

DOKUMENTATION DER DATENSÄTZE

„Rohdaten_FoPra_2000-2001.sav“ UND „Datensatz_FoPra_2000-2001.sav“

Der Hintergrund der Befragung

Die Datensätze „Rohdaten_FoPra_2000-2001.sav“ und „Datensatz_FoPra_2000-2001.sav“ wurden im Rahmen des soziologischen Forschungspraktikums 2000 / 2001 erstellt – ersterer ist ungelabelt, letzterer ist mit Variablen- und Wertelabeln versehen. Thema des Forschungspraktikums war „Lebensraum Stadt und seine Gestaltung. Städtevergleich Bamberg, Erlangen, Forchheim, Nürnberg“. Die Praktikums Teilnehmer erstellten Projektberichte zu folgenden Themen:

- Die Wohnsituation in den Bamberger Stadtteilen im Wandel
- Wohnen im Wandel: Entwicklung der Wohnsituation in Forchheim seit 1950
- Kommunalpolitische Wandlungsprozesse in Bamberg am Beispiel der Wohnqualität 1952 - 1997
- Die Kommunalpolitik der Stadt Forchheim im Wandel
- Modernisierungsdruck auf die Kommunalverwaltung - die Gestalter passen sich an
- Lebensqualität in Städten - subjektive Lebensqualität und deren Wandel
- Lebensqualität im Wandel
- Die Gestaltung des Zusammenlebens: Das Verhältnis von Stadt und Kirche am Beispiel der Domstadt Bamberg
- Lebensraum Stadt: Der Wandel der städtischen Wirtschaftsstrukturen in Bamberg 1950 - 2000
- Die Gestaltung des Lebensraumes Stadt aus wirtschaftspolitischer Perspektive am Beispiel der Stadt Bamberg
- Wandel der Wirtschaftsstrukturen in Forchheim 1945 - 2000

Im Laufe des Projektes führten die Praktikums Teilnehmer in den vier Städten eine standardisierte Umfrage zu verschiedenen Themen des Lebens in der Stadt durch. Am Beispiel dieses Datensatzes illustriere ich im folgenden, wie Probleme des Datenerhebungsprozesses diskutiert werden können.

Stichprobenprobleme

Der Datensatz entstand im Rahmen eines Lehrprojekts. Viele Forschungsschritte wurden aus Übungszwecken durchgeführt. Er hat deshalb einige methodische Mängel, die für solche Projekte typisch sind. Der wichtigste Mangel ist, dass es sich bei der Stichprobe um keine Zufallsstichprobe handelt. Dies hat systematische Fehler zur Folge.

Hinweis:

In den Kapiteln des Buches sowie in den meisten Aufgaben wird zu Übungszwecken angenommen, dass es sich bei dem Datensatz um eine repräsentative Zufallsstichprobe aus der Grundgesamtheit der Bevölkerung der Städte Bamberg, Erlangen, Forchheim und Nürnberg handelt. Methodisch gesehen ist dies falsch, weshalb diese Annahme nicht getroffen werden darf, wenn die Daten für wissenschaftliche Veröffentlichungen verwendet werden.

Art der Stichprobe

Der Datensatz enthält N = 493 Fälle, d. h. es wurden insgesamt 493 Personen befragt. Die Befragung im Rahmen des soziologischen Forschungspraktikums 2000 / 2001 bestand aus zwei Teilen:

- 1) Straßenbefragung (n = 413 Fälle): Im Frühjahr 2001 befragten die Praktikums Teilnehmer auf belebten Plätzen in den Innenstädten von Bamberg, Erlangen, Forchheim und Nürnberg Passanten, die in der jeweiligen Stadt wohnten. Die Befragung fand zu verschiedenen Tageszeiten statt, wobei versucht wurde, die Befragungszeit möglichst stark zu streuen. Der Beginn der Befragung wurde im Fragebogen festgehalten. Die Interviewer hatten einen Quotenplan zu erfüllen:

Geschlecht	männlich	weiblich	<i>Gesamt</i>
Alter			
bis ca. 40 Jahre	25 %	25 %	50 %
älter als 40 Jahre	25 %	25 %	50 %
<i>Gesamt</i>	50 %	50 %	100 %

- 2) Telefonbefragung (n = 80 Fälle): Um überprüfen zu können, inwieweit die Ergebnisse der Straßenbefragung verzerrt sind, wurde im Vorfeld der Straßenbefragung (Januar / Februar 2001) eine telefonische Befragung durchgeführt. Hierzu wurde eine Zufallsstichprobe aus den örtlichen Telefonbüchern gezogen.

Ausfallprotokoll

Um Anhaltspunkte für die Art der Verzerrung des Datensatzes zu erhalten, führten die Interviewer ein Ausfallprotokoll, in dem sie folgende Merkmale der Personen festhielten, die nicht an der Befragung teilnehmen wollten:

- Geschlecht (bei beiden Erhebungsarten)
- geschätztes Alter (nur bei Straßenbefragung)
- Ausfallgrund (nur bei Straßenbefragung)

Es ergab sich folgendes Bild:

Tab. 1: Ausfälle (Straße):

Alter / Geschlecht	männlich	weiblich	gesamt
bis ca. 40 Jahre	139 19,0%	151 20,6%	290 39,6%
älter als 40 Jahre	216 29,5%	226 30,9%	442 60,4%
gesamt	355 48,5%	377 51,5%	732 100%
Wohnt nicht in betreffender Stadt	247		

Tab. 2: Ausfälle (Telefon):

Geschlecht:	Ausfälle :
männlich	27 43,5%
weiblich	35 56,5%
gesamt	62 100%

Tab. 3: Ausfallgründe (Straße)

Gründe:	Anzahl der Nennungen :
Wetter	17 29,8%
Tageszeit	12 21,1%
Alter	9 15,7%
Länge d. Fragebogens	6 10,5%
Ausländer	3 5,2%
Sonstige Umfragen	2 3,5%
Sonstiges (Ort; Wochentag, -ende;	8 14,0%
Gesamt	57 100%

Allgemeine Mängel

Insgesamt verweigerten mehr Frauen als Männer die Teilnahme an der Untersuchung. Dies könnte mehrere Gründe haben:

- Frauen könnten kritischer und ängstlicher als Männer sein. Wäre dies der Fall, würde dies bedeuten, dass im Datensatz eher die „mutigen“ und „unkritischen“ Personen vertreten sind. Dies könnte sich beispielsweise auf Fragen auswirken, die Aspekte wie die Sicherheit des Wohnumfeldes oder die Effizienz der Kommunalpolitik berühren.
- Es könnte aber auch sein, dass Frauen sowohl zu Hause (Telefonbefragung) als auch in der Innenstadt (Straßenbefragung) eher anzutreffen waren und dadurch insgesamt eher befragt wurden. Dies ist durchaus plausibel, weil nach wie vor die geschlechtliche Arbeitsteilung existiert, dass Männer eher Vollzeit erwerbstätig sind und Frauen eher Teilzeit oder gar nicht erwerbstätig sind und stattdessen Hausarbeit und Besorgungen erledigen. Sie sind also tagsüber sowohl zu Hause als auch in der Stadt überrepräsentiert. Wenn Frauen die größere Gruppe der Befragten darstellen, dann müssten bei gleichem Verweigerungsverhalten von Männern und Frauen mehr Frauen die Antwort verweigern als Männer.

Außerdem sind in jeder Untersuchung Personen überrepräsentiert, die bestimmte Persönlichkeitseigenschaften haben (mitteilsam, aktiv, nicht sehr misstrauisch, sprachlich gewandt). Diese Persönlichkeitseigenschaften wirken sich aber nicht nur auf die Bereitschaft aus, an einer Untersuchung teilzunehmen, sondern auch auf Einstellungen, Meinungen, Werte und das soziale Umfeld.

Mängel der Straßenbefragung

Eine Straßenbefragung ist keine Zufallsstichprobe. Bestimmte Personengruppen trifft man gar nicht auf der Straße an:

- (Chronisch) kranke Menschen
- Sehr alte und gebrechliche Menschen
- Extrem ängstliche Menschen
- Personen, die in geschlossenen Anstalten untergebracht sind (Gefängnisse, psychiatrische Kliniken u.ä.)

Andere Personengruppen sind auf der Straße überrepräsentiert:

- Obdachlose
- Personen, deren Beruf einen Aufenthalt auf der Straße erfordert (Polizisten, Straßenverkäufer, Straßenmusikanten etc.)
- Studenten und Schüler (weil sie mehr Freizeit haben und sich ihre Freizeit frei einteilen können)
- Personen mit Freizeitbeschäftigungen, die man außerhalb des Hauses macht, sind überrepräsentiert.

Je nach Tageszeit, Wochentag und Jahreszeit sind bestimmte Personengruppen unter- andere überrepräsentiert:

- Frauen sind vor- und nachmittags (siehe oben), Männer abends und nachts überrepräsentiert
- Schüler sind vormittags unter-, abends eher überrepräsentiert.
- Erwerbstätige sind tagsüber unterrepräsentiert.
- Ältere Menschen sind abends unterrepräsentiert.

Dass bestimmte Personengruppen auf der Straße unterrepräsentiert sind, hätte zur Folge gehabt, dass bei einer zufälligen Befragung der Passanten diese Personengruppen auch in der Untersuchung unterrepräsentiert gewesen wären. Deshalb ist in diesem Fall eine Quotenstichprobe sogar von Vorteil, denn viele der oben genannten Aspekte hängen mit dem Alter und dem Geschlecht der Befragten zusammen. Der Quotenplan stellte sicher, dass zumindest ungefähr ebenso viele Männer wie Frauen und ebenso viele ältere Menschen wie jüngere befragt wurden. Ein weiterer Vorteil des Quotenplans ist, dass die unterschiedliche Bereitschaft verschiedener Personengruppen, an der Untersuchung teilzunehmen, ausgeglichen wurde: Nicht nur verweigerten eher die Teilnahme an der Untersuchung, sondern auch ältere Personen (siehe oben). Im Ausfallprotokoll hielten die Interviewer auch fest, warum bestimmte Personen nicht an der Untersuchung teilnehmen wollten:

Tab. 3: Ausfallgründe (Straße)

Gründe:	Anzahl der Nennungen :
Wetter	17 29,8%
Tageszeit	12 21,1%
Alter	9 15,7%
Länge d. Fragebogens	6 10,5%
Ausländer	3 5,2%
Sonstige Umfragen	2 3,5%
Sonstiges (Ort; Wochentag, -ende;	8 14,0%
Gesamt	57 100%

Die genannten Gründe dürften die Untersuchungsergebnisse in unterschiedlichem Maße verzerren. 247 Personen wohnten nicht in der betreffenden Stadt (vgl. Tab. 1). Da diese nicht Gegenstand der Untersuchung waren, wirkt sich diese Zahl auch nicht auf das Untersuchungsergebnis aus. Die Ausfallgründe „Wetter“, „Länge des Fragebogens“ und „Tageszeit“ dürften die Untersuchung insofern verzerren, dass mitteilssame Personen noch stärker überrepräsentiert sind (also Personen, die sich trotz schlechten Wetters, des Zeitaufwandes, der für die Teilnahme an der Untersuchung notwendig ist, und Hektik, die für sie zu einer bestimmten Tageszeit entsteht, nicht abschrecken lassen, an einer Untersuchung teilzunehmen).

Im Ausfallgrund „sonstige Umfragen“ drückt sich eine allgemeine Tendenz aus: Immer weniger Menschen sind bereit, an sozialwissenschaftlichen Untersuchungen teilzunehmen. Dies hat mehrere Gründe: Es werden immer mehr Befragungen durchgeführt. Menschen vertrauen immer weniger blind in die Wissenschaft und fühlen sich deshalb auch nicht mehr unbedingt geehrt, an einem wissenschaftlichen Projekt teilnehmen zu dürfen. Die Diskussion um

Datenschutzfragen erhöht das Misstrauen gegen sozialwissenschaftliche Untersuchungen und schmälert dadurch die Bereitschaft, an einer Untersuchung teilzunehmen. Zu erwarten ist also,

dass im Datensatz eher kritische und misstrauische Personen unterrepräsentiert sind.

Die Ausfallgründe „Alter“ und „Ausländer“ hängen mit fast allen Untersuchungszielen zusammen, weil Menschen unterschiedlichen Alters und unterschiedlicher Herkunft typischerweise unterschiedliche Lebenseinstellungen haben. Die geringere Teilnahmebereitschaft älterer Menschen wurde ein Stück weit durch den Quotenplan ausgeglichen. Es kann allerdings sein, dass sehr alte Menschen unterrepräsentiert sind (was Sie noch im Lauf Ihrer Analysen überprüfen müssen).

Mängel der Telefonbefragung

Auch die telefonische Befragung ist keine Zufallsstichprobe: Wird die Stichprobe mit Hilfe des Telefonbuchs gezogen, sind Obdachlose und Personen ohne Telefon gar nicht repräsentiert. Von letzteren ist anzunehmen, dass sie eher folgende Eigenschaften aufweisen: geringeres Einkommen (d. h. sie können sich kein Telefon leisten) oder sozial isoliert (d. h. sie brauchen kein Telefon, weil sie niemand anruft).

Zu bestimmten Befragungszeiten sind ebenso wie auf der Straße bestimmte Personen unterrepräsentiert oder gar nicht vertreten, andere überrepräsentiert:

- Hausfrauen sind tagsüber überrepräsentiert.
- Erwerbstätige, die zu Hause arbeiten, sind überrepräsentiert.
- Erwerbstätige, die nicht zu Hause arbeiten, sind tagsüber nicht anwesend.
- Schüler sind vormittags nicht anwesend.
- Studenten sind tagsüber überrepräsentiert, weil sie oft zu Hause lernen.
- Ältere Menschen sind den ganzen Tag überrepräsentiert, weil sie seltener fortgehen.
- Personen mit Freizeitbeschäftigungen, die man zu Hause macht, sind überrepräsentiert.

Ein weiterer Mangel bei der Telefonbefragung ist, dass sich die Interviewer unterschiedliche verhielten, wenn eine Person nicht zu Hause anwesend war: Einige versuchten mehrere weitere Male, die Person erneut zu erreichen, andere nicht.

Abschätzung systematischer Fehler

Um die Verzerrung des Datensatzes abschätzen zu können, ist es also notwendig, Hypothesen aufzustellen, welche Personengruppen warum nicht befragt werden konnten. Diese Hypothesen sollten nicht „aus dem hohlen Bauch“ stammen sondern idealerweise empirisch fundiert sein. Um Anhaltspunkte zu gewinnen, welche Personenkreise warum verweigern, führen Forscher deshalb immer wieder Studien zum Befragungs- und Verweigerungsverhalten bei verschiedenen Befragungs- und Stichprobenmethoden durch, auf die man zurückgreifen kann. Eine gute Quelle ist die Zeitschrift „ZUMA-Nachrichten“, die bei der GESIS (www.gesis.org) bezogen werden kann. Dort werden regelmäßig solche Studien veröffentlicht.

Um abschätzen zu können, inwiefern diese allgemeinen Probleme auf das eigene Forschungsprojekt übertragbar sind, ist es des weiteren notwendig, den eigenen Datenerhebungsprozess so gut wie möglich zu dokumentieren. Insbesondere sollten Ausfälle protokolliert werden. Wie ich oben gezeigt habe, erlauben die Ausfallprotokolle, die Hypothesen mit dem Fragebogen zu verknüpfen.

Inwieweit fehlende Werte und Ausfälle das Ergebnis verzerren, muss von Frage zu Frage entschieden werden. Wie man systematische Fehler abschätzt, beschreiben z. B. *Schulze* (2004) oder *Behnke, Behnke und Baur* (2004).

Literatur

Behnke, Joachim / Behnke, Nathalie / Baur, Nina (2004): Empirische Methoden der Politikwissenschaft. Paderborn: Ferdinand Schöningh. Im Erscheinen

Schulze, Gerhard (2004): Einführung in die Methoden der empirischen Sozialforschung. Bamberger Beiträge zur empirischen Sozialforschung, 2., überarbeitete Auflage. Band 1. Bamberg