

Wie kommt man von den Ergebnissen der Faktorenanalyse zu Dimensionsvariablen?

Eine Einführung in die Dimensionsbildung
mit SPSS für Windows

2., korrigierte Auflage

Nina Baur

Bamberger Beiträge zur empirischen Sozialforschung

Herausgeber:

Gerhard Schulze und Nina Baur

Nr. 13, 2003

ISSN 1611-583X

Bamberger Beiträge zur empirischen Sozialforschung

1	Einführung in die Methoden der empirischen Sozialforschung	Gerhard Schulze	2002
2	Einführung in die Methoden der empirischen Sozialforschung. Übungsaufgaben und Lösungen	Gerhard Schulze	2002
3	Die biographische Methode. Ein Verfahren zur qualitativen Analyse individueller Verlaufsmuster in den Sozialwissenschaften	Nina Baur	2003
4-1	Sozialwissenschaftliche Datenanalyse am PC für Fortgeschrittene. Ein Arbeitsbuch mit SPSS für Windows	Nina Baur	2003
4-2	Sozialwissenschaftliche Datenanalyse am PC für Fortgeschrittene. Daten-CD	Nina Baur	In Vorb.
5	Takeoff der Auswertung. Zur Vorbereitung statistischer Analysen	Nina Baur	2003
6	Typenbildung, Umgang mit fehlenden Werten und Bilden neuer Variablen. Eine Einführung in die Datenmodifikation und -selektion in SPSS für Win- dows	Sabine Fromm	In Vorb.
7	Datenbereinigung und Matchen. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	In Vorb.
8	Univariate Statistik. Eine Einführung in den Umgang mit eindimensionalen Häufigkeitsvertei- lungen mit SPSS für Windows	Nina Baur	2003
9	Bivariate Statistik, Drittvariablenkontrolle und das Ordinalskalenproblem. Eine Einführung in die Kausalanalyse und in den Umgang mit zweidimen- sionalen Häufigkeitsverteilungen mit SPSS für Windows (2., korrigier- te Auflage)	Nina Baur	2003
10	Multiple lineare Regressionsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	2003
11	Logistische Regressionsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	In Vorb.
12	Faktorenanalyse und Reliabilitätsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	In Vorb.
13	Wie kommt man von den Ergebnissen der Faktorenanalyse zu Dimensi- onsvariablen? Eine Einführung in die Dimensionsbildung mit SPSS für Windows	Nina Baur	2003
14	Diskriminanzanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	In Vorb.
15	Clusteranalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	In Vorb.
16	Korrespondenzanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Sabine Fromm	In Vorb.
17	Wissenschaftstheorie für Sozialwissenschaftler	Gerhard Schulze	In Vorb.
18	Mittelwertvergleiche, sequentielle Gruppenaufteilung und Varianzanalyse Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Nina Baur	In Vorb.
	Quantitative Analyse zeitlicher Veränderung		
19-1	Band 1: Überblick und theoretische Grundlage	Nina Baur (Hg.)	In Vorb.
19-2	Band 2: Panelanalyse	Nina Baur (Hg.)	In Vorb.
19-3	Band 3: Kohortenanalyse	Nina Baur (Hg.)	In Vorb.
19-4	Band 4: Zeitreihenanalyse	Nina Baur (Hg.)	In Vorb.
19-5	Band 5: Ereignisanalyse	Nina Baur (Hg.)	In Vorb.
20	Präsentation statistischer Daten Eine Einführung für Sozialwissenschaftler mit SPSS für Windows	Simone Zdro- jewski / Jan D. Engelhardt	2003
21	Datenaufbereitung. Arbeitsschritte zwischen Erhebung und Auswertung quantitativer Daten	Detlev Lück	2003

ISSN: 1611-583X

Herausgegeben von Gerhard Schulze und Nina Baur
 Professur für Methoden der empirischen Sozialforschung
 Otto-Friedrich-Universität Bamberg
 96045 Bamberg
<http://www.uni-bamberg.de/sowi/empirie/Download/bbzesf.html>

Inhalt

Einleitung.....	4
Bildung einer Dimensionsvariable.....	7
Dimensionen in der sozialen Realität und Variablen im Fragebogen	7
Verstehen / Erfassen der dimensionalen Struktur von Variablen.....	8
Schritt 1: Bestimmung der dimensionalen Struktur der Variablen.....	9
Schritt 2: Skalenbildung	14
Vorgehen in SPSS am Beispiel der Dimension „Einkaufsverhalten in der Nachbarschaft“	19

EINLEITUNG

In vielen sozialwissenschaftlichen Studiengängen erwerben Studierende in den ersten Studiensemestern Kenntnisse in Wissenschaftstheorie, Methoden der empirischen Sozialforschung, in Statistik, in soziologischer Theorie und in speziellen Soziologien. Parallel dazu eignen sie sich den Umgang mit diversen Programmpaketen an. Typischerweise sind die nächsten Fragen: Was mache ich nun mit diesen Kenntnissen? Wie verbinde ich sie? Diese Wissensbereiche sind totes Wissen, solange es nicht gelingt, sie im praktischen Forschungsprozess zu verknüpfen.

Genau diese Verknüpfung ist das Ziel des Kurses „Datenanalyse am PC für Fortgeschrittene“. Dieser Kurs ist als einsemestriger Kurs konzipiert. Die Studierenden lösen jede Woche Aufgaben. Dabei vertiefen und erweitern sie parallel die bereits vorhandenen Methoden-, Statistik- und Theoriekenntnisse. Sie bearbeiten konkrete Fragestellungen mit einem Datensatz. Das Endziel ist, dass die Kursteilnehmer multiple lineare Regressionsanalyse und Faktorenanalysen selbständig in SPSS durchführen und interpretieren können. Der Kurs setzt also bereits verschiedene Grundkenntnisse voraus. Die im Anhang an dieses Kapitel aufgelistete Literatur ist geeignet, etwaige Lücken in den Bereichen Wissenschaftstheorie, Methoden und Dimensionsanalyse zu schließen. Eine ausführliche Literaturliste finden Sie im Band 4-1 dieser Reihe. Aufbauend auf diesen Kenntnissen besteht der Kurs aus folgenden Materialien:

BBS Nr.	Autor	Titel	Inhalt
4-1	Nina Baur	<i>Sozialwissenschaftliche Datenanalyse am PC für Fortgeschrittene. Ein Arbeitsbuch mit SPSS für Windows</i>	Vorschlag für einen Arbeitsplan und Aufgabensammlung
4-2	Nina Baur	<i>Sozialwissenschaftliche Datenanalyse am PC für Fortgeschrittene. Daten-CD</i>	Datensatz
5	Nina Baur	<i>Takeoff der Auswertung</i>	Dokumentation des Datensatzes; Die Rolle von SPSS im Forschungsprozess; Vom Fragebogen zum Rohdatensatz zu statistischen Maßzahlen; Datenvor- und -aufbereitung; Umgang mit statistischen Daten; Nützliche Software; Fundorte für Daten; Allgemeine Hinweise zur Datenanalyse am PC
21	Detlev Lück	<i>Datenaufbereitung. Arbeitsschritte zwischen Erhebung und Auswertung quantitativer Daten</i>	Detaillierte Darstellung, wie man Daten für die statistische Analyse aufbereitet
6	Nina Baur	<i>Univariate Statistik</i>	Explorative Datenanalyse; Analyse von eindimensionalen Häufigkeitsverteilungen
9	Nina Baur	<i>Bivariate Statistik, Drittvariablenkontrolle und das Ordinalskalensproblem</i>	Einführung in die Kausalanalyse und in den Umgang mit zweidimensionalen Häufigkeitsverteilungen; Umgang mit ordinalskalierten Daten
10	Sabine Fromm	<i>Multiple lineare Regressionsanalyse</i>	Multiple lineare Regressionsanalyse als Beispiel für kausalanalytische Verfahren
12	Sabine Fromm	<i>Faktorenanalyse und Reliabilitätsanalyse</i>	Faktorenanalyse als Beispiel für dimensionsanalytische Verfahren: Erkennen der Zahl der Dimensionen in einem Datensatz; Entscheidung, welche Variablen zu einer Dimension gehören. Dimensionsbildung
13	Nina Baur	<i>Wie kommt man von den Ergebnissen der Faktorenanalyse zu Dimensionsvariablen?</i>	
20	Simone Zdrojewski / Jan D. Engelhardt	<i>Präsentation statistischer Daten. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows</i>	Was muss man bei der Präsentation statistischer Daten beachten? Was muss man bei der Gestaltung von Tabellen und Grafiken beachten?

Dieser Band (Band 13) baut unmittelbar auf Band 12 auf. Am Beispiel der Faktorenanalyse geht er folgenden Fragen nach:

- Welche Analyseziele verfolgt die Dimensionsanalyse?
- Wie erkennt man eine Dimension in den Daten?
- Welche Variablen messen die Dimension?
- Welche Ausprägung hat ein Individuum auf der Dimension?

Einführungstexte: Wissenschaftstheorie

Schulze (2002a) und Behnke und Behnke (2003a) richten sich an Studienanfänger und andere Personen, die sich noch nie mit Wissenschaftstheorie beschäftigt haben. Weiterführende Darstellungen zum Verhältnis von Wissenschaftstheorie und Soziologie, zur Soziologie der Forschung und dazu, wie man als Soziologe aus Daten Theorien konstruiert, finden Sie in den Darstellungen von Gerhard Schulze, die nach und nach in dieser Reihe erscheinen werden.

Behnke, Joachim / Behnke, Nathalie (2003a): **Kurs zur Wissenschaftstheorie**. Kurs Nr. 33206 der Fernuniversität Hagen

Schulze, Gerhard (2002a): **Einführung in die Methoden der empirischen Sozialforschung**. Bamberger Beiträge zur empirischen Sozialforschung. Band 1. Bamberg

Einführungstexte: Methoden der empirischen Sozialforschung

Baur, Behnke und Behnke (2003) sowie Schulze (2002a) führen in die empirische Sozialforschung ein. Sie erklären das nötige Basiswissen, um eigene Forschungsprojekte durchzuführen. Stattdessen können auch andere Einführungen in die Methoden

der empirischen Sozialforschung und Wissenschaftstheorie gelesen werden. Aufbauend auf diesen Kenntnissen, erläutert Baur (2003a), wie man von soziologischen Fragen mit Hilfe standardisierter Daten zu theoretischen Schlüssen kommt.

Atteslander, Peter u.a. (2000): **Methoden der empirischen Sozialforschung**. Berlin: De Gruyter

Baur, Nina / Behnke, Joachim / Behnke, Nathalie (2003): **Empirische Methoden der Politikwissenschaft**. Paderborn: Ferdinand Schöningh

Baur, Nina (2003a): **Takeoff der Auswertung. Zur Vorbereitung statistischer Analysen**. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Band 5

Diekmann, Andreas (2000): **Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen**. Reinbek: Rowohlt

Esser, Hartmut / Hill, Paul B. / Schnell, Rainer (1999): **Methoden der empirischen Sozialforschung**. München: Oldenbourg

Kromrey, Helmut (2000): **Empirische Sozialforschung**. Stuttgart: UTB

Schulze, Gerhard (2002a): **Einführung in die Methoden der empirischen Sozialforschung**. Bamberger Beiträge zur empirischen Sozialforschung. Band 1. Bamberg

Einführungstexte: SPSS für Windows

Angele (2003) und Wittenberg / Cramer (2000) schreiben für Personen, die noch nie mit Statistik-Programmpaketen zu tun hatten. Angele (2003) konzentriert sich auf die wichtigsten Konzepte und gibt einen schnellen Überblick. Wittenberg / Cramer (2000) stellen einzelne Aspekte von SPSS ausführlicher dar. Sie erklären Dinge, die in anderen Büchern nicht erklärt werden. Baur (2003a) erklärt, wie die Daten in den Datensatz kommen und wie man vom Rohdatensatz zu statistischen Maßzahlen kommt. Fromm (2003a,2003b) erläutert, wie man den Datensatz so aufbereitet, dass man die gewünschten statistischen Maße berechnen kann. Backhaus u.a. (2000) geben einen Überblick über eine große Bandbreite multivariater Analyseverfahren mit SPSS. Sie beschränken sich dabei auf die Syntax-Befehle. Jeder Autor erklärt auf knapp 50 Seiten die Grundlagen eines statistischen Verfahrens und wie man es mit SPSS umsetzt. Brosius (2002) ist dagegen für diejenigen geeignet, die das Menü bevorzugen. In dieser Reihe erscheinen außerdem nach und nach Texte, die in multivariate Analyseverfahren mit SPSS einführen.

Angele, German (2003): **SPSS 11.5 für Windows. Eine Einführung**. Bamberg: Schriftenreihe des Rechenzentrums der Otto-Friedrich-Universität Bamberg. <http://www.uni-bamberg.de/urz/netze/spss/skript/index.htm>

Baur, Nina (2003a): **Takeoff der Auswertung. Zur Vorbereitung statistischer Analysen**. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Band 5

Backhaus, Klaus / Erichson, Bernd / Plinke, Wulff / Weiber, Rolf (Hg.) (2000): **Multivariate Analysemethoden. Eine anwendungsorientierte Einführung**. 9., überarbeitete und erweiterte Auflage. Berlin / Heidelberg / New York u.a.: Springer

Brosius, Felix (2002): **SPSS 11. Professionelle Statistik unter Windows**. Bonn: MITP-Verlag

Fromm, Sabine (2003a): **Typenbildung, Umgang mit fehlenden Werten und Bilden neuer Variablen. Eine Einführung in die Datenmodifikation und -selektion in SPSS für Windows**. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Band 6. In Vorbereitung

Fromm, Sabine (2003b): **Datenbereinigung und Matchen. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows**. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Band 7. In Vorbereitung

Reihe: **Bamberger Beiträge zur empirischen Sozialforschung**. ISSN: 1611-583X. Herausgeber: Gerhard Schulze und Nina Baur

SPSS Inc. (2001): **SPSS 11.0 Syntax Reference Guide for SPSS Base, SPSS Regression Models, SPSS Advanced Models**. [Den Syntax-Guide finden Sie im Menü „Hilfe“ von SPSS. Dort sind alle Befehle, über die SPSS verfügt, aufgeführt und erklärt.]

Wittenberg, Reinhard / Cramer, Hans (2000): **Datenanalyse mit SPSS für Windows**. 2., neubearbeitete Auflage. Reihe: Handbuch für computergestützte Datenanalyse. Band IX. Herausgegeben von Reinhard Wittenberg. Stuttgart: Lucius & Lucius

Einführungstexte: Dimensionsanalyse

Die Faktorenanalyse ist ein Verfahren zur Dimensionsanalyse. Lesen Sie das Wort „Faktor“ als Synonym für „Dimension“ bzw. „Disposition“ im Sinne von Schulze (2002a). Wiederholen Sie deshalb auch das Kapitel „Dimensionsanalyse“ in Schulze (2002a). Beginnen Sie mit Fromm (2003d) und Schulze (2002c; 2002d). Ziehen Sie die anderen Texte hinzu, wenn Sie mit diesen Texten Schwierigkeiten haben.

Fromm, Sabine (2003d): **Faktorenanalyse und Reliabilitätsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows**. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Band 12. In Vorbereitung

Schulze, Gerhard (2002c): **Faktorenanalyse in Grundzügen**. Paper zum soziologischen Forschungspraktikum 2002/2003 an der Otto-Friedrich-Universität Bamberg. Bamberg 2002

Backhaus, Klaus / Erichson, Bernd / Plinke, Wulff / Weiber, Rolf (Hg.) (2000): **Multivariate Analysemethoden. Eine anwendungsorientierte Einführung**. 9. Auflage. Berlin / Heidelberg u.a.: Springer-Verlag. S. 1-69; 628

Überla, Karl (1977): **Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler**. 2. Auflage. Berlin / Heidelberg: Springer-Verlag. S. 2-5; 7; 22-36; 43-81; 92-100; 155-157; 165-185; 207-214; 235-237; 241-248; 310-317; 355-363

Schulze, Gerhard (2002d): **Das Modell der klassischen Testtheorie in Grundzügen**. Paper zum soziologischen Forschungspraktikum 2002/2003 an der Otto-Friedrich-Universität Bamberg. Bamberg 2002

Zusätzliche Literatur Statistik

Krämer (2001) schreibt für diejenigen, die mit Mathematik und Statistik schon immer auf Kriegsfuss standen. Er erklärt die wichtigsten Konzepte der Statistik, damit weiterführende Literatur nicht wie ein Buch mit sieben Siegeln erscheint. Bortz (1999) richtet sich an Sozialwissenschaftler und beschreibt Konzepte, die in den meisten weiterführenden Statistikbüchern nicht erklärt werden. Hartung u.a. (2002) und Hartung / Elpelt (1999) decken fundiert die meisten statistischen Verfahren auf. Die Bücher eignen sich also für diejenigen, die es gerne genauer wissen. Eine Alternative hierzu (wenn auch auf Englisch) sind die Bücher aus der Reihe „Quantitative Applications in the Social Sciences“, die im Sage-Verlag erscheint. Jedes Buch führt in ein einziges statistisches Verfahren ein. Auf jeweils 80 bis 120 Seiten werden anschaulich, leicht verständlich und mit vielen Beispielen Fragestellungen, Probleme und Konzepte des Verfahrens dargestellt.

Bortz, Jürgen (1999): **Statistik für Sozialwissenschaftler**. 5., vollständig überarbeitete und aktualisierte Auflage. Berlin / Heidelberg: Springer

Hartung, Joachim / Elpelt, Bärbel (1999): **Multivariate Statistik. Lehr- und Handbuch der angewandten Statistik**. München: Oldebourg

Hartung, Joachim / Elpelt, Bärbel / Köseener, Karl-Heinz (2002): **Statistik**. München: Oldenbourg
Krämer, Walter (2001): **Statistik verstehen. Eine Gebrauchsanweisung**. München / Zürich: Piper
Reihe: Quantitative Applications in the Social Sciences. Erschienen bei Sage. Verschiedene Herausgeber

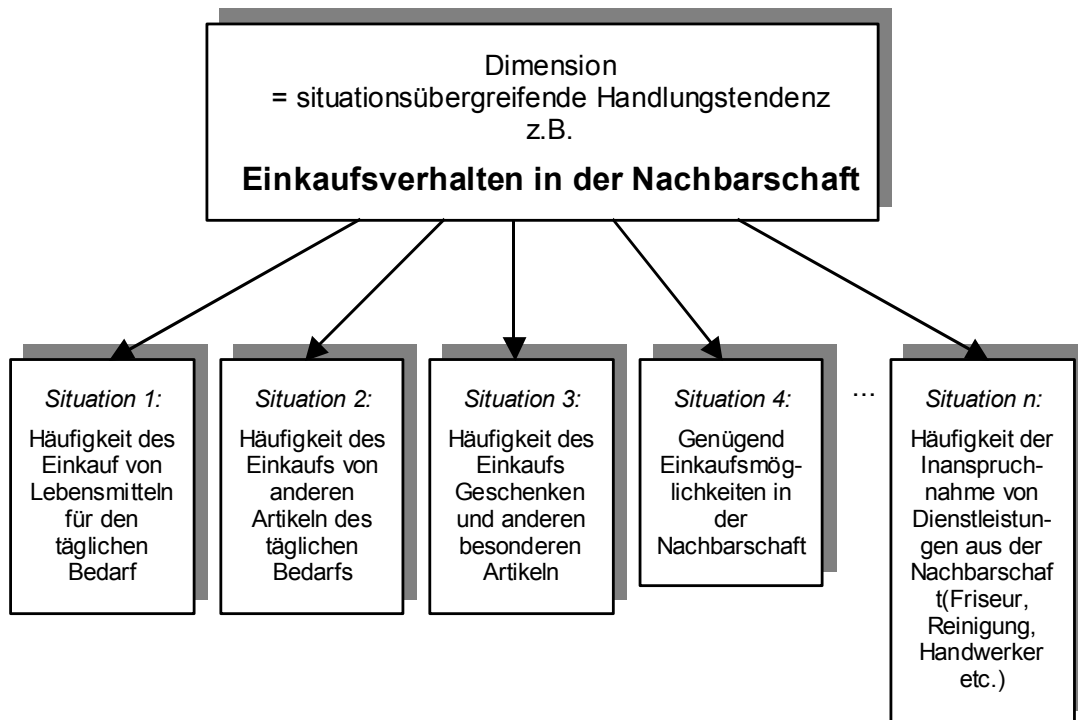
BILDUNG EINER DIMENSIONS VARIABLE

Dimensionen (= Dispositionen, Indikatoren, Skalen) in der sozialen Realität und Variablen im Fragebogen

Dimensionen

In der sozialen Realität kommen sehr häufig Dispositionen (= Dimensionen) vor. Dispositionen sind situationsübergreifende Handlungstendenzen. Dies bedeutet, dass viele Menschen in vielen verschiedenen Situationen ähnlich handeln und sich so Muster im Handeln und Denken ergeben, an denen wir uns orientieren können (vgl. hierzu z. B. Schulze (2002c)).

Beispiel: Das „Einkaufsverhalten in der Nachbarschaft“ zeigt sich nicht in einer einzigen Situation, sondern es schlägt sich im Verhalten in zahlreichen verschiedenen Situationen nieder (siehe Grafik unten): wie oft man beim Bäcker, Metzger, im Tante Emma-Laden oder Supermarkt direkt um die Ecke Lebensmittel einkauft; wie oft man Drogerie-Artikel oder andere Artikel des täglichen Bedarfs in der Nachbarschaft einkauft; wie oft man außergewöhnliche Artikel wie Geschenke oder Delikatessen in der Nachbarschaft kauft; wie oft man Dienstleistungen wie Handwerker, die Reinigung, die Bank oder den Friseur in der Nachbarschaft in Anspruch nimmt; was man genau kauft; wie viel Geld man in der Nachbarschaft ausgibt; an welchen Wochentagen man dort einkauft; ob man findet, dass es genügend Einkaufsmöglichkeiten in der Nachbarschaft gibt, oder ob man lieber in die Innenstadt fährt; ob man sich Zeit zum Einkauf in der Nachbarschaft lässt und auch mal mit den Verkäufern plauscht oder ob man schnell ein paar Besorgungen macht; usw.



Diese zahlreichen Situationen fassen wir im Alltagsleben zum Denkschema „Einkaufsverhalten in der Nachbarschaft“ zusammen. Wir reden davon, dass jemand oft oder selten in der Nachbarschaft einkauft. An solchen Denkschemata orientieren wir unser Denken und Handeln: Wenn wir wissen, dass jemand oft in der Nachbarschaft einkauft, fragen wir ihn vielleicht, ob er morgens vor der Arbeit kurz etwas besorgen könnte. Oder wir fragen ihn, ob er einen netten Geschenkladen oder einen guten Friseur in seiner Nähe kennt usw. Sozialwissenschaftliche Analyse hat deshalb unter anderem zwei Ziele:

- (1) Aufspüren von Dispositionen in Kollektiven.
- (2) Messung dieser Dispositionen bei einzelnen Personen

Von Dimensionshypothesen zu Variablen im Fragebogen

Wenn man Dispositionen mit Hilfe eines Fragebogens messen will, greift man aus der Vielzahl der Situationen, die die Dimension erfassen, einige heraus und formuliert hierzu eine Frage. Weil es sich bei Dispositionen um situationsübergreifende Handlungstendenzen handelt, muss man mindestens zwei Situationen im Fragebogen abfragen, um eine Dimension messen zu können.

Beispiel: Im Rahmen des soziologischen Forschungspraktikums 2000 /2001 sollte die Dimension „Einkaufsverhalten in der Nachbarschaft“ gemessen werden.¹ Aus den in der Grafik auf S. 7 unten angeführten Situationen wurden drei herausgegriffen:

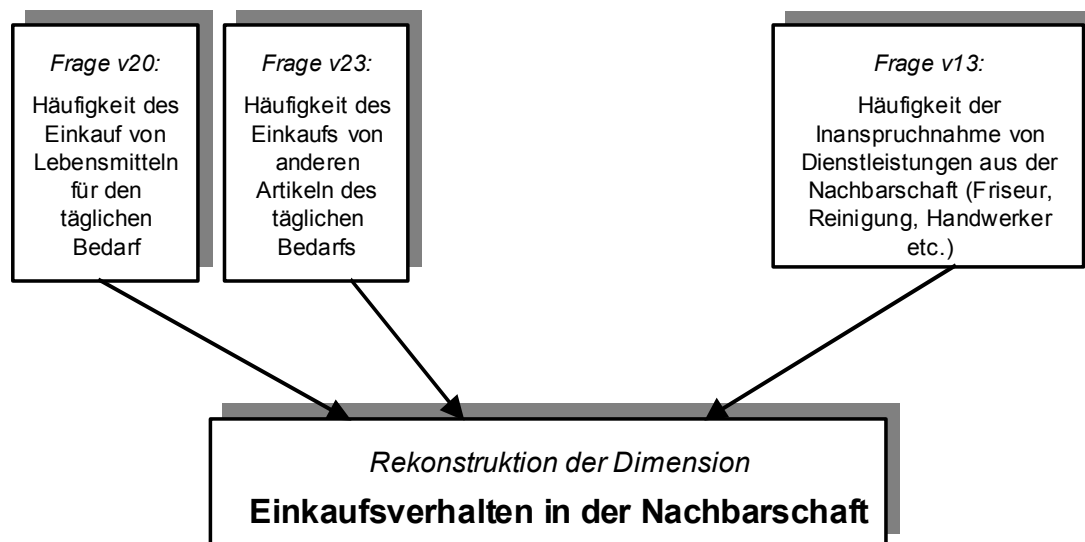
- Situation 1: die Häufigkeit des Einkaufs von Lebensmitteln für den täglichen Bedarf in Läden in der Nachbarschaft (v20);
- Situation 2: die Häufigkeit des Einkaufs von anderen Artikeln, die die Befragten regelmäßig brauchen (z. B. Hygiene- und Drogerie-Artikel) in Läden in der Nachbarschaft (v23);
- Situation 3: ob eine Person findet, dass es zu wenig Einkaufsmöglichkeiten für den täglichen Bedarf in der Nähe gibt (v13).

Oft ist man sich nicht sicher, ob die in einem Kollektiv verankerte Disposition überhaupt existiert und welche der Situation besonders gut zur Messung der Dimension geeignet sind. In diesem Fall sollte man lieber mehr als weniger Fragen in den Fragebogen aufnehmen – man kann schlecht geeignete Variablen später immer noch weglassen. Wenn man aber vergessen hat, etwas Wichtiges zu fragen, sind die Informationen nicht verfügbar und man kann die Dimension vielleicht nicht (so gut) messen.

Bei der Entscheidung, welche Variablen man weglassen sollte und welche zur Messung der Dimension gut geeignet sind, helfen die statistischen Verfahren, die im Rahmen dieses Papers besprochen werden. Häufig nimmt man deshalb in den Fragebogen der Voruntersuchung sehr viele Variablen zur Messung einer Dimension auf. Nach der Auswertung der Voruntersuchung entscheidet man dann, welche Variablen man in den Fragebogen der Hauptuntersuchung aufnimmt, um eine Dimension zu messen.

Von Variablen im Fragebogen zu Dimensionen

Analysiert man einen Fragebogen, liegen meist nur wenige Variablen vor, die eine bestimmte Dimension messen (könnten). Mit Hilfe dieser Variablen versucht man, die Ausprägung, die eine Person auf der Dimension hat, zu rekonstruieren:



Verstehen / Erfassen der dimensional Struktur von Variablen

Dimension und Korrelation

Wenn eine situationsübergreifende Handlungstendenz existiert, müssen sich Personen in ähnlichen Situationen konsistent verhalten, d. h. jemand, der häufig in der Nachbarschaft einkauft, wird wohl öfters als eine Person, die selten in der Nachbarschaft einkauft, Lebensmittel, andere Artikel des täglichen Bedarfs und außergewöhnliche Artikel kaufen. Er wird eher Dienstleistungen in der Nachbarschaft in Anspruch nehmen und er wird eher finden, dass es genug Einkaufsstätten in der Nachbarschaft gibt, als dies Personen tun, die selten in der Nachbarschaft einkaufen. Dies schlägt sich empirisch so nieder, dass es einen starken statistischen Zusammenhang (z. B. hohe Korrelationen) zwischen dem Verhalten von Personen in der einen Situation und dem Verhalten derselben Personen in der anderen Situation gibt: Leute, die oft Lebensmittel in der Nachbarschaft einkaufen, kaufen auch oft andere Artikel des täglichen Bedarfs in der Nachbarschaft ein. Leute, die selten Lebensmittel in der Nachbarschaft einkaufen, kaufen auch selten andere Artikel des täglichen Bedarfs in der Nachbarschaft ein.

¹ Hinweis: Den Datensatz zum soziologischen Forschungspraktikum 2000 / 2001 finden Sie in Baur (2003a). Er ist in Baur (2003b) dokumentiert.

Wichtiger Hinweis zum Thema „Ordinalskalen“

Ich behandle in diesem Band Variablen wie das Einkaufsverhalten in der Nachbarschaft als metrisch. Dies verweist auf das Ordinalskalenskalenproblem, das in Baur (2003c) ausführlich besprochen wird. Dass ich diese Variablen als metrisch behandle, bedeutet nicht, dass ich für die eine oder andere Form der Lösung des Ordinalskalenskalenproblems plädiere – diesbezüglich müssen Sie Ihre eigene Haltung finden. Wie auch immer Sie sich entscheiden: Sie sollten Ihre Entscheidung begründen und jedes Mal überlegen, welche Fehler Sie durch diese Entscheidung im konkreten Fall machen. Der Grund, warum ich solche Variablen als intervallskaliert interpretiere, ist didaktischer Natur: Fast alle sozialwissenschaftlichen Variablen sind empirisch ordinalskaliert. Gleichzeitig gehören Kenntnisse zahlreicher statistischer Verfahren heute zu den Basisqualifikationen von Sozialwissenschaftlern. Die Faktorenanalyse setzt ein metrisches Skalenniveau voraus. Verfahren für nominale und ordinale Daten bauen meist auf Verfahren für metrische Daten auf, d. h. es ist sinnvoll, erst die Verfahren zur Analyse metrischer Daten zu erlernen, weil es dann leichter fällt, die anderen Verfahren zu erlernen. Der sinnvollste Weg, diese Verfahren zu erlernen (und insbesondere ihre Schwächen kennen zu lernen) ist, Sie praktisch zu üben. Hierzu benötigt man aber Datensätze – die meist hauptsächlich ordinalskalierte Daten enthalten. Die Katze beißt sich in den Schwanz. Zu Übungszwecken behandle ich deshalb in diesem Band solche Variablen als intervallskaliert. Damit ich mich auf das Wesentliche – die Erläuterung der statistischen Verfahren – konzentrieren kann, führe ich nicht meine Überlegungen über das Fehlerrisiko aus. Wenn Sie in einem Forschungsbericht mit ordinalskalierten Daten arbeiten, sollten Sie dagegen – soweit möglich – ein Verfahren für ordinalskalierte Variablen vorziehen. Ist dies nicht möglich, müssen Sie sich entscheiden, wie Sie weiter verfahren, diese Entscheidung begründen und bei der Analyse stets das Fehlerrisiko im Auge behalten. Nähere Informationen hierzu finden Sie in: Baur (2003c)

Was Korrelationen inhaltlich bedeuten können.

Allerdings kann man von hohen Korrelationen zwischen zwei Variablen im Fragebogen nicht automatisch schließen, dass sie zu einer bestimmten gemeinsamen Dimension gehören. Korrelationen zwischen Variablen im Fragebogen können folgendes bedeuten:

- (1) Die Variablen messen eine gemeinsame Dimension, und es ist auch die Dimension, die man zu messen beabsichtigt.
- (2) Die Variablen messen eine gemeinsame Dimension, aber nicht die Dimension, die man zu messen beabsichtigt.
- (3) Es besteht ein Kausalzusammenhang zwischen den Variablen.
- (4) Drittvariablen beeinflussen den Zusammenhang zwischen den Variablen.
- (5) Mischformen zwischen (1) bis (4).

Wenn zwei Variablen hoch miteinander korrelieren, stellen sich folgende Fragen: Welche der fünf Alternativen trifft zu? Und wie kann man überprüfen, ob man die richtige Entscheidung getroffen hat?

In den meisten Fällen hilft kein statistisches Verfahren, diese Fragen zu beantworten. Stattdessen muss der Sozialforscher hier seine Entscheidung mit Hilfe seiner theoretischen Kenntnisse begründen. Sinnvoll ist hierbei ein falsifikatorisches Vorgehen, d. h. der Forscher überlegt, ob die Zusammenhänge auch etwas anderes bedeuten könnten.

Beispiel Nachbarschaftsverhalten: Es erscheint durchaus plausibel, dass es so etwas wie ein Einkaufsverhalten in der Nachbarschaft gibt. Auch scheint es möglich, dass die genannten Variablen v20, v23 und v13 das Einkaufsverhalten in der Nachbarschaft erfassen. Allerdings gibt es zwei wichtige Gegenthesen:

- (1) Im Fragebogen wurden auch Fragen zum Einkaufsverhalten außerhalb der Nachbarschaft gestellt. Zu Anfang dieses Papers wurde die These aufgestellt, dass das Einkaufsverhalten in der Nachbarschaft nicht unbedingt etwas mit dem Einkaufsverhalten außerhalb der Nachbarschaft zu tun hat, d. h. dass Leute, die oft in der Nachbarschaft einkaufen, nicht unbedingt gleichzeitig selten außerhalb der Nachbarschaft einkaufen. Deshalb wird eine Dimension „Einkaufsverhalten in der Nachbarschaft“ und eine Dimension „Einkaufsverhalten außerhalb der Nachbarschaft“ vermutet. Es könnte aber auch sein, dass beide zusammen eine übergeordnete Dimension bilden, d. h. dass Leute entweder oft in der Nachbarschaft oder oft außerhalb der Nachbarschaft einkaufen. Ein erstes Indiz dafür, dass Letzteres nicht der Fall ist, ist die geringere Korrelation zwischen den Variablen, die das Einkaufsverhalten in der Nachbarschaft messen, und den Variablen, die das Einkaufsverhalten außerhalb der Nachbarschaft messen. Dennoch sollte man diese Gegenthese bei der weiteren Analyse im Auge behalten.
- (2) Es könnte sein, dass die Variable v13 („Zu wenig Einkaufsmöglichkeiten für den täglichen Bedarf in der Nähe“) gar nicht zur Dimension „Einkaufsverhalten in der Nachbarschaft“ gehört, sondern dass ein kausaler Zusammenhang zwischen den Variablen existiert: Wenn die Befragten finden, dass es zu wenig Einkaufsmöglichkeiten in der Nachbarschaft gibt, kaufen sie seltener in der Nachbarschaft ein. Auch diese Gegenthese sollte man bei der weiteren Analyse zu überprüfen versuchen.

Schritt 1: Bestimmung der dimensional Struktur der Variablen

Umsortierung der Korrelationsmatrix

Vermutet ein Forscher einen dimensional Zusammenhang zwischen verschiedenen Variablen, überprüft er zuerst, ob die Variablen überhaupt zusammenhängen. Dies kann er machen, indem er einfach die Korrelationsmatrix analysiert und umsortiert. Beispielsweise finden Sie auf S. 11 eine Korrelationsmatrix einer Reihe von Variablen aus dem Datensatz des

Soziologischen Forschungspraktikums 2000/2001. Diese wurde bereits umsortiert:

Beispiel für eine Korrelationsmatrix, in der dimensionale Strukturen vermutet werden

Sortierte und gekürzte Korrelationsmatrix

	Einkauf in der Nachbarschaft				Einkauf an einem anderem Ort als dem Wohnort				Kontaktfreudigkeit / Geselligkeit				
	Kauf Lebensmittel Nachbarschaft	Kauf and. Artikel Nachbarschaft	Einkaufsmöglichkeiten	Kauf Lebensmittel Innenstadt	Kauf and. Artikel Innenstadt	Kauf Lebensmittel Stadtrand	Kauf and. Artikel Stadtrand	Kauf and. Artikel Stadtrand	Nachbarn grüßen	mit Nachbarn unterhalten	Nachbarn um Gefälligkeit bitten	Wichtigkeit Kontakt mit Nachbarn	gewünschte Kontakthäufigkeit
Einkauf in der Nachbarschaft	1,000	,596	,250										
Kauf Lebensm. Nachbar.													
Kauf and. Artikel Nachb.	,596	1,000	,234										
Einkaufsmöglichkeiten	,250	,234	1,000										
Einkauf an einem anderen Ort als dem Wohnort													
Kauf Lebensm. Innenst.	-,029	,099	,069	1,000	-,176	,489	-,141						
Kauf Lebensm. Stadtrand	-,193	-,224	-,079	-,176	1,000	-,163	,699						
Kauf and. Artikel Innenst.	-,054	-,155	-,039	,489	-,163	1,000	-,277						
Kauf and. Artikel Stadtr.	-,100	-,120	-,136	-,141	,699	-,277	1,000						
Geselligkeit / Kontaktfreudigkeit													
Nachbarn grüßen	,034	-,048	,019	-,065	,050	,010	,051	1,000	,593	,387	,201	-,095	
mit Nachbarn unterhalten	,054	,027	-,013	,007	,080	,057	,043	,593	1,000	,456	,279	-,200	
Nachbarn um Gefallen	,047	-,003	-,100	-,053	,056	-,020	,099	,387	,456	1,000	,293	-,180	
Wichtigkeit Kontakt	,079	,119	,085	,140	-,005	,035	-,012	,201	,279	,293	1,000	,105	
Wunsch Kontakthfgkt.	,051	,077	,053	-,013	-,043	-,108	-,023	-,095	-,200	-,180	,105	1,000	
Geschlecht													
Alter in Jahren	-,026	-,024	-,027	,011	,077	-,035	,067	-,051	-,107	-,042	-,046	-,009	
Wohndauer im Stadtteil	,052	,077	,057	,145	-,127	,059	-,123	,233	,243	,231	,185	-,045	
Haushaltsgröße	,108	,112	,024	-,001	-,054	,048	-,029	,201	,177	,106	,115	-,082	
Bewohner pro Haus	-,018	,017	-,107	-,107	,214	,005	,229	-,235	-,041	-,001	-,078	-,193	
Probleme mit Nachbarn	,133	,121	,184	,133	-,148	,109	-,149	-,195	-,194	-,182	-,135	,062	
Berufstätig, Lehrling, usw.	,003	-,015	-,022	-,125	,026	-,093	,088	-,159	-,153	-,110	-,060	,091	
Entfernung Arbeitsplatz	-,066	-,081	-,067	-,117	,138	-,091	,118	-,112	-,119	-,111	-,129	,091	
Freizeit- / Kulturangebot	,009	-,054	-,034	-,081	,006	,045	-,013	-,034	-,002	-,013	-,017	-,032	
Unsichere Gegenden	-,088	-,085	-,060	-,065	,051	-,024	,070	-,106	-,094	-,044	-,059	,129	
	-,023	-,047	,034	,016	,012	,079	,010	,040	,034	,118	,101	-,023	

Folgendes kann man aus dieser Korrelationsmatrix schließen:

- (1) Es scheint, als ob jeweils mehrere Variablen zu einer von drei Dimensionen gehören:
 - eine Dimension, die das *Einkaufsverhalten in der Nachbarschaft* erfasst. Diese Dimension scheint aus folgenden Variablen zu bestehen:
 - Häufigkeit des Einkaufs von Lebensmitteln für den täglichen Bedarf in Läden in der Nachbarschaft (v20 bzw. v20a);
 - Häufigkeit des Einkaufs von anderen Artikeln, die die Befragten regelmäßig brauchen (z. B. Hygiene- und Drogerie-Artikel) in Läden in der Nachbarschaft (v23 bzw. v23a);
 - Zu wenig Einkaufsmöglichkeiten für den täglichen Bedarf in der Nähe (v13 bzw. *kproblem*).
 - eine Dimension, die das *Einkaufsverhalten außerhalb der Nachbarschaft* erfasst. Diese Dimension scheint aus folgenden Variablen zu bestehen:
 - Häufigkeit des Einkaufs von Lebensmitteln für den täglichen Bedarf in Geschäften oder Kaufhäusern in der Innenstadt (v21 bzw. v21a);
 - Häufigkeit des Einkaufs von Lebensmitteln für den täglichen Bedarf in Großmärkten und Einkaufszentren am Stadtrand (v22 bzw. v22a);
 - Häufigkeit des Einkaufs von anderen Artikeln, die die Befragten regelmäßig brauchen (z. B. Hygiene- und Drogerie-Artikel) in Geschäften oder Kaufhäusern in der Innenstadt (v24 bzw. v24a);
 - Häufigkeit des Einkaufs von anderen Artikeln, die die Befragten regelmäßig brauchen (z. B. Hygiene- und Drogerie-Artikel) in Großmärkten und Einkaufszentren am Stadtrand (v25 bzw. v25a).
 - eine Dimension, die so etwas wie die *Geselligkeit, Kontakthäufigkeit mit den Nachbarn oder Kontaktfreudigkeit* erfasst. Diese Dimension scheint aus folgenden Variablen zu bestehen:
 - Häufigkeit mit der die Befragten normalerweise einem ihrer Nachbarn begegnen und einen Gruß wechseln (v26 bzw. v26a);
 - Häufigkeit mit der die Befragten normalerweise mit einem ihrer Nachbarn eine kleine Unterhaltung haben (v27 bzw. v27a);
 - Häufigkeit mit der die Befragten normalerweise von einem ihrer Nachbarn um eine Gefälligkeit gebeten werden (v28 bzw. v28a);
 - Zufriedenheit der Befragten mit der Art des Kontaktes zu ihren Nachbarn (v29 bzw. *kontzufr*);
 - Allgemeine Wichtigkeit von Kontakten zu den Bewohnern des Stadtviertels für die Befragten (v30 bzw. *kontwich*).
- (2) Die übrigen statistischen Zusammenhänge lassen eher Kausalzusammenhänge vermuten.
- (3) Variablen, die zu einer vermuteten Dimension gehören, korrelieren fast immer mit denselben Variablen außerhalb der Dimension. *Beispiel:* Die Variable „Wohndauer im Stadtteil“ (v02) hängt mit den Variablen v26a, v27a, v28a und *kontzufr* zusammen. Bis auf die Variable *kontwich* hängt die Variable v02 damit mit allen Variablen der vermuteten Dimension „Geselligkeit, Kontakthäufigkeit mit den Nachbarn oder Kontaktfreudigkeit“ zusammen. Es ist plausibel, hier einen Kausalzusammenhang zu vermuten: Vermutlich haben die Befragten umso mehr Kontakt mit ihren Nachbarn (v26a, v27a, v28a und *kontzufr* = abhängige Variable), je länger sie im Stadtteil wohnen (v02 = unabhängige Variable). Man könnte nun für jedes Variablenpaar (v02 und v26a; v02 und v27a; v02 und v28a; v02 und *kontzufr*) untersuchen, ob ein kausaler Zusammenhang existiert. Dies wären vier gesonderte Kausalanalysen. Wesentlich zeitsparender wäre es, wenn man nur untersucht, ob ein kausaler Zusammenhang zwischen der Variable „Wohndauer im Stadtteil“ (v02) und der Dimension „Geselligkeit, Kontakthäufigkeit mit den Nachbarn oder Kontaktfreudigkeit“ existiert. Gleichzeitig wäre dies für die Theoriebildung vorteilhaft, weil man ein höheres Abstraktionsniveau erreichen würde.

Faktorenanalyse

Häufig – wie auch in diesem Fall – ist es wegen der großen Anzahl der betrachteten Variablen schwer zu entscheiden, ob statistische Zusammenhänge auf dimensionale Strukturen oder Kausalzusammenhänge hinweisen. Deshalb gibt es statistische Verfahren, die es erleichtern, dimensionale Strukturen in Daten zu erkennen. Die Faktorenanalyse ist ein solches Verfahren. Im folgenden gehe ich davon aus, dass Sie wissen, wie die Faktorenanalyse funktioniert und wie man sie in SPSS umsetzt (hierzu z. B. Backhaus / Erichson / Plinke / Weiber (2002); Fromm (2003); Schulze (2002a) und Überla (1977)).

Auch in der Faktorenanalyse kristallisieren sich die drei Dimensionen mit den entsprechenden Variablen heraus:

Rotierte Komponentenmatrix^a

	Komponente		
	Kontaktfreudigkeit / Geselligkeit	Einkauf an einem anderem Ort als dem Wohnort	Einkauf in der Nachbarschaft
Häufigkeit einer kleinen Unterhaltung mit Nachbarn	,863	6,006E-02	3,211E-02
Häufigkeit der Begegnung mit / des Grüßens von Nachbarn	,773	4,551E-04	-1,854E-02
Häufigkeit, mit der Nachbarn um Gefälligkeit bitten	,729	9,128E-02	-3,825E-02
Wichtigkeit der allgemeinen Kontakte zu den Bewohnern des Stadtviertels	,590	-2,517E-02	,112
Probleme mit den Nachbarn	-,362	,290	,117
Einkauf von anderen Artikeln des täglichen Bedarfs in der Innenstadt	3,518E-02	-,756	-,220
Einkauf von anderen Artikeln des täglichen Bedarfs in Großmärkten / Einkaufszentren am Stadtrand	,114	,720	-,257
Einkauf von Lebensmitteln in der Innenstadt	2,497E-02	-,664	-2,726E-02
Einkauf von Lebensmitteln in Großmärkten / Einkaufszentren am Stadtrand	8,174E-02	,646	-,460
Einkauf von Lebensmitteln in der Nachbarschaft	4,649E-02	-4,451E-02	,838
Einkauf von anderen Artikeln des täglichen Bedarfs in der Nachbarschaft	5,854E-02	-1,829E-02	,837
zu wenig Einkaufsmöglichkeiten für den täglichen Bedarf in der Nähe	8,827E-03	-2,998E-02	-,493
Zufriedenheit mit Kontakthäufigkeit mit Nachbarn	,111	8,329E-02	-,115

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 5 Iterationen konvergiert.

Das dreifaktorielle Modell erklärt etwa 50 % der Streuung zwischen den dreizehn Variablen. Jeder Faktor allein erklärt etwa ein Sechstel der Streuung zwischen den dreizehn Variablen:

Erklärte Gesamtvarianz

Komponente	Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,446	18,817	18,817	2,393	18,411	18,411
2	2,304	17,722	36,539	2,054	15,800	34,211
3	1,711	13,163	49,703	2,014	15,492	49,703

Extraktionsmethode: Hauptkomponentenanalyse.

Die Variable „Zufriedenheit mit der Kontakthäufigkeit mit den Nachbarn“ fällt aus der Reihe: Sie würde zwar inhaltlich sehr gut zu der Dimension „Kontaktfreudigkeit“ passen, wird aber von keiner der drei Faktoren gut abgebildet:

Kommunalitäten

	Extraktion
Probleme mit den Nachbarn	,229
zu wenig Einkaufsmöglichkeiten für den täglichen Bedarf in der Nähe	,244
Einkauf von Lebensmitteln in der Nachbarschaft	,706
Einkauf von Lebensmitteln in der Innenstadt	,442
Einkauf von Lebensmitteln in Großmärkten / Einkaufszentren am Stadtrand	,635
Einkauf von anderen Artikeln des täglichen Bedarfs in der Nachbarschaft	,704
Einkauf von anderen Artikeln des täglichen Bedarfs in der Innenstadt	,621
Einkauf von anderen Artikeln des täglichen Bedarfs in Großmärkten / Einkaufszentren am Stadtrand	,597
Häufigkeit der Begegnung mit / des Grüßens von Nachbarn	,598
Häufigkeit einer kleinen Unterhaltung mit Nachbarn	,750
Häufigkeit, mit der Nachbarn um Gefälligkeit bitten	,542
Wichtigkeit der allgemeinen Kontakte zu den Bewohnern des Stadtviertels	,361
Zufriedenheit mit Kontakthäufigkeit mit Nachbarn	3,255E-02

Extraktionsmethode: Hauptkomponentenanalyse.

Diese Variable war bereits theoretisch schwer einzuordnen: Bildet sie mit den anderen Variablen eine gemeinsame Dimension oder besteht zwischen dieser Variablen und der Dimension „Kontaktfreudigkeit“ ein Kausalzusammenhang? Die Ergebnisse der Faktorenanalyse sprechen eher für Letzteres. Wir müssen also diese Variable bei der weiteren Analyse besonders im Auge zu halten.

Fassen wir zusammen: Die bisherige Analyse zielte darauf ab zu klären, ob im Kollektiv bestimmte Dimensionen vorliegen (Frage 1 auf S. 7). Drei Dimensionsvariablen scheinen sich herauszuschälen: das „Einkaufsverhalten in der Nachbarschaft“, das „Einkaufsverhalten außerhalb der Nachbarschaft“ und die „Geselligkeit, Kontakthäufigkeit mit den Nachbarn oder Kontaktfreudigkeit“. Wohlgedacht: Diese Lösung ist nie eindeutig. Andere Lösungen sind oft denkbar. Beispielsweise könnte sich die Dimension „Einkaufsverhalten außerhalb der Nachbarschaft“ weiter unterteilen in eine Dimension „Einkaufsverhalten in der Innenstadt“ und „Einkaufsverhalten am Stadtrand“. Ich habe hier diese Lösung gewählt, weil das Modell recht ökonomisch ist, d. h. dreizehn Variablen werden auf drei Dimension reduziert, weil es trotz der geringen Faktorenzahl noch relativ viel Streuung erfasst (etwa 50 % der Gesamtstreuung) und vor allem, weil es inhaltlich und theoretisch plausibel erscheint.

Schritt 2: Skalenbildung

Das restliche Paper widmet sich deshalb der Frage 2 von S. 7: Wie bildet man eine Dimensionsvariable (= einen Indikator; eine Skala)? Wie schätzt man also ein, welche Ausprägung eine einzelne Person auf einer Dimension hat. Das Vorgehen bei der Dimensionsbildung führe ich am Beispiel der vermuteten Dimension „Einkaufsverhalten in der Nachbarschaft“ aus. Für die Dimensionsbildung hat das Konzept der Reliabilität eine zentrale Bedeutung. Deshalb gehe ich zunächst der Frage nach: Was bedeutet Reliabilität?

Bestimmung der Reliabilität einer Dimension

Irrtumskomponente bei der Messung

Jedes Mal, wenn man eine Dimension misst, verzerren andere Faktoren das Ergebnis:

- (1) *Andere Dimensionen*: Ein Ereignis misst normalerweise nicht nur eine einzige Dimension, sondern mehrere verschiedene Dimensionen gleichzeitig. Beispielsweise spielt in die Häufigkeit, mit der ein Befragter Lebensmittel in der Nachbarschaft kauft, nicht nur das Einkaufsverhalten in der Nachbarschaft hinein, sondern auch Faktoren wie beispielsweise:
 - Vorliebe für bestimmte Nahrungsmittel: Wenn jemand nur frisches Gemüse oder asiatische Lebensmittel mag, geht er in Geschäfte, die diese Lebensmittel anbieten, egal wo sie liegen.
 - Vorliebe für bestimmte Lebensmittelgeschäfte: Manche Leute gehen aus Prinzip nur zu Aldi, Rewe, Marktkauf oder dem kleinen Gemüsehändler.
 - Kosten der Lebensmittel: Wenn jemand wenig Geld hat, kann er es sich vielleicht nicht leisten, in bestimmten Läden in der Nachbarschaft einzukaufen, auch wenn er gerne würde.
 - Qualität der Lebensmittel: Wenn um die Ecke nur ein bestimmter Supermarkt ist, der altes Gemüse, billige Konserven und ranzige Butter verkauft, fährt der Befragte vielleicht lieber in einen anderen Stadtteil, um dort bessere Lebensmittel zu bekommen.
 - Bequemlichkeit des Einkaufs: Wenn jemand einen Supermarkt direkt neben dem Arbeitsplatz hat und dort in der Mittagspause kurz einkaufen kann, während die Lebensmittelgeschäfte in der Nachbarschaft geschlossen haben, wenn er heimkommt, kauft er diese vielleicht eher im Supermarkt am Arbeitsplatz ein.
- (2) *Zufällige Einflüsse*. Zufallsfehler haben zahlreiche Ursachen, die sich je nach Erhebungsmethode unterscheiden. Bei der standardisierten Befragung können Zufallsfehler beispielsweise verursacht werden durch Fehler beim Kodieren; durch missverständliche Frageformulierungen oder Intervieweranweisungen; durch unterschiedliche Betonung desselben Wortes während des Interviews; durch Ermüdung des Interviewers usw.

Wenn man eine Dimension misst, hat die Messung also immer eine Irrtumskomponente. Variablen, die zur Messung von Dimensionen verwendet werden, unterscheiden sich deshalb nicht, ob sie durch eine Irrtumskomponente verzerrt sind oder nicht – sie sind immer durch eine Irrtumskomponente verzerrt. Vielmehr unterscheiden sich diese Variablen dadurch, wie stark sie durch die Irrtumskomponente verzerrt werden. Irrtumskomponenten haben einen unsystematischen Charakter: Bei manchen Personen wird die Ausprägung der Dimension unterschätzt, bei anderen überschätzt.

Reliabilität („Reliability“)

Die Reliabilität bezeichnet das Ausmaß, mit dem ein Messverfahren dieselben Ergebnisse liefert, wenn man es wiederholt. Die Reliabilität ist umso höher, je geringer die Irrtumskomponente des Messverfahrens ist. Weil jeder Messvorgang eine Irrtumskomponente hat, können Messverfahren nie vollständig reliabel sein.

*Klassische Testtheorie (= Likert-Skalierung; „classical test score theory“, „classical test theory“, „test theory“):
Einschätzung der Ausprägung einer Dimension bei einer Person*

Die klassische Testtheorie ist ein statistisches Modell. Jeder beobachtete Wert setzt sich laut der klassischen Testtheorie aus folgenden Elementen zusammen:

$$\begin{array}{rcccl} \text{Beobachteter Wert} & = & \text{Wahrer, aber unbekannter Wert} & + & \text{Irrtumskomponente} \\ X & = & t & + & e \end{array}$$

Der wahre Wert, den eine Person A auf der Dimension hat, kann nicht direkt gemessen werden. Wenn man allerdings bei derselben Person unendlich oft die Dimension misst, kann man folgende Annahmen treffen:

- (1) Nimmt man zahlreiche Messungen vor, gleichen sich die Irrtumskomponenten im Mittel aus. Mathematisch ausgedrückt bedeutet dies, dass der Erwartungswert des Zufallsfehlers Null ist: $E(e_A) = 0$.
- (2) Der wahre Wert und die Irrtumskomponente hängen nicht zusammen. Die Irrtumskomponente wirkt gleichermaßen bei Personen, die hohe Ausprägungen auf der Dimension haben, und Personen, die niedrige Ausprägungen auf der Dimension haben. Die Korrelation zwischen wahrem Wert und der Irrtumskomponente ist also Null: $\rho_{(t,e)} = 0$
- (3) Die Irrtumskomponente bei einer Messung beeinflusst nicht den wahren Wert bei einer zweiten Messung. Die Irrtumskomponente bewirkt also nicht, dass sich der wahre Wert verändert. Die Korrelation zwischen der Irrtumskomponente bei der Messung zum Zeitpunkt 1 und wahren Wert bei der Messung zum Zeitpunkt 2 ist also Null: $\rho_{(e_1,t_2)} = 0$
- (4) Es gibt keinen Zusammenhang zwischen den Irrtumskomponenten bei unterschiedlichen Messungen. Die Korrelation der Irrtumskomponenten bei verschiedenen Messungen ist also Null: $\rho_{(e_1,e_2)} = 0$

Wenn die Annahmen nicht erfüllt sind, kann man die Reliabilität auch nicht einschätzen. Wenn die gerade genannten Annahmen aber gelten und man unendlich oft die Dimension X bei der Person A misst, lässt sich zeigen, dass Folgendes gilt:

$$\begin{array}{rcccl} \text{Erwartungswert des} & = & \text{Erwartungswert des} & + & \text{Erwartungswert der} \\ \text{beobachteten Werts} & = & \text{wahren, aber unbekannten Werts} & + & \text{Irrtumskomponente} \\ E(X_A) & = & E(t_A) & + & E(e_A) \\ \mu_{X_A} & = & \mu_{t_A} & + & \mu_{e_A} \end{array}$$

Weil aber nach der obigen Annahme (1) der Erwartungswert der Irrtumskomponente Null ist [$E(e) = 0$], gilt:

$$E(X_A) = E(t_A) + 0$$

Über unendlich viele Messungen betrachtet gleichen sich also die Irrtumskomponenten gegenseitig aus. Der Erwartungswert $E(X_A) = \mu_{X_1}$ ist der wahre Wert von Person A auf der Dimension X.

Beispiel: Person A kauft gelegentlich in der Nachbarschaft ein. Allerdings liegt direkt gegenüber dem Arbeitsplatz von Person A eine Drogerie, weshalb Person A dort alle Drogerieartikel einkauft und nicht in der Nachbarschaft. Gleichzeitig gibt es direkt vor der Haustür von Person A einen Gemüsehändler. Person A mag gerne frisches Gemüse und kauft deshalb nur dort Gemüse. Bei jedem Produkt, das Person A kauft, spielen solche Faktoren mit in das Kaufverhalten hinein. Manchmal unterschätzt und manchmal überschätzt man das Einkaufsverhalten in der Nachbarschaft. Wenn man unendlich oft messen würde, würde der Durchschnitt aller dieser Messungen das wahre Einkaufsverhalten von Person A in der Nachbarschaft widerspiegeln, weil es sich bei jedem Produkt, das Person A kauft, um andere Faktoren handelt, die das Ergebnis verzerren. Man bekäme heraus, dass Person A gelegentlich in der Nachbarschaft einkauft.

Allerdings ist es unmöglich, unendlich oft zu messen. Deshalb bleibt der Erwartungswert μ_{X_A} immer ein hypothetischer Wert und unterscheidet sich von dem Durchschnitt $\overline{x_A}$, den man erhält, wenn man mehrere Mal die Ausprägung der Dimension misst.

Einschätzung der Reliabilität einer Dimension

Der Erwartungswert erlaubt, mit Hilfe von wiederholten Messungen den wahren, aber unbekannten Wert einer einzigen Person abzuschätzen. Reliabilität bedeutet aber, dass man wissen will, ob wiederholte Messungen bei *allen* Personen der Grundgesamtheit verlässliche Ergebnisse liefern. Wie kann man von einer einzelnen Person auf alle Personen der Grundgesamtheit verallgemeinern?

Als Hilfsmittel nimmt man die Varianz: Die Varianz erfasst die Streuung der Ausprägungen, die die Personen der Grundgesamtheit haben. Da laut der obigen Annahme (2) der wahre Wert und der Irrtumskomponenten nicht zusammenhängen, gilt bei unendlich vielen Messungen:

$$\begin{array}{rclcl} \text{Varianz des} & = & \text{Varianz des} & + & \text{Varianz der} \\ \text{beobachteten Werts} & & \text{wahren, aber unbekannten Werts} & & \text{Irrtumskomponente} \\ \text{VAR}(X) & = & \text{VAR}(t) & + & \text{VAR}(e) \end{array}$$

Wenn man diese Gleichung umformt, erhält man die Reliabilität der Variable X als Maß des wahren, aber unbekannten Wertes t:

$$\text{Reliabilität} = \frac{\text{Varianz des wahren, aber unbekannten Wertes}}{\text{Varianz des beobachteten Wertes}}$$

$$\rho_x = \frac{\text{VAR}(t)}{\text{VAR}(X)} = 1 - \frac{\text{VAR}(e)}{\text{VAR}(X)}$$

Die Reliabilität ρ_x gibt an, wie gut die gemessene Variable X – über alle Befragten hinweg gesehen – die Dimension erfasst. Die Reliabilität schwankt zwischen Null und Eins. Wenn die Reliabilität Null ist ($\rho_x = 0$), dann werden die unterschiedlichen Antworten der Befragten auf die Frage X vollständig durch die Irrtumskomponente verursacht. Wenn die Reliabilität Eins ist ($\rho_x = 1$), dann werden die unterschiedlichen Antworten der Befragten auf die Frage X vollständig durch die unterschiedliche Ausprägungen bei der Dimension verursacht. Je höher die Reliabilität, desto besser ist eine Variable also zu Messung der Dimension geeignet. Wie aber schätzt man die Reliabilität ab?

Parallele Messungen

Oben haben wir festgelegt, dass wiederholte Messungen notwendig sind, um Reliabilität feststellen zu können. Was aber macht man, wenn man nur eine einzige Untersuchung vorliegen hat? Hier greift man auf die Bedeutung des Begriffs „Dimension“ zurück: Dimensionen wirken in verschiedenen Situationen auf ähnliche Weise. Man kann also eine Dimension feststellen, wenn ein Befragter immer wieder in dieselbe Situation kommt und immer wieder ähnlich reagiert – er kauft beispielsweise immer wieder in der Nachbarschaft Lebensmittel ein und das schon seit Jahren. Wenn man also alle paar Monate fragt, wie oft er in der Nachbarschaft Lebensmittel einkauft, bekommt man immer wieder die Antwort „oft“, bis auf den August, weil er da im Urlaub ist.

Man kann aber auch eine Dimension abschätzen, indem man bei einer einzigen Untersuchung simultan parallele Messungen vornimmt. Parallele Messungen sind Messungen, bei denen die wahren Werte t gleich sind und die Varianz der gemessenen Variablen gleich ist. Der Befragte verhält sich in ähnlichen Situationen immer wieder ähnlich. Statt wiederholten Messungen erhebt man in einem einzigen Fragebogen also eine Mehrzahl von Items. *Beispiel:* Die Befragten wurden gefragt:

- (1) wie häufig sie Lebensmittel für den täglichen Bedarf in Läden in der Nachbarschaft einkaufen (v20 bzw. v20a);
- (2) wie häufig sie andere Artikel, die sie regelmäßig brauchen (z. B. Hygiene- und Drogerie-Artikel) in Läden in der Nachbarschaft einkaufen (v23 bzw. v23a);
- (3) ob es zu wenig Einkaufsmöglichkeiten für den täglichen Bedarf in der Nähe ihrer Wohnung gibt (v13 bzw. kproblem).

Typischerweise wird ein Befragter, der insgesamt oft in der Nachbarschaft einkauft, oft Lebensmittel und oft andere Artikel in der Nachbarschaft einkaufen und eher finden, dass es genug Einkaufsmöglichkeiten in der Nachbarschaft gibt. Ein Befragter dagegen, der insgesamt selten in der Nachbarschaft einkauft, wird selten Lebensmittel und selten andere Artikel in der Nachbarschaft einkaufen und eher finden, dass es zu wenig Einkaufsmöglichkeiten in der Nachbarschaft gibt. Durch den Einfluss anderer Faktoren können sich Abweichungen von diesem typischen Antwortschema ergeben.

Wenn die verschiedenen Antworten einer Person auf zwei Fragen A und B nur aufgrund von Irrtumskomponenten variieren, kann man sie als parallele Messungen sehen. In diesem Fall entspricht die Korrelation zwischen den Variablen A und B der Reliabilität: $\rho_{AB} = \rho_x$.

Dies bedeutet, dass man mindestens zwei Variablen benötigt, um die Reliabilität einzuschätzen. Je mehr Variablen man zur Messung der Dimension heranzieht, desto höher wird die Reliabilität. Aller-

dings erhöht jede zusätzliche Variable die Reliabilität weniger als die vorhergegangenen – ab einer bestimmten Variablenzahl macht es nur noch wenig Sinn, zusätzliche Variablen zur Messung der Dimension heranzuziehen.

Die Methode der internen Konsistenz („Internal Consistency Method“)

Bislang wurde theoretisch dargestellt, wie man die Reliabilität einer Dimension messen könnte, nämlich entweder über wiederholte Messung derselben Variable oder parallele Messungen von Variablen, die alle die Dimension ausdrücken. Nun stellen sich zwei Folgefragen:

- (1) Wie schätzt man die Reliabilität in einem konkreten Fall ein?
- (2) Wie viele Messungen braucht man, um eine Dimension verlässlich zu messen? Wünschenswert ist eine möglichst kleine Zahl von Messungen, also eine möglichst geringe Itemzahl.

Es gibt verschiedene Möglichkeiten, parallele Messungen vorzunehmen. Die für die Sozialwissenschaften wichtigste ist die *Methode der internen Konsistenz*: Wenn mehrere Variablen in einem Fragebogen dieselbe Dimension messen, kann man auch ein einziges Maß verwenden, um die Reliabilität des Messverfahrens zu bestimmen. Diese Maße nennt man *Maße der internen Konsistenz*. Das gebräuchlichste dieser Maße ist *Cronbachs α* (nicht zu verwechseln mit dem α , das beim Signifikanztesten verwendet wird).

Schätzung der Reliabilität

Ausgangspunkt für die Berechnung von Cronbachs α ist die Korrelationsmatrix. Für den im Rahmen des soziologischen Forschungspraktikum verwendeten Datensatz und die Variablen zum Thema „Einkaufsverhalten in der Nachbarschaft“ sieht diese folgendermaßen aus:

Korrelationsmatrix nach der Umpolung

	<i>Einkauf von Lebensmitteln in der Nachbarschaft</i>	<i>Einkauf von anderen Artikeln des täglichen Bedarfs in der Nachbarschaft</i>	<i>genug Einkaufsmöglichkeiten i.d. Nähe</i>
<i>Einkauf von Lebensmitteln in der Nachbarschaft</i>	1,000	,652	,267
<i>Einkauf von anderen Artikeln des täglichen</i>	,652	1,000	,215
<i>genug Einkaufsmöglichkeiten i.d. Nähe</i>	,267	,215	1,000

Cronbachs α berechnet man mit Hilfe der Korrelationsmatrix folgendermaßen:

$$\text{Alpha} = \frac{\text{Anzahl der zur Dimensionsbildung herangezogenen Variablen} \cdot \text{durchschnittliche Korrelation dieser Variablen}}{1 + \text{durchschnittliche Korrelation dieser Variablen} \cdot (\text{Anzahl der zur Dimensionsbildung herangezogenen Variablen} - 1)}$$

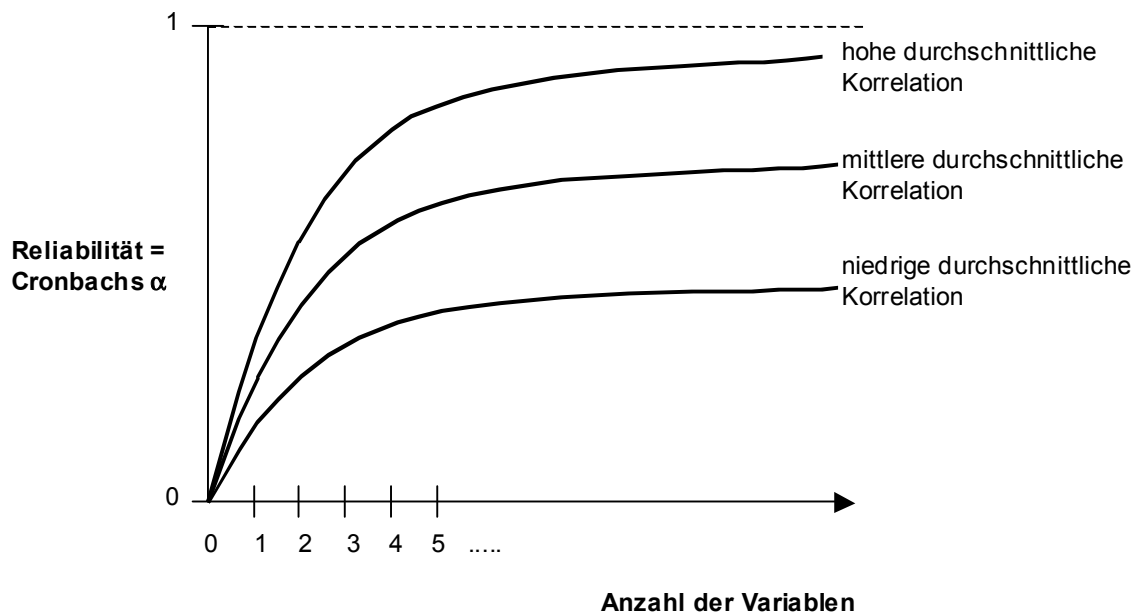
$$\alpha = \frac{i \cdot \bar{r}}{1 + \bar{r} \cdot (i - 1)}$$

Für die Variablen zum Thema „Einkaufsverhalten in der Nachbarschaft“ berechnet man Cronbachs α beispielsweise folgendermaßen:

$$\bar{r} = \frac{0,652 + 0,267 + 0,215}{3} = 0,378$$

$$\alpha = \frac{3 \cdot 0,378}{1 + 0,378 \cdot (3 - 1)} = \frac{1,134}{1,756} = 0,65$$

Wie groß Cronbachs α ist, hängt nicht nur von der Variablenzahl ab, sondern auch davon, wie hoch die durchschnittliche Korrelation ist: Je höher die durchschnittliche Korrelation ist, desto höher ist Cronbachs α :



Welche Variablen sollte man zur Dimensionsbildung heranziehen?

Es kann vorkommen, dass eine Variable eine Dimension zwar misst, aber stärker als andere Variablen durch eine Irrtumskomponente beeinflusst ist. Die Korrelation einer solchen Variablen mit den anderen Variablen ist im Vergleich zu den Korrelationen der übrigen Variablen untereinander sehr niedrig. Dadurch sinkt die durchschnittliche Korrelation, wenn man diese Variable mit zur Dimensionsbildung heranzieht. Dies wiederum hat zur Folge, dass die Reliabilität sinkt, wenn man diese Variable mit zur Dimensionsbildung heranzieht. In einem solchen Fall ist es sinnvoller, die betreffende Variable bei der Bildung der Dimensionsvariable wegzulassen.

Beispiel: Bei Variablen zum Thema „Einkaufsverhalten in der Nachbarschaft“ ist auffällig, dass die Variable „Nicht genügend Einkaufsmöglichkeiten für den täglichen Bedarf in der Nähe des Wohnortes“ schwächer mit den anderen Variablen zusammenhängt als diese beiden Variablen untereinander zusammenhängen. Lässt man diese Variable bei der Berechnung von Cronbachs α weg, erhält man folgendes Ergebnis:

$$\bar{r} = \frac{0,652}{1} = 0,652$$

$$\alpha = \frac{2 * 0,652}{1 + 0,652 * (2 - 1)} = \frac{1,304}{1,652} = 0,79$$

Gleichzeitig wurde bereits erwähnt, dass die Variable vielleicht nicht zur Dimension „Einkaufsverhalten in der Nachbarschaft“ gehört, sondern dass eventuell ein Kausalzusammenhang zwischen dieser Variable und dem Einkaufsverhalten existiert. Es erscheint also sinnvoll, nur die Variablen „Einkauf von Lebensmitteln in der Nachbarschaft“ und „Einkauf von anderen Artikeln in der Nachbarschaft“ zur Dimensionsbildung heranzuziehen.

Beschließt man, eine Variable *nicht* zur Dimensionsbildung heranzuziehen, sollte es – wie in diesem Beispiel – nicht nur statistische Argumente dafür geben, sondern die Entscheidung sollte auch inhaltlich-theoretisch begründbar sein.

Wie groß sollte α sein?

Idealerweise sollte Cronbachs α mindestens 0,8 groß sein. Allerdings ist es in der Forschungspraxis aus Zeit- und Kostengründen oft nicht möglich, einen Fragebogen zu konstruieren, der genügend Fragen enthält, damit man dieses Maß der Reliabilität erhält. Deshalb begnügt man sich oft mit niedrigeren α -Werten. Wichtig ist vor allem, dass man im Forschungsbericht dokumentiert, welche Variablen zur Bildung der Dimensionsvariablen herangezogen wurden und wie hoch α war,

Bildung der neuen Dimensionsvariable

Hat man sich entschieden, welche Variablen man zur Dimensionsbildung heranzieht, bildet man im nächsten Schritt die neue Dimensionsvariable. Es gibt verschiedene Methoden, Dimensionsvariablen zu bilden. In den meisten Fällen liefern diese verschiedenen Methoden relativ ähnliche Ergebnisse. Hier sollen zwei Alternativen besprochen werden: die Addition der Punktwerte und die regressionsanalytische Schätzung der Faktoren.

Addition der Punktwerte

Die einfachste Methode, eine neue Dimensionsvariable zu bilden, ist, die Werte, die ein Befragter bei den einzelnen Variablen, die zur Dimensionsbildung herangezogen werden, einfach mit Hilfe eines „*COMPUTE*“-Befehls zu addieren.

Beispiel: Zur Bildung der Dimensionsvariable „Häufigkeit des Einkaufs in der Nachbarschaft“ addiert man die Variablen *v20* und *v23*. Da beide Variablen die Ausprägungen 1, 2 und 3 haben, schwankt die neue Variable zwischen folgenden Werten:

- *niedrigster Wert* = (Zahl der Variablen) \times (niedrigste Ausprägung) = $2 \times 1 = 2$. Da „1“ sowohl bei *v20* als auch bei *v23* für „häufiges Einkaufen“ steht, kaufen die Befragten am häufigsten in der Nachbarschaft ein, die bei der neuen Variablen den Wert „2“ aufweisen.
- *höchster Wert* = (Zahl der Variablen) \times (höchste Ausprägung) = $2 \times 3 = 6$. Da „3“ sowohl bei *v20* als auch bei *v23* für „seltenes Einkaufen“ steht, kaufen die Befragten am seltensten in der Nachbarschaft ein, die bei der neuen Variablen den Wert „6“ aufweisen.

Für zwei fiktive Befragte A und B ist die neue Variable folgendermaßen zu interpretieren.

- Befragter A hat auf die Frage *v20* „oft“ (1) geantwortet und auf die Frage *v23* „gelegentlich“ (2). Er hat auf der Dimension „Häufigkeit des Einkaufs in der Nachbarschaft“ den Wert $1 + 2 = 3$. Er kauft also relativ häufig in der Nachbarschaft ein, weil die „3“ näher an der „2“ liegt als an der „6“.
- Befragter B hat auf die Frage *v20* „gelegentlich“ (2) geantwortet und auf die Frage *v23* „selten / nie“ (3). Er kauft also seltener in der Nachbarschaft ein als Befragter A. Dies spiegelt sich auch im Wert wider, den er auf der Dimension „Häufigkeit des Einkaufs in der Nachbarschaft“ hat: $2 + 3 = 5$.

Regressionsanalytische Schätzung

Die Alternative zur Addition der Punktwerte ist, die neuen Dimensionen regressionsanalytisch zu schätzen, d. h. man führt eine Regressionsanalyse durch, bei der die Faktoren die abhängige und die ursprünglichen Variablen die unabhängigen Variablen sind. SPSS macht dies automatisch, wenn man bei der Faktorenanalyse den Unterbefehl angibt „*PRINT= fscore*“ angibt. Das Ergebnis findet sich in der Ausgabe „Koeffizientenmatrix der Komponentenwerte“. Mit Hilfe dieser Ausgabe und eines „*COMPUTE*“-Befehls kann man die neuen Dimensionen bilden.

Vorgehen in SPSS am Beispiel der Dimension „Einkaufsverhalten in der Nachbarschaft“

Schritt 1: Analyse der Korrelationsmatrix

Vermutet man, dass mehrere Variablen Ausdruck einer gemeinsamen Dimension sind, muss sich dies in der Korrelationsmatrix niederschlagen: Alle Korrelationen zwischen den Variablen müssen relativ hoch sein. Deshalb betrachtet man zunächst die Korrelationsmatrix, um zu überprüfen, ob dies auch der Fall ist. Der SPSS-Befehl für die Variablen *v20*, *v23* und *v12* lautet:

```
** SCHRITT 1: KORRELATIONSMATRIX
```

```
CORRELATIONS VARIABLES = v20, v23, v13
/MISSING = PAIRWISE.
```

SPSS macht dann folgende Ausgabe:

Korrelationsmatrix (Annahme des Intervallskalenniveaus für ordinalskalierte Variablen)

	<i>Kauf Lebensmittel Nachbarschaft</i>	<i>Kauf and. Artikel Nachbarschaft</i>	<i>Einkaufsmöglichkeiten</i>
<i>Kauf Lebensmittel Nachbarschaft</i>	1,000	,652	-,267
<i>Kauf and. Artikel Nachbarschaft</i>	,652	1,000	-,215
<i>Einkaufsmöglichkeiten</i>	-,267	-,215	1,000

Tatsächlich korrelieren die drei Variablen relativ hoch miteinander. Vermutlich sind sie Ausdruck einer gemeinsamen Dimension, die die Häufigkeit des Einkaufs in der Nachbarschaft messen. Um die Dimensionsvariable bilden zu können, müssen allerdings alle Variablen gleich *gepolt* sein: Wenn bei einer Variable eine niedrige Zahl eine niedrige Ausprägung der Dimension bedeutet und eine hohe Zahl eine hohe Ausprägung der Dimension, muss dies auch bei allen anderen Variablen gelten.

Beispiel: Die extremen Ausprägungen der Dimension „Häufigkeit des Einkaufs in der Nachbarschaft“ sind „selten / nie“ und „oft“. Wenn man die Dimensionsvariable bildet, kann man

- entweder für das Extrem „selten / nie“ eine niedrige Zahl vergeben und für das Extrem „oft“ eine hohe Zahl vergeben
- oder für das Extrem „selten / nie“ eine hohe Zahl vergeben und für das Extrem „oft“ eine niedrige Zahl vergeben.

Die Variablen v20, v23 und v12 sind bezüglich der „Häufigkeit des Einkaufs in der Nachbarschaft“ folgendermaßen gepolt:

- v20: eine hohe Zahl (3) für das Extrem „selten / nie“ und eine niedrige Zahl für das Extrem „oft“ (1)
- v23: eine hohe Zahl (3) für das Extrem „selten / nie“ und eine niedrige Zahl für das Extrem „oft“ (1)
- v13: eine niedrige Zahl (1) für das Extrem „stimmt“ und eine hohe Zahl für das Extrem „stimmt nicht“ (3)

Die Variable v13 ist also genau umgekehrt gepolt wie die anderen beiden Variablen. Dadurch sind die Werte der Variablen nicht direkt vergleichbar. Manche der Korrelationen in der Korrelationsmatrix sind negativ und manche positiv. Deshalb müssen die Variablen umgepolt werden, und zwar

- muss entweder die Variable v13 so rekodiert werden, dass eine hohe Zahl (3) für das Extrem „stimmt“ und eine niedrige Zahl für das Extrem „stimmt nicht“ (1) steht
- oder die Variablen v20 und v23 müssen so rekodiert werden, dass eine niedrige Zahl (1) für das Extrem „selten / nie“ und eine hohe Zahl für das Extrem „oft“ (3) steht.

Da es weniger Arbeit macht, nur eine Variable zu rekodieren, wird im folgenden die erste Variante gewählt.

Schritt 2 / 3: Polung der Variablen

Die Variable v13 wird mit folgendem Befehl umgepolt:

```
** SCHRITT 2: POLUNG DER VARIABLEN
** Variable v13 muss umgepolt werden
```

```
RECODE v13 (1 = 3) (2 = 2) (3 = 1) INTO v13r.
EXECUTE.
```

```
VARIABLE LABELS v13r 'genug Einkaufsmöglichkeiten i.d. Nähe'.
VALUE LABELS v13r 1 'stimmt'
                  2 'stimmt teilweise'
                  3 'stimmt nicht'.
```

```
FREQUENCIES v13 v13r
/STATISTICS = NONE.
```

Vergleicht man die Häufigkeitsverteilungen der ursprünglichen Variable (*v13*) und der neuen Variable (*v13r*), so macht SPSS folgende Ausgabe:

Die neue und die alte Variable haben exakt dieselbe Verteilung, nur dass die Werte vertauscht sind.

Betrachtet man nun noch einmal die Korrelationsmatrix, so müssten die Korrelationen nun alle positiv sein:

```
** SCHRITT 3: ERNEUTE  
** KORRELATIONSMATRIX
```

```
CORRELATIONS  
VARIABLES = v20, v23,  
v13r  
/MISSING = PAIRWISE.
```

Die SPSS-Ausgabe zeigt, dass dies auch der Fall ist:

zu wenig Einkaufsmöglichkeiten für den täglichen Bedarf in der Nähe

	<i>Absolute Häufigkeiten</i>	<i>Relative Häufigkeiten in %</i>
<i>stimmt</i>	90	15,6
<i>stimmt teilweise</i>	98	17,0
<i>stimmt nicht</i>	389	67,4
<i>Gesamt</i>	577	100,0

genug Einkaufsmöglichkeiten i.d. Nähe

	<i>Absolute Häufigkeiten</i>	<i>Relative Häufigkeiten in %</i>
<i>stimmt</i>	389	67,4
<i>stimmt teilweise</i>	98	17,0
<i>stimmt nicht</i>	90	15,6
<i>Gesamt</i>	577	100,0

Korrelationsmatrix nach der Umpolung

	<i>Einkauf von Lebensmitteln in der Nachbarschaft</i>	<i>Einkauf von anderen Artikeln des täglichen Bedarfs in der Nachbarschaft</i>	<i>genug Einkaufsmöglichkeiten i.d. Nähe</i>
<i>Einkauf von Lebensmitteln in der Nachbarschaft</i>	1,000	,652	,267
<i>Einkauf von anderen Artikeln des täglichen</i>	,652	1,000	,215
<i>genug Einkaufsmöglichkeiten i.d. Nähe</i>	,267	,215	1,000

Schritt 4: Berechnung der Reliabilität

Der nächste Schritt der Analyse ist nun, Cronbachs α für die drei Variablen zu messen, um die Reliabilität bestimmen zu können. Die SPSS-Syntax hierfür lautet folgendermaßen:

```
** SCHRITT 4: BERECHNUNG DER RELIABILITÄT
```

```
RELIABILITY  
  /VARIABLES = v20 v23 v13r  
  /FORMAT = LABELS  
  /SCALE(ALPHA) = ALL  
  /MODEL = ALPHA  
  /SUMMARY = TOTAL.
```

SPSS macht dann folgende Ausgabe:

```
***** Method 1 (space saver) will be used for this analysis *****

  R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   ( A L P H A )

  1.      V20           Einkauf von Lebensmitteln in der Nachbarschaft
  2.      V23           Einkauf von anderen Artikeln des täglich. Bed. i.d.N.
  3.      V13R          genug Einkaufsmöglichkeiten i.d. Nähe

Item-total Statistics

              Scale          Scale          Corrected          Alpha
              Mean          Variance        Item-              if Item
              if Item        if Item        Total              Deleted
              Deleted        Deleted        Correlation

V20              3,4489          1,6305          ,6028          ,3424
V23              3,1761          1,5880          ,5529          ,4123
V13R             3,6743          2,4422          ,2584          ,7843

Reliability Coefficients

N of Cases =      568,0                      N of Items =      3

Alpha =      ,6495
```

Die **Reliabilität** beträgt $\alpha = 0,65$, wenn man alle drei Variablen zur Dimensionsbildung heranzieht. Die Reliabilität ist recht hoch – es lohnt sich also durchaus, eine Dimensionsvariable zu bilden.

Die **Trennschärfe** misst, wie das einzelne Item mit den übrigen Items zusammenhängt. Sie berechnet sich aus der Korrelation des Items mit der durchschnittlichen Korrelation der übrigen Items. *v13r* hängt also nur schwach mit den anderen Items zusammen.

Es könnte durchaus sein, dass eine oder mehrere Variablen die Dimension besser erfassen als andere. Schlechte Variablen verringern die Reliabilität. Deshalb berechnet SPSS für jede Variable, wie hoch die Reliabilität wäre, wenn man diese Variable weglassen würde und nur die anderen Variablen zur Dimensionsbildung heranziehen würde. Es sind zwei Ergebnisse möglich:

- Bei Variablen, die die Dimension gut erfassen, verschlechtert sich die Reliabilität – α sinkt. Diese Variablen sollten unbedingt zur Dimensionsbildung mitverwendet werden. Dies ist bei den Variablen *v20* und *v23* der Fall.
- Bei Variablen, die die Dimension schlecht erfassen, erhöht sich die Reliabilität – α steigt. Diese Variablen sollte man *nicht* zur Dimensionsbildung heranziehen. Dies ist bei der Variablen *v13r* der Fall. Wie bereits mehrfach erwähnt, sprechen auch theoretische Überlegungen dafür, diese Variable nicht zur Dimensionsbildung heranzuziehen.

Schritt 5: Bildung der neuen Dimensionsvariable (Addition der Punktwerte)

Die SPSS-Syntax für die Bildung der Dimensionsvariablen lautet folgendermaßen:

```
** SCHRITT 5: BERECHNUNG DER NEUEN DIMENSION
** (ohne v13r, weil Alpha steigt, wenn v13r nicht dabei ist)

COMPUTE kaufnah = v20 + v23.
EXECUTE.
```

```
VARIABLE LABELS kaufnah 'Einkauf in der Nachbarschaft'.
VALUE LABELS kaufnah 2 '2 oft'
6 '6 selten oder nie'.
```

```
FREQUENCIES kaufnah
/STATISTICS = NONE.
```

SPSS macht dann folgende Ausgabe:

Einkauf in der Nachbarschaft

	<i>Absolute Häufigkeiten</i>	<i>Relative Häufigkeiten in %</i>
2 oft	209	36,2
3,00	71	12,3
4,00	110	19,1
5,00	59	10,2
6 selten oder nie	128	22,2
Gesamt	577	100,0

Nun ergibt sich ein weiteres Problem:

Nicht immer kommen alle Werte der neuen Dimension auch bei den Befragten vor. Dies gilt insbesondere für die Extremwerte.

Außerdem bildet man oft nicht nur eine Dimensionsvariable sondern mehrere. Beispielsweise vermutet man in diesem Datensatz noch mindestens die Dimensionen „Einkaufsverhalten außerhalb der Nachbarschaft“ und „Geselligkeit, Kontakthäufigkeit mit den Nachbarn oder Kontaktfreudigkeit“. Wenn die Dimensionsvariablen aus unterschiedlich vielen Variablen gebildet werden, sind die Wertebereiche unterschiedlich. Dies erschwert die Vergleichbarkeit der Variablen.

Deshalb ist es oft sinnvoll, die gebildeten Dimensionsvariablen zu klassieren. Da die sehr viele Variablen im Beispieldatensatz drei Ausprägungen haben, sollen auch die neuen Dimensionsvariablen so klassiert werden, dass sie drei Ausprägungen haben. Dabei soll so klassiert werden, dass alle Kategorien ungefähr gleich stark besetzt sind.

Schritt 6: Klassierung der Dimensionsvariable

Die SPSS-Syntax für die Klassierung der Dimensionsvariable lautet:

```
** SCHRITT 6: NEUE DIMENSION KLASSIEREN
```

```
RECODE kaufnah (2 = 1) (3,4 = 2) (5,6 = 3) INTO kaufnah2.
EXECUTE.
```

```
VARIABLE LABELS kaufnah2 'Einkauf in der Nachbarschaft'.
VALUE LABELS kaufnah2 1 'oft'
2 'gelegentlich'
3 'selten oder nie'.
```

```
FREQUENCIES kaufnah2
/STATISTICS = NONE.
```

Die Häufigkeitsverteilung der neuen Dimensionsvariable sieht dann folgendermaßen aus:

Einkauf in der Nachbarschaft

	<i>Absolute Häufigkeiten</i>	<i>Relative Häufigkeiten in %</i>
oft	209	36,2
gelegentlich	181	31,4
selten oder nie	187	32,4
Gesamt	577	100,0

Weiterführende Literatur:

Carmines, Edward G. / Zeller, Richard A. (1982): **Reliability and Validity Assessment**. Reihe: Sage University Papers.

Quantitative Applications in Social Sciences. Herausgegeben von John L. Sullivan. Band 17. Beverly Hills / London: Sage

Publications [Signatur in der Bamberger Bibliothek: 31/MR 2100 WG 2193-17]

Cronbach, L.J. (1951): **Coefficient Alpha and the Internal Structure of Tests**. In: *Psychometrika. A Journal of Quantitative Psychology*. Band 16. S. 297-334 [Signatur in der Bamberger Bibliothek: 20/CZ 6460]

Gigerenzer, Gerd (1981): **Messung und Modellbildung in der Psychologie**. München / Basel: Ernst Reinhardt Verlag

Lord, Frederic M. / Novick, Melvin R. (1968): **Statistical Theories of Mental Test Scores**. Reading: Addison-Wesley [Signatur in der Bamberger Bibliothek: 30/CM 600 BY 618]

Nunnally, Jum C. (1978): **Psychometric Theory**. New York: McGraw-Hill [Signatur in der Bamberger Bibliothek: 30/CN 220 BW 521]

Schulze, Gerhard (2002b): **Das Modell der klassischen Testtheorie in Grundzügen**. Paper zum soziologischen Forschungspraktikum 2002/2003 an der Otto-Friedrich-Universität Bamberg

Zur Erstellung dieses Papers verwendete Literatur

Backhaus, Klaus / Erichson, Bernd / Plinke, Wulff / Weiber, Rolf (2002): **Multivariate Analysemethoden. Eine anwendungsorientierte Einführung**. 9. Auflage. Berlin / Heidelberg u.a.: Springer-Verlag. 1-69; 628

Baur, Nina (2003a): **Sozialwissenschaftliche Datenanalyse am PC für Fortgeschrittene. Daten-CD**. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Band 4-2

Baur, Nina (2003b): **Takeoff der Auswertung. Zur Vorbereitung statistischer Analysen**. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Band 5

Baur, Nina (2003c): **Bivariate Statistik, Drittvariablenkontrolle und das Ordinalskalensproblem. Eine Einführung in die Kausalanalyse und in den Umgang mit zweidimensionalen Häufigkeitsverteilungen mit SPSS für Windows**. Reihe: Bamberger Beiträge zur empirischen Sozialforschung. Band 9.

Carmines, Edward G. / Zeller, Richard A. (1982): **Reliability and Validity Assessment**. Reihe: Sage University Papers. Quantitative Applications in Social Sciences. Herausgegeben von John L. Sullivan. Band 17. Beverly Hills / London: Sage Publications

Fromm, Sabine (2000): **Faktorenanalyse**. Paper zum Hauptseminar „Datenanalyse am PC“ an der Otto-Friedrich-Universität Bamberg. Bamberg

Kim, Jae-On / Mueller, Charles W. (1978): **Factor Analysis. Statistical Methods and Practical Issues**. Reihe: Sage University Papers. Quantitative Applications in Social Sciences. Herausgegeben von Michael S. Lewis-Beck. Band 14. 16. Auflage (1990). Newbury Park / London / New Delhi: Sage Publications. S. 69-73

Stanley, J. C. (1971): Reliability. In: Thorndike, Robert L. (Hg.) (1971): **Educational Measurement**. 2. Auflage. Washington D.C.: American Council of Education. S. 256 – 442 [Signatur in der Bibliothek:25+30/DD 500 BB 379]

Kim, Jae-On / Mueller, Charles W. (1978): **Introduction to Factor Analysis. What It Is and How To Do It**. Reihe: Sage University Papers. Quantitative Applications in Social Sciences. Herausgegeben von Michael S. Lewis-Beck. Band 13. 19. Auflage (1991). Newbury Park / London / New Delhi: Sage Publications. S. 50-51

Schulze, Gerhard (2001): **Kollektiv verankerte Dispositionen**. Paper 3 (2. Fassung) zum HS „Daten und Theorie. Teil I: Grundlagen“. WS 2001/2002. Otto-Friedrich-Universität Bamberg: Unveröffentlichtes Seminarpaper

Schulze, Gerhard (2002a): **Faktorenanalyse in Grundzügen**. Paper zum soziologischen Forschungspraktikum 2002/2003 an der Otto-Friedrich-Universität Bamberg. Bamberg

Schulze, Gerhard (2002b): **Das Modell der klassischen Testtheorie in Grundzügen**. Paper zum soziologischen Forschungspraktikum 2002/2003 an der Otto-Friedrich-Universität Bamberg

Schulze, Gerhard (2002c): **Einführung in die Methoden der empirischen Sozialforschung**. Bamberger Beiträge zur empirischen Sozialforschung. Band 1. Bamberg

SPSS Inc. (1999): **SPSS 10.0 Syntax Reference Guide für SPSS Base, SPSS Regression Models, SPSS Advanced Models**. 1002-1011 („Reliability“) [findet sich unter „Hilfe“ in SPSS]

Überla, Karl (1977): **Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler**. 2. Auflage. Berlin / Heidelberg: Springer-Verlag, 2-5; 7; 22-36; 43-81; 92-100; 155-157; 165-185; 207-214; 235-237; 241-248; 310-317; 355-363