x_j$, wobei

die Summationen $\sum_{S(i)}$ sich auf die Variablenwerte y_j bzw. x_j derjenigen Personen beziehen, die in der Stichprobe S(i) im i-ten Stimmbezirk der Stichprobe von Stimmbezirken der Größe k enthalten sind - beim ALLBUS 1982 war k=630. Man erhält durch (3) sofort den ungewichteten bzw. mit der reduzierten Haushaltsgröße gewichteten Verhältnisschätzer, wenn man in folgender Weise spezialisiert:

$$x_j = \begin{cases} 1, & \text{Befragter hat für die Variable Y eine Antwort gegeben} \\ 0, & \text{sonst} \end{cases}$$

$y_j =$ Wert, den der Befragte angegeben hat,

bzw.

$$x_j = \begin{cases} h_j, & \text{reduzierte Haushaltsgröße bzgl. des Befragten, der eine Antwort auf Y gab} \\ 0, & \text{sonst} \end{cases}$$

$y_j =$ [Wert, den der Befragte angegeben hat] * [reduzierte HH'größe].

Im ersten Fall ergibt sich also $\hat{T}=\sum_{\zeta} y_j/n$, im zweiten $\hat{T}=\sum_{\zeta} y_j h_j / \sum_{\zeta} h_j$, wobei die Summation sich über die gesamte realisierte (bzgl. Y) Stichprobe ζ von Personen erstreckt und n die Anzahl der Personen in ζ ist.[6)]

Verwendet man Schätzfunktionen vom Typ $\hat{T}(X,Y)$, so sollen diese hinreichend genau den Populationsmittelwert $\bar{Y}$ schätzen und zwar für möglichst viele Variablen Y. Für sich genommen ist $\hat{T}(X,Y)$ selbstverständlich zunächst nur eine Schätzung für den eigenen Erwartungswert $E\hat{T}(X,Y)$, also diejenige Zahl, die man erhielte, wenn man $\sum_{\zeta} P(\zeta)\hat{T}(X,Y,\zeta)=E\hat{T}(X,Y,\zeta)$ genau berechnen könnte. Dies ist natürlich unmöglich; man kann diese Größe lediglich approximativ und unter zusätzlichen Voraussetzungen bestimmen. Im Prinzip leistet $\hat{T}(X,Y)$ jedoch genau das, was es leisten soll, wenn $\bar{Y} \simeq E\hat{T}(X,Y)$ und wenn $MSE(\hat{T}(X,Y))$ für viele der Variablen klein ist.

Es sei zunächst für $\hat{T}=\sum_{\zeta} y_j/n$ angegeben, welche Bedingungen die 'Erwartungstreue' dieses ungewichteten Schätzers garantieren.

Zentral für 'gutes' Verhalten von $\hat{T}$ ist naheliegenderweise 'gutes' Verhalten der Zufallsexperimente auf Primäreinheitenebene. Die Personenstichprobe in einem Stimmbezirk sollte also trotz der kleinen Fallzahl eine erwartungstreue Schätzung des Mittelwerts möglichst vieler Variablen Y zulassen. Ausgedrückt mit den bereitgestellten Formalismen bedeutet dies, daß für eine beliebige Primäreinheit mit assoziiertem Zufallsexperiment S(i) = Zufallsstichprobe von Personen, gelten soll

$$(4) \qquad E\,\{[g_i/\hat{V}_i(S(i))] * \sum_{S(i)} h_j y_j\} = M_i \bar{Y}_i$$

wobei wieder die zufällige Größe $\hat{V}_i(S(i))$ die Anzahl der in der i-ten Primäreinheit realisierten Interviews beschreibt; g_i ist das Bedeutungsgewicht = geschätzte Anzahl der Privathaushalte, y_j und h_j sind die Ausprägungen der Variablen Y bzw. der Variablen "reduzierte Haushaltsgröße" in S(i); M_i ist die Anzahl der Personen der Grundgesamtheit in der i-ten Primäreinheit und $\bar{Y}_i$ stellt den dortigen Mittelwert von Y dar. Falls die zufällige Größe $\hat{V}_i$ (zufällig) Null ist, wird der Ausdruck innerhalb der geschweiften Klammern in (4) insgesamt gleich Null gesetzt.

Inhaltlich bedeutet (4), daß man $M_i\bar{Y}_i$ auf der Basis einer einzelnen Stichprobe S(i) so schätzt, daß die Summe der mit der reduzierten Haushaltsgröße inflationierten Werte nochmals über die 'sampling fraction' auf Haushaltsebene inflationiert wird, um in die 'richtige' Größenordnung - nämlich die des Primäreinheitentotals von Y - zu kommen (vgl. das Substitutionsprinzip in KIRSCHNER, 1980). Die mit der Wahrscheinlichkeit des Auftretens der einzelnen Stichproben S(i) gewichtete Summe der einzelnen Schätzwerte über alle möglichen Stichproben innerhalb der Primäreinheit sollte dann den Wert $M_i\bar{Y}_i$ haben. Etwas weniger streng formuliert wird also angenommen, daß, wenn man nur hinreichend oft im Stimmbezirk das Zufallsexperiment mit dem Ausgang S(i) durchführt, man für viele Variablen Y im Schnitt $M_i\bar{Y}_i$ 'ganz gut' schätzen kann. Zu beachten ist hierbei jedoch, daß wegen der geringen Fallzahl = $\hat{V}_i(S_i)$ die Varianz von $[g_i/\hat{V}_i] * \sum_{S(i)} h_j y_j$ recht groß sein wird - der Mittelwert der $\hat{V}_i$, genommen über alle 607 Primäreinheiten mit wenigstens einem Interview, liegt für den ALLBUS 1982 bei 5.

Die Annahme (4) ist in bezug auf die vier, die Stichprobe S(i) generierenden Randomisierungseffekte so einzuordnen, daß bei vernachlässigbaren Implementierungsprozessen der Ausdruck in den geschweiften Klammern gerade so gewählt ist, daß (4) 'automatisch' aus stichprobentheoretischen Gründen gilt. Beeinflussen jedoch diese Implementationsprozesse die Erhebbarkeit einer Variablen Y, so kann (4) schon deshalb schwerlich gelten, weil sich diese Randomisierungseffekte nicht im Formelausdruck niederschlagen.

Würde man z.B. die Anzahl der 'eligibles' M_i schätzen wollen - Y wäre dann identisch Eins - wäre $\sum_{S(i)} h_j y_j$ tendenziell zu groß, da Einpersonenhaushalte bekanntlich im Schnitt schwieriger zu kontaktieren sind. Die Konsequenz wäre eine Überschätzung von M_i.

Würde Y - dichotom in 0-1-Codierung - die Zugehörigkeit zur Kohorte der mindestes 70jährigen Frauen beschreiben, wäre wegen der allgemein bekannten Zugangsproblematik die Summe $\sum_{S(i)} h_j y_j$ tendenziell zu klein, was notwendig zu einer Unterschätzung führt.

Allgemein ist also die Voraussetzung (4) für solche Variablen Y realistisch, deren Assoziation zu Variablen, die für die Implementierungsprozesse konstituierend sind, vernachlässigbar klein ist.

Von ähnlich inhaltlicher Bedeutung wie (4) sind die beiden folgenden Annahmen, wiederum bezogen auf eine beliebig herausgegriffene Primäreinheit:

(5.1) Die Voraussetzung (4) gelte auch für die Variable Y/H, mit H = reduzierte Haushaltsgröße;

(5.2) $\mathrm{Cov}(\underset{\sim}{\bar{y}}_\nu, \underset{\sim}{h}_\nu) = 0 = \sum_\nu (\bar{y}_\nu - \frac{1}{M_i}\sum_\nu \bar{y}_\nu)(h_\nu - \frac{1}{M_i}\sum_\nu h_\nu)$
wobei $\underset{\sim}{\bar{y}}_\nu$ den Vektor der Mittelwerte von Y in den einzelnen Haushalten der Primäreinheit bezeichnet und $\underset{\sim}{h}_\nu$ der entsprechende Vektor der reduzierten Haushaltsgröße ist.

Sowohl (5.1) als auch (5.2) sind in Ergänzung zu (4) zu verstehen und unterstreichen nochmals, daß 'Y mit H möglichst wenig zu tun haben sollte'.

Von mehr technischer Natur sind

(5.3) $E\hat{V}_i = m$ für jede Primäreinheit ($E\hat{V}_i$ soll also konstant sein);
$\mathrm{Var}\hat{V}_i$ sei klein für jede der Primäreinheiten.

Die Annahme (5.3) spiegelt das Bemühen wieder, bei Umfragen mit Hilfe des ADM-Stichprobenplans die Anzahl der realisierten Interviews pro Primäreinheit möglichst konstant zu halten; (5.3) ist somit recht realistisch.

Insgesamt folgt schließlich für eine Variable Y aus den Annahmen (4), (5.1) - (5.3) und einer zusätzlichen technischen Voraussetzung über das systematische Ziehen der Primäreinheiten, die hier nicht näher erläutert werden kann, daß in erster Näherung gilt

(5.4) $$E\hat{T}(Y) = E\{\frac{1}{n}\sum_\zeta y_j\} = \bar{Y} + \text{Restterm.}$$

Der Beweis dieser Aussage soll hier nicht geführt werden. Es sei lediglich angemerkt, daß dabei neben den bisher genannten Voraussetzungen eine große Rolle spielt, daß der Quotient M_i/g_i, also die durchschnittliche reduzierte

Haushaltsgröße in der Primäreinheit, eine sehr geringe Variation aufweist, d.h. sehr eng um den Wert 2 streut.

Eine Abschätzung des Restterms soll ebenfalls hier nicht ausgeführt werden. Man kann jedoch zeigen, daß er unter vernünftigen Zusatzannahmen relativ zu $\bar{Y}$ sehr klein wird.

Die ungewichtete Schätzung $\sum_{\zeta} y_j/n$ leistet also unter nicht allzu starken Annahmen das, was sie leisten soll, nämlich $\bar{Y}$ erwartungstreu zu schätzen.

Für die Schätzfunktion $\sum_{\zeta} y_j h_j / \sum_{\zeta} h_j$ läßt sich mit den eingeführten Bezeichnungen nun ebenfalls formulieren, unter welchen Bedingungen sie erwartungstreu ist:

Für eine Variable Y, die der Voraussetzung (4) genügt, gilt

$$(5.5) \qquad E\{\sum_{\zeta} y_j h_j \,/\, \sum_{\zeta} h_j\} = \bar{Y} + \text{Restterm},$$

wenn nur zusätzlich die Annahme (5.3) erfüllt ist, und wenn des weiteren gilt, daß für jede Primäreinheit $E[\sum_{S_i} h_j] = M_i * m/g_i$.

Wiederum können in diesem Rahmen weder der Beweis der Aussage selbst noch eine Diskussion des Restterms ausgeführt werden. Zu letzterem gilt jedoch eine analoge Aussage wie zum Restterm des Erwartungswerts der zuerst behandelten Schätzfunktion. Auf die Voraussetzungen (5.1) bzw. (5.2) kann im übrigen verzichtet werden, weil die mit der reduzierten Haushaltsgröße gewichtete Schätzfunktion ihrer Struktur nach bereits vollständig durch Voraussetzung (4) erfaßt wird, wie man per Augenschein unmittelbar sieht. Die Annahme $E[\sum_{S_i} h_j] = M_j * m/g_j$ ist ein Analogon zu der für den ungewichteten Schätzer gemachten Annahme $E\hat{V}_i = m$.

Zusammenfassend ist also festzuhalten, daß der in (3) angegebene Typ von Mittelwertschätzung für die Spezialfälle der gänzlich ungewichteten Berechnung und der mit der reduzierten Haushaltsgröße gewichteten Berechnung im Rahmen des ADM-Stichprobenplans unter bestimmten Bedingungen bis auf kleinere Verzerrungen tatsächlich den Populationsmittelwert einer Variablen schätzt. Von entscheidender Bedeutung dafür ist, daß Zähler und Nenner von $\hat{T}$ selber Summen von Schätzern auf Primäreinheitenebene sind und daß man für diese Summanden unterstellen darf, daß sie als Schätzer auf Primäreinheitenebene 'vernünftig' sind.

Die potentielle Variabilität von $\hat{T}$ oder - formal - die mittlere quadratische Abweichung von $\hat{T}$ wird offensichtlich über die Variation der einzelnen Summanden und durch deren Wechselwirkung miteinander erzeugt. Wir werden im folgenden Abschnitt sehen, daß dieser Gedanke zentral ist sowohl für die theoretische Behandlung von MSE($\hat{T}$) als auch deren Schätzung aus dem Datenmaterial.

5. Die Schätzung der mittleren quadratischen Abweichung - Theorie

Sind Schätzfunktionen der Gestalt wie in (3) gegeben, läßt sich bei gegebener Variablen Y die mittlere quadratische Abweichung in erster Näherung wie folgt schreiben:

$$E(\hat{T}-\bar{Y})^2 \doteq \frac{\sigma^2(\hat{U}) + \bar{Y}^2\sigma^2(\hat{V}) - 2\bar{Y}\sigma(\hat{U})\sigma(\hat{V})r(\hat{U},\hat{V})}{(E\hat{V})^2} \tag{6}$$

mit $\hat{U}=\sum_{i=1}^{k}\hat{T}_i(S_i,Y)$, $\sigma^2(\hat{U})=\mathrm{Var}\hat{U}, \sigma^2(\hat{V})=\mathrm{Var}\hat{V}$ und $r(\hat{U},\hat{V})$=Pearsonscher Korrelationskoeffizient bzgl. $\hat{U}$ und $\hat{V}$.

Voraussetzung für diese Schreibweise ist, daß $E\hat{U}/E\hat{V}=\bar{Y}$. Dieses wiederum folgt für ungewichtete und gewichtete Schätzer aus den entsprechenden Voraussetzungen des vorigen Abschnitts (zur Darstellung in (6) vgl. DES RAJ, 1968:125).

Nach Formel (6) reduziert sich also die Schätzung von MSE($\hat{T}$) auf die Schätzungen von $\sigma^2(\hat{U})$, $\sigma^2(\hat{V})$, $r(\hat{U},\hat{V})$, $E\hat{V}$ und $\bar{Y}$. Zentraler Term dabei ist $\sigma^2(\hat{U})/(E\hat{V})^2$, da $\bar{Y}^2 * \sigma^2(\hat{V})$ (=$\bar{Y}$ * Varianz der Größe der realisierten Stichprobe oder Varianz der Summe der reduzierten Haushaltsgrößen aus der realisierten Stichprobe) relativ zu $(E\hat{V})^2$ sehr klein ist und so in der Tat die Größe $\sigma^2(\hat{U})/(E\hat{V})^2$ dominierend wird. $r(\hat{U},\hat{V})$ ist in aller Regel ebenfalls sehr klein.

$\sigma^2(\hat{U})$ als die Varianz der Summe von - möglicherweise abhängigen - zufälligen Größen ist jedoch vermöge der folgenden Identität einer Schätzung zugänglich:

$$\text{(7)} \qquad E\{\sum_{i=1}^{k}(\hat{T}_i - \frac{1}{k}\sum_{j=1}^{k}\hat{T}_j)^2\} = \sigma^2(\hat{U})\{\frac{\sum_i \sigma^2(\hat{T}_i)}{\sigma^2(\hat{U})} - \frac{1}{k}\} + \sum_{i=1}^{k}(E\hat{T}_i - \frac{1}{k}\sum_{j=1}^{k}E\hat{T}_j)^2$$

Man beachte, daß für den Fall der Unkorreliertheit der Größen $\hat{T}_i$, also wenn der Ausdruck in geschweiften Klammern rechts vom Gleichheitszeichen gleich $(k-1)/k$ ist, und der Gleichheit aller Erwartungswerte $E\hat{T}_i$ und Varianzen $\sigma^2(\hat{T}_i)$, (7) die bekannte Tatsache beschreibt, daß $\sum_{i=1}^{k}(\hat{T}_i-\bar{\hat{T}}.)^2/k-1$ eine erwartungstreue Schätzung von $\sigma^2(\hat{T}_i)$ ist.

Dieser Spezialfall beschreibt im Prinzip schon die Argumentationsweise, die Formel (7) für die Schätzung von $MSE(\hat{T})$ nutzbar macht: Die zufälligen Größen $\hat{T}_i$ sind nicht voneinander unabhängig, weil sie über die systematische Ziehung der Primäreinheiten miteinander verbunden sind. Jedoch läßt sich für eine Variable Y, die in einem eher technischen Sinne auf dem Aggregationsniveau 'Primäreinheit' nicht mit den Ziehungswahrscheinlichkeiten kovariiert, zeigen, daß die $\hat{T}_i(S_i,Y)$ paarweise nahezu unkorreliert sind. Zusätzlich gilt zusammen mit den Voraussetzungen aus dem vorigen Abschnitt, daß auch die Mittelwertskomponente in (7) sowohl für gewichteten als ungewichteten Schätzer nahe bei Null liegt. $\sigma^2(\hat{U})$ wird also (fast erwartungstreu) unter diesen Umständen durch die Summe $\sum_{i=1}^{k}(\hat{T}_i-\bar{\hat{T}}.)^2$ geschätzt. Zu berechnen sind also pro Primäreinheit in der Stichprobe die Werte von $\hat{T}_i$ und $\hat{V}_i$, und damit sind dann die entsprechenden Schätzungen aufzubauen.

Grundsätzlich ist somit ein Algorithmus gegeben, der es erlaubt, $MSE(\hat{T})$ zu schätzen. Bei der konkreten Anwendung im Forschungsprozeß in den Sozialwissenschaften ist jedoch zusätzlich zu berücksichtigen, daß zum einen eher 'ie Größenordnung von $MSE(\hat{T})$ interessiert und weniger ein bis auf mehrere chkommastellen angegebener Wert, und zum anderen die Rechnung selber einh sein muß und vor allem nicht den extensiven Gebrauch eines Taschenners erzwingen darf. Es wird daher im folgenden Abschnitt 6 an Beispielen demonstriert werden, daß für bestimmte Variablen aus dem ALLBUS 1982 die Schätzung für $\sigma^2(\hat{U})$ in (6) in der Tat der zentrale Term ist, und so nicht die ganze Formel (6) ausgerechnet werden muß, indem man Schätzwerte einsetzt. Dies wiederum bedeutet, daß sehr einfach mit den Hilfsmitteln von SPSS 9 unmittelbar ein Eindruck von der Größenordnung von $MSE(\hat{Y})$ gewonnen werden kann.

Die Möglichkeit der Schätzung von $\sigma^2(\hat{U})/E\hat{V}$ mit Hilfe von (7) - für den Nenner setzt man n bzw. $\sum_\zeta h_j$ ein - ist, wie gezeigt wurde, an ganz bestimmte Voraussetzungen gebunden. D.h. man weiß so, für welche Konstellationen Schätzungen dieser Art mit Vorsicht zu betrachten sind. Es ist in diesem Zusammenhang von großem Interesse, daß man $\{\sum_{i=1}^{k}(T_i - \frac{1}{k}\sum_{j=1}^{k}T_j)^2\}^{1/2}/E\hat{V}$ als alleinigen Schätzer für $\sigma(\hat{T})$ erhält, wenn man ein recht grobes mathematisches Modell für den ADM-Stichprobenplan verwendet. Ein Modell also, daß weit weniger als das hier vorgestellte Einsichten darüber erlaubt, ob das, was man tatsächlich numerisch ausrechnet, dasjenige ist, was man eigentlich ausrechnen, oder genauer 'schätzen' möchte. Grob gesagt unterstellt dieses vereinfachte Modell, daß die Ziehung der Primäreinheiten mit Zurücklegen erfolgt und daß die Wahrscheinlichkeit, für den i-ten Zug die i-te Primäreinheit zu bekommen, proportional zu g_i, also deren Bedeutungsgewicht, ist. Des weiteren wird unterstellt, daß die Nenner von $\hat{T}=\sum_\zeta y_j/n$ bzw. $\hat{T}=\sum_\zeta h_j y_j/\sum_\zeta h_j$ nicht zufallsabhängig sind. Die Voraussetzungen für die Gültigkeit von (5.4) und (5.5) implizieren dann, daß z.B. für den ungewichteten Mittelwertschätzer $\sum_\zeta y_j/n$ die Größe

$$\text{(8)} \qquad \{\sum_{i=1}^{k}(\sum_{S_i} y_j - \frac{1}{k}\sum_{j=1}^{k}\sum_{S_j} y_j)^2\}^{1/2}/n$$

als eine Schätzung von $\mathrm{Var}\hat{T}$ aufgefaßt werden kann; k ist dabei die Anzahl der Stimmbezirke, in denen wenigstens ein Interview realisiert wurde.

Der Gewinn, der also mit einem dem wirklichen Geschehen näheren Modellansatz zu erzielen ist, besteht hier darin zu erkennen, daß Formel (8) lediglich eine grobe Näherung darstellt und wie man (8) verfeinern kann.

Schätzungen vom Typ (8) finden sich im übrigen bei SUDMAN (1976:174f.) und DES RAJ (1968:120f.). Dort steht allerdings die Idee der Replikation, die hier über die Unterstellung des Ziehens mit Zurücklegen eher implizit eingeführt wird, im Vordergrund.

6. Die Schätzung der mittleren quadratischen Abweichung - Praxis und Beispiele

In diesem abschließenden Abschnitt soll für drei Variablengruppen demonstriert werden, wie man mit geringem Aufwand Schätzungen für die mittlere quadratische Abweichung erhält und welche Werte diese Schätzungen haben. Des weiteren werden die zugehörigen Design-Effekte angegeben und diskutiert.

Die erste Variablengruppe besteht aus 7 Variablen, die inhaltlich mit Hilfe einer 7-Punkte-Skala beschreiben, für wie wichtig der Befragte die Lebensbereiche 'Eigene Familie und Kinder' (WFAM), 'Beruf und Arbeit' (WBER), 'Freizeit und Erholung' (WFREI), 'Freunde und Bekannte' (WFREU), 'Verwandtschaft' (WVERW), 'Religion und Kirche' (WREL) sowie 'Politik und öffentliches Leben' (WPOL) hält (vgl. Codebuch ALLBUS 1982). Diese Variablen werden bei statistischen Analysen in aller Regel so benutzt, als ob sie zumindest ordinal skaliert wären. Von vorrangigem Interesse sind also weniger die Stichprobenfehler der relativen Häufigkeiten der einzelnen Skalenpunkte, sondern der Stichprobenfehler des Mittelwerts, den diese Variablen in der Population haben.

Zur Berechnung des Design-Effekts benötigt man nach Formel (2) eine Schätzung für S_Y^2. Da keine ohne Zurücklegen gezogene Vergleichsstichprobe aus der Gesamtpopulation existiert, mit deren Daten eine solche Schätzung möglich wäre, wird hier für S_Y^2 der Mittelwert der jeweiligen Standardabweichungen für diese sieben Variablen bzgl. der Umfragen ALLBUS 1980 und ALLBUS 1982 verwendet (vgl. auch FRANKEL, 1971:171).

Über den Grad an Assoziation der Variablen WFAM usw. zur Variablen 'reduzierte Haushaltsgröße' läßt sich aus der Stichprobe z.B. berechnen, daß der Pearsonsche Korrelationskoeffizient höchstens von der Größenordnung 0.2 ist. Dies möge als ein möglicher Indikator für Assoziation stehen (vgl. die

Anmerkung nach Formel 5.2)). Vom Inhalt der Fragen her erkennt man allerdings sofort und ohne Berechnungen anzustellen, daß - wenn überhaupt - WFAM am ehesten mit der Variablen 'reduzierte Haushaltsgröße' kovariieren wird und daß dies für die anderen sechs Variablen in viel geringerem Maße zutrifft. Man darf also hoffen, daß ungewichtete und mit der reduzierten Haushaltsgröße gewichtete Schätzer im Schnitt das schätzen, was sie schätzen sollen, nämlich die 'wahren' Mittelwerte, von denen wir hier selbstverständlich undiskutiert unterstellen müssen, daß es sie überhaupt gibt.

Die zweite Variablengruppe besteht aus 12 dichotomen Variablen, die bestimmte Alters- und Geschlechtskohorten beschreiben (vgl. Tabelle 3). Von Interesse sind hier die Stichprobenfehler der 12 einzelnen Anteilswerte bzw. Prozentzahlen. Im Unterschied zur ersten Variablengruppe sind jetzt die 'wahren' Mittelwerte aus dem Mikrozensus (1980) ziemlich genau bekannt, so daß der Nenner für den Design-Effekt unproblematisch zu berechnen ist. In bezug auf die 'Erwartungstreue' der Schätzungen für die Anteile muß allerdings angemerkt werden, daß mit großer Sicherheit für die Kohorte der mindestens 70jährigen Frauen der mit der reduzierten Haushaltsgröße gewichtete Schätzer den Anteil systematisch unterschätzt (vgl. auch KIRSCHNER, 1984:167). Der zentrale Term $\widehat{\sigma(\hat{U})}/ \sum_{\zeta} h_j$ gibt also streng genommen Auskunft über die potentielle Variabilität von $\hat{T}$ um $E\hat{T}$ 'herum'; wollte man $E(\hat{T}-\bar{Y})^2$ berechnen, müßte man also zu $\sigma(\hat{U})^2/ \sum_{\zeta} h_j$ die (unbekannte) Größe $(E\hat{T}-\bar{Y})^2$ hinzuaddieren. In den Tabellen 3 und 4 werden aus rein pragmatischen Gründen alle 12 Variablen jedoch gleichartig dargestellt in der Erwartung, daß der Leser die ausgewiesenen Schätzwerte ohnehin im Sinne dieser Diskussion einer kritischen Würdigung unterziehen wird.

Die dritte Variablengruppe von 6 Variablen entstand durch Dichotomisierung der Variablen mit dem Fragetext: 'Wie oft verfolgen Sie kirchliche Sendungen im Radio oder Fernsehen?' (vgl. Codebuch ALLBUS 1982). Die 6 Antwortvorgaben gehen von 'mehr als einmal in der Woche' bis hin zu 'nie'.

Die Dichotomisierung dieser Variablen und die Berechnung von Stichprobenfehlern für die sechs resultierenden dichotomen Variablen wurde gewählt, da an dieser Stelle nicht die Problematik des Stichprobenfehlers einer (Multinomial-)Verteilung diskutiert werden kann.

Zur Berechnung der Design-Effekte müssen, da die Frage nach dem Verfolgen kirchlicher Sendungen nicht im ALLBUS 1980 vertreten war, zur Schätzung von S_Y^2 die aus der Stichprobe des ALLBUS 1982 gewonnenen Anteilswerte herangezogen werden.

Die 'Ferne' der sechs dichotomen Variablen zur Variablen 'reduzierte Haushaltsgröße' äußert sich darin, daß gewichtet und ungewichtet berechnete Anteilswerts sich kaum unterscheiden, wie man den Tabellen 5 und 6 entnehmen kann. Man wird also unterstellen dürfen, daß die Anwendung von (6) insofern gerechtfertigt ist, als man in der Tat die Variation von T um Y , den 'wahren' Wert, mißt.

Technisch gesehen bedeutet die Anwendung von (6) und (7) die Durchführung des folgenden Algorithmus: Gegeben sei eine Variable Y, für die eine Mittelwertbildung über ihre Werte sinnvoll ist.

I) Man bestimme für jede der Stichproben S_i aus der i-ten Primäreinheit $\sum_{S_i} y_j$ bzw. $\sum_{S_i} h_j y_j$, wobei sich natürlich die Summation nur auf die gültigen Werte von Y bezieht - dies kann bedeuten, daß $\sum_{S_i} y_j$ u.U. gleich Null ist.
Man bestimme ebenso $\sum_{S_i} 1$, also die Anzahl der gültigen Werte von Y in S_i sowie $\sum_{S_i} h_j$, die Summe der zu gültigen Werte von Y in S_i gehörigen reduzierten Haushaltsgrößen.

II) Man bilde im Sinne von (7) die folgenden Summen von Quadraten:

$$\sum_{i=1}^{k} \Big(\sum_{S_i} y_\nu - \frac{1}{k} \sum_{j=1}^{k} \sum_{S_j} y_\mu \Big)^2 = \widehat{\sigma^2 \Big(\sum_{i=1}^{k} \sum_{S_i} y_j \Big)} = \widehat{\sigma^2} (\hat{U})$$

$$\sum_{i=1}^{k} \Big(\sum_{S_i} h_\nu y_\nu - \frac{1}{k} \sum_{j=1}^{k} \sum_{S_j} h_\mu y_\mu \Big)^2 = \widehat{\sigma^2 \Big(\sum_{i=1}^{k} \sum_{S_i} h_j y_j \Big)} = \widehat{\sigma^2 (\hat{U})}$$

$$\sum_{i=1}^{k} \Big(\sum_{S_i} 1 - \frac{n}{k} \Big)^2 = \sigma^2 \widehat{\Big(\sum_{i=1}^{k} \sum_{S_i} 1 \Big)} \quad \text{und}$$

$$\sum_{i=1}^{k} \Big(\sum_{S_i} h_\nu - \frac{1}{k} \sum_{j=1}^{k} \sum_{S_j} h_\mu \Big)^2 = \sigma^2 \widehat{\Big(\sum_{i=1}^{k} \sum_{S_i} h_j \Big)}. \qquad 7)$$

Schritt II) bedeutet offensichtlich die Bildung der üblichen Summen von Abweichungsquadraten auf dem Aggregationsniveau der Primäreinheiten. Die Summation erstreckt sich dabei über alle k, ursprünglich für die Feldarbeit bereitgestellten Primäreinheiten - beim ALLBUS 1982 waren dies 630.

Zieht man SPSS - Version 9 - zur Berechnung dieser Summen heran, heißt dies:

$\tilde{\mathrm{I}}$) a) Sortierung der Datei nach der Identifikationsvariablen für die Primäreinheiten.

b) Anwendung der Aggregationsstatistik 'SUM' auf die Variablen Y, Y * [reduzierte Haushaltsgröße] sowie die Variablen 'identisch Eins' und 'reduzierte Haushaltsgröße'. Dabei darf sich für die beiden letztgenannten Variablen die Summation nur über die Fälle erstrekken, die zu gültigen Werten von Y gehören.

$\tilde{\mathrm{II}}$) Man ermittelt vermöge der mit AGGREGATE erzeugten Datei die Summen von Quadraten in II, indem man auf die mit der Wurzel aus der um Eins verminderten Anzahl der im Datensatz vertretenen Primäreinheiten multiplizierten Variablen die Statistik-Prozedur CONDESCRIPTIVE anwendet. Die unter VARIANCE ausgewiesenen Werte sind dann die gesuchten Größen.

Die Schritte $\tilde{\mathrm{I}}$) und $\tilde{\mathrm{II}}$) können mit Hilfe eines kommentierten SPSS-Setup im Anhang dieser Arbeit nachvollzogen werden.

$\tilde{\mathrm{I}}$) und $\tilde{\mathrm{II}}$) stellen für sich allein zweifelsohne sehr einfache EDV-Prozeduren dar. Zu zeigen bleibt allerdings, daß man damit auch $E(\hat{T}-\bar{Y})^2$ schätzen kann. Denn:

1) Die Summation erstreckt sich nicht über alle Primäreinheiten, sondern nur über die in der realisierten Stichprobe = Datensatz vertretenen - beim ALLBUS 1982 waren dies 607.
2) Unbekannt sind auf der rechten Seite des Gleichheitszeichens in (6) (in der Regel) $\bar{Y}$ - der 'wahre' Mittelwert - $E\hat{V}$ und $r(\hat{U},\hat{V})$.
3) Selbst wenn man - pragmatisch - für $\bar{Y}, E\hat{V}$ und $r(\hat{U},\hat{V})$ in (6) Werte einsetzt, bedeutet dies immer noch den Einsatz des Taschenrechners, um mit den in $\tilde{\mathrm{II}}$ beschriebenen Summen von Quadraten zu einer Schätzung von MSE zu kommen.

Einwand 1) ist zwar berechtigt, jedoch zeigt sich empirisch, daß die leichte Vergrößerung der Schätzungen für $\sigma^2(\hat{U})$ bzw. $\sigma^2(\hat{V})$ durch Summation über alle k (=630) Primäreinheiten - natürlich unter Verwendung des Mittelwerts dieser 630 Einheiten - keinen wesentlichen Effekt auf die Schätzung des MSE hat.[8] Man kann es also bei der bequemen Kombination von AGGREGATE und CONDESCRIPTIVE belassen, ohne die Schätzungen wesentlich gegenüber dem Resultat aus der 'Idealprozedur' zu verschlechtern.

Dem Einwand 2) muß pragmatisch Rechnung getragen werden. Falls keine 'wahre' Größe $\bar{Y}$ - sei es Anteil oder Mittelwert - vorliegt, wird man den Schätzwert aus der Stichprobe für $\vec{Y}$ in (6) einsetzen. Dies ist insofern nicht sehr gravierend, weil, wie im vorigen Abschnitt betont, der Einfluß von $\bar{Y}^2\sigma^2(\hat{V})/(E\hat{V})^2$ auf das Gesamtergebnis nicht entscheidend ist. Für $E\hat{V}$ wird man die Anzahl der zu gültigen Werten von Y gehörenden Fälle einsetzen bzw. die zugehörige Summe über die reduzierten Haushaltsgrößen. Da ohnehin eine geringe Varianz von $\hat{V}$ vorausgesetzt werden kann - beim ALLBUS 1982 liegt die Schätzung von $\sigma(\hat{V})$ größenordnungsmäßig bei 40, wobei n=2991 - erscheint die Vorgehensweise berechtigt. Für die im folgenden wiedergegebenen Tabellen wurde für $r(\hat{U},\hat{V})$ durchgehend bei den Berechnungen für den MSE der Wert 0.1 eingesetzt. Wegen der geringen Varianz von $\hat{V}$ ist $r(\hat{U},\hat{V})$ vermutlich zumindest für die drei angegebenen Variablengruppen noch kleiner, jedoch selbst die Verwendung eines kleineren Wertes als 0.1 würde die Schätzung für den MSE nicht wesentlich ansteigen lassen.

Das Faktum der vergleichsweise geringen Varianz von $\hat{V}$ führt schließlich zu einem empirischen Argument gegen Einwand 3): Alle sechs folgenden Tabellen zeigen, daß die alleinige Berechnung der Schätzung für $\sigma(\hat{U})/E\hat{V}$ bereits einen guten Eindruck von den Größenordnungen der MSEs vermittelt, daß man also $(\bar{Y}^2\sigma^2(\hat{V})-2\bar{Y}\sigma(\hat{U})\sigma(\hat{V})r(\hat{U},\hat{V}))/(E\hat{V})^2$ nicht unbedingt mitzuschätzen hat, um einen groben Überblick über die Stichprobenfehler bestimmter Variablen zu bekommen. Das am Ende des vorigen Abschnitts besprochene 'einfache Modell', das zur Schätzung eben von $\sigma(\hat{U})/E\hat{V}$ führt, reicht also für diesen Zweck aus. Man erhält so zwar Unterschätzungen des MSE für Variablen, die metrischen Charakter haben - dies zeigen Tabelle 1 und Tabelle 2 deutlich - jedoch finden sich bei dichotomen Variablen, die im Resultat Anteilswerte bzw. Prozentzahlen beschreiben, nur geringe Abweichungen zwischen den Spalten für $\widehat{\sqrt{MSE}}$ und $\widehat{\sigma(\hat{U})/E\hat{V}}$ in den Tabellen 3 bis 6.

Insgesamt beschreiben also $\tilde{\mathrm{I}}$ und $\tilde{\mathrm{II}}$ trotz aller vereinfachenden Annahmen einen Weg, um auf einfache Weise durch Ablesen einer für STD DEV ausgewiesenen Größe in einem CONDESCRIPTIVE-output und die anschließende Division durch die Fallzahl des Datensatzes bzw. die Summe der im Datensatz erscheinenden reduzierten Haushaltsgrößen - auch dieses ließe sich natürlich maschinell gleich mit erledigen - die Größenordnungen von Stichprobenfehlern unter dem ADM-Stichprobenplan zu ermitteln. Die errechneten Zahlen müssen dann allerdings unter Berücksichtigung aller in den vorangegangenen Abschnitten diskutierten einschränkenden Annahmen sehr sorgfälgig interpretiert und nach Möglichkeit stets in bezug zum DEFF gesetzt werden.

Die Tabellen 1 und 2 geben einen Überblick über die Stichprobenfehler für die Variablengruppe 'Wichtigkeit von Lebensbereichen' (Die Erklärungen für die Abkürzungen der Variablennamen finden sich zu Beginn dieses Abschnitts). In Tabelle 1, also mit $\hat{T}= \sum_{\varsigma} y_j/n$, wurde MSE=$E(\hat{T}-\bar{Y})^2$ mit Hilfe von (6) geschätzt, wobei folgende Werte eingesetzt wurden: $E\hat{V}=2991$, $r=0.1, \bar{Y}=\hat{T}$.
In allen Tabellen wird zur Berechung von $\sqrt{DEFF}$ der in der jeweiligen Spalte für $\sqrt{MSE}$ ausgewiesene Wert herangezogen.

Es sei daran erinnert, daß für den Nenner von $\sqrt{DEFF}$(MSE) das arithmetische Mittel der Standardabweichungen aus dem ALLBUS 1980 und dem ALLBUS 1982 herangezogen wird. Diese Größen wie auch diejenigen für $\hat{T}$ liest man z.B. unmittelbar ab, indem man für die Variablen WFAM usw. auf den entsprechenden Datensatz die SPSS-Prozedur CONDESCRIPTIVE anwendet (unter Ausschluß der fehlenden Werte).

Man sieht, daß mit zunehmend extremer Lage des geschätzten Mittelwerts auf der Skala von 1 bis 7 der Design-Effekt tendenziell zunimmt. Obwohl hier im einzelnen nicht der Raum ist, dies zu diskutieren, sei dennoch angemerkt, daß z.B. in bezug auf die Variable WFAM viele der Primäreinheiten recht homogen bei höheren Skalenwerten liegen werden, so daß sich der notwendig ergebende Klumpungseffekt im DEFF deutlich niederschlägt.

Tabelle 1: Wichtigkeit von Lebensbereichen / ungewichtet

	$\hat{T}$	$\sqrt{MSE}$	$\widehat{\sigma(\hat{U})/E\hat{V}}$	$\sqrt{DEFF}$ (MSE)
WFAM	6.08	0.116	0.091	4.085
WBER	5.39	0.105	0.084	3.386
WFREI	5.59	0.106	0.083	4.186
WFREU	5.36	0.102	0.080	3.848
WVERW	4.57	0.091	0.074	2.890
WREL	3.98	0.086	0.073	2.346
WPOL	4.16	0.086	0.071	2.671

Datenbasis: ALLBUS 1982.

$\sqrt{DEFF}$ (MSE) bedeutet, daß $\sqrt{DEFF}$ auf der Basis der Zahlen in der Spalte für $\sqrt{MSE}$ berechnet wurde.

Bei Tabelle 2, für die $\hat{T}=\sum_{\zeta} h_j y_j / \sum_{\zeta} h_j$, wurde wie folgt eingesetzt:

$E\hat{V}=5968$, $r=0.1$, $\bar{Y}=\hat{T}$.

Tabelle 2: Wichtigkeit von Lebensbereichen / gewichtet

	$\hat{T}$	$\sqrt{MSE}$	$\widehat{\sigma(\hat{U})/E\hat{V}}$	$\sqrt{DEFF}$ (MSE)
WFAM	6.22	0.148	0.115	5.624
WBER	5.52	0.136	0.109	4.578
WFREI	5.62	0.133	0.103	5.301
WFREU	5.33	0.127	0.099	4.832
WVERW	4.59	0.114	0.091	3.675
WREL	3.98	0.105	0.088	2.927
WPOL	4.19	0.105	0.085	3.320

Datenbasis: ALLBUS 1982.

$\sqrt{DEFF}$ (MSE) bedeutet, daß $\sqrt{DEFF}$ auf der Basis der Zahlen in der Spalte für $\sqrt{MSE}$ berechnet wurde.

In Tabelle 2 fällt auf, daß $\sqrt{MSE}$ durchweg größer ist als in Tabelle 1; des weiteren sind die Unterschiede von gewichtetem zu ungewichtetem $\hat{T}$ nicht gravierend - auch wenn sie teilweise in der Größenordnung von $\sqrt{MSE}$ liegen.

Der erste Befund ist einfach dadurch zu erklären, daß in $\hat{T}$ durch die Gewichtung mit der reduzierten Haushaltsgröße die Größenordnung der Variabilität des Zählers angehoben wird, während $\sum_{\zeta} h_j$ = Nenner ebenso wie n eine relativ geringe Variabilität aufweist. Die sehr geringen Unterschiede zwischen gewichteter und ungewichteter Berechnung für die Variablen WFREI, WFREU, WVERW, WREL, WPOL liegen in deren bereits angedeuteter 'Ferne' zur Variablen 'reduzierte Haushaltsgröße'. Daß für WFAM und WBER die Unterschiede etwas größer sind, begründet sich in deren um mehr als das Doppelte größeren Pearsonschen Korrelationen zu dieser Variablen.

Sowohl in Tabelle 1 als auch in Tabelle 2 liegt der Quotient $\hat{T}/\sqrt{MSE}$ in Prozenten ausgedrückt stets unter 2.65%, also in Größenordnungen, die üblicherweise im Rahmen der Sozialwissenschaften in aller Regel für sich allein genommen für nicht sehr interpretationsfähig gehalten werden. Wie im Abschnitt 2 sei jedoch hier nochmals betont, daß weniger die absolute Größe von MSE oder der 'Variationskoeffizient' $\hat{T}/\sqrt{MSE}$ in diesem Zusammenhang von vorrangigem Interesse sind; man weiß ohnehin, daß diese Zahlen bei einem n von 3000 nicht sehr groß sein können. Vielmehr bietet sich zur Interpretation der Design-Effekt an. So sieht man in den Tabellen 3 bis 6, daß für die dort untersuchten dichotomen Variablen die Werte für $\sqrt{DEFF}$ deutlich niedriger liegen, mithin also die nicht unplausible Interpretation naheliegt, daß weniger an Klumpen = Primäreinheiten - Homogenität vorliegt als bei der ersten Variablengruppe - daß also z.B. in bezug auf Alters- und Geschlechtskohorten bereits auf Primäreinheitenebene eine gute 'Durchmischung' vorliegt.

Im einzelnen wurden für die Tabellen 3 und 4 in Formel (6) die folgenden Zahlen eingesetzt: $E\hat{V}=2991$ bzw. $E\hat{V}=5968$ (bei Gewichtung); r=0.1 sowohl für Tabelle 3 als auch Tabelle 4; für $\bar{Y}$ wurden die in beiden Tabellen ausgewiesenen Zahlen aus einer Sonderauszählung des Mikrozensus 1980 verwendet.

In den Tabellen 3 und 4 sind insgesamt sieben Werte mit Eins ausgewiesen. Dies bedeutet nicht, daß $\sqrt{MSE}$ und $\sqrt{P(1-P)/2991}$ genau gleich sind. Vielmehr war in diesen Fällen $\sqrt{MSE}$ kleiner als der Vergleichswert. Üblicherweise wird dann für den Quotienten beider Werte eine Eins gesetzt. Ohne dieses Verfahren hier näher diskutieren zu wollen, sei angemerkt, daß alleine der potentiellen Variabilität wegen, die auch der Schätzung von $E(\hat{T}-\bar{Y})^2$ innewohnt, ein Überschreiten des Wertes von $\sqrt{MSE}$ durch $\sqrt{P(1-P)/n}$ durchaus vorkommen kann, obwohl unter dem ADM-Design und seiner Implementierung kaum vorstellbar ist, daß man 'besser liegt' als mit einer Zufallsstichprobe, die durch Ziehen ohne Zurücklegen entstanden ist.

Tabelle 3: Alters- und Geschlechtskohorten / ungewichtet

		$\hat{T}$[%]	MZ80	$\sqrt{MSE}$[%]	$\widehat{\sigma(\hat{U})/E\hat{V}}$ [%]	$\sqrt{DEFF}$ (MSE)
	18-29	8.9	11.2	0.544	0.539	1.000
	30-39	8.5	8.0	0.531	0.530	1.070
männ-	40-49	9.6	9.3	0.558	0.557	1.051
lich	50-59	7.3	7.4	0.478	0.477	1.000
	60-69	6.2	5.0	0.450	0.451	1.187
	70-	4.3	5.1	0.397	0.398	1.000
	18-29	10.1	10.8	0.577	0.575	1.017
	30-39	9.5	8.1	0.570	0.570	1.143
weib-	40-49	10.0	9.3	0.563	0.561	1.060
lich	50-59	8.3	9.2	0.519	0.517	1.000
	60-69	8.1	7.7	0.517	0.517	1.061
	70-	9.2	8.9	0.604	0.604	1.160

Datenbasis: ALLBUS 1982

$\sqrt{DEFF}$ (MSE) bedeutet, daß $\sqrt{DEFF}$ auf der Basis der Zahlen in der Spalte für $\sqrt{MSE}$ berechnet wurde. [%] zeigt an, daß die in der betreffenden Spalte angegebenen Zahlen die mit 100 multiplizierten ursprünglichen Anteilswerte sind.

Tabelle 4: Alters- und Geschlechtskohorten / gewichtet

		$\hat{T}$[%]	MZ80	$\sqrt{MSE}$ [%]	$\widehat{\sigma(\hat{U})/E\hat{V}}$ [%]	$\sqrt{DEFF}$ (MSE)
	18-29	9.9	11.2	0.665	0.657	1.153
	30-39	7.7	8.0	0.509	0.504	1.026
männ-	40-49	10.9	9.3	0.680	0.678	1.280
lich	50-59	8.7	7.4	0.612	0.612	1.279
	60-69	6.6	5.0	0.500	0.502	1.255
	70-	3.8	5.1	0.386	0.385	1.000
	18-29	10.8	10.8	0.707	0.702	1.246
	30-39	9.0	8.1	0.580	0.577	1.163
weib-	40-49	11.4	9.3	0.703	0.701	1.324
lich	50-59	8.8	9.2	0.625	0.621	1.183
	60-69	8.5	7.7	0.458	0.453	1.000
	70-	5.9	8.9	0.414	0.401	1.000

Datenbasis: ALLBUS 1982

$\sqrt{DEFF}$ (MSE) bedeutet, daß $\sqrt{DEFF}$ auf der Basis der Zahlen in der Spalte für $\sqrt{MSE}$ berechnet wurde. [%] zeigt an, daß die in der betreffenden Spalte angegebenen Zahlen die mit 100 multiplizierten ursprünglichen Anteilswerte sind.

Für die beiden letzten Tabellen 5 und 6 wurden zur Berechnung von MSE gemäß Formel (6) die folgenden Werte verwendet: $E\hat{V}$=2980 bzw. $E\hat{V}$=5947 (bei Gewichtung); r=0.1 sowohl für Tabelle 5 als auch Tabelle 6; $\bar{Y}$ wurde wie bei der ersten Variablengruppe wieder gleich $\hat{T}$ gesetzt.

Tabelle 5: Kirchliche Sendungen / ungewichtet

	$\hat{T}$[%]	$\sqrt{MSE}$ [%]	$\widehat{\sigma(\hat{U})/E\hat{V}}$ [%]	$\sqrt{DEFF}$ (MSE)
1	5.0	0.442	0.443	1.107
2	11.8	0.672	0.669	1.137
3	11.5	0.702	0.700	1.201
4	15.6	0.776	0.768	1.167
5	27.7	0.975	0.940	1.189
6	28.4	1.067	1.036	1.292

Datenbasis: ALLBUS 1982

$\sqrt{DEFF}$ (MSE) bedeutet, daß $\sqrt{DEFF}$ auf der Basis der Zahlen in der Spalte für $\sqrt{MSE}$ berechnet wurde. [%] zeigt an, daß die in der betreffenden Spalte angegebenen Zahlen die mit 100 multiplizierten urpsrünglichen Anteilswerte sind.

1: mehr als einmal in der Woche
2: einmal in der Woche
3: ein- bis dreimal im Monat
4. mehrmals im Jahr
5: seltener
6: nie

Tabelle 6: Kirchliche Sendungen / gewichtet

	$\hat{T}$ [%]	$\sqrt{MSE}$ [%]	$\widehat{\sigma(\hat{U})/E\hat{V}}$ [%]	$\sqrt{DEFF}$ (MSE)
1	4.5	0.431	0.432	1.135
2	11.4	0.725	0.719	1.245
3	11.0	0.744	0.739	1.298
4	16.0	0.883	0.868	1.315
5	28.6	1.120	1.060	1.353
6	28.5	1.173	1.119	1.419

Datenbasis: ALLBUS 1982

$\sqrt{DEFF}$ (MSE) bedeutet, daß $\sqrt{DEFF}$ auf der Basis der Zahlen in der Spalte für $\sqrt{MSE}$ berechnet wurde. [%] zeigt an, daß die in der betreffenden Spalte angegebenen Zahlen die mit 100 multiplizierten urpsrünglichen Anteilswerte sind.

Kodierung s. Fußzeile in Tabelle 5

7. Schlußbemerkungen

In den vorangegangenen Abschnitten wurde gezeigt, daß sich für den weithin in der Forschungspraxis gebräuchlichen ADM-Stichprobenplan unter ganz bestimmten Voraussetzungen sinnvolle Stichprobenfehlerberechnungen durchführen lassen, und zwar bezogen auf ungewichtete bzw. auf mit der Anzahl der Personen der Grundgesamtheit im Haushalt gewichtete Berechnung von Mittelwerten. Es ergaben sich insbesondere in bezug auf Design-Effekte interpretationsfähige Resultate. Die Methode der Berechnungen nutzte aus, daß bei einer nach dem ADM-Design gezogenen Stichprobe die Varianz der auf Stimmbezirks(=Primäreinheiten)-Ebene hochaggregierten Daten von großer Aussagekraft ist.

Zur praktischen Durchführung der Berechnungen wurde eine Prozedur angegeben, die sehr leicht mit den Hilfsmitteln von SPSS 9 durchführbar ist und die zudem Resultate liefert, die sowohl im Sinne einer 'Daumenregel' zur groben Abschätzung von mittleren quadratischen Abweichungen benutzt werden können, als auch als Basis für genauere Berechnungen dienen, die ihrerseits zu den Werten für Design-Effekte führen.

Diese Arbeit wurde in der Absicht geschrieben, dem methodisch interessierten Leser einen Vorschlag zu schildern und zu begründen, wie er mit einfachen Mitteln zu Abschätzungen für Stichprobenfehler kommt. Demgemäß wurde durchgehend auf strenge stichprobentheoretische Begründungen zu zentralen Aussagen insbesondere in Abschnitt 4 verzichtet. Es ist allerdings geplant, eine in theoretischer Hinsicht ausführlichere Darstellung an anderer Stelle zu veröffentlichen.

Diese Arbeit wurde von Hans-Peter Kirschner verfaßt, der bei ZUMA für Stichprobenfragen zuständig ist.

Anmerkungen

1) Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982 dient dem Ziel, der Forschung und der Lehre in den Sozialwissenschaften aktuelle Daten über Einstellungen und Verhaltensweisen der Bevölkerung in der Bundesrepublik zur Verfügung zu stellen. Das Projekt wurde gefördert von der Deuschen Forschungsgemeinschaft.
Der ALLBUS wird in enger Zusammenarbeit mit dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) e.V. in Mannheim und dem Zentralarchiv für empirische Sozialforschung der Universität zu Köln realisiert. Für den ALLBUS 1982 lag die ZUMA-Projektleitung bei Karl Ulrich Mayer (jetzt Berlin).
2) Man beachte, daß $\binom{N}{n}$ für einen Querschnitt wie den ALLBUS 1982 mit $N \simeq 44\ 000\ 000$ und $n \simeq 3\ 000$ eine unvorstellbar große Zahl ist. Sie ist größer als eine 1 mit 12 000 anhängenden Nullen.
3) Eine solche Variable ist etwa im ALLBUS 1982 die Frage nach der Befragteneinstufung der Wichtigkeit des Lebensbereichs Familie.
4) Zum Beweis nutzt man aus, daß für Zahlen a(1), ...a(n), $n \geq 2$, gilt:

$$\frac{1}{n-1} \sum_{i=1}^{n} (a(i) - \bar{a})^2 \leq \frac{n}{n-1} \frac{R^2(a(i))}{4},$$

mit R(a(i))=(kleinstes der a(i))-(größtes der a(i))
= range der Zahlen a(i).
5) Man beachte aber, daß z.B. für das ADM-Design die Wahrscheinlichkeit dafür nur überaus aufwendig berechnet werden kann, daß zwei bestimmte Stimmbezirke in der Stichprobe enthalten sind.
6) In dieser Darstellung werden alle Auswertungen auf die Ebene von Personen bezogen, da die Behandlung von Stichprobenfehlern auch auf Haushaltsebene den gegebenen Rahmen sprengen würde. Dies umso mehr, als diese Arbeit nicht umfassend, sondern lediglich in exemplarischer Weise die Berechnung von Stichprobenfehlern demonstrieren soll.
7) Es sei darauf hingewiesen, daß unter Voraussetzung 5.3 bzw. unter der Voraussetzung $E[\sum_{S_i} h_j] = M_i \cdot m/g_i$ - vgl. Abschnitt 2 - die Größen

$$\widehat{\sigma^2(\sum_{i=1}^{k} \sum_{S_i} 1)} \quad \text{bzw.} \quad \widehat{\sigma^2(\sum_{i=1}^{k} \sum_{S_i} h_j)}$$

'gute' Schätzungen für die Varianz des Nenners von $\hat{T}$ sind.
8) Für Zahlen $x_1, \ldots, x_{\tilde{k}}, x_{\tilde{k}+1}, \ldots, x_k$, $0 = x_{\tilde{k}+1} = \ldots = x_k$,

zeigt man leicht, daß

$$\sum_{i=1}^{k} (x_i - \frac{1}{k} \sum_{j=1}^{k} x_j)^2 = \sum_{i=1}^{\tilde{k}} (x_i - \frac{1}{\tilde{k}} \sum_{j=1}^{\tilde{k}} x_j)^2 + \left[\frac{1}{\tilde{k}} \sum_{i=1}^{\tilde{k}} x_i\right]^2 \frac{\tilde{k}(k-\tilde{k})}{k}.$$

Literatur

Codebuch ALLBUS 1982. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften - ALLBUS 1982,Codebuch mit Methodenbericht und Vergleichsdaten, hrsg. vom Zentralarchiv für empirische Sozialforschung Köln und dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) e.V., Mannheim, 1984.

DES RAJ. Sampling Theory. New York: McGraw-Hill, 1968.

FRANKEL, M.R. Inference from survey samples. An empirical investigation. Michigan: Institute for social research, 1971.

KIRSCHNER, H.-P. Komplexe Bevölkerungsstichproben. Eine Fallstudie. In: H. STENGER (Hrsg.), Praktische Anwendung von Stichprobenverfahren. Sonderhefte zum Allgemeinen Statistischen Archiv, Heft 17. Göttingen, 1980, 69-84.

KIRSCHNER, H.-P. ALLBUS 1980: Stichprobenplan und Gewichtung. In: K.U. MAYER & P. SCHMIDT (Hrsg.), Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Beiträge zu methodischen Problemen des ALLBUS 1980. Monographien Sozialwissenschftliche Methoden, Band 5. Frankfurt: Campus, 1984, 114-182.

KISH, L. A procedure for objective respondent selection within the household. Journal of the American Statistical Association, 44, 1949, 380-387.

SUDMAN,S. Applied Sampling. New York: Academic Press, 1976.

VINTER, S. Survey sampling errors with OSIRIS IV. In: M.M. BARNITT & D. WISHART (eds.), COMPSTAT 1980. Wien: Physika, 1980, 72-77.

Anhang

Im folgenden werden zu den Punkten Ĩ und ĨĨ aus Abschnitt 6 zwei SPSS-Setups konkret angegeben, die eine (mögliche) technische Umsetzung darstellen. Auf die Angabe von JCL wurde bewußt verzichtet, da diese stark vom jeweils verwendeten Computermodell abhängt. Die Variablennummern im ersten Setup stimmen im übrigen nicht mit denjenigen des OSIRIS-Datensatzes zum ALLBUS 1982 überein, der beim Zentralarchiv in Köln zu beziehen ist.

Beide Setups wurden auf der SIEMENS 7541 des Rechenzentrums der Universität Mannheim erstellt und gerechnet.

SETUP für AGGREGATE

```
OSIRIS VARS      V14,V17 TO V23,V77,V103,V141,V245,V291,V337,V394,V395
INPUT MEDIUM     DISK
ALLOCATE         TRANSPACE=...
COMPUTE          LFDNR=V395
COMPUTE          NET=V394

COMMENT          LFDNR UND NET ZUSAMMEN BILDEN DIE
COMMENT          KENNUNG DER PRIMÄREINHEITEN

DO REPEAT        A=WFAM,WBER,WFREI,WFREU,WVERW,WREL,WPOL,
                 KS1 TO KS6,
                 AS1 TO AS12,
COMPUTE          A=0
END REPEAT

COMMENT          WFAM USW. S. ABSCHNITT 5
COMMENT          KS1 TO KS6: KIRCHLICHE SENDUNGEN
COMMENT          AS1 TO AS12: ALTERS- UND GESCHLECHTSKOHORTEN/
COMMENT          SETZE VARIABLEN AUF DEFAULTWERT NULL

DO REPEAT        B=CWFAM,CWBER,CWFREI,CWFREU,CWVERW,CWREL,CWPOL,
                 CKS,CAS/
                 C=HWFAM,HWBER,HWFREI,HWFREU,HWVERW,HWREL,HWPOL,
                 HKS,HAS
COMPUTE          B=1
COMPUTE          C=V291
END REPEAT

COMMENT          CWFAM USW. WERDEN KONSTANT AUF 1 GESETZT
COMMENT          HWFAM USW. WERDEN AUF DEN WERT DER REDUZIERTEN
COMMENT          HAUSHALTSGROESSE GESETZT

DO REPEAT        VAR=WFAM,WBER,WFREI,WFREU,WVERW,WREL,WPOL/
                 CVAR=CWFAM,CWBER,CWFREI,CWFREU,CWVERW,CWREL,CWPOL/
                 HVAR=HWFAM,HWBER,HWFREI,HWFREU,HWVERW,HWREL,HWPOL/
                 V=V17 TO V23,V103,V141
IF               (V LT 99) VAR=V
```

```
IF              (V GE 99) CVAR=0
IF              (V GE 99) HVAR=0
END REPEAT

COMMENT         WFAM USW. HABEN NUR FUER GUELTIGE WERTE DER
COMMENT         AUSGANGSVARIABLEN DEREN WERT SONST GILT DER
COMMENT         DEFAULTWERT NULL
COMMENT         CWFAM, HWFAM USW. SIND NUR FUER GUELTIGE
COMMENT         WERTE VON WFAM. HWFAM USW. VON NULL VERSCHIEDEN

IF              (V77 GE 9) CKS=0
IF              (V77 GE 9) HKS=0

COMMENT         CKS, HKS SIND NUR FUER GUELTIGE WERTE VON V27
COMMENT         VON NULL VERSCHIEDEN

DO REPEAT       ASEX=AS1 TO AS12/
                UBND=1964,1952,1942,1932,1922,1912,
                     1964,1952,1942,1932,1922,1912/
                LBND=1953,1943,1933,1923,1913,1800,
                     1953,1943,1933,1923,1913,1800/
                SEX=1,1,1,1,1,1,2,2,2,2,2,2,
IF              ((V245 GE LBND AND LE UBND) AND V337 EQ SEX) ASEX=1
END REPEAT

DO REPEAT       K=KS1 TO KS6/NR=1,2,3,4,5,6
IF              (V77 EQ NR) K=1
END REPEAT

COMMENT         ERSTELLUNG DICHOTOMER VARIABLEN
COMMENT         V245=GEBURTSJAHR
COMMENT         V337=GESCHLECHT
COMMENT         V77=KIRCHLICHE SENDUNGEN

DO REPEAT       WFAM,WBER,WFREI,WFREU,WVERW,WREL,WPOL,
                KS1 TO KS6,
                AS1 TO AS12/
                G=GWFAM,GWBER,GWFREI,GWFREU,GWVERW,GWREL,GWPOL,
                GKS1 TO GKS6,
                GAS1 TO GAS12
COMPUTE         G=F*V291
END REPEAT

COMMENT         GWFAM USW.: GEWICHTUNG MIT DER REDUZIERTEN
COMMENT         HAUSHALTSGROESSE

SORT CASES      NET,LFDNR

COMMENT         SORTIERUNG DER DATEI NACH DER
COMMENT         KENNUNG DER PRIMAEREINHEITEN

AGGREGATE       GROUPVARS=NET,LFDNR/
                VARIABLES=WFAM,WBER,WFREI,WFREU,WVERW,WREL,WPOL,
                GWFAM,GWBER,GWFREI,GWFREU,GWVERW,GWREL,GWPOL,
                CWFAM,CWBER,CWFREI,CWFREU,CWVERW,CWREL,CWPOL,
```

```
                HWFAM,HWBER,HWFREI,HWFREU,HWVERW,HWREL,HWPOL,
                KS1 TO KS6,GKS1 TO GKS6,CKS,HKS,
                AS1 TO AS12,GAS1 TO GAS12,CAS,HAS/
                AGGSTATS=SUM
STATISTICS      3
FINISH
```

SETUP für CONDESCRIPTIVE

```
VARIABLE LIST   GROUPNO,GROUPSZE,
                WFAM,WBER,WFREI,WFREU,WVERW,WREL,WPOL,
                GWFAM,GWBER,GWFREI,GWFREU,GWVERW,GWREL,GWPOL,
                CWFAM,CWBER,CWFREI,CWFREU,CWVERW,CWREL,CWPOL,
                HWFAM,HWBER,HWFREI,HWFREU,HWVERW,HWREL,HWPOL,
                KS1 TO KS6,GKS1 TO GKS6,CKS,HKS,
                AS1 TO AS12,GAS1 TO GAS12,CAS,HAS
INPUT FORMAT    BINARY(20V/20V/20V/20V/14V)

COMMENT         VARIABLENNAMEN S. SETUP FUER AGGREGATE

INPUT MEDIUM    DISK
ALLOCATE        TRANSPACE=...
COMPUTE         CONST=SQRT(606)
DO REPEAT       VAR=WFAM,WBER,WFREI,WFREU,WVERW,WREL,WPOL
                GWFAM,GWBER,GWFREI,GWFREU,GWVERW,GWREL,GWPOL,
                CWFAM,CWBER,CWFREI,CWFREU,CWVERW,CWREL,CWPOL,
                HWFAM,HWBER,HWFREI,HWFREU,HWVERW,HWREL,HWPOL,
                KS1 TO KS6,GKS1 TO GKS6,CKS,HKS,
                AS1 TO AS12,GAS1 TO GAS12,CAS,HAS
COMPUTE         VAR=VAR*CONST
END REPEAT

COMMENT         VAR=VAR*CONST STELLT SICHER,
COMMENT         DASS VARIANCE BZW. STD DEV
COMMENT         IM OUTPUT VON CONDESCRIPTIVE IN
COMMENT         DER RICHTIGEN GROESSENORDNUNG
COMMENT         ERSCHEINEN

CONDESCRIPTIVE  ALL
STATISTICS      ALL
FINISH
```