

1 Einführung: Sozialwissenschaftliche Datenanalyse

Henning Best und Christof Wolf

^a Universität Mannheim

^b GESIS – Leibniz-Institut für Sozialwissenschaften und Universität Mannheim

1 Moderne Sozialwissenschaft und die Bedeutung der Statistik

Sozialwissenschaft, wie wir sie heute kennen, ist vor allem durch ihre starke empirische Ausrichtung gekennzeichnet. Bereits 1967 bemerkte René König, dass Wissenschaft, und damit auch die Soziologie, „letztlich nur als empirische Forschung möglich“ ist (König 1967, S. 3). Mittlerweile hat sich diese Einsicht durchgesetzt und die quantitativ-empirische Forschung ist zum Standard in den Sozialwissenschaften geworden. Fleck (2010) untersucht die Verbreitung der empirischen Forschung seit Ende des zweiten Weltkrieges am Beispiel von Veröffentlichungen in der *American Sociological Review* (*ASR*) und der *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (*KZfSS*) und berichtet, dass seit den 1990er Jahren ca. 90 % der Beiträge auf – vornehmlich quantitativer – empirischer Forschung beruhen. Für das Jahr 2008 findet er sogar 100 % empirische Beiträge in der *Kölner Zeitschrift* und immerhin 98 % in der *ASR*. Um moderne sozialwissenschaftliche Forschung zu verstehen und betreiben zu können, sind folglich hinreichende Kenntnisse in quantitativen Methoden und modernen statistischen Verfahren notwendig.

Vor diesem Hintergrund ist es erfreulich, dass in der universitären Lehre Methoden und Statistik einen hohen Stellenwert haben. Die Deutsche Gesellschaft für Soziologie empfiehlt in ihrem Standardcurriculum für Bachelor-Studiengänge Kurse zu Statistik. In nahezu allen sozialwissenschaftlichen Studiengängen ist zumindest ein Einführungskurs in Statistik verpflichtend, und viele Fakultäten bieten fortgeschrittene Veranstaltungen in multivariater Statistik an.

Die beschriebene Entwicklung hat auch dazu geführt, dass – international und auf dem deutschsprachigen Markt – eine große Zahl an Statistikbüchern erhältlich ist. Bei den meisten deutschsprachigen Veröffentlichungen zur sozialwissenschaftlichen Statistik handelt es sich jedoch um Lehrbücher für den Einstieg. So steht mittlerweile eine große Auswahl an grundständigen Einführungen in die Statistik zur Verfügung (vgl. z. B. Gehring & Weins 2009; Benninghaus 2007; Kühnel & Krebs 2007; Bortz 2005; Fahrmeir et al. 2009), und jede Dozentin oder jeder Student wird ein entsprechendes Einführungswerk finden, das seinem oder ihrem Geschmack entspricht. In Bezug auf fortgeschrittene multivariate Verfahren ist die Situation hingegen gänzlich anders, zumindest im deutschsprachigen Raum. Zwar gibt es ausreichend Monografien zu einzelnen Verfahren, etwa zur Regressionsanalyse, zur Analyse kategorialer Daten, zur Mehrebenenanalyse, zu Strukturgleichungsmodellen und vielem mehr (vgl. z. B. Urban

& Mayerl 2006; Andreß et al. 1997; Langer 2004; Reinecke 2005). Auch wenn sich diese Bücher mitunter hervorragend eignen, um sich intensiv mit einem bestimmten Verfahren zu beschäftigen, sind sie oftmals weniger dienlich, wenn man einen Überblick erhalten möchte, einen ersten, kurzen Einstieg in eine Methode sucht oder sich über mehrere Verfahren im Vergleich informieren möchte. Hierfür wären insbesondere Überblickswerke zu multivariaten Analyseverfahren geeignet – je nach Zweck Nachschlagewerke, Handbücher, praxisorientierte Darstellungen oder fortgeschrittene Lehrbücher. Diese liegen jedoch deutschsprachig kaum vor. Wichtige Ausnahmen sind insbesondere das von Andreas Diekmann herausgegebene Sonderheft der *KZfSS* „Methoden der Sozialforschung“ (Diekmann 2006), das sich allerdings vornehmlich mit neueren Entwicklungen beschäftigt und daher grundlegende, häufig verwendete Verfahren ausklammert, ein kleiner aber dennoch sehr nützlicher Sammelband zu „Methoden der Politikwissenschaft“ von Behnke et al. (2006) und schließlich das aus der Marketingforschung stammende Lehrbuch von Backhaus et al. (2008), dessen starker Anwendungsbezug („Kochbuch“) es bei Studierenden zwar beliebt macht, das jedoch keine ausreichende Darstellung der jeweiligen mathematisch-statistischen Grundlagen bietet. All diese Bücher sind zudem, was die Zahl der dargestellten Verfahren angeht, weniger umfangreich als das vorliegende Handbuch.

Angesichts der in den letzten Jahren und Jahrzehnten immer weiter gestiegenen Bedeutung der empirischen Sozialforschung, und damit auch der multivariaten Datenanalyse, ist der konstatierte Mangel an übergreifenden Darstellungen verschiedener, häufig verwendeter multivariater Analysemethoden eklatant. Wir hoffen, dass dieser Mangel mit dem hier vorgelegten Handbuch zumindest teilweise beseitigt wird.

2 Ziele und Zielgruppe des Handbuchs

Das *Handbuch der sozialwissenschaftlichen Datenanalyse* verfolgt mehrere Ziele und spricht folglich unterschiedliche Zielgruppen an: Einerseits dient es als Überblicks- und Nachschlagewerk für eine große Anzahl von multivariaten Analyseverfahren und ist insofern für alle Anwender empirischer Methoden von großem Interesse. Sie finden jedes Verfahren ausführlich diskutiert und mit seinen mathematisch-statistischen Grundlagen dargestellt. Andererseits bietet das Handbuch durch den Aufbau der Beiträge auch die Möglichkeit, Analyseverfahren neu kennen zu lernen und einen Einstieg in die multivariate Datenanalyse zu finden. Als Zielgruppe kommen insofern auch fortgeschrittene Studierende der Sozialwissenschaften, Doktoranden und Mitarbeiter in empirischen Forschungsprojekten in Betracht.

Um diesen Zielen – Einführungs-, Überblicks- und Nachschlagewerk – gleichzeitig gerecht werden zu können, folgen die Beiträge des Handbuchs, mit gegebenenfalls notwendigen Anpassungen, einer vorgegebenen Gliederung: Jedes Verfahren wird zunächst allgemeinverständlich beschrieben, bevor in einem zweiten Schritt eine strenger formalisierte mathematische Darstellung erfolgt und fortgeschrittene Aspekte des Analyseverfahrens dargestellt werden. Ein besonderer Schwerpunkt liegt, drittens, auf einer anwendungsorientierten Darstellung des Verfahrens anhand von Beispielanalysen.

Soweit möglich basieren die Beispielanalysen auf dem kumulierten ALLBUS 1980–2006¹ oder dem SOEP². Um eine Replikation der Beispiele zu ermöglichen und die praktische Anwendung der Verfahren zu erleichtern, dokumentieren wir die Analysesyntax (meist als Stata- oder SPSS-Code) auf der begleitenden Homepage des Handbuchs unter <http://www.handbuch-datenanalyse.de>. Die Beiträge enden mit einem kurzen Abschnitt zu „häufigen Fehlern“ und kommentierten Literaturempfehlungen.

Durch diesen Aufbau hoffen wir, den vielfältigen und hohen Ansprüchen an dieses Handbuch gerecht werden zu können: Während die erfahrene Wissenschaftlerin, die etwas nachschlagen möchte, insbesondere die weiterführenden Abschnitte der Beiträge interessieren werden, wird der Doktorand, der den ersten Kontakt zu einem Verfahren sucht, sicherlich mit dem allgemeinverständlichen, einführenden Teil beginnen, den formalisierten Abschnitt u. U. zunächst überblättern und mit dem Beispiel und „häufigen Fehlern“ fortfahren. Selbstverständlich kann die letztere Vorgehensweise nicht die intensive Beschäftigung mit einer Analyseverfahren ersetzen; sie kann aber den Einstieg erleichtern und in der Folge eine Vertiefung stimulieren.

Eine angemessene Struktur garantiert jedoch noch nicht die Qualität der Beiträge. Diese hängt vielmehr vom Kenntnisstand und der Erfahrung der Autorinnen und Autoren ab. Daher ist es für uns besonders erfreulich, dass es gelungen ist, Kolleginnen und Kollegen aus ganz verschiedenen sozialwissenschaftlichen Disziplinen zur Mitarbeit zu gewinnen, die in ihren Fächern, teilweise weit über die Fachgrenzen hinaus, zu anerkannten Vertretern der Datenanalyse gehören. Zur Qualität der Beiträge hat schließlich auch ein umfassender Begutachtungsprozess beigetragen.

Die Expertise der Autorinnen und Autoren und die didaktisch motivierte Struktur der einzelnen Kapitel sollte das Handbuch für eine breite Zielgruppe zu einem interessanten und nützlichen Begleiter bei der Datenanalyse werden lassen.

3 Themenauswahl

Aus den Zielen des Handbuches folgt unmittelbar die Frage nach der Auswahl der Analyseverfahren, die behandelt werden sollen. So soll ein Handbuch, wie beschrieben, einen Überblick über das behandelte Feld geben. Eine vollständige Darstellung ist jedoch aufgrund der Vielzahl verschiedener Verfahren leider nicht möglich, zumindest, wenn man sich auf *ein* Buch beschränken möchte (das Handbook of Econometrics beispielsweise hat sechs Bände in sieben Büchern). In unserer Auswahl haben wir versucht, einerseits die ganze Breite quantitativer Methoden abzudecken, uns andererseits aber auf häufig genutzte und damit (in einem praktischen Sinne) wichtige Verfahren zu konzentrieren. Da wir außerdem der Auffassung sind, dass empirische Sozialforschung theorieprüfend und, zumindest perspektivisch, kausalanalytisch sein sollte, freuen wir uns besonders, dass Karl-Dieter Opp diesen Band mit einem wissenschaftstheoretischen Beitrag zu Kausalität und multivariater Statistik eröffnet. Hierdurch machen wir deutlich, dass Sozialforschung keine reine Deskription sozialer Tatbestände oder, in Adornos

¹ Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ZA-Nr. 4241), siehe <http://www.gesis.org/allbus>

² Sozio-oekonomisches Panel, siehe <http://www.diw.de/soep>

Worten, eine „verdinglichte Apperzeption des Dinghaften“ (Adorno 1969, S. 89) sein kann, sondern vielmehr einer theoretischen und epistemologischen Grundlage bedarf.

Obwohl wir mit 40 Kapiteln auf ca. 1100 Seiten ein sehr umfassendes Werk vorlegen, wird sicher der eine oder die andere ein besonders wichtiges Lieblingsverfahren vermissen. Auch wir hätten gerne noch weitere Themen in das Handbuch aufgenommen. Nichts desto trotz gehen wir davon aus, dass wir den Bedarf richtig eingeschätzt und die wichtigsten Methoden abgedeckt haben.

4 Gliederung und Inhalt

Das Handbuch beginnt mit einem Beitrag zu wissenschaftstheoretischen Grundlagen der Datenanalyse und der Verbindung von Kausalanalyse und multivariater Statistik. Der zweite Teil versammelt Aufsätze zu grundlegenden datenanalytischen Verfahren wie bivariater Statistik und explorativen grafischen Verfahren, statistischer Inferenz und Schätzverfahren. Behandelt werden auch Methoden zum Umgang mit fehlenden Werten und Verfahren zur Gewichtung. Hierauf folgen Beiträge zu Skalierungsverfahren, beginnend mit einer Diskussion grundlegender Gütekriterien und klassischer Methoden wie Likert-, Guttman- und Mokkenskalisierung oder der Hauptkomponentenanalyse. Darauf aufbauend werden modernere Ansätze präsentiert (Item Response Theorie, Multidimensionale Skalierung und Korrespondenzanalyse). Im dritten Teil wird die statistische Analyse von Häufigkeiten und Gruppen behandelt. Dazu werden Verfahren wie Diskriminanz- und Varianzanalyse, log-lineare Modelle, die Analyse latenter Klassen oder die Netzwerkanalyse dargestellt. Mit über 500 Seiten liegt der Schwerpunkt des Handbuchs jedoch auf den beiden folgenden Blöcken zur multivariaten (Regressions-)Analyse von Quer- und Längsschnittdaten. Zunächst werden hier in mehreren Beiträgen ausführlich verschiedene Aspekte der linearen Regression behandelt. Hierauf aufbauend werden dann fortgeschrittenere Verfahren wie logistische Regression, Strukturgleichungsmodelle, Mehrebenenanalyse oder Propensity Score Matching diskutiert. Da zeitbezogenen Daten eine wachsende Bedeutung zukommt, schließen wir den Band mit Beiträgen zur Ereignisdatenanalyse, Panelregression, Zeitreihen- und Sequenzanalyse sowie weiteren Verfahren.

5 Ausblick

Wir hoffen, mit diesem Handbuch einen guten Überblick über statistische Analyseverfahren zu geben und einen Beitrag dazu zu leisten, erstens die Methodenausbildung im deutschsprachigen Raum zu erleichtern und zweitens die informierte, kenntnisreiche und korrekte Anwendung komplexer statistischer Methoden zu fördern.

Hierfür möchten wir jedoch um Ihre Unterstützung bitten: Trotz aller Mühe, die wir uns mit dem Satz und der Redaktion dieses Handbuches gegeben haben, ist leider damit zu rechnen, dass sowohl der Text als auch die Gleichungen einzelne Fehler enthalten. Diese Errata möchten wir auf der Internetseite des Handbuchs (<http://www.handbuch-datenanalyse.de>) dokumentieren bzw. korrigieren. Sollte

Ihnen ein Fehler auffallen, der noch nicht dokumentiert ist, teilen Sie uns dies bitte über das Kontaktformular auf der Homepage mit.

Literaturverzeichnis

- Adorno, T. W. (1969). Soziologie und empirische Forschung. In T. W. Adorno, R. Dahrendorf, H. Pilot, H. Albert, J. Habermas, & K. R. Popper (Hg.), *Der Positivismusstreit in der deutschen Soziologie* (S. 81–101). Darmstadt: Luchterhand.
- Andreß, H.-J., Hagenars, J. A., & Kühnel, S. (1997). *Analyse von Tabellen und kategorialen Daten. Log-lineare Modelle, latente Klassenanalyse, logistische Regression und GSK-Ansatz*. Berlin: Springer.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2008). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Berlin: Springer, 12. Auflage.
- Behnke, J., Gschwend, T., Schindler, D., & Schnapp, K.-U. (2006). *Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren*. Baden-Baden: Nomos.
- Benninghaus, H. (2007). *Deskriptive Statistik. Eine Einführung für Sozialwissenschaftler*. Wiesbaden: VS Verlag für Sozialwissenschaften, 11. Auflage.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer Medizin Verlag, 6. Auflage.
- Diekmann, A., Hg. (2006). *Methoden der Sozialforschung*. Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 44. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fahrmeir, L., Künstler, R., Pigeot, I., & Tutz, G. (2009). *Statistik: Der Weg zur Datenanalyse*. Berlin: Springer, 7. Auflage.
- Fleck, C. (2010). 60 Jahre Empirische Sozialforschung in vergleichender Perspektive. In F. Faulbaum & C. Wolf (Hg.), *Gesellschaftliche Entwicklungen im Spiegel der empirischen Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gehring, U. & Weins, C. (2009). *Grundkurs Statistik für Politologen und Soziologen*. Wiesbaden: VS Verlag für Sozialwissenschaften, 5. Auflage.
- König, R., Hg. (1967). *Handbuch der empirischen Sozialforschung*, Band 1 und 2. Stuttgart: Enke.
- Kühnel, S.-M. & Krebs, D. (2007). *Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen*. Reinbek: Rowohlt, 4. Auflage.
- Langer, W. (2004). *Mehrebenenanalyse. Eine Einführung für Forschung und Praxis*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg Verlag.
- Urban, D. & Mayerl, J. (2006). *Regressionsanalyse: Theorie, Technik und Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften, 2. Auflage.