

Zur Qualität der sozialwissenschaftlichen Methodenausbildung: am Beispiel statistischer Datenanalyse

Ludwig-Mayerhofer, Wolfgang

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Ludwig-Mayerhofer, W. (2003). Zur Qualität der sozialwissenschaftlichen Methodenausbildung: am Beispiel statistischer Datenanalyse. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 53, 144-155. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198894>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Zur Qualität der sozialwissenschaftlichen Methodenausbildung – am Beispiel statistischer Datenanalyse

von Wolfgang Ludwig-Mayerhofer¹

Zusammenfassung

In jüngster Zeit werden Vorschläge für die Methodenausbildung gemacht, die auf der Grundlage von empirischen Erhebungen die in der Forschungspraxis verwendeten Verfahren zum Maßstab einer gelingenden Methodenausbildung machen. Die Qualität der Ausbildung kann sich aber nicht nur daran messen lassen, dass die in der beruflichen Praxis erforderlichen Verfahren mehr oder weniger oberflächlich und routinemäßig beherrscht werden. Der Artikel nennt einige Aspekte, bei denen in der gegenwärtigen Praxis der statistischen Datenanalyse Qualitätsdefizite auszumachen sind; ebenso werden einige Probleme angeführt, die sich aufgrund unzulänglicher Darstellungen in Lehr- und Handbüchern in der Ausbildung ergeben.

Abstract

Recent contributions to debates about university training argue that the current practice of social research should serve as a guideline to determine the contents of research methods training. However, the quality of training cannot only be measured by the degree to which it is in line with the typical superficial routine application of a certain array of established procedures. The article outlines a number of features concerning which current research practice doesn't always follow highest standards and thus cannot be used as a criterion against which to judge the training in research methods; also, some problems are outlined that result from inadequate presentations in introductions to data analysis and computer software.

¹ Dr. **Wolfgang Ludwig-Mayerhofer** ist Professor für Soziologie – Empirische Sozialforschung am Fachbereich 1 der Universität Siegen, Adolf-Reichwein-Str. 2, 57068 Siegen, Telefon: 0271/740-4384, E-Mail: ludwig-mayerhofer@soziologie.uni-siegen.de. Mein Dank gilt **Uta Liebeskind** für Vorschläge zur Verbesserung des Textes. Unzulänglichkeiten desselben gehen selbstverständlich alleine auf mein Konto.

1 Soziologische Methodenausbildung und Praxis

In den jüngsten Diskussionen um die soziologische Methodenausbildung wird zunehmend geltend gemacht, dass die zu vermittelnden Inhalte noch besser als bisher an der Praxis sozialwissenschaftlicher Forschung ausgerichtet werden sollten. Ein Beispiel hierfür ist der Aufsatz „Konträr und ungenügend? Ansprüche an Inhalt und Qualität einer sozialwissenschaftlichen Methodenausbildung“ von **Manuela Pötschke** und **Julia Simonson** (**Pötschke** und **Simonson** 2003) im letzten Heft der ZA-Information.² Grundlage ihrer Argumentation ist eine Umfrage unter „Sozialforschern“ (S. 76; wie ich vermute, auch einigen Sozialforscherinnen), die entweder der „Wissenschaft“ oder der „kommerziellen Markt- und Meinungsforschung“ angehörten (dazu S. 77), sowie eine Inhaltsanalyse dreier deutschsprachiger soziologischer Fachzeitschriften hinsichtlich der in den empirischen Beiträgen verwendeten Verfahren der Datenerhebung, Stichprobengewinnung und Datenanalyse. **Pötschke** und **Simonson** stellen einerseits in vielen Bereichen eine Übereinstimmung zwischen ihren Ergebnissen und den (mutmaßlich) in der Methodenausbildung vermittelten Inhalten fest, andererseits können sie auf Punkte hinweisen, die wohl zumindest teilweise in der gegenwärtig praktizierten Methodenausbildung und auch in Lehrbüchern zu wenig berücksichtigt werden. Zu letzteren gehören etwa die besonderen Bedürfnisse der Markt- und Meinungsforschung (S. 90 f.) oder der Umgang mit Sekundärdaten (SOEP, ALLBUS etc. [S. 91]).

Auf dieser Grundlage sind – so die Autorinnen – Vorschläge zu den Inhalten der Methodenausbildung möglich, die nicht nur von den individuellen Vorlieben einzelner Personen abhängen, sondern möglichst umfassend das Spektrum der in der Forschungspraxis einzusetzenden Verfahren widerspiegeln. Die Auseinandersetzungen um einen „Kanon“ für die Methodenausbildung sind damit sicher nicht abgeschlossen, die Übersicht über die in der Berufspraxis häufig verwendeten Verfahren gibt aber zumindest eine Grundlage für weitere Diskussionen ab. Zu wenig Beachtung findet jedoch meines Erachtens, dass die Orientierung an der gegenwärtigen Praxis auch problematisch sein kann: Auch wenn feststeht, welche methodischen Verfahren in der beruflichen Praxis angewandt werden, so ist damit noch nichts über die *Qualität* der Anwendung gesagt. Es kommt ja nicht nur darauf an, dass etwa im Bereich der statistischen Datenanalyse – auf den ich mich im Folgenden

2 Dieser Aufsatz fasst Ergebnisse zusammen, die in ähnlicher Form bereits in einem von **Uwe Engel** (**Engel** 2002a) herausgegebenen Sammelband veröffentlicht wurden (**Engel** 2002b; **Simonson** und **Pötschke** 2002a, b),

hauptsächlich beziehen werde³ – Verfahren der deskriptiven Statistik oder der multivariaten Analyse wie die OLS-Regression gelehrt werden, sondern dass die Anwendung dieser Verfahren *möglichst hohen Standards* entspricht.

Genau dies ist meines Erachtens nicht der Fall. Die Datenauswertungsverfahren, die ich, wie erwähnt, hier ins Zentrum stellen möchte, werden nur teilweise nach dem aktuellen Stand des Wissens durchgeführt (oder publiziert) und geben in diesem Sinne nicht immer ein gutes Vorbild für die Ausbildung in diesen Verfahren. Hinzu kommt – ein Aspekt, der für die Qualität der Ausbildung gleichfalls von großer Bedeutung ist –, dass auch einschlägige *Lehrbücher* in manchen Aspekten defizitär sind. Wohlgermerkt: defizitär, also keineswegs rundum schlecht. Wenn aber diese Defizite sich in den publizierten Forschungsergebnissen und/oder in den Lehrbüchern bemerkbar machen, so sollte die Methodenausbildung darauf drängen, diese Defizite künftig zu verringern oder aufzuheben. Den Maßstab hierfür müssen aber Qualitätsstandards bilden, die sich gerade nicht aus den gegenwärtig praktizierten Vorgehensweisen, sondern aus dem (selbstverständlich mindestens teilweise umstrittenen und ständig im Fluss befindlichen) Stand der einschlägigen Fachwissenschaften ableiten lassen – im Bereich der Analyse standardisierter Daten also der Statistik.

Mit meiner Kritik stehe ich keineswegs alleine. Beispielsweise hat **Rainer Schnell** (*Schnell* 2002) die Ausbildung in Statistik als „veraltet und simplizistisch“ qualifiziert. In der Wortwahl scheint mir dies etwas überzogen, in der Richtung aber zutreffend. Es ist daher erforderlich, genauer die Aspekte herauszuarbeiten, in denen die Ausbildung zu verbessern ist. Meine folgenden Ausführungen können sich allerdings nicht auf empirische Erkenntnisse stützen, was in der Methodenausbildung faktisch gelehrt wird. Daher gehe ich einen indirekten Weg: Wenn empirische Publikationen in sozialwissenschaftlichen Fachzeitschriften oder Büchern nicht immer den vorhandenen Standards genügen, so ist anzunehmen, dass die entsprechenden Standards auch in der Ausbildung zu kurz gekommen sind.⁴ Was die Lehrbücher angeht, stütze ich mich auf eigene Erfahrungen in der Methodenausbildung, denen

3 Damit will ich nicht sagen, dass hinsichtlich der Datenerhebungsverfahren und der Stichprobenziehung keine Probleme in Ausbildung und Anwendungspraxis festzustellen wären. Ich habe jedoch den Eindruck, dass die Defizite deutlicher gesehen und thematisiert werden.

4 Alternative Erklärungen wären denkbar: Erstens könnte es sein, dass die Standards in der Analysepraxis zwar verfolgt werden, die Darstellung der Ergebnisse für die Publikation jedoch auf jenes Maß „zurechtgestutzt“ wird, das der gängigen Publikationspraxis entspricht. Zweitens wäre möglich, dass Sozialwissenschaftler und Sozialwissenschaftlerinnen die Standards zwar kennen, aber ihnen nicht immer im vollen Umfang folgen, weil sie den Eindruck haben, dass die Erfüllung der Standards gar nicht erwartet wird. Vor allem die zweite Möglichkeit halte ich für nicht gänzlich unplausibel. Umso wichtiger wäre aber, schon in der Ausbildung auf die Einhaltung der Standards zu drängen.

zufolge Studierende in den Lehrbüchern wenig Anhaltspunkte finden, wie bei der Datenanalyse ganz konkret vorzugehen ist – und in manchen Werken, nicht zuletzt solchen, die sich als ergänzende Handbücher zu verbreiteter Statistik-Software verstehen, durch die enge Anlehnung an die Vorgaben der jeweiligen Programme teilweise sogar in die Irre geführt werden.

Im Folgenden möchte ich einige wichtige Schwachstellen der statistischen Datenauswertung ansprechen; mein Anspruch ist nicht, sämtliche wichtigen Probleme zu erörtern oder diese gar ausführlich zu diskutieren, sondern nur durch eine kurze Auflistung einen ersten Anstoß zur Diskussion zu geben. Auch möchte ich besonders strittige oder schwierige Fragen wie den Umgang mit fehlenden Werten (Item-Non-Response) oder die Korrektur von Selbst- oder Fremdselektionseffekten nach **Heckman** (1979) nur an dieser Stelle kurz erwähnen, aber im Folgenden nicht weiter besprechen – nicht weil diese unwichtig wären, sondern weil hier der Stand der Forschung wenig eindeutig ist (siehe etwa **Stolzenberg** und **Relles** 1997).

Um einem denkbaren Missverständnis entgegen zu treten, möchte ich betonen, dass es hier darum geht, die Überlegungen von **Pötschke** und **Simonson** zu erweitern und zu präzisieren. Mein Beitrag darf nicht so verstanden werden, als hätten sich die Autorinnen *gegen* Qualitätsstandards oder die nachfolgend angesprochenen Verfahrensweisen ausgesprochen. Auffällig ist allerdings, dass im Titel ihres Aufsatzes explizit „Inhalt und Qualität“ der Methodenausbildung genannt werden, der Aspekt der Qualität jedoch im Text keine Rolle spielt. Ob **Pötschke** und **Simonson** der Auffassung sind, dass die aktuelle Praxis den wichtigsten Qualitätsstandards entspricht oder nicht, geht dementsprechend aus ihren Ausführungen nicht hervor.

2 Probleme der Praxis statistischer Datenanalyse und ihrer Vermittlung in der Lehre

1. Zunächst ist auf die weitgehende Vernachlässigung (numerischer und graphischer) *explorativer* Verfahren der Datenanalyse hinzuweisen. **Rainer Schnell** hat in dieser Hinsicht bereits mehrfach – wenn auch, wie es scheint, weitgehend ungehört – das Nötige gesagt (**Schnell** 1994, 2000). Daher nur in wenigen Worten: Mit der auf **Tukey** (1977) zurückgehenden und inzwischen auch in Publikationen mit Lehrbuchcharakter (**Erickson** und **Nosanchuk** 1992; **Marsh** 1988) zugänglichen explorativen Datenanalyse liegen Werkzeuge vor, mit denen Daten nach auffälligen Strukturen gemustert werden können. Ferner können diese Verfahren wichtige Hilfestellung geben, wenn es darum geht, Daten zu transformieren und statistische Modelle auszuwählen, die der Datenstruktur entsprechen (oder anders formuliert: wenn zu prüfen ist, ob die Voraussetzungen für die Anwendung bestimmter statisti-

scher Modelle gegeben sind). In der Forschungspraxis, soweit sie sich in sozialwissenschaftlichen Publikationen niederschlägt, überwiegt jedoch ein Vorgehen bei der Datenanalyse, das sich auf die routinemäßige (aber ungeprüfte, siehe unten) Anwendung einiger statistischer Modelle reduziert.

2. Weiterhin sind Lehrbücher und Forschungspraxis größtenteils hinsichtlich der *Verbindung von Stichprobenziehung und statistischer Datenanalyse* blind. Ich beziehe mich auf die Tatsache, dass sozialwissenschaftliche Datensätze häufig nicht auf einfachen Zufallsstichproben beruhen, dass aber die in der gängigen Software zur Datenanalyse implementierten Signifikanztests entweder durchgängig oder zumindest als Voreinstellung genau solche Stichproben unterstellen.⁵ Angesichts der häufigen Verwendung von Klumpenstichproben bzw. mehrfach geschichteter Auswahlen werden so die Standardfehler (und damit die Konfidenzintervalle) in der Regel deutlich unterschätzt (sog. Design-Effekt, siehe *Kish* 1995).⁶

3. Auch sonst finden Fragen der *statistischen Signifikanztests* zu wenig Aufmerksamkeit. Wozu quälen wir in der Statistikausbildung Studierende mit dem Problem *einseitiger oder zweiseitiger Hypothesen*, wenn in der Forschungspraxis im Allgemeinen mehr oder weniger blind die Signifikanztests der Softwarepakete übernommen werden, die standardmäßig stets auf zweiseitige Fragestellungen bezogen sind? Dabei liegen häufig einseitige Fragestellungen vor, auch wenn diese nicht immer explizit als solche benannt werden; beispielsweise wird wohl derzeit kaum ein Forscher oder eine Forscherin, der oder die sich mit Lohnunterschieden von Frauen und Männern beschäftigt, die zweiseitige Hypothese prüfen, dass die Löhne der beiden Gruppen sich in nicht vorhersehbarer Richtung unterscheiden.

4. Während Verfahren der Stichprobenziehung in den Lehrbüchern ausführlich dargestellt werden, wird verblüffenderweise Fragen der *Stichprobengröße* wenig Aufmerksamkeit geschenkt (mit Ausnahme etwa allgemeiner Hinweise auf die Abhängigkeit statistischer Schlussfolgerungen vom Umfang der Stichprobe gemäß dem „Wurzel-n-Gesetz“, nach dem der Standardfehler von Schätzern nicht proportional zum Stichprobenumfang n abnimmt, sondern nur proportional zu \sqrt{n}). Die Folge: Die Forschungspraxis frönt einer Art Zahlenmagie – denn wie sonst ist zu erklären, dass als Stichprobenumfang so häufig ein ganzzahliges Vielfaches von 1 000 festge-

5 STATA ist eines der wenigen Programme, das Nutzern und Nutzerinnen Optionen anbietet, der „Clustering“ von Fällen im Datensatz gerecht zu werden. – Positiv hervorzuheben ist hier das Lehrbuch von *Kühnel* und *Krebs* (2001), welches immerhin auf dieses Problem wiederholt hinweist.

6 *Rainer Schnell* hat zu solchen Fragestellungen zwei Forschungsprojekte durchgeführt – aber bislang bedauerlicherweise wenig Ergebnisse publiziert.

legt wird? Ein rationaler Umgang mit Problemen der Stichprobengröße würde sich auf Überlegungen hinsichtlich der zu erwartenden Effektgröße und des zu tolerierenden Fehlers 2. Art stützen, wie sie unter dem Begriff der statistischen „Power“ oder Teststärke behandelt werden (*Cohen* 1988; *Buchner, Erdfelder* und *Faul* 1996). Gewiss sind entsprechende Überlegungen, die sich häufig auf experimentelle Forschungsdesigns beziehen, nicht ohne weiteres auf die meist komplexen Datenstrukturen der Soziologie zu übertragen, doch kann dies kein Grund sein, diesen Aspekt überhaupt nicht zu thematisieren.

5. Dieser Punkt führt zu grundlegenden Diskussionen, die seit langer Zeit (siehe schon *Morrison* und *Henkel* 1970) über die Rolle von Signifikanztests geführt werden und mit überzeugenden Gründen die gängige Praxis der Signifikanztests überhaupt, zumindest aber deren unreflektierte Routineanwendung ablehnen (viele neuere Beiträge sind in *Harlow, Mulaik* und *Steiger* 1997 versammelt). Selbst wenn man sich diesen Argumenten⁷ nicht oder nicht vollständig anschließt, wäre aus dieser Diskussion zumindest die Anregung zu übernehmen, in den Veröffentlichungen nicht Signifikanzniveaus, sondern Konfidenzintervalle (oder wenigstens Standardfehler) auszuweisen. So könnte man auch die übliche, aber dennoch statistisch gesehen sinnlose Praxis der Kennzeichnung verschiedener Signifikanzniveaus durch eine unterschiedliche Zahl von „Sternchen“ einstellen (dazu schon *Guttman* 1977). Gleichzeitig würde man die für die auch in der Soziologie zukünftig wohl wichtiger werdenden Meta-Analysen erforderlichen Informationen zur Verfügung stellen.

6. Den meisten Sozialwissenschaftlern und Sozialwissenschaftlerinnen, die die gängigen multivariaten Verfahren (insbesondere die verschiedenen Arten der Regressionsanalyse) anwenden, dürfte bewusst sein, dass es sich dabei um *Modelle* handelt. Es kann also nicht ohne weiteres davon ausgegangen werden, dass die Struktur eines gegebenen Datensatzes den durch die Wahl eines bestimmten Modells vorgegebenen Annahmen entspricht. Mit Verfahren der Modelldiagnostik kann zumindest ansatzweise geprüft werden, ob das gewählte Modell mit den Daten einigermaßen kompatibel ist. Im OLS-Regressionsmodell wären dies etwa: Prüfung der Normalverteilung der Residuen, Kontrolle auf Vorliegen von Autokorrelation sowie Heteroskedastizität, Ermittlung von Multikollinearität und Prüfung ihrer möglichen Auswirkungen, Prüfung von partiellen Residuenplots auf Linearität der Zusammenhänge

7 Zu den wichtigsten Gründen zählt neben der zumeist fehlenden Berücksichtigung der Teststärke (und damit des Fehlers 2. Art), dass (a) die den statistischen Tests zugrunde liegende Nullhypothese (die „nil hypothesis“, so *Cohen* 1997, S. 27) im Allgemeinen unplausibel ist, so dass deren Ablehnung kaum Aussagekraft besitzt, und (b) das schematische Festhalten an einer vorgegebenen Irrtumswahrscheinlichkeit (meist 0,05) zu problematischen Entweder-Oder-Entscheidungen – nämlich „Beibehalten“ oder „Verwerfen“ von Hypothesen – führt.

und mögliche Ausreißer bzw. einflussreiche Fälle, Test auf solche Fälle anhand von Statistiken wie Leverage und Cook's D (siehe im Detail **Brüderl** 2000; **Fox** 1991, 1997).⁸ In nahezu allen Veröffentlichungen, die Ergebnisse einschlägiger Modellschätzungen berichten, wird jedoch nicht einmal angegeben, ob solche Verfahren angewandt wurden, geschweige denn, dass Näheres über deren Ergebnisse mitgeteilt würde.

7. Schließlich möchte ich auf den nicht unbedingt zentralen, aber auch nicht zu vernachlässigenden Aspekt der *Darstellung* von Daten und Analyseergebnissen eingehen. Nur zwei Beispiele: Es lassen sich (erstens) immer noch abschreckende Beispiele für *Präsentationsgraphiken* finden – und dies nicht nur, aber auch in Lehrbüchern zur Datenauswertung. Dass mir hier Beispiele aus der akademischen Forschungspraxis weniger geläufig sind, liegt vielleicht auch daran, dass Präsentationsgraphiken dort wenig eingesetzt werden – m. E. *zu wenig*, zumindest wenn man (wie ich) die Grenzen zwischen explorativer und Präsentationsgraphik weniger scharf sieht und berücksichtigt, dass sich z. B. bivariate Zusammenhänge ausgezeichnet zur graphischen Darstellung eignen. Was (zweitens) die *numerische* Darstellung deskriptiver Daten etwa in Kreuztabellen angeht,⁹ so möchte ich der Einfachheit halber auf den Beitrag von **Pötschke** und **Simonson** (etwa Tabelle 1, S. 79) zurückgreifen; die im folgenden angeführten Probleme ließen sich auch an vielen anderen Beispielen verdeutlichen. Einerseits wird in der genannten Tabelle das Prozentzeichen hinter jeder einzelnen Prozentzahl wiederholt (was nicht zwingend nötig scheint und die Tabelle etwas schwer zu lesen macht), andererseits findet sich kein Hinweis, welche der Prozentwerte Spalten- und welche Zeilenprozentwerte wiedergeben (was sich schnell herausfinden lässt, aber auch als Legende zur Tabelle angegeben werden könnte). Auch werden die Prozentwerte mit einer Nachkommastelle angegeben; da die Daten (angesichts der methodischen Vorgehensweise verständlicherweise) auf einer relativ kleinen Stichprobe mit nicht gut zu beurteilender Stichprobenqualität beruhen – wie die Autorinnen selbst zu Recht bemerken –, handelt es sich hier möglicherweise um Scheinexaktheit. Ferner scheint es nicht nötig, in allen Zellen zusätzlich zu den Prozentwerten die absoluten Zahlen auszuweisen; eine Beschränkung auf die Randverteilungen wäre hier ausreichend. Schließlich

8 Insbesondere Verfahren der Residuenanalyse (die eine erste Auskunft darüber geben, ob das gewählte Modell den Daten nicht allzu „fremd“ ist) sowie der Analyse einflussreicher Fälle liegen auch für andere häufig verwendete statistische Modelle vor, etwa die logistische Regression (**Pregibon** 1981) oder die Analyse von Verlaufsdaten (**Collett** 1994).

9 Siehe dazu auch **Ludwig-Mayerhofer** (1994).

könnte man darauf verzichten, sämtliche Zellen der Tabelle durch horizontale und vertikale Linien voneinander zu trennen.¹⁰

Abschließend möchte ich noch auf einige Probleme eingehen, die sich aus fehlenden oder unzulänglichen Darstellungen in den Lehrbüchern insbesondere im Bereich der statistischen Modellierung ergeben. Es stehen hier also weniger Qualitätsmängel der Forschungspraxis im Vordergrund, sondern die Tatsache, dass die genannten Fragen in den Lehrbüchern nicht ausreichend oder falsch erörtert, in wissenschaftlichen Publikationen jedoch meist schon vorausgesetzt werden, so dass es in der Lehre oftmals eines erheblichen Aufwandes bedarf, bis die angeführten Punkte wenigstens einem Teil der Studierenden hinlänglich verdeutlicht worden sind.

8. *Bildung und Interpretation von Dummy-Variablen*: Zumindest Studierenden (gelegentlich aber auch Forschungspraktikern) ist nicht ohne weiteres bewusst, wie sehr die auf Dummy-Variablen bezogenen Testergebnisse von der Wahl der Referenzkategorie abhängig sind. Eine ungeschickte Wahl der Referenzkategorie kann die inferenzstatistischen Schlussfolgerungen beeinträchtigen (ist die Referenzkategorie nur sehr schwach besetzt, führt dies zu hoher Multikollinearität unter den Dummy-Variablen und zu entsprechend hohen Standardfehlern). Auch muss immer wieder verdeutlicht werden, dass die Koeffizienten für die Dummy-Variablen nur mit Blick auf die Referenzkategorie Aussagekraft haben. Nicht zuletzt ist es wichtig zu betonen, dass ein Set von Dummy-Variablen Einfluss haben kann, auch wenn die einzelnen Koeffizienten nicht signifikant von null verschieden sind – nämlich dann, wenn sich die Schätzwerte für die Regressionskoeffizienten zwar nicht signifikant von null (d.h. der Referenzkategorie), aber untereinander unterscheiden.

9. *Nicht-lineare Zusammenhänge*: Einige nicht-lineare Zusammenhänge können relativ einfach durch polynomische Terme (meistens bereits durch Aufnahme eines quadrierten Terms) angenähert werden; ein typisches Beispiel ist der Befund, dass der Arbeitslohn nicht linear vom Alter oder der Betriebszugehörigkeitsdauer abhängt, da irgendwann Sättigungs- oder andere Effekte eintreten, die die Einkommenszuwächse verlangsamen oder sogar anhalten. Obwohl die grundlegenden mathematischen Zusammenhänge nicht gerade kompliziert sind, scheinen sie doch für viele Studierende eine große Herausforderung darzustellen. Hilfe in den Lehrbüchern ist jedoch nicht zu finden.

¹⁰ Insgesamt scheint sich die Tabelle stark am Vorbild der Outputs eines populären Datenanalysepaketes zu orientieren. Aber wenn sogar Lehrbücher solche Tabellen einfach unkommentiert und unkritisiert aus dem Programm-Output kopieren (siehe etwa *Atteslander* 2003, S. 349), ist es kein Wunder, wenn sich solche Praktiken auch bei anderen finden.

10. *Praxis der Modellierung*: Studierende, die sich bis zur multivariaten Modellierung vortasten, wählen immer wieder Verfahren der automatischen Variablenselektion wie etwa den schrittweisen Einschluss von Variablen. Automatisierte Verfahren mögen in Einzelfällen gerechtfertigt sein, wenn etwa die Wahl eines „besten Modells“ besonders schwierig ist – sofern im Anschluss eine sorgfältige Abwägung unter verschiedenen Modellschätzungen erfolgt. Verfahren des schrittweisen Einschlusses von Variablen gehören aber sicher nicht zu den besser geeigneten Selektionsalgorithmen; sie werden, soweit ersichtlich, vor allem deshalb gewählt, weil einschlägige Einführungen in den Umgang mit SPSS diese Verfahren erwähnen oder gar empfehlen. Bei der Anwendung des schrittweisen Einschlusses besteht die Gefahr, dass wichtige Prädiktoren irrtümlich nicht in das Modell aufgenommen werden; Gründe hierfür können sein, dass sie wegen (möglicherweise leicht zu behandelnder) Multikollinearität das den Einschlussalgorithmen zugrunde liegende Signifikanzkriterium nicht überschreiten, oder dass zwei Merkmale nur gemeinsam Einfluss haben, der Selektionsalgorithmus jedoch alle Variablen nur einzeln hinsichtlich ihres Erklärungsbeitrages prüft. Eine wesentlich bessere Modellierungsstrategie besteht m. E. darin, vom vollen Modell (dem Modell mit allen als hypothetisch relevant ausgewählten Variablen) auszugehen und nach Prüfung, ob die Anwendungsvoraussetzungen dieses Modells gegeben sind, Variablen ohne Erklärungsbeitrag (möglichst einzeln) aus dem Modell auszuschließen.¹¹

3 Einige abschließende Bemerkungen

Meine kurzen Anmerkungen sollten verdeutlichen, dass eine Ermittlung des Bedarfs an bestimmten Verfahren nur eine erste – gewiss sehr wichtige – Orientierung für eine Verbesserung der Methodenausbildung darstellt. Die *Qualität* der Ausbildung kann sich jedenfalls nicht nur daran messen lassen, dass die in der beruflichen Praxis erforderlichen Verfahren mehr oder weniger oberflächlich und routinemäßig beherrscht werden; zumindest für die *wissenschaftliche* Praxis kann das *nicht* ausreichen.¹² Wenn die Praxis sich als ungenügend erweist, ist eine Orientierung aus-

11 Die in sozialwissenschaftlichen Veröffentlichungen häufig zu findende Wiedergabe ausschließlich des vollen Modells ist ebenfalls problematisch, insbesondere wenn weitere Modellprüfungen unterbleiben. Irrelevante Prädiktoren können infolge von Multikollinearität die Standardfehler anderer, möglicherweise relevanter Prädiktoren in die Höhe treiben, so dass letztere als nicht signifikant ausgewiesen werden, obwohl sie nach Entfernung der irrelevanten Prädiktoren tatsächlich signifikant wären.

12 Selbstverständlich sollte – und dürfte – auch die Markt- und Meinungsforschung hohe Qualitätsstandards anstreben. Deren Schwerpunkte liegen aber möglicherweise anderswo, so wie nach den Ergebnissen von *Pötschke* und *Simonson* dort auch vielfach andere Datenanalysever-

schließlich an deren etablierten Verfahrensweisen für die Weiterentwicklung sozialwissenschaftlicher Ausbildung eher problematisch. Ich habe an einigen Beispielen aus dem Bereich der statistischen Datenanalyse zu verdeutlichen versucht, dass tatsächlich sowohl die Praxis – wie sie sich in sozialwissenschaftlichen Publikationen niederschlägt – als auch manche Lehrbücher den Studierenden (aber auch den bereits als Sozialforschern und Sozialforscherinnen Tätigen) nicht immer hilfreiche Anhaltspunkte, sondern manchmal zu wenig oder gar irreführende Anleitung bieten.¹³

Es sollte offenkundig sein, dass die genannten (und denkbare weitere) Probleme einen Anlass darstellen, die Ausbildung in Verfahren der statistischen Datenanalyse eher zu intensivieren – wie dies ja in letzter Zeit vielfach gefordert wurde (vgl. die meisten Beiträge in *Engel* 2002a). Abschließend sei jedoch darauf hingewiesen, dass dies natürlich auch für Verfahren qualitativer Datenanalyse gilt. Ein von der AG (inzwischen Sektion) Methoden der qualitativen Sozialforschung in der DGS eingesetzter Arbeitskreis hat bekanntlich Vorschläge für eine Methodenausbildung gemacht, die qualitativen Verfahren einen Platz einräumt, der dem der standardisierten Verfahren gleichwertig ist (AG Methoden der qualitativen Sozialforschung 2002). Diese Vorschläge wurden von den Vertretern der standardisierten Forschung (meist in der Methoden-Sektion der DGS versammelt) nicht zuletzt mit dem Hinweis auf den geringen Stellenwert abgewertet, den diese Verfahren in der Forschungspraxis haben. Auch hier liegt eine (erstaunliche) Gleichsetzung falsch verstandener Praxisrelevanz mit Qualität vor.¹⁴ Denn auch wenn qualitative Verfahren möglicherweise weniger zum Einsatz kommen als solche der standardisierten

fahren zum Einsatz kommen. Mangels ausreichender Bekanntschaft mit diesem Berufsfeld kann ich mich hierzu nicht kompetent äußern.

- 13 Das kann in Einzelfällen auch Bücher betreffen, die zu Recht weite Verbreitung und Anerkennung finden. Beispielsweise haben *Pötschke* und *Simonson* bei ihrer Inhaltsanalyse von Zeitschriftenaufsätzen die Intercoderreliabilität geprüft, dabei aber dem Lehrbuch von *Andreas Diekmann* (1995) eine Maßzahl entnommen, deren Problematik von ihnen selbst gesehen wurde (S. 80, Fn. 10). Ein besser geeignetes Maß – Kappa –, welches einen Korrekturfaktor für jene Übereinstimmungen enthält, die sich alleine aus den Randverteilungen ergeben, wurde bereits vor über 40 Jahren von *J. Cohen* (*Cohen* 1960) entwickelt, es ist seit langer Zeit in Übersichtsartikeln in deutscher (*Wegner* 1976) und englischer Sprache (*Tinsley* und *Weiss* 1975; *Bartko* und *Carpenter* 1976) ausführlich dargestellt und auch seit vielen Jahren etwa im Programm SPSS implementiert – aber in der deutschen Sozialforschung offensichtlich immer noch weitgehend unbekannt.
- 14 Aus wissenschaftssoziologischer Sicht müsste man natürlich noch weiter gehen und prüfen, ob die geringe Bedeutung qualitativer Verfahren unter den Forschungsberichten, die in den wichtigsten Zeitschriften publiziert werden, nicht Resultat von Prozessen der Bildung wissenschaftlicher Schulen und der Auseinandersetzung zwischen diesen ist. Ebenso könnte es sein, dass deshalb weniger Arbeiten publiziert werden, die auf qualitativen Verfahren beruhen, weil die Ausbildung in diesen Verfahren (und dementsprechend die Qualität der Ergebnisse) unzulänglich ist. Beide Möglichkeiten verweisen auf weitere Grenzen einer starken Orientierung an der bestehenden Forschungs- und Publikationspraxis als Maßstab.

Datenauswertung, so müssen diejenigen, die sie anwenden sollen, darin ebenso gut ausgebildet werden wie jene, die standardisierte Verfahren einsetzen.

Literatur

Arbeitsgruppe Methoden der qualitativen Sozialforschung 2002: Jahresbericht 2001/2002. In: Soziologie. Forum der Deutschen Gesellschaft für Soziologie 3/2002, S. 94-106.

Atteslander, Peter unter Mitarbeit von *Jürgen Cromm, Busso Grabow, Harald Klein, Andrea Maurer, Gabriele Siegert* 2003: Methoden der empirischen Sozialforschung. 10. neu bearbeitete und erweiterte Auflage. Berlin, New York: de Gruyter.

Bartko, J. J.; Carpenter, W. T. 1976: On the methods and theory of reliability. In: Journal of Nervous and Mental Disease 163, S. 307-317.

Brüderl, Josef 2000: Regressionsverfahren in der Bevölkerungswissenschaft, in: *Ulrich Mueller; Bernhard Nauck; Andreas Diekmann* (Hg.), Handbuch der Demographie 1. Modelle und Methoden, Berlin: Springer, S. 589-642.

Buchner, Axel; Erdfelder, Edgar; Faul, Franz 1996: Teststärkeanalysen, in: *Edgar Erdfelder; Rainer Mausfeld; Thorsten Meiser; Georg Rudinger* (Hg.), Handbuch Quantitative Methoden, Weinheim: Beltz – PsychologieVerlagsUnion, S. 123-136.

Cohen, Jacob 1960: A coefficient of agreement for nominal scales. In: Educational and Psychological Measurement 20, S. 37-46.

Cohen, Jacob 1988: Statistical power analysis for the behavioral sciences (2. Aufl.). Hillsdale: Erlbaum.

Cohen, Jacob 1997: The earth is round ($p < .05$), in: *Lisa L. Harlow; Stanley A. Mulaik; James H. Steiger* (Hg.), What if there were no significance tests?, Mahwah, New Jersey; London: Lawrence Erlbaum Associates, S. 21-35.

Collett, David 1994: Modelling Survival Data in Medical Research. London: Chapman & Hall.

Diekmann, Andreas 1995: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt.

Engel, Uwe (Hg.) 2002a: Praxisrelevanz der Methodenausbildung (Sozialwissenschaftliche Tagungsberichte Band 5), Bonn: Informationszentrum Sozialwissenschaften.

Engel, Uwe 2002b: Methodenlehre empirischer Sozialforschung. Bedarf und Reformperspektiven, in: *Uwe Engel* (Hg.), Praxisrelevanz der Methodenausbildung (Sozialwissenschaftliche Tagungsberichte Band 5), Bonn: Informationszentrum Sozialwissenschaften, S. 19-34.

Erickson, Bonnie H.; Nosanchuk, Terry A. 1992: Understanding Data. Maidenhead, Philadelphia: Open University Press.

Fox, John 1991: Regression Diagnostics (Series: Quantitative Applications in the Social Sciences). London: Sage.

Fox, John 1997: Applied Regression Analysis, Linear Models, and Related Methods. Newbury Park: Sage.

Guttman, Louis 1977: What is not what in statistics. In: The Statistician 26, S. 81-107.

Harlow, Lisa L.; Mulaik, Stanley A.; Steiger, James H. (Hg.) 1997: What If There Were No Significance Tests? Mahwah, NJ: Erlbaum.

Heckman, James J. 1979: Sample selection bias as a specification error. In: Econometrica 47, S. 153-161.

Kish, Leslie 1995: Methods for design effects. In: Journal of Official Statistics 11, S. 55-77.

- Kühnel, Steffen; Krebs, Dagmar** 2001: Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt.
- Ludwig-Mayerhofer, Wolfgang** 1994: Kleine Anmerkung, die Verbesserung der Darstellung von Kreuztabellen betreffend. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 46, S. 12-129.
- Marsh, Catherine** 1988: Exploring Data. Cambridge: Polity Press.
- Morrison, Denton E.; Henkel, Ramon E.** (Hg.) 1970: The Significance Test Controversy, Chicago: Aldine.
- Pötschke, Manuela; Simonson, Julia** 2003: Konträr und ungenügend? Ansprüche an Inhalte und Qualität einer sozialwissenschaftlichen Methodenausbildung. In: ZA-Information 52, S. 72-92.
- Pregibon, D.** 1981: Logistic regression diagnostics. In: The Annals of Statistics 9, S. 705-724.
- Schnell, Rainer** 1994: Graphisch gestützte Datenanalyse. München, Wien: Oldenbourg.
- Schnell, Rainer** 2000: Less frequently asked questions: Nutzen und Notwendigkeit grafisch gestützter Datenanalyse. In: Österreichische Zeitschrift für Soziologie 25, S. 5-25.
- Schnell, Rainer** 2002: Ausmaß und Ursachen des Mangels an quantitativ qualifizierten Absolventen sozialwissenschaftlicher Studiengänge, in: **Uwe Engel** (Hg.), Praxisrelevanz der Methodenausbildung (Sozialwissenschaftliche Tagungsberichte Band 5), Bonn: Informationszentrum Sozialwissenschaften, S. 35-44.
- Simonson, Julia; Pötschke, Manuela** 2002a: Zur Praxisrelevanz der Methodenausbildung: Ergebnisse einer Online-Befragung, in: **Uwe Engel** (Hg.), Praxisrelevanz der Methodenausbildung (Sozialwissenschaftliche Tagungsberichte Band 5), Bonn: Informationszentrum Sozialwissenschaften, S. 131-158.
- Simonson, Julia; Pötschke, Manuela** 2002b: Zur Bedeutung empirischer Methoden im Spiegel ausgewählter soziologischer Fachzeitschriften, in: **Uwe Engel** (Hg.), Praxisrelevanz der Methodenausbildung (Sozialwissenschaftliche Tagungsberichte Band 5), Bonn: Informationszentrum Sozialwissenschaften, S. 171-182.
- Stolzenberg, Ross M.; Relles, Daniel A.** 1997: Tools for intuition about sample selection bias and its corrections. In: American Sociological Review, 62, S. 494-507.
- Tinsley, H. E. A.; Weiss, D. J.** 1975: Interrater reliability and agreement of subjective judgments. In: Journal of Counseling Psychology 22, S. 358-376.
- Tukey, John W.** 1977: Exploratory data analysis. Reading, MA: Addison & Wesley.
- Wegner, Reinhard** 1976: Ratingmethoden, in: **Jürgen van Koolwijk; Maria Wieken-Mayser** (Hg.), Techniken der empirischen Sozialforschung, München: Oldenbourg, S. 103-130.