Materialien zum Modul *Methoden der Demographie, Wirtschafts- und Sozialstatistik*

G. Rohwer

Version 3 **September 2008**

Fakultät für Sozialwissenschaft
Ruhr-Universität Bochum, GB 1
44780 Bochum

goetz.rohwer@ruhr-uni-bochum.de

met-3.tex September 2008

Dieser Text enthält Materialien zu den Teilen I und II des Moduls „Methoden der Demographie, Wirtschafts- und Sozialstatistik".

Hinweise zum Text

- Wie im Inhaltsverzeichnis angegeben wird, gliedert sich der Text in Kapitel, die meisten von ihnen auch in Abschnitte. Eine weitere Untergliederung in Paragraphen wird zu Beginn jedes Kapitels angegeben.
- Einfache Anführungszeichen werden zur Kennzeichnung sprachlicher Ausdrücke verwendet, doppelte Anführungszeichen werden verwendet, um Zitate kenntlich zu machen oder um anzudeuten, dass ein Ausdruck unklar ist und/oder metaphorisch verwendet wird. Innerhalb von Zitaten wird versucht, die im Original verwendeten Anführungszeichen zu reproduzieren. Wenn nicht anders angegeben, folgen Hervorhebungen in Zitaten stets dem Original; eigene Zusätze, Änderungen und Auslassungen werden durch eckige Klammern kenntlich gemacht.
- Wir unterscheiden die Zeichen '=' und ':='. Ein Gleichheitszeichen mit vorangestelltem Doppelpunkt wird verwendet, um anzudeuten, dass eine definitorische Gleichsetzung vorgenommen wird, d.h. der Ausdruck auf der linken Seite wird durch den Ausdruck auf der rechten Seite definiert. Dagegen dient ein einfaches Gleichheitszeichen zur Formulierung einer Gleichheitsbehauptung und setzt deshalb voraus, dass beide Seiten schon definiert sind.
- Als Dezimalpunkt wird ein Punkt und nicht, wie im Deutschen üblich, ein Komma verwendet.
- Bei den Notationen aus der Mengenlehre und zum Funktionsbegriff folgen wir den Ausführungen bei Rohwer und Pötter (2001, S. 21ff.).

Inhalt

Kapitel 1

Einleitung

1. Verwendung statistischer Methoden.
2. Ein abstrakter Gesellschaftsbegriff.
3. Abgrenzung von Gesellschaften.
4. Konstruktionen gesellschaftlicher Verhältnisse.
5. Modelle und modale Fragestellungen.

1. Verwendung statistischer Methoden. In der empirischen Sozialforschung[1] werden vornehmlich statistische Methoden verwendet. Der wichtigste Grund dafür ist, dass die zu erforschenden gesellschaftlichen Verhältnisse nicht unmittelbar empirisch zugänglich sind.[2] Wohl kann man sich empirisch, also in beobachtender und kommunikativer Anteilnahme, mit den Lebensverhältnissen jeweils einzelner Menschen oder kleiner Gruppen beschäftigen, aber eine Vergegenwärtigung von Aspekten der gesellschaftlichen Verhältnisse der gesamten Bevölkerung eines größeren Gebietes kann bestenfalls aus Einzelbeobachtungen konstruiert werden. Statistische Methoden werden verwendet, um solche Konstruktionen in methodisch kontrollierter Weise durchführen zu können.[3]

Somit dienen statistische Methoden in der Sozialforschung zunächst dem Zweck, Aspekte gesellschaftlicher Verhältnisse *darzustellen*. Dafür gibt es zahlreiche unterschiedliche Möglichkeiten, insbesondere dann, wenn man versuchen möchte, auch Aspekte sozialen Wandels, der Veränderung gesellschaftlicher Verhältnisse, zu erfassen. In späteren Kapiteln dieses Textes werden wir uns damit ausführlich beschäftigen.

Schwieriger ist die Frage, ob bzw. wie mit statistischen Methoden ermittelt werden kann, wie Menschen von ihren gesellschaftlichen Verhältnissen abhängig sind. Insbesondere mit Methoden der Regressionsrechnung können zwar Zusammenhänge zwischen Merkmalen, die den jeweiligen Untersuchungseinheiten (Individuen, Haushalte, Unternehmen) zurechenbar sind, ermittelt und dargestellt werden; zum Beispiel zwischen Bildungsniveau und Einkommenshöhe bei abhängig Beschäftigten oder zwischen dem Alter arbeitsloser Personen und der Dauer ihrer Arbeitslosigkeit. Bei solchen Zusammenhängen handelt es sich aber offensichtlich nicht um „Naturgesetze“, sondern sie sind wiederum als Aspekte gesellschaftlicher Verhältnisse aufzufassen. Somit kann und sollte man nicht nur untersuchen, wie sie sich historisch verändern; es stellt sich vor allem auch die theoretische Frage, durch welche Prozesse die statistisch ermittelbaren Zusammenhänge zustande kommen und sich verändern.

[1] Wir gehen in diesem Text von einem weiten Verständnis des Begriffs ‘empirische Sozialforschung’ aus, so dass zu ihrem Gegenstand grundsätzlich alle, insbesondere auch wirtschaftliche Aspekte des gesellschaftlichen Lebens von Menschen gehören.

[2] Nicht einmal „kleine Gemeinden“ sind, wie noch René König (1958a: 43) unterstellt hat, unmittelbar überschaubar.

[3] Der theoretische Ansatz statistischer Begriffsbildungen wird im Abschnitt 2.1 besprochen.

2. Ein abstrakter Gesellschaftsbegriff. Im Unterschied zu dem problematischen und nicht ohne weiteres klaren Ausdruck ‘gesellschaftliche Verhältnisse’ verwenden wir in diesem Text den Gesellschaftsbegriff in einer bestimmten Bedeutung: zum Verweis auf (irgendwie räumlich und zeitlich abgegrenzte) Gesamtheiten von Menschen.[4] Zur Rechtfertigung dieses Gesellschaftsbegriffs, der im wesentlichen mit dem einer Bevölkerung identisch ist,[5] sei kurz auf einige Schwierigkeiten hingewiesen, die auftreten, wenn man sogleich mit „sozialen Beziehungen“ *beginnen* möchte. Zum Beispiel findet sich in einer Arbeit von Heinrich Popitz über „Die normative Konstruktion von Gesellschaft“ (1980: 1) folgende Bemerkung:

> „‘Gesellschaft’ – was wir mit diesem Begriff alles meinen und meinen könnten, ist uferlos. Mindestens aber unterstellen wir, daß mehrere Menschen aufeinander bezogen sind, indem sie ihr Verhalten aneinander orientieren.“

Dem ersten Satz kann man sofort zustimmen;[6] verschafft aber die dann folgende Überlegung einen geeigneten Ausgangspunkt? Dass Menschen ihr Verhalten in vielen Situationen aneinander orientieren, ist sicherlich von grundlegender Bedeutung. Aber sobald man bei dem Wort ‘Gesellschaft’ an eine größere Anzahl von Menschen denkt, zum Beispiel an die Gesamtheit der Menschen, die gegenwärtig in Deutschland leben, merkt man, dass eine solche Gesamtheit nicht dadurch definiert werden kann, dass ihre Mitglieder ihr Verhalten aneinander orientieren. Denn die meisten von ihnen sind sich noch nie begegnet und werden sich auch in Zukunft nicht begegnen, und sie können deshalb – selbst wenn sie wollten – ihr Verhalten nicht aneinander orientieren.[7] Offenbar genügt es nicht, bei der allgemeinen Idee

[4] Dies entspricht nach Theodor Geiger (1931: 202) einer wörtlichen Bedeutung: „Gesellschaft bedeutet wörtlich den Inbegriff räumlich vereint lebender oder vorübergehend auf einem Raum vereinter Personen.“

[5] Zum Beispiel versteht Jürgen Friedrichs (1977: 51) unter einer „Bevölkerung [. . .] die Gesamtheit aller auf einem abgrenzbaren Gebiet vorhandenen oder lebenden Individuen.“

[6] Unterschiedliche Ausführungen zum Gesellschaftsbegriff findet man etwa bei T. Geiger (1931), B. Nikles und J. Weiß (1975), H. Esser (1993: 323ff.), H.P. Bahrdt (1994: 181ff.) und J. Ritsert (2000).

[7] Hier muss eine Ambivalenz beachtet werden, die sich anhand folgender Bemerkung von Claude Lévi-Strauss (1953: 536) erläutern lässt: „A society consists of individuals and groups which communicate with one another.“ Man kann dabei an eine Menge von Menschen denken, in der jedes Mitglied entweder mit jedem anderen oder mit mindestens einem anderen Mitglied kommuniziert. In beiden Fällen gelangt man offenbar nicht zu Abgrenzungen zwischen Gesellschaften. Im ersten Fall gelangt man zu einer

eines gesellschaftlichen Zusammenlebens von Menschen nur an unmittelbare Interaktionsprozesse zu denken, sondern es muss berücksichtigt werden, dass Menschen auch auf indirekte und nicht unmittelbar durchschaubare Weise voneinander abhängig sein können.

Deshalb eignet sich auch ein etwas abstrakter ansetzender Gedanke von Georg Simmel (1908: 4) nicht: „Ich gehe [...] von der weitesten, den Streit um Definition möglichst vermeidenden Vorstellung der Gesellschaft aus: daß sie da existiert, wo mehrere Individuen in Wechselwirkung treten.“ Denn die meisten Mitglieder einer Gesellschaft treten überhaupt nicht „in Wechselwirkung“ miteinander. Einen sinnvollen Anknüpfungspunkt findet man jedoch etwas später (S. 8), wo Simmel die Idee einer „Wechselwirkung“ zunächst zurückstellt und stattdessen ausführt:

> „Der Begriff der Gesellschaft deckt zwei, für die wissenschaftliche Behandlung streng auseinander zu haltende Bedeutungen. Sie ist einmal der Komplex vergesellschafteter Individuen, das gesellschaftliche geformte Menschenmaterial, wie es die ganze historische Wirklichkeit ausmacht. Dann aber ist ‘Gesellschaft’ auch die Summe jener Beziehungsformen, vermöge deren aus den Individuen eben die Gesellschaft im ersten Sinne wird.“

Man erkennt nämlich, dass sich nur die erste dieser beiden Bedeutungen eignet, *um zu beginnen.* Denn um von Beziehungen (oder noch abstrakter von Beziehungsformen) sprechen zu können, benötigt man zunächst Vorstellungen über eine Mehrzahl von Menschen, die sich irgendwie in Beziehungen befinden (können). Das gilt im übrigen nicht nur für Gesellschaften, die aus einer großen Anzahl von Menschen bestehen, sondern auch für kleine Gruppen. Um zum Beispiel die Beziehungen in einer Familie darzustellen, muss man zunächst die Familie als eine Gesamtheit von Personen bestimmen.[8]

großen Anzahl sich überschneidender „Cliquen“ (im Sinne der Netzwerkanalyse), und im zweiten Fall gelangt man zur Gesamtheit aller jeweils lebenden Menschen. – Auf letzteres wurde bereits von A. R. Radcliffe-Brown (1940: 193) hingewiesen: „It is rarely that we find a community that is absolutely isolated, having no outside contact. At the present moment of history, the network of social relations spreads over the whole world, without any absolute solution of continuity anywhere. This gives rise to a difficulty which I do not think that sociologists have really faced, the difficulty of defining what is meant by the term ‘a society’.“ Diese Schwierigkeit tritt allerdings nur auf, wenn man versucht, zur Definition einer Gesellschaft gedanklich auf soziale Beziehungen zwischen ihren Mitgliedern zurückzugreifen.

[8] Diese Reihenfolge, nämlich zunächst von Menschen auszugehen, bevor man in einem zweiten Schritt über mögliche Beziehungen nachdenkt, erlaubt es auch, einen bei Simmel naheliegenden Fehler zu vermeiden, der darin besteht, Beziehungsformen als Bedingungen für das Verhalten der jeweils Beteiligten aufzufassen. Simmels oben zitierte Rede von Beziehungsformen, „vermöge deren aus den Individuen eben die Gesellschaft im ersten Sinne wird“, ist offenbar zweideutig. – Allerdings war sich Simmel des Problems durchaus bewusst, denn an anderer Stelle bemerkt er: „Es gehört zu den häufigsten Ausartungen des menschlichen Kausaltriebes, formale Bedingungen, ohne die bestimmte Ereignisse nicht stattfinden können, für positive, produktive Ursachen derselben zu halten.“ (Simmel 1903/1983: 221)

Jedenfalls für die empirische Sozialforschung erscheint es deshalb sinnvoll, nicht mit Beziehungen zu beginnen – weder konkret mit „Wechselwirkungen“ oder „Sich-aneinander-Orientieren“, noch abstrakt mit „Beziehungsformen“ –, sondern mit Gesamtheiten von Menschen. Diese Überlegung führt zu einem einfachen Gesellschaftsbegriff: Eine (menschliche) Gesellschaft ist eine Menge von Menschen. Der Zusatz ‘menschlich’ ist gegebenenfalls erforderlich, weil man in dieser allgemeinen Bedeutung auch bei anderen Lebewesen von Gesellschaften sprechen kann. Wenn im Folgenden ohne Zusatz von Gesellschaften gesprochen wird, sind jedoch stets menschliche Gesellschaften gemeint.

Ich nenne dies den *statistischen Gesellschaftsbegriff,* da sich die Definition auf die Vorstellung einer statistischen Gesamtheit, also einer *Menge* von Menschen beschränkt (das wird im Abschnitt 2.1 näher ausgeführt). Das ist eine abstrakte Vorstellung, inbesondere in folgenden Hinsichten:

– Es wird von Beziehungen zwischen den Mitgliedern der Gesellschaft abstrahiert.[9] Natürlich wird durch die Begriffsbildung nicht ausgeschlossen, solche Beziehungen zu ermitteln und in ihrer Relevanz für das Verhalten der Gesellschaftsmitglieder zu untersuchen.

– Es wird von der räumlichen Umwelt abstrahiert, in der sich das Leben der Menschen abspielt, also insbesondere von der gesamten materiellen Kultur, wie etwa Straßen und Häuser, die sich Menschen als ihren Lebensraum geschaffen haben. Aber das ist eigentlich keine Abstraktion, sondern eine Unterscheidung. Es wird ja nicht bestritten, dass man sich auf einen räumlichen Kontext beziehen muss, sobald man in empirisch bestimmter Weise von einer Gesellschaft sprechen möchte.

– Es wird unterschieden zwischen einer Gesellschaft im Sinne einer Gesamtheit von Menschen und den „gesellschaftlichen Verhältnissen“, in denen diese Menschen leben. Das ist indessen schon deshalb sinnvoll, weil sich ein bestimmter Begriff gesellschaftlicher Verhältnisse nicht ohne weiteres definieren lässt.

Schließlich ist bemerkenswert, dass der statistische Gesellschaftsbegiff eine klare Antwort auf die Frage erlaubt, *worüber* man spricht, wenn man über eine Gesellschaft spricht: nämlich über die Menschen, die der Gesellschaft als Mitglieder angehören. Wie bereits gesagt wurde, impliziert die Begriffsbildung nicht, dass man sich auf statistische Aussagen über eine Gesellschaft beschränkt.

Es sei auch betont, dass mit diesem Definitionsvorschlag nicht nahegelegt werden soll, dass es in der empirischen Sozialforschung ausschließlich oder auch nur hauptsächlich um Aussagen über Gesamtheiten von Menschen geht oder – in einer an Lebensverläufen orientierten Sozialfor-

[9] Und zwar: vollständig; wir verwenden also den Ausdruck ‘Gesellschaft’ auch nicht als „terminus technicus für [irgend-] eine bestimmte Integrationsebene des Universums“, wie einmal von Norbert Elias (1996: 79) vorgeschlagen wurde.

schung – um Aussagen über Aspekte individueller Lebensverläufe. Thema der empirischen Sozialforschung sind allgemein gesellschaftliche Verhältnisse, die auf viele unterschiedliche Weisen konzeptualisiert, beschrieben und modelliert werden können, insbesondere durch eine Betrachtung von Institutionen und Organisationen.

3. Abgrenzung von Gesellschaften. Wenn man von „der menschlichen Gesellschaft“ spricht, ist nach dem hier vertretenen begrifflichen Ansatz die Gesamtheit aller Menschen gemeint, die gegenwärtig leben oder in einem bestimmten Zeitraum gelebt haben. Gleichwohl kann es oft zweckmäßig oder aus Gründen der Verfügbarkeit von Daten unvermeidlich sein, kleinere Teilgesamtheiten als spezifische Gesellschaften zu fixieren. Offenbar erlaubt die im vorangegangenen Paragraphen gegebene Definition, bei der Bildung solcher Teilgesamtheiten beliebig vorzugehen; insbesondere wird für den Gesellschaftsbegriff nicht gefordert, dass man in irgendeinem Sinne von einer „realen Einheit“ der zu einer Gesellschaft zusammengefassten Menschen sprechen kann.[10]

Unbeschadet dieser grundsätzlichen Freiheit bei der Bildung von Teilgesellschaften kann es oft sinnvoll sein, sich an vorgegebenen Abgrenzungen und Unterscheidungen zu orientieren. Dafür kommen in erster Linie politische und verwaltungstechnische Abgrenzungen in Betracht, die durch staatliche Institutionen vorgenommen werden.[11] So kann man etwa staatlich organisierte Gesellschaften unterscheiden und z.B. von Deutschland, Dänemark, Polen usw. sprechen. Es ist jedoch wichtig, gleichwohl an der begrifflichen Unterscheidung zwischen Gesellschaften und den staatlichen Institutionen, die zur Abgrenzung verwendet werden, festzuhalten. Wenn man etwa die gegenwärtig in Deutschland lebenden Menschen gedanklich zu einer Gesellschaft zusammenfasst, entsteht begrifflich eine *Menge* von Menschen, aber kein Staat, der vielmehr als eine Gesamtheit von Institutionen definiert werden müsste.[12]

Auch wird, um dies noch einmal zu wiederholen, durch eine solche Abgrenzung einer Gesamtheit keine in irgendeiner Weise „substantielle Einheit“ ihrer Mitglieder postuliert; und zwar unabhängig davon, ob und ggf. wie Menschen *ihre* Zugehörigkeit zu einer Gesellschaft jeweils selbst auffassen. Der unterschiedliche Ansatz wird in folgender Bemerkung von H. Popitz (1995: 126f.) deutlich:

> „Gesellschaften im hier [von Popitz] gemeinten Sinne sind *soziale Einheiten.* [. . .] Soziale Einheiten sind leicht zu erkennen, weil sie Wert darauf legen, sich erkennbar zu machen. Sie ziehen Grenzen zwischen Innen und Außen, Drinsein und Draußensein. Wer als Zugehöriger anerkannt wird, muß über bestimmte Qualitäten verfügen, angeborene wie Geschlecht und Herkunft, oder erworbene wie bestimmte Leistungen oder Bewährungen. Vergesellschaftung bedeutet, daß Menschen in Strukturen von Zugehörigkeiten leben. Also in Ein- und Ausgrenzungen. Das zellenbildende Prinzip der Vergesellschaftung ist ein Prinzip der Grenzziehung. Soziale Einheiten sind das Produkt solcher Alternativen.“

Diese Überlegung führt jedoch in eine Sackgasse; zunächst schon rein formal durch ihre Grammatik, die „Gesellschaften“ bzw. „soziale Einheiten“ zu Subjekten ihrer eigenen Abgrenzung macht. Wollte man aber versuchen, sie empirisch zu wenden, würde man schließlich nur irgendwelche partikularen Ideologien finden, die zwar von Sozialwissenschaftlern ideologiekritisch thematisiert, nicht aber mit ihren eigenen Begriffsbildungen vermengt werden sollten.

4. Konstruktionen gesellschaftlicher Verhältnisse. Wir werden nicht versuchen, auch einen bestimmten Begriff gesellschaftlicher Verhältnisse zu definieren, denn gesellschaftliche Verhältnisse können auf unterschiedliche Weisen konzeptualisiert werden. In erster Näherung lassen sich drei solche Möglichkeiten zur Konstruktion von *Aspekten* gesellschaftlicher Verhältnisse unterscheiden:

– Man kann statistische Strukturen verwenden, um Aspekte gesellschaftlicher Verhältnisse zu definieren;

[10] Die Vorstellung, dass Gesellschaften in irgendeinem Sinn „reale Einheiten“ sind, durchzieht die Geschichte der Soziologie; z.B. heißt es bei Emile Durkheim (1888: 41), „daß die Gesellschaft nicht eine einfache Ansammlung von Individuen ist, sondern ein Sein, das seine Besonderheit, sein Leben, sein Bewußtsein, seine Interessen und seine Geschichte hat.“ Dagegen impliziert der Mengenbegriff, den wir zur Definition von Gesellschaften verwenden, nur eine gedankliche Einheit (ohne jedoch weiterhin zu implizieren, dass Gesellschaften „einfache Ansammlungen von Individuen“ sind).

[11] Von Soziologen ist eine solche Orientierung oft als „äußerlich“ kritisiert und gefordert worden, stattdessen von der psychischen Verfassung der jeweils beteiligten Menschen auszugehen. Als Beispiel sei hier auf René Königs Überlegungen zur Soziologie der Gemeinde (1958a) hingewiesen, in denen er „die grundsätzliche Verwechslung zwischen der Gemeinde als Verwaltungseinheit und der Gemeinde als sozialer Wirklichkeit“ (S. 7) kritisiert. Begrifflicher Ausgangspunkt ist für König die Idee einer „globalen Gesellschaft“: „eine mehr oder weniger große lokale und gesellschaftliche Einheit, in der Menschen zusammenwirken, um ihr wirtschaftliches, soziales und kulturelles Leben gemeinsam zu fristen.“ (S. 26) Gemeinden werden dann als besondere Erscheinungsformen „globaler Gesellschaften“ bestimmt (ebda.), und weiterhin „als ein <soziales System>, d.h. als ein Zusammenhang, der sich unter anderem dadurch auszeichnet, daß alle Menschen, die in ihn einbeschlossen sind, ein Bewußtsein dieses Zusammenhangs sowie seiner Grenzen und seiner Verschiedenheit von anderen ähnlichen Zusammenhängen haben.“ (S. 29) Wie bereits von Hans Linde (1972: 19ff.) kritisiert wurde, ist es offenbar fragwürdig, Gemeindegrenzen durch Zusammengehörigkeitsgefühle zu bestimmen; ganz abgesehen davon, ob und in welchen Erscheinungsformen es solche Gefühle überhaupt gibt (im Unterschied zu einer Kenntnis von formalen Zugehörigkeiten zu Gemeinden, die deren vorgängige, typischerweise verwaltungstechnische Definition voraussetzt). Aber auch Königs Begriff „globaler Gesellschaften“ eignet sich nicht, um Teilgesellschaften abzugrenzen. Denn einerseits bleibt vollständig unklar, was in diesem Zusammenhang „gemeinsames Zusammenwirken“ bedeuten könnte; denkt man andererseits an staatlich oder verwaltungstechnisch abgegrenzte Teilgesellschaften, gibt es wohl immer auch Beziehungen, die über die jeweiligen Grenzen hinausgehen.

[12] Im Übrigen sprechen nicht nur begriffliche, sondern auch politische Gründe für eine Unterscheidung von Gesellschaften und Staaten, wie z.B. von E.-W. Böckenförde (1972) ausgeführt wurde.

- man kann (stattdessen oder ergänzend) von Beziehungen ausgehen und soziale Netzwerke konstruieren;
- man kann sich auf Institutionen und Organisationen beziehen und sie als Bedingungen des Lebens von Menschen in einer Gesellschaft thematisieren.

Wir werden später diese drei Möglichkeiten genauer besprechen und auch versuchen, einige Zusammenhänge herzustellen.

5. Modelle und modale Fragestellungen. Schließlich soll noch erläutert werden, wie in diesem Text von Modellen gesprochen wird. Denn in der Literatur findet man oft die Auffassung, dass Modelle als „vereinfachende Beschreibungen“ von Ausschnitten der menschlichen Erfahrungswelt verstanden werden können; hier sind zwei typische Formulierungen:

„A *scientific model* is an abstract and simplified description of a given phenomena.“ (Olkin, Gleser und Derman 1980: 2) „A model of any set of phenomena is a formal representation thereof in which certain features are abstracted while others are ignored with the intent of providing a simpler description of the salient aspects of the chosen phenomena.“ (Hendry und Richard 1982: 4)

Sehr ähnlich sind Formulierungen, in denen von „Abbildungen“ gesprochen wird, zum Beispiel:

„Modelle können wir uns in erster Näherung denken als begriffliche Konstrukte zur ‘Abbildung’ realer Systeme oder zum Umgang mit solchen.“ (Balzer 1997: 16) „Ein Modell ist wohl immer aufzufassen als eine Abbildung. Die Frage ist nur, was abgebildet wird, und wie die Abbildungsfunktion aussieht.“ (Frey 1961: 89) „Ein Modell ist eine Abbildung von für die jeweilige Fragestellung bedeutsamen Teilaspekten der Wirklichkeit zu einem vereinfachten System.“ (Wirth 1979: 130f.) „Ein Modell ist immer eine vereinfachte Abbildung eines interessierenden Realitätsausschnitts.“ (Bossel 1992: 27)

In vielen, vermutlich sogar den meisten Fällen haben jedoch Modelle, wie sie in der sozialwissenschaftlichen Literatur konstruiert und diskutiert werden, nicht die Aufgabe, Ausschnitte unserer Erfahrungswelt zu beschreiben oder „abzubilden“. Beispielsweise kann man an demographische Modelle denken, die dem Zweck dienen, *mögliche* Bevölkerungsentwicklungen vorstellbar zu machen, die aus hypothetischen Annahmen über Geburten, Todesfälle und Migrationen ableitbar sind. Offenbar kann man nicht sagen, dass durch Modelle dieser Art reale historische Prozesse beschrieben werden.[13] Dieses Beispiel liefert auch einen allgemeinen Gesichtspunkt: In vielen Fällen dienen Modelle dem Nachdenken über Möglichkeiten und Wahrscheinlichkeiten; oder in einer kurzen Formulierung: *Modelle sind Hilfsmittel zur Reflexion modaler Fragestellungen,* wobei sich diese Fragestellungen sowohl auf zukünftige Möglichkeiten als auch auf die Beschaffenheit bereits realisierter Sachverhalte, über die nur unzureichende Informationen verfügbar sind, beziehen können. Die als Leitfaden dienenden modalen Fragestellungen können natürlich sehr unterschiedlich sein und müssen bei der Konstruktion von Modellen erläutert werden.

[13] Hierbei setze ich ein bestimmtes Verständnis des Worts *‘beschreiben’* voraus, insbesondere: dass nur Sachverhalte beschrieben werden können, die es in der menschlichen Erfahrungswelt tatsächlich gibt oder gegeben hat, so dass empirische Feststellungen möglich sind. Allerdings wird das Wort oft in einer weiteren Bedeutung verwendet, so dass auch Darstellungen fiktiver Sachverhalte als Beschreibungen bezeichnet werden können; dann sollte ggf. zwischen empirischen und fiktiven Beschreibungen unterschieden werden.

Kapitel 2

Statistische Begriffsbildungen

In diesem Kapitel beschäftigen wir uns zunächst mit dem gedanklichen Ansatz der in der empirischen Sozialforschung verwendeten statistischen Begriffsbildungen. Dann folgen Überlegungen zu statistischen (Sozial-) Strukturbegriffen.

2.1 Statistische Variablen und Verteilungen

1. Bezugnahme auf Gesamtheiten. Die Entwicklung der Statistik kann als eine Folge des Wunsches verstanden werden, empirisch explizierbare Vorstellungen über Gesamtheiten zu gewinnen, die nicht unmittelbar überschaubar sind. Ursprünglich ging es in erster Linie um eine Erfassung von Bevölkerungen (Populationen). Etwa seit der zweiten Hälfte des 19. Jahrhunderts hat sich die Statistik zu einer abstrakten Methodenwissenschaft entwickelt, deren Begriffsbildungen auf beliebige Gesamtheiten anwendbar sind. Dass in irgendeiner Weise eine Bezugnahme auf Gesamtheiten erfolgt, ist jedoch in jedem Fall relevant, um statistische Aussagen zu verstehen.[1] Maurice Kendall und Alan Stuart haben das zu Beginn ihrer „Advanced Theory of Statistics“ (1977: 1) so ausgedrückt:

[1] Allerdings gibt es auch eine konkurrierende Idee: wiederholbare Verfahren. An dieser Idee orientieren sich die Wahrscheinlichkeitstheorie und die *probabilistische* Statistik.

„The fundamental notion in statistical theory is that of the group or aggregate, a concept for which statisticians use a special word – “population”. This term will be generally employed to denote any collection of objects under consideration, whether animate or inanimate; for example, we shall consider populations of men, of plants, of mistakes in reading a scale, of barometric heights on different days, and even populations of ideas, such as that of the possible ways in which a hand of cards might be dealt. [...] The science of Statistics deals with the properties of populations. In considering a population of men we are not interested, statistically speaking, in whether some particular individual has brown eyes or is a forger, but rather in how many of the individuals have brown eyes or are forgers, and whether the possession of brown eyes goes with a propensity to forgery in the population. We are, so to speak, concerned with the properties of the population itself. Such a standpoint can occur in physics as well as in demographic sciences.“

In diesem Zitat wird auch schon darauf hingewiesen, dass sich statistische Aussagen *in spezifischer Weise* auf Gesamtheiten beziehen; das wird weiter unten (in § 6) genauer besprochen. Bereits an dieser Stelle kann aber festgestellt werden, dass Aussagen über Gesamtheiten von Aussagen über ihre individuellen Mitglieder zu unterscheiden sind.

2. Gesamtheiten als Mengen. Wenn in der Statistik von Gesamtheiten gesprochen wird, sind Mengen im Sinne der Mengenlehre gemeint, d.h. Zusammenfassungen von Elementen zu einer gedanklichen Einheit, wobei von allen möglicherweise vorhandenen Beziehungen zwischen den Elementen abstrahiert wird. Der Begründer der Mengenlehre, Georg Cantor (1845–1918), hat einmal folgende Definition gegeben:

„Unter einer „Menge“ verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten unsrer Anschauung oder unseres Denkens (welche die „Elemente“ von M genannt werden) zu einem Ganzen.“ (Cantor 1962: 282)

Der gedankliche Ansatz ist allgemein und abstrakt. Es gibt keinerlei Einschränkungen hinsichtlich der Arten von Objekten, die man gedanklich zu einer Menge zusammenfassen kann. Es muss sich auch nicht unbedingt um materielle Objekte im umgangssprachlichen Sinn dieses Worts handeln, zum Beispiel können auch Zahlen, Eigenschaften, Ereignisse und Gebiete eines Raums zu Mengen zusammengefasst werden, und auch Mengen selbst können wiederum als Elemente zur Definition neuer Mengen verwendet werden.[2]

Wichtig ist auch, dass mit dem Mengenbegriff nur eine gedankliche Einheit der jeweils in Betracht gezogenen Elemente gemeint ist (von der Frage, ob und ggf. in welcher Weise der Menge auch eine „reale Einheit“ entspricht, wird also abgesehen). Ebenfalls wird von allen Beziehungen

[2] In diesem Text wird eine Kenntnis der Grundbegriffe der Mengenlehre vorausgesetzt. Kurze Erläuterungen der für die Statistik relevanten Begriffsbildungen findet man bei Rohwer und Pötter (2001: 21ff.).

abstrahiert, die möglicherweise zwischen den zu einer Menge zusammengefassten Elementen bestehen bzw. hergestellt werden können. Infolgedessen wird auch von räumlichen oder zeitlichen Anordnungen der Elemente, soweit man ggf. davon sprechen kann, abgesehen. Insbesondere spielt die Reihenfolge, in der die Elemente einer Menge vorgestellt oder aufgeschrieben werden, keine Rolle.[3]

3. Repräsentation von Gesamtheiten. Im Rahmen statistischer Überlegungen werden Gesamtheiten immer als Mengen aufgefasst, und wenn in diesem Text ohne weiteren Zusatz von Gesamtheiten gesprochen wird, sind deshalb stets Mengen gemeint. Bei ihren Elementen kann es sich um real existierende oder um fiktive Objekte handeln. Im ersten Fall gibt es die Objekte in der menschlichen Erfahrungswelt (einschließlich der empirisch zugänglichen Vergangenheit), im zweiten Fall gibt es sie nur in der Vorstellungswelt eines oder mehrerer Menschen. Dementsprechend kann man im ersten Fall von *empirischen*, im zweiten Fall von *fiktiven Gesamtheiten* sprechen. Fiktive Gesamtheiten können endlich oder unendlich viele Elemente enthalten, empirische Gesamtheiten haben jedoch immer nur endlich viele Elemente. Weiterhin gilt natürlich auch, dass empirische Gesamtheiten nur Elemente enthalten können, die es in der *bisherigen* Erfahrungswelt von Menschen tatsächlich gibt oder gegeben hat, also insbesondere keine möglicherweise in der Zukunft existierenden Objekte.

In der *Sozialstatistik* beschäftigt man sich mit empirischen Gesamtheiten.[4] Dabei ist zu berücksichtigen, dass man sich oft nicht unmittelbar und vollständig auf alle Elemente einer intendierten empirischen Gesamtheit beziehen kann. Als Beispiel kann man an die Gesamtheit der Menschen

[3] Dies muss auch deshalb betont werden, weil in der soziologischen Literatur (und in der Umgangssprache) der Mengenbegriff gelegentlich anders verwendet wird. Zum Beispiel schreiben R. Boudon und F. Bourricaud in ihren „Soziologischen Stichworten“ (1992: 184): „Eine Menge und eine Masse sind nicht dasselbe. Die beiden Bezeichnungen beziehen sich auf unterschiedliche soziale Situationen. In einer Menge, die einem Fußballspiel beiwohnt, stehen die Beteiligten in Interaktionsbeziehungen zueinander. Die einen pfeifen, die anderen klatschen; und in beiden Lagern entwickelt sich eine Solidarität sowie – je nach dem Grad ihrer Begeisterung – eine Differenzierung zwischen den Fans. [. . .] Die *Masse* derjenigen dagegen, die ein Fernsehprogramm verfolgen oder eine Zeitung lesen, hat kaum Möglichkeiten, miteinander in Kontakt zu treten. Außerdem kommen die Beziehungen zwischen ihnen nur durch die Vermittlung der ausgestrahlten Sendung oder der Druckseite zustande. Ihre Gemeinsamkeiten beschränken sich darauf, daß sie Leser derselben Zeitung oder Zuschauer desselben Programms sind.“ Offenbar wird hier von einer „Menge“ im Unterschied zu einer „Masse“ dann gesprochen, wenn es zwischen ihren Elemente gewisse Interaktionsbeziehungen gibt. Es ist natürlich zulässig, das Wort ‘Menge’ auch in dieser Bedeutung zu verwenden; es ist aber wichtig zu wissen, dass das Wort in der Mengenlehre und der sich an sie anschließenden Statistik anders verwendet wird, nämlich in einer Bedeutung, die von allen möglicherweise vorhandenen Beziehungen zwischen den Elementen abstrahiert.

[4] In der Literatur findet man gelegentlich die Wortkombination *Wirtschafts- und Sozialstatistik*. In diesem Text wird von Sozialstatistik in einem umfassenden Sinn gesprochen, der Bezugnahmen auf wirtschaftliche Sachverhalte einschließt.

denken, die in Deutschland im September 2003 arbeitslos gewesen sind. Offenbar ist eine empirische Gesamtheit gemeint, die in diesem Beispiel aus Menschen besteht, die im angegebenen Zeitraum tatsächlich gelebt haben.[5] Aber es ist auch klar, dass diese Gesamtheit nicht unmittelbar beobachtet werden kann. Deshalb ist man gezwungen, sich in irgendeiner Form eine Repräsentation der Gesamtheit, über die man sprechen möchte, zu verschaffen.

Mit Repräsentationen sind in diesem Zusammenhang gegenständliche oder sprachliche Hilfsmittel gemeint, die es erlauben sollen, sich die Elemente einer nicht unmittelbar überschaubaren Gesamtheit zu vergegenwärtigen. Je nach dem verfügbaren Vorwissen gibt es dafür unterschiedliche Möglichkeiten. Als Beispiel kann man an eine Kartei im Personalbüro eines Unternehmens denken, die für jede in dem Unternehmen beschäftigte Person eine Karteikarte mit Informationen über die Person enthält. Eine solche Kartei repräsentiert dann (im hier gemeinten Sinn) die Belegschaft des Unternehmens, wobei es gleichgültig ist, in welchen technischen Formen die Kartei existiert (etwa in Form eines Karteikastens mit Karteikarten oder in Gestalt einer Datei in einem Computer).

Eine minimale Anforderung an eine Repräsentation besteht darin, dass es für die Elemente der intendierten Gesamtheit Namen gibt, die ihre Unterscheidung ermöglichen. Solche Namen benötigt man auch dann, wenn eine empirische Identifikation noch gar nicht stattgefunden hat, sondern zunächst nur als Möglichkeit vorstellbar ist (wie zum Beispiel bei den Menschen, die im September 2003 in Deutschland arbeitslos gewesen sind). Folgende allgemeine Notation eignet sich sowohl für empirische als auch für fiktive Gesamtheiten: $\Omega := \{\omega_1, \ldots, \omega_n\}$. In dieser Notation sind die Symbole $\omega_1, \ldots, \omega_n$ Namen der Elemente (wobei die natürliche Zahl n auf die Anzahl der Namen verweist), und die Mengenklammern geben an, dass sie (die Namen bzw. die durch sie repräsentierten Elemente) zu einer Menge zusammengefasst werden sollen. Schließlich erhält diese Menge per Definition den Namen Ω.

4. Statistische Variablen. Statistische Aussagen über Gesamtheiten gehen von deren Elementen aus. Die einfachste Aussage stellt nur fest, wieviele Elemente die Gesamtheit enthält. Alle weiteren statistischen Aussagen über Gesamtheiten gehen von Eigenschaften aus, die sich zunächst ihren Elementen zurechnen lassen. Solche Eigenschaften werden durch *statistische Variablen* repräsentiert, die allgemein als *Funktionen im mathematischen Sinn* definiert sind. Zur Erläuterung kann folgendes Schema verwendet werden:

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

[5] Ersichtlich sind sowohl zeitliche als auch räumliche Bezugnahmen erforderlich. Eine explizite Berücksichtigung der zeitlichen Bezüge ist insbesondere dann wichtig, wenn Prozesse dargestellt werden sollen; das wird in Kapitel 4 besprochen.

Hierbei ist Ω eine statistische Gesamtheit (wir sprechen oft von einer *Objektmenge* oder auch von einer *Referenzmenge*), und $\tilde{\mathcal{X}}$ ist der *Merkmalsraum* der Variablen, d.h. eine Menge von Attributen, so dass jedes Element von Ω durch genau eines dieser Attribute charakterisiert werden kann. Schließlich ist X der Name der statistischen Variablen, also der Funktion, die jedem Element das ihm entsprechende Attribut zuordnet.

Als Beispiel kann man an eine statistische Variable denken, die jedem Mitglied einer Gesamtheit von Menschen sein Geschlecht zuordnet, also entweder das Attribut 'männlich' oder das Attribut 'weiblich'. Offenbar kann man diese Attribute auch durch Zahlen repräsentieren, also etwa einen Merkmalsraum $\tilde{\mathcal{X}} := \{0, 1\}$ verwenden und vereinbaren, dass die Zahl 0 das Attribut 'männlich' und die Zahl 1 das Attribut 'weiblich' bedeuten soll. In diesem Beispiel handelt es sich um einen *qualitativen* Merkmalsraum, womit gemeint ist, dass es für die Elemente des Merkmalsraums keine sinnvolle lineare Ordnung gibt. Dagegen sind *quantitative* Merkmalsräume dadurch definiert, dass es für ihre Elemente eine sinnvolle lineare Ordnung gibt; als Beispiel kann man an einen Merkmalsraum $\tilde{\mathcal{Y}} := \{0, 1, 2, 3, \ldots\}$ denken, dessen Elemente zur Feststellung des Alters von Menschen (in diesem Beispiel in vollendeten Lebensjahren) verwendet werden können.[6]

Es sei betont, dass statistische Variablen Funktionen sind und von *logischen Variablen* (Leerstellen in Aussageformen) unterschieden werden müssen.[7] Außerdem dürfen statistische Variablen nicht mit ihren Merkmalsräumen verwechselt werden, wie dies gelegentlich in der Methodenliteratur geschieht.[8] Man kann natürlich abkürzend ohne Zusatz von Variablen sprechen, wenn aus dem Kontext hervorgeht, ob statistische oder logische Variablen gemeint sind.

Zum Verständnis ist auch zu beachten, dass das Wort 'Funktion' in unterschiedlichen Bedeutungen verwendet werden kann. Hauptsächlich sind zwei Verwendungsmöglichkeiten zu unterscheiden. Einerseits eine Verwendung, in der das Wort 'Funktion' auf einen Zweck, eine Leistung oder eine Aufgabe verweisen soll; andererseits die mathematische Verwendung des Funktionsbegriffs, in der das Wort die Zuordnung der Elemente einer Menge zu Elementen derselben oder einer anderen Menge meint. In Teilen der soziologischen Literatur (insbesondere im Umkreis sogenannter „funktionalistischer" Theorieansätze) wird das Wort in der ersten dieser beiden Bedeutungen verwendet;[9] wir werden das Wort in diesem Text jedoch ausschließlich in seiner mathematischen Bedeutung verwenden.[10]

[6] Ausführlichere Überlegungen zu unterschiedlichen Arten von Merkmalsräumen findet man bei Rohwer und Pötter (2002a, Kap. 4).

[7] Dazu ausführlich Rohwer und Pötter (2002b, Kap. 9).

[8] Unklare Verwendungen des Variablenbegriffs in der Methodenliteratur werden bei Rohwer und Pötter (2002a: 14ff.) besprochen.

[9] Zum Beispiel schreibt H. Joas in einer Einführung für ein Lehrbuch der Soziologie (2001: 21): „Der Ausdruck „Funktion" bezeichnet den Beitrag, den jede soziale Beziehung, Position, Organisation, jeder Wert oder jede Eigenschaft einer Gesellschaft für das soziale System als Ganzes leistet. [...] So besteht die Funktion von Schulen darin, Schüler auszubilden, die über die von den Unternehmen geforderten Fertigkeiten verfügen und am öffentlichen Leben als Bürger ihres Landes teilnehmen können." Eine Besprechung unterschiedlicher Verwendungsweisen des Funktionsbegriffs in der soziologischen Literatur findet man bei R. K. Merton (1957: 20ff.).

Tabelle 2.1-1 Fiktive Daten für eine eindimensionale statistische Variable X (links) und eine zweidimensionale statistische Variable (X, Y) (rechts).

ω	$X(\omega)$
ω_1	0
ω_2	1
ω_3	0
ω_4	0
ω_5	1
ω_6	0
ω_7	1
ω_8	1
ω_9	0
ω_{10}	0

ω	$X(\omega)$	$Y(\omega)$
ω_1	0	22
ω_2	1	29
ω_3	0	26
ω_4	0	25
ω_5	1	26
ω_6	0	24
ω_7	1	22
ω_8	1	25
ω_9	0	25
ω_{10}	0	23

Es sei auch angemerkt, dass eine Funktion im mathematischen Sinn nicht mit der Vorstellung eines „funktionalen Zusammenhangs" verwechselt werden darf. Bereits zur Interpretation statistischer Variablen passt eine solche Vorstellung offenbar nicht.

5. Mehrdimensionale statistische Variablen. In vielen Fällen ist es möglich und oft von besonderem Interesse, die Elemente einer Gesamtheit gleichzeitig durch zwei oder mehr Arten von Merkmalen zu charakterisieren. Man spricht dann von *mehrdimensionalen statistischen Variablen*, wobei jeder einzelne Merkmalsraum als eine „Dimension" (in einem rein formalen, nicht räumlich aufzufassenden Sinn) zählt. Denkt man zur Illustration wieder an eine Personengesamtheit, könnte jeder Person gleichzeitig ein Geschlecht und ein Alter zugeordnet werden. Dem entspricht dann eine zweidimensionale statistische Variable

$$(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$$

wobei sich der Merkmalsraum $\tilde{\mathcal{X}} := \{0, 1\}$ auf das Geschlecht und der Merkmalsraum $\tilde{\mathcal{Y}} := \{0, 1, 2, 3, \ldots\}$ auf das Alter bezieht. Dementsprechend wäre $(X, Y)(\omega) = (1, 25)$ so zu verstehen, dass ω der Name einer 25jährigen Frau ist.

[10] In der Notation und Terminologie folgen wir den Ausführungen bei Rohwer und Pötter (2001: 24ff.). Zur Geschichte des mathematischen Funktionsbegriff vgl. man H.-G. Steiner (1969).

Tabelle 2.1-1 illustriert die Begriffsbildungen mit fiktiven Daten. Die linke Hälfte illustriert eine eindimensionale, die rechte Hälfte eine zweidimensionale statistische Variable. Die Personengesamtheit ist in beiden Fällen identisch und besteht aus 10 Personen. Die eindimensionale Variable ordnet jeder Person ein Geschlecht zu, die zweidimensionale Variable ordnet jeder Person außerdem ein Alter zu.

6. Der statistische Verteilungsbegriff. Mit dem Begriff einer statistischen Variablen steht ein sehr allgemeines Schema zur Repräsentation von Objekten und ihrer Merkmale zur Verfügung. Kennt man eine statistische Variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, kennt man auch für jedes Element $\omega \in \Omega$ den Merkmalswert $X(\omega)$. Das statistische Erkenntnisinteresse zielt jedoch gar nicht auf ein solches Wissen über die individuellen Mitglieder der jeweiligen Gesamtheit, sondern nur auf das Ausmaß, in dem bestimmte Merkmalswerte in der Gesamtheit vorkommen. Als Beispiel können die Daten für die Variable X in Tabelle 2.1-1 dienen. Aus statistischer Sicht interessiert nicht, dass ω_1 der Name einer männlichen und ω_2 der Name einer weiblichen Person ist, sondern dass es in der Gesamtheit sechs männliche und vier weibliche Personen gibt; oder in relativen Häufigkeiten ausgedrückt: 60 % sind männlich und 40 % sind weiblich.

Diesem spezifischen Erkenntnisinteresse dient der Begriff einer *statistischen Verteilung*.[11] Wie bei statistischen Variablen handelt es sich um Funktionen; aber – und darin kommt der statistische Perspektivenwechsel zum Ausdruck – als Definitionsbereich der Funktion dient jetzt nicht die Objektmenge Ω, sondern die Gesamtheit aller möglichen Merkmalskombinationen, also die Potenzmenge $\mathcal{P}(\tilde{\mathcal{X}})$ des Merkmalsraums $\tilde{\mathcal{X}}$.[12] Also kann folgende Definition gegeben werden:

> Die *Verteilung einer statistischen Variablen* $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ ist eine Funktion, die jeder Teilmenge $\tilde{X}$ des Merkmalsraums $\tilde{\mathcal{X}}$ die (absolute oder relative) Häufigkeit derjenigen Objekte in Ω zuordnet, die einen Merkmalswert in $\tilde{X}$ aufweisen.

Zur Notation verwenden wir $\mathrm{P}^*[X]$, wenn auf absolute Häufigkeiten Bezug genommen wird, und $\mathrm{P}[X]$, wenn auf relative Häufigkeiten Bezug genommen wird. In eckigen Klammern steht der Name der Variablen, deren Verteilung bezeichnet werden soll.[13] Somit gelangt man zu den Definitionen

$$\mathrm{P}^*[X](\tilde{X}) := |\{\omega \in \Omega \,|\, X(\omega) \in \tilde{X}\}| \quad \text{und} \quad \mathrm{P}[X](\tilde{X}) := \frac{\mathrm{P}^*[X](\tilde{X})}{|\Omega|}$$

wobei $\tilde{X}$ eine beliebige Teilmenge von $\tilde{\mathcal{X}}$ ist.[14] Als Konvention wird vereinbart, dass, wenn ohne Zusatz von Häufigkeiten gesprochen wird, stets relative Häufigkeiten gemeint sind. Dies soll analog auch für das Reden von statistischen Verteilungen gelten.

Zur Illustration beziehen wir uns wieder auf die Variable X in Tabelle 2.1-1. In diesem Beispiel ist der Merkmalsraum $\tilde{\mathcal{X}} = \{0, 1\}$, es gibt also vier Teilmengen mit folgenden absoluten bzw. relativen Häufigkeiten:

$\tilde{X}$	$\mathrm{P}^*[X](\tilde{X})$	$\mathrm{P}[X](\tilde{X})$
$\emptyset$	0	0.0
$\{0\}$	6	0.6
$\{1\}$	4	0.4
$\tilde{\mathcal{X}}$	10	1.0

Es sollte beachtet werden, dass als Argumente einer Häufigkeitsfunktion $\mathrm{P}_{[X]}$ nicht Elemente, sondern Teilmengen des Merkmalsraums der Variablen X verwendet werden; solche Teilmengen werden auch *Merkmalsmengen* genannt. Die Berücksichtigung der leeren Menge $\emptyset$ und der Gesamtmenge $\tilde{\mathcal{X}}$ dient natürlich nur der formalen Vollständigkeit.

7. Statistische Aussagen über Gesamtheiten. Bereits zu Beginn dieses Abschnitts wurde betont, dass sich statistische Aussagen stets auf Gesamtheiten beziehen; jetzt kann genauer gesagt werden, dass es sich stets um *Aussagen über statistische Verteilungen* handelt. Dass es sich um eine spezifische Art von Aussagen über Gesamtheiten handelt, wird deutlich, wenn man darauf achtet, dass unsere Sprache zweideutig ist, wenn im Plural über die Mitglieder irgendeiner Gesamtheit gesprochen wird. Eine Aussage der Art „Für die Mitglieder der Gesamtheit Ω gilt ...“ kann bedeuten:

(1) Für jedes Mitglied aus Ω gilt ...; oder

(2) Für die Gesamtheit der Mitglieder aus Ω, also für Ω gilt ...

Statistische Aussagen, die vom Begriff einer statistischen Verteilung ausgehen, sind stets vom Typ (2), nicht vom Typ (1).

[11] In gleicher Bedeutung spricht man auch von *Häufigkeits-* und *Merkmalsverteilungen*, oder auch kurz von *der Verteilung* (einer statistischen Variablen).

[12] Wenn M irgendeine Menge ist, bezeichnet $\mathcal{P}(M)$ ihre *Potenzmenge*, d.h. die Menge aller Teilmengen von M.

[13] Diese eckigen Klammern bilden einen Teil des Namens der Funktion und dürfen nicht mit Argumenten verwechselt werden, die in runden Klammern angehängt werden. Natürlich kann die Angabe in den eckigen Klammern entfallen, wenn aus dem Kontext deutlich wird, auf welche Variablen Bezug genommen wird.

[14] Wenn M eine endliche Menge ist, soll $|M|$ die Anzahl ihrer Elemente bedeuten.

Natürlich müssen zunächst Daten über individuelle Mitglieder einer Gesamtheit erhoben werden, bevor eine statistische Verteilung gebildet werden kann. Insofern bezieht sich die Erhebung statistischer Daten auf individuelle Objekte. Ein Perspektivenwechsel findet jedoch statt, sobald man statistische Verteilungen betrachtet. Die Aufmerksamkeit richtet sich dann auf die Gesamtheit, nicht mehr auf ihre individuellen Mitglieder, anhand derer die Daten gewonnen worden sind. Diese der statistischen Methode eigentümliche Abstraktion wurde vom *International Statistical Institute* (1986: 238) in einer „Declaration of Professional Ethics" folgendermaßen formuliert:

„Statistical data are unconcerned with individual identities. They are collected to answer questions such as 'how many?' or 'what proportions?', not 'who?'. The identities and records of co-operating (or non-cooperating) subjects should therefore be kept confidential, whether or not confidentiality has been explicitly pledged."

8. Mehrdimensionale Verteilungen. Die Idee einer Häufigkeitsfunktion kann leicht für mehrdimensionale Variablen verallgemeinert werden. Als Beispiel verwenden wir die in Tabelle 2.1-1 angegebene Variable (X, Y), bei der sich X auf das Geschlecht und Y auf das Alter der Mitglieder einer aus 10 Personen bestehenden Gesamtheit Ω bezieht. Als Merkmalsmengen kommen jetzt alle Teilmengen des Merkmalsraums von (X, Y), also des kombinierten Merkmalsraums $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$, in Betracht. Die Häufigkeitsfunktion von (X, Y), für die die Notation $\mathrm{P}[X, Y]$ verwendet wird (oder $\mathrm{P}^*[X, Y]$, wenn auf absolute Häufigkeiten Bezug genommen werden soll), kann also durch folgendes Schema verdeutlicht werden:

$$\mathrm{P}[X,Y] : \mathcal{P}(\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}) \longrightarrow [0,1]$$

Ist M irgendeine Merkmalsmenge, d.h. eine Teilmenge von $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ bzw. ein Element der Potenzmenge von $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$, wird ihr durch die Häufigkeitsfunktion eine Zahl $\mathrm{P}[X, Y](M)$ im Intervall von 0 bis 1 zugeordnet, die den Anteil der Mitglieder von Ω angibt, die Merkmalswerte in M aufweisen:

$$\mathrm{P}[X,Y](M) = \frac{|\{\omega \in \Omega \,|\, (X,Y)(\omega) \in M\}|}{|\,\Omega\,|}$$

Als Beispiel sei etwa $M := \{1\} \times \{20, \ldots, 25\}$. Dann ist $\mathrm{P}[X, Y](M) = 2/10$, d.h. 20 % der Personen in der Referenzmenge Ω sind weiblich und 20 bis 25 Jahre alt.

2.2 Statistische Strukturbegriffe

Von „Strukturen" wird in unterschiedlichen Bedeutungen gesprochen. In diesem Abschnitt werden statistische Strukturbegriffe, insbesondere der statistische Sozialstrukturbegriff Peter Blaus, besprochen. Außerdem werden zur Reflexion der Frage, wie statistische Stukturen und Sachverhalte entstehen, substantielle und datenerzeugende Prozesse unterschieden.

1. Statistische Strukturen und Sachverhalte. Ein in der Sozialstrukturforschung verbreiteter Sprachgebrauch verwendet den Strukturbegriff im wesentlichen gleichbedeutend mit dem Begriff einer statistischen Häufigkeitsverteilung. Ich nenne dies den *statistischen Strukturbegriff.*[15] So wird z.B. von einer Altersstruktur der Bevölkerung oder von einer Berufsstruktur der Erwerbstätigen gesprochen.[16] Ulrich Mueller hat diesen statistischen Strukturbegriff folgendermaßen erläutert:

„Die Struktur einer bestimmten Bevölkerung wird beschrieben durch die absolute Zahl der Einheiten sowie die Verteilung der jeweils interessierenden Merkmalsausprägungen bei den Einheiten dieser Bevölkerung zu einem bestimmten Zeitpunkt t." (Mueller 1993: 2)

Diese Formulierung stammt aus einer Einführung in die Bevölkerungsstatistik. Aber der statistische Strukturbegriff ist nicht nur in der Demographie verbreitet,[17] sondern spielt in den meisten Varianten der Sozialstrukturforschung eine wichtige Rolle. Dies gilt insbesondere für die zahlreichen Ansätze, die eine zentrale Aufgabe der Sozialstrukturforschung darin sehen, die Mitglieder einer Gesellschaft in Klassen oder Schichten einzuteilen. Typischerweise meint der Sozialstrukturbegriff dann eine statistische Verteilung der Bevölkerung auf die zuvor konstruierten Klassen bzw. Schichten. Das ist von einigen Autoren als „oberflächlich" kritisiert worden,[18]

[15] Es sollte beachtet werden, dass auch innerhalb der statistischen Literatur noch in anderen Bedeutungen von „Struktur" gesprochen wird.

[16] Zahlreiche Illustrationen findet man u.a. bei E. Bodzenta (1979) und K.-E. Edinger (1998: 7ff.).

[17] Hier noch eine Formulierung des Demographen R. Pressat (1972: 1): „Demography is the discipline that seeks a statistical description of human populations with respect to (1) their demographic structure (the number of the population; its composition by sex, age and marital status; statistics of families, and so on) at a given date, and (2) the demographic events (births, deaths, marriages and terminations of marriages) that take place in them." Mit „demographischen Strukturen" sind hier offenbar Varianten statistischer Strukturen gemeint.

[18] Eine durchaus typische Variante dieser Kritik kommt etwa in folgenden Bemerkungen von Friedrich Fürstenberg (1966: 443) zum Ausdruck: „Es gibt eine Reihe von Autoren, die „Sozialstruktur" als *statistisches Klassifikationssystem* interpretieren. Sie setzen damit den „Gliederungsaspekt" des Begriffes, auf den *Karl Martin Bolte* hingewiesen hat, absolut. Autoren mit dieser Blickrichtung stehen häufig der wirtschaftsstatistischen Sichtweise nahe und haben als reine Empiriker ein unreflektiertes Verhältnis zur soziologischen Theorie." Man vgl. auch die Kritik von René König (1958b: 259) an „reinen

hier interessiert uns jedoch der Begriff selbst und erst im Anschluss auch die Frage, was mit ihm erreicht werden kann.

Noch eine terminologische Bemerkung: Da wir den statistischen Strukturbegriff synonym mit dem Begriff einer statistischen Verteilung verwenden, können Feststellungen statistischer Strukturen auch als statistische Sachverhalte bezeichnet werden. Allgemein verstehen wir unter *statistischen Sachverhalten* Feststellungen, die sich sowohl auf statistische Verteilungen als auch auf aus diesen abgeleitete Charakterisierungen (wie z.B. Mittelwerte, Streuungen, Anteilswerte, Raten und Regressionskoeffizienten) beziehen können.

2. Einige Besonderheiten des statistischen Strukturbegriffs. Es sollte beachtet werden, dass einige mit dem Strukturbegriff oft verbundene Vorstellungen nicht zu seiner statistischen Verwendung passen. Dies betrifft zunächst die Annahme, dass mit dem Strukturbegriff stets auf Beziehungen zwischen irgendwelchen Elementen verwiesen werden soll. George C. Homans (1976: 54) hat das so formuliert:

„[M]any sociologists use “social structure” to refer to some kind of social whole, which can be divided, at least conceptually, into parts, and in which the parts are in some way interdependent, at least in the sense that a change in some of them will be associated with changes in some of the others.“

In dieser Formulierung erinnert Homans an einen relationalen Strukturbegriff. Mit einigen Varianten solcher Vorstellungen beschäftigen wir uns in Kapitel 3. Hier soll zunächst darauf hingewiesen werden, dass der statistische Strukturbegriff vollständig unabhängig von irgendwelchen Vorstellungen über Beziehungen konzipiert ist.[19]

Eine zweite Differenz betrifft die Annahme, dass mit dem Strukturbegriff auf Sachverhalte verwiesen wird, die besonders dauerhaft sind oder jedenfalls dauerhafter als Vorgänge oder Prozesse, die sich im Rahmen gegebener Strukturen abspielen. Auch auf diese Konnotation des Strukturbegriffs wird von Homans (1976: 54) hingewiesen. Stefan Hradil (1987: 14) bemerkt dazu: „Es wird, wie immer, wenn der Strukturbegriff Anwendung findet, eine relativ beständige Anordnung von Elementen angesprochen.“[20]

Inventarisierungen einer Bevölkerung“. – Von Vorwürfen dieser Art sind Überlegungen zu unterscheiden, die gegen statistische Strukturbegriffe einwenden, dass soziale Beziehungen unberücksichtigt bleiben. Darauf wird in Abschnitt 3.3 näher eingegangen.

[19] Leider wird diese Unterscheidung nicht immer beachtet. Zum Beispiel verwendet P. M. Blau, mit dessen Ansatz wir uns weiter unten genauer beschäftigen werden, im wesentlichen einen statistischen Sozialstrukturbegriff; scheinbar darauf Bezug nehmend, gibt es jedoch immer wieder Formulierungen, die eigentlich einen relationalen Strukturbegriff voraussetzen (z.B. Blau 1977: 26ff.).

[20] So auch E. K. Scheuch und T. Kutsch (1975: 215): „Struktur bezeichnet das Dauerhafte an einem Gefüge von Elementen.“ Hier setzt auch ein leicht irreführender Kontrast zum Prozessbegriff an, z.B. in einer Formulierung von J. M. Blaut (1971: 19): „The relatively static events are often referred to as ‘structure’; the relative mobile ones as ‘process’ or ‘function’.“

Wiederum muss jedoch beachtet werden, dass der statistische Strukturbegriff diese Vorstellung nicht beinhaltet. Die Begriffsbildung hat keinerlei Implikationen für die Frage, wie sich eine statistische Verteilung im Zeitablauf entwickelt. Man kann deshalb auch ganz unproblematisch die Frage stellen, wie sich statistische Strukturen (z.B. Haushaltsstrukturen) im Zeitablauf verändern.

Schließlich sollte auch beachtet werden, dass statistisch definierte Strukturen, also Häufigkeitsverteilungen, nicht mit der Vorstellung eines „Musters“ in Verbindung gebracht werden können. Zwar kann man sinnvoll von der *Form einer statistischen Verteilung* sprechen (insbesondere dann, wenn ein quantitativer Merkmalsraum gegeben ist, so dass es eine lineare Ordnung der Merkmalswerte gibt); bekanntlich können Verteilungen graphisch dargestellt werden und liefern dadurch eine direkte Anschauung ihrer Form,[21] die auch durch statistische Kennzahlen charakterisiert werden kann. Aber solche graphischen Darstellungen vermitteln nicht die Vorstellung eines mit Regelmäßigkeiten assoziierbaren Musters.

3. Unterschiedliche Sozialstrukturbegriffe. Der statistische Strukturbegriff ist so allgemein, dass in zahlreichen Varianten – oder vielleicht besser: Aspekten – von „Sozialstruktur“ gesprochen werden kann. Unterschiede kann es sowohl in den Arten der Objekte geben, auf die man sich bezieht (etwa Personen, Haushalte, Unternehmen oder Regionen), als auch bei den Merkmalsräumen, die zur Charakterisierung der Objekte verwendet werden (etwa Alter, Einkommen und Bildung bei Personen oder Beschäftigtenzahl, Umsatz und Wirtschaftszweig bei Unternehmen).

Eine weitere Unterscheidung kann im Anschluss an folgende Ausführungen von Wolfgang Zapf erläutert werden:

„Unter *Sozialstruktur* kann man mindestens dreierlei verstehen. Erstens die demographische Grundgliederung der Bevölkerung und die Verteilung zentraler Ressourcen wie Bildung, Beruf und Einkommen. [. . .]

Zweitens kann man unter Sozialstruktur – unter Einschluß von Werten und Mentalitäten – die Zusammenfassung dieser Gliederungen in soziale Klassen und Schichten verstehen [. . .].

Drittens gibt es den anspruchsvolleren Begriff von Sozialstruktur als dem jeweils historisch ausgeprägten System gesellschaftlicher Ordnungen oder Grundinstitutionen [. . .].“ (Zapf 1995: 187)

Offenbar handelt es sich in den ersten beiden Fällen um Varianten des statistischen Strukturbegriffs (erst Zapfs dritte Variante führt zu einem grundsätzlich anderen Zugang zur Idee einer Sozialstruktur). In der ersten Variante sind statistische Verteilungen gemeint, deren Merkmalsräume unmittelbar auf in der gesellschaftlichen Praxis übliche Unterscheidungen verweisen; in der zweiten Variante setzt der Sozialstrukturbegriff die vorgängige Konstruktion eines Schemas zur Klassifikation der Mitglieder einer Ge-

[21] Als Beispiel kann man die Altersverteilung in Abbildung 5.2-6 betrachten.

sellschaft voraus. Beide Varianten werden in den folgenden Paragraphen etwas näher besprochen.

4. Der Sozialstrukturbegriff bei Peter M. Blau. Der ersten Variante lassen sich die meisten statistischen Beiträge zur Beschreibung gesellschaftlicher Verhältnisse zurechnen. Wegen seines theoretischen Anspruchs ist in diesem Zusammenhang besonders der Forschungsansatz von Peter M. Blau, von ihm selbst als „makro-strukturell“ bezeichnet, von Interesse:

„Macrostructural concepts refer to people's distribution in various dimensions and the degrees to which these dimensions of social differences among people are related. Macrosociology is concerned primarily with large populations – composed of many thousands or even millions of persons. My endeavor is to develop a systematic theoretical scheme for the study of macrostructures and their impact on social life.“ (Blau 1994a: 1)

Bei den „Dimensionen“ kann es sich um beliebige Attribute (Merkmalsräume) handeln, durch die sich die Mitglieder einer Gesellschaft unterscheiden lassen.[22] Wichtig ist der Hinweis, dass Beziehungen zwischen diesen „Dimensionen“ ermittelt werden sollen, denn das setzt voraus, dass man sie nicht sogleich zu Klassifizierungen zusammenfasst. Dem entspricht folgende Definition:

„Social structure can be conceptualized as a multidimensional space of social positions among which a population is distributed.“ (Blau 1994a: 4)

Diese Begriffsbildung kann gut mithilfe einer mehrdimensionalen statistischen Variablen ausgedrückt werden:

$$(X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

In dieser Formulierung bezieht sich Ω auf eine Population (eine irgendwie abgegrenzte Gesellschaft), und die Komponenten $X_1, \ldots, X_m$ der mehrdimensionalen Variablen erfassen Eigenschaften der Mitglieder von Ω in den m „Dimensionen“ (Merkmalsräumen) $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$. Was Blau „Sozialstruktur“ nennt, entspricht formal der Verteilung dieser m-dimensionalen statistischen Variablen. Anhand dieser Formulierung wird auch deutlich, was damit gemeint ist, Beziehungen zwischen den „Dimensionen“ der Sozialstruktur zu ermitteln. Es geht um Charakterisierungen gemeinsamer Verteilungen der Komponenten von $(X_1, \ldots, X_m)$, etwa durch Korrelationen (Blau 1994a: 5) oder allgemeiner durch Regressionsfunktionen.[23]

5. Bezugseinheiten statistisch definierter Sozialstrukturen. Blaus Sozialstrukturbegriff bezieht sich zunächst auf Populationen, also irgendwie abgegrenzte Gesamtheiten von Menschen. Vollständig analog kann man aber

[22] Blau spricht in diesem Zusammenhang von „sozialen Positionen“, meint aber „any difference among people in terms of which they make social distinctions among themselves in social intercourse“ (Blau 1994a: 3).

[23] Regressionsfunktionen werden in Abschnitt 15.1 besprochen.

auch bei anderen Bezugseinheiten Aussagen über statistische Strukturen (Merkmalsverteilungen) machen; insbesondere kann man sich auf Haushalte, Unternehmen und Regionen beziehen.

In diesem Zusammenhang sollte auch auf eine Ambivalenz im Sprachgebrauch geachtet werden, die sich anhand der folgenden Bemerkung von Blau (1974: 615f.) erläutern lässt:

„The concept of social structure is used widely in sociology, often broadly, and with a variety of meanings. [...] A generic difference is whether social structure is conceived explicitly as being composed of different elements and their interrelations or abstractly as a theoretical construct or model. [...[24]] If one adopts the first view, as I do, that social structure refers to the differentiated interrelated parts in a collectivity, not to theories about them, the fundamental question is how these parts and their connections are conceived.

My concept of social structure starts with simple and concrete definitions of the component parts and their relations. The parts are groups or classes of people, such as men and women, ethnic groups, or socioeconomic strata; more precisely, they are the positions of people in different groups and strata. The connections among as well as within the parts are the social relations of people that find expression in their social interaction and communication.“

Zunächst erscheint diese Aussage als ein Widerspruch zur oben in § 4 zitierten Bezugnahme auf Populationen, also Gesamtheiten von Individuen. Der Widerspruch verschwindet jedoch, wenn man zwischen Individuen (oder allgemeiner irgendwelchen Bezugseinheiten) und Positionen (im Sinne Blaus) unterscheidet. Unser Begriff einer statistischen Variablen macht das deutlich. Denkt man an das Schema $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, repräsentiert Ω die Population, und die Elemente des Merkmalsraums $\tilde{\mathcal{X}}$ sind die Positionen, die zur Charakterisierung der Mitglieder von Ω verwendet werden sollen. Die statistische Variable X induziert nun außerdem eine Partition der Population: Jeder Position $\tilde{x} \in \tilde{\mathcal{X}}$ entspricht eine Menge

$$X^{-1}(\{\tilde{x}\}) = \{\omega \in \Omega \,|\, X(\omega) = \tilde{x}\}$$

die aus denjenigen Mitgliedern von Ω besteht, die die Position $\tilde{x}$ haben. Zwar muss man, wie Blau selbst bemerkt, zwischen der Position $\tilde{x}$ und der Teilpopulation $X^{-1}(\{\tilde{x}\})$ unterscheiden; aber die Häufigkeit der Position $\tilde{x}$ in der statistischen Struktur entspricht natürlich dem Umfang der korrespondierenden Teilpopulation.

Eine weitere Quelle für Unklarheiten ist allerdings Blaus Bemerkung, dass sich sein statistischer Strukturbegriff auf „component parts *and their relations*“ bezieht. Wie bereits in § 2 bemerkt worden ist, passt diese Rhetorik nicht zur Verwendung eines statistischen Strukturbegriffs. Dass zwei

[24] Hier erwähnt Blau kurz die Auseinandersetzung zwischen Radcliffe-Brown und Lévi-Strauss über den theoretischen Status des (Sozial-) Strukturbegriffs. Eine ausführliche Diskussion findet man bei Michael Oppitz (1975: 33ff.).

oder mehr Personen irgendeine statistisch erfassbare Eigenschaft gemeinsam haben, kann zwar möglicherweise zur Feststellung von Ähnlichkeiten dienen, begründet aber keine substantielle Beziehung zwischen diesen Personen.[25] Umgekehrt verweist Blaus Rede von „social interaction and communication“ zwar auf substantielle Beziehungen (und also indirekt auf einen relationalen Sozialstrukturbegriff), auf *diese* Beziehungen nimmt aber Blaus statistischer Sozialstrukturbegriff gar keinen Bezug.

6. Wie entstehen statistische Sachverhalte? Dieser Frage können unterschiedliche Bedeutungen gegeben werden. Zur Erläuterung betrachten wir folgendes Bild:

○ ○ ● ○ ● ● ○ ○ ○ ●

Das Bild stellt einen Sachverhalt dar, der aus 10 Objekten besteht, von denen 4 schwarz, die übrigen nicht schwarz sind. Offenbar handelt es sich in dieser Darstellung nicht bereits um einen statistischen Sachverhalt. Ein statistischer Sachverhalt *entsteht erst durch eine spezifische Konzeptualisierung*, die drei wesentliche Schritte umfasst:

– die Konzeption einer statistischen Variablen;
– einen (realen oder fiktiven) *datenerzeugenden Prozess*, der Informationen über die Werte der Variablen liefert; und
– rechnerische Operationen, die es erlauben, sich gedanklich auf die Verteilung der statistischen Variablen zu beziehen.

In unserem Beispiel können diese drei Schritte offenbar problemlos ausgeführt werden. Zunächst kann man die 10 Objekte durch eine Objektmenge $\Omega := \{\omega_1, \ldots, \omega_{10}\}$ repräsentieren (wobei die Zuordnung der Namen beliebig erfolgen kann); und es kann ein Merkmalsraum $\tilde{\mathcal{Y}} := \{0, 1\}$ festgelegt werden (wobei etwa 1 für schwarz und 0 für nicht schwarz steht), so dass schließlich der erste Schritt durch die Definition einer statistischen Variablen $Y : \Omega \longrightarrow \tilde{\mathcal{Y}}$ abgeschlossen werden kann, durch die jedem der 10 Objekte ein Wert im Merkmalsraum $\tilde{\mathcal{Y}}$ zugeordet wird. In einem zweiten Schritt können dann Daten erzeugt werden. Das ist in diesem Beispiel direkt und vollständig möglich und liefert eine Tabelle:

ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
$Y(\omega)$	0	0	1	0	1	1	0	0	0	1

Diese Tabelle enthält das Datenmaterial für den dritten Schritt, in dem die Verteilung der Variablen Y berechnet wird. Auch dieser Schritt ist in diesem Beispiel direkt durchführbar und erfordert nicht einmal eine

[25] Eine ausführliche Darlegung dieser Kritik findet man bei F. L. Bates und W. G. Peacock (1989).

tabellarische oder graphische Darstellung; es genügt die Angabe, dass 40 % der Objekte schwarz, die übrigen nicht schwarz sind.

Diese Aussage beschreibt nun einen statistischen Sachverhalt, und somit illustriert das Beispiel *eine* Antwort auf unsere Ausgangsfrage: Ein statistischer Sachverhalt entsteht durch eine gedankliche und praktische Konstruktion, die in den drei genannten Schritten abläuft. Diese Antwort macht deutlich, dass statistische Sachverhalte durch spezifische gedankliche und praktische Konstruktionen entstehen.[26]

7. Datenerzeugende und substantielle Prozesse. Außerdem macht die Antwort deutlich, dass die Konstruktion eines statistischen Sachverhalts voraussetzt, dass die Mikro-Sachverhalte, auf die sich ein datenerzeugender Prozess beziehen kann, bereits existieren. Infolgedessen können zwei Arten von Fragen unterschieden werden:

– Wie entstehen diese Mikro-Sachverhalte bzw. wie sind sie entstanden? (Wie sind z.B. die schwarzen und nicht-schwarzen Objekte in dem oben in § 6 angeführten Bild entstanden?)
– Wie entsteht ausgehend von bereits bestehenden Mikro-Sachverhalten ein statistischer Sachverhalt?

Das folgende Bild veranschaulicht die Unterscheidung:

$$\left.\begin{array}{r} \dashrightarrow Y(\omega_1) \\ \vdots \\ \dashrightarrow Y(\omega_n) \end{array}\right\} \Longrightarrow \mathrm{P}[Y] \tag{2.1}$$

Durch $Y(\omega_1), \ldots, Y(\omega_n)$ werden die Mikro-Sachverhalte angedeutet,[27] wobei jedem dieser Mikro-Sachverhalte ein Prozesspfeil $\dashrightarrow$ zugeordnet ist, der einen *substantiellen Prozess* andeuten soll, durch den der jeweilige Mikro-Sachverhalt entstanden ist. Andererseits wird durch den Pfeil $\Longrightarrow$ auf den *Konstruktionsprozess* verwiesen, durch den der statistische Sachverhalt $\mathrm{P}[Y]$ entsteht, wobei – wie das Bild zeigt – die Mikro-Sachverhalte vorausgesetzt werden. Es müssen also insgesamt drei Arten von Prozessen unterschieden werden:

– *Substantielle Prozesse,* durch die den Elementen einer Objektmenge individuell zurechenbare Mikro-Sachverhalte entstehen;
– *datenerzeugende Prozesse,* durch die empirische Informationen (die

[26] Sicherlich könnte man das Bild, von dem wir in diesem Beispiel ausgegangen sind, noch auf andere Weisen beschreiben. Insbesondere könnte man auf Aspekte achten, wie etwa räumliche Anordnungen, von denen bei der Konstruktion statistischer Sachverhalte abstrahiert wird.

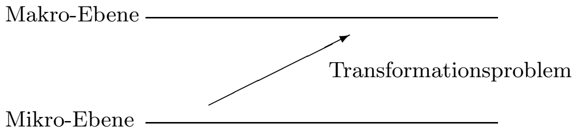
[27] Dies ist offenbar eine verkürzte Darstellung, da es sich bei $Y(\omega_1), \ldots, Y(\omega_n)$ nicht um Sachverhalte, sondern um Elemente des Merkmalsraums von Y handelt. Eine explizte Notation für die korrespondierenden Mikro-Sachverhalte findet man bei Rohwer und Pötter (2002b: 211ff.).

Daten) über die für die Datenerzeugung vorauszusetzenden Mikro-Sachverhalte entstehen;[28] und

– *Rechenprozesse,* durch die ausgehend von jeweils gegebenen Mengen von Daten (die als Resultat eines datenerzeugenden Prozesses entstanden sind) statistische Sachverhalte konstruiert werden.

Offenbar interessieren in erster Linie die substantiellen Prozesse, durch die die Mikro-Sachverhalte in der sozialen Realität entstehen; man möchte verstehen, wie diese Prozesse ablaufen und wodurch sie bedingt werden.

8. Statistische Sachverhalte im Mikro-Makro-Schema. Eine wichtige Frage betrifft die theoretische Konzeptualisierung substantieller Prozesse. Hier soll zunächst darauf hingewiesen werden, dass eine Einordnung statistischer Sachverhalte in ein Mikro-Makro-Schema dafür kaum hilfreich ist. Zur Erläuterung kann folgendes Bild dienen:



In diesem Bild gibt es eine Mikro-Ebene, mit der auf individuelle Menschen, ihre Tätigkeiten und ihnen zurechenbare Mikro-Sachverhalte verwiesen werden soll, und eine Makro-Ebene, die sich zunächst allgemein auf „kollektive Phänomene“ bezieht. Außerdem gibt es einen Pfeil, der von der Mikro- zur Makro-Ebene führt und die Vorstellung andeuten soll, dass die „kollektiven Phänomene“ der Makro-Ebene irgendwie aus auf der Mikro-Ebene fixierbaren Sachverhalten und Vorgängen resultieren.[29] Die Frage, wie dies geschieht, wird im Anschluss an Siegwart Lindenberg (1977) oft als *Transformationsproblem* bezeichnet.[30]

Zu überlegen ist, welche Arten von Sachverhalten auf der Makro-Ebene dieses Schemas verortet werden können. In der Literatur findet man Hinweise auf sehr unterschiedliche Arten von Sachverhalten. So spricht etwa Lindenberg (1977: 49) von „kollektiven Phänomenen“ „wie z.B. kollektive Handlungen (etwa Streiks), Strukturen (etwa Statusstrukturen), Verteilungen (etwa Einkommensverteilungen), Institutionen (etwa Institutionalisierung von Konflikten)“. Offenbar wird hier einerseits auf statistische Sachverhalte Bezug genommen,[31] andererseits aber auch auf zahlreiche andere Arten „kollektiver Phänomene“. Infolgedessen muss jedoch auch der Pfeil, der von der Mikro- zur Makro-Ebene führt, jeweils unterschiedlich interpretiert werden. Handelt es sich z.B. um einen Streik, kann man sinnvoll von einem Sachverhalt sprechen, der aus den Tätigkeiten einer Mehrzahl beteiligter Akteure resultiert. Handelt es sich dagegen um statistische Sachverhalte, kann man dies nicht sagen. Man kann zwar in vielen Fällen sinnvoll davon sprechen, dass die dem statistischen Sachverhalt korrespondierenden Mikro-Sachverhalte aus Tätigkeiten von Akteuren resultieren. Sowohl diese Tätigkeiten wie auch die aus ihnen resultierenden Mikro-Sachverhalte gehören jedoch zur Mikro-Ebene des Schemas. Wenn man also statistische Sachverhalte auf der Makro-Ebene ansiedelt, entspricht dem Pfeil, der von der Mikro- zur Makro-Ebene führt, auch kein substantieller Prozess, sondern nur der statistische Konstruktionsprozess, der aus *bereits entstandenen* Mikro-Sachverhalten eine spezifische Art ihrer Beschreibung erzeugt.

[28] Es sei angemerkt, dass der Begriff eines datenerzeugenden Prozesses in der Literatur gelegentlich auch anders verwendet wird, nämlich als Verweis auf die unter (a) genannten substantiellen Prozesse; man vgl. dazu die Hinweise und Literaturangaben bei Rohwer und Pötter (2002b: 19ff.).

[29] In einem vollständigen Mikro-Makro-Schema gibt es auch einen Pfeil, der von der Makro- zur Mikro-Ebene führt und die Vorstellung andeuten soll, dass die auf der Mikro-Ebene fixierbaren Sachverhalte und Vorgänge auch von Sachverhalten auf der Makro-Ebene abhängig sind. Davon kann hier jedoch zunächst abgesehen werden.

[30] Nicht zu verwechseln mit dem ökonomischen Transformationsproblem, das sich auf den Zusammenhang zwischen Werten und Produktionspreisen bezieht.

[31] Ebenso sprechen G. Büschges, M. Abraham und W. Funk (1998: 18) von einer Makro-Ebene, „die kollektive Phänomene wie statistische Verteilungen (z.B. die Quote der Frauenerwerbsbeteiligung in einer Gesellschaft) oder kollektives Verhalten (wie Demonstrationen vieler Individuen) abbildet.“

Kapitel 3

Relationale Begriffsbildungen

Die im vorangegangenen Kapitel besprochenen Begriffe bilden die Grundlage einer *statistischen Betrachtungsweise* von Gesamtheiten: Man geht von Eigenschaften aus, die sich den Elementen einer Gesamtheit jeweils individuell zurechnen lassen, und betrachtet dann deren Häufigkeitsverteilungen. In einem gewissen Spannungsverhältnis dazu stehen *relationale Betrachtungsweisen*, die von Beziehungen zwischen den Elementen einer Gesamtheit ausgehen. Damit beschäftigen wir uns in diesem Kapitel. Da oft der Systembegriff verwendet wird, um eine relationale Betrachtungsweise anzudeuten, beginnen wir mit einer kurzen Besprechung unterschiedlicher Systembegriffe. Dann werden einige formale Begriffsbildungen besprochen, woran sich Überlegungen zu relationalen Strukturbegriffen anschließen.

3.1 Unterschiedliche Systembegriffe

1. Ein abstrakter Systembegriff. In der Literatur wird oft in einer sehr allgemeinen und abstrakten Weise von Systemen gesprochen. Dem entspricht etwa folgende an den Mengenbegriff anknüpfende Definition, die von A. D. Hall und R. E. Fagen gegeben wurde:

„A system is a set of objects together with relationships between the objects and between their attributes.“ (Hall und Fagen 1956: 18)

Ich nennen dies den *abstrakten Systembegriff,* da die Definition vollständig offen lässt, wie Systeme von einer Umwelt abgegrenzt werden können. Vielmehr wird vorausgesetzt, dass man sich auf eine irgendwie gegebene Menge von Elementen beziehen kann: Systeme werden als Mengen konzipiert, bei denen man von einer Beziehung zwischen mindestens zwei Elementen sprechen kann. Orientiert man sich an diesem abstrakten Systembegriff, kann man offenbar *alle* Gegenstände auch als Systeme betrachten, denn immer lassen sich mindestens zwei Aspekte unterscheiden und in irgendeine, ggf. beliebig ausgedachte Beziehung bringen. Demgegenüber gibt es jedoch auch durchaus spezifischere Verwendungsweisen des Systembegriffs.

2. Der traditionelle Systembegriff: ein geordnetes Ganzes. Hinweise zur geschichtlichen Entwicklung des Systembegriffs findet man bei A. von der Stein (1968).[1] Der Begriff stammt aus der griechischen Antike und hat auch dort bereits die Bedeutung gewonnen, an der sich der allgemeine, insbesondere in den technischen Wissenschaften verbreitete Sprachgebrauch bis heute meistens orientiert. Stein schreibt (S. 5):

„Zusammenfassend läßt sich nun der griechische Sprachgebrauch folgendermaßen charakterisieren: 1. $\sigma\acute{\upsilon}\sigma\tau\eta\mu\alpha$ [systema] ist nach einer wenn auch nicht kontinuierlichen Entwicklung im griechischen Sprachgebrauch selbst zu einem Allgemeinbegriff geworden, der sowohl auf künstliche wie auch auf natürliche Objekte angewandt wird. 2. Seine Bedeutung ist: ein Gebilde, das irgendein Ganzes ausmacht und dessen einzelne Teile in ihrer Verknüpfung irgendeine Ordnung aufweisen. Damit hat der griechische Sprachgebrauch schon die vollständige Basis für einen Systembegriff nach heutigem Verständnis geschaffen.“

Ich nenne dies den *traditionellen Systembegriff*. Im Unterschied zur oben zitierten Definition von Hall und Fagen besteht der Leitgedanke darin, das es sich bei einem System um ein geordnetes Ganzes handelt. Um irgendeine Menge von Objekten als ein System bezeichnen zu können, genügt es also nicht, auf irgendwelche Beziehungen zu verweisen, sondern man muss in der Lage sein, die Idee eines geordneten Zusammenhangs zu erläutern. Der traditionelle Systembegriff hat deshalb eine explizierbare Bedeutung zunächst dort, wo Menschen geordnete Zusammenhänge herstellen. Infolgedessen ist dieser traditionelle Systembegriff auch nicht von der Vorstellung ablösbar,

[1] Noch ausführlicher informiert M. Riedel (1990).

dass Systeme einem Zweck (oder einer Mehrzahl möglicher Zwecke) dienen. Für Johann Heinrich Lambert war dies noch ganz selbstverständlich.[2] In seinem „Fragment einer Systematologie“ schreibt Lambert:

„Zu einem System werden also Teile, und zwar mehrere erfordert. Diese müssen auseinandergesetzt, jedes für sich kenntlich, mit Absicht gestellt oder geordnet und alle miteinander so verbunden sein, daß sie gerade das der vorgesetzten Absicht gemäße Ganze ausmachen, und dieses muß, so gut es angeht oder so lange es die Absicht erfordert, fortdauern können, es sei daß es unverändert bleibe oder seiner Absicht gemäße Veränderungen leide.“ (Lambert 1988: 126)

Diese Ausführungen liefern eine gute Erläuterung des traditionellen Systembegriffs. Zugleich werden auch die Sinnvoraussetzungen und Sinngrenzen des Begriffs deutlich. Der traditionelle Systembegriff verlangt einen gedanklichen Rückgriff auf Akteure, die ein geordnetes Ganzes herstellen und auf deren Absichten man sich für ein Verständnis des Systems beziehen kann. Infolgedessen setzt der traditionelle Systembegriff zunächst voraus, dass man sich auf eine Praxis beziehen kann, in der Menschen geordnete Zusammenhänge herstellen, sei es in gegenständlicher Form oder in der Form einer Gestaltung der materiellen und sozialen Bedingungen ihrer Tätigkeiten. Dabei ist ‘Ordnung’ kein absoluter Begriff, sondern setzt voraus, dass man sich gedanklich auf Zwecke beziehen kann, um derentwillen eine Ordnung gebildet worden ist. Der traditionelle Systembegriff verlangt also, dass man der Idee, dass eine Ordnung *hergestellt* wird, einen explizierbaren Sinn geben kann. Insofern ist dieser Systembegriff an eine – zunächst menschliche – Praxis gebunden, und weiterhin an die Annahme, dass die Akteure, die eine Ordnung herstellen, die dafür erforderlichen Fähigkeiten besitzen.

3. Reale und epistemische Systeme. Eine gewisse Komplikation entsteht daraus, dass man sich mit dem traditionellen Systembegriff sowohl auf Gegenstände und Einrichtungen der menschlichen Erfahrungswelt als auch auf Ausarbeitungen von Überlegungen beziehen kann. Im ersten Fall kann man von *realen Systemen* sprechen, womit dann also Gegenstände oder Einrichtungen gemeint sind, die eine materielle Existenz in der menschlichen Erfahrungswelt aufweisen; zum Beispiel ein Karteikasten, dessen Elemente (Karteikarten) in einer bestimmten Weise geordnet sind. Im zweiten Fall kann man von *epistemischen Systemen* sprechen, womit dann eine Menge von (meistens in schriftlicher Form vorliegenden) Überlegungen gemeint ist, die von ihrem Autor in einer bestimmten Weise angeordnet worden sind.

Offenbar deckt der traditionelle Systembegriff beide Möglichkeiten ab. Bei der epistemischen Verwendung des Systembegriffs entsteht jedoch eine leicht irreführende Ambivalenz. Wenn man von einer Menge von Überlegungen sagt, dass sie „ein System bilden“ (ein „Gedankengebäude“), ist

[2]Lambert lebte 1728 – 1777 und war einer der ersten Philosophen, die sich explizit mit dem Systembegriff beschäftigt haben. Ausführliche Hinweise gibt Siegwart (1988).

zunächst nur gemeint, dass sie von ihrem Autor auf eine systematische, d.h. geordnete Weise in einen Zusammenhang gebracht worden sind. Eine solche Charakterisierung setzt offenbar nicht voraus, dass es sich bei dem Gegenstand, auf den sich die Überlegungen beziehen, selbst um ein System handelt. Aber ein Missverständnis liegt natürlich nahe. Man denke zum Beispiel an das Reden von einem „Planetensystem“. Gemeint ist einerseits ein Modell, also eine Menge von Überlegungen, die von Menschen ausgearbeitet worden sind, um sich ein geordnetes, für Voraussagen nützliches Bild gewisser Himmelserscheinungen zu machen; insofern bezieht sich das Wort auf ein epistemisches System, dem – wenn man den traditionellen Systembegriff ernst nimmt – kein reales System entspricht. (Ein reales System wäre dagegen ein Planetarium, das von Menschen hergestellt worden ist, um ihre Modellvorstellungen anschaulich zu machen.) Andererseits liegt es natürlich nahe, auch die Sonne mit ihren Planeten selbst als ein reales System aufzufassen.

4. Poietischer und reflektierender Systembegriff. So gelangt man zu der Frage, unter welchen Bedingungen sinnvoll von einem System gesprochen werden kann bzw. was Systeme von anderen Arten von Gegenständen oder Aspekten unserer Erfahrungswelt unterscheidet. Offenbar erlaubt der abstrakte Systembegriff keine bzw. nur eine willkürliche Antwort, weil alles Beliebige „als ein System aufgefasst“ werden kann.

Orientiert man sich am traditionellen Systembegriff, kann von einem System gesprochen werden, wenn und insoweit geordnete Verhältnisse *hergestellt* worden sind. Dem entspricht die Idee, dass ein System (im traditionellen Verständnis des Worts) eine *teleologische Betrachtungsweise* erlaubt. Damit ist gemeint, dass ein System im Hinblick auf Zwecke betrachtet wird, denen es dienen kann.[3] Eine solche Betrachtungsweise ist offenbar zunächst dann sinnvoll möglich, wenn und insoweit Menschen Systeme herstellen, damit sie bestimmten Zwecken dienen können.

An dieser Stelle entsteht allerdings die Möglichkeit, den traditionellen Systembegriff auf eine problematische Weise zu verallgemeinern. Denn einerseits erlaubt zwar das Vorhandensein eines Systems eine teleologische Betrachtungsweise; andererseits setzt eine solche Betrachtungsweise nicht voraus, dass man es tatsächlich mit einem System (im traditionellen Wortverständnis) zu tun hat. Denn im Prinzip können beliebige Ge-

[3]Der Zweckbegriff bedarf einer ähnlichen Erläuterung wie der oben verwendete Ordnungsbegriff. Zwecke existieren nur als Vergegenwärtigungen eines reflektierenden Bewusstseins, nicht als objektivierbare Sachverhalte, die den Gegenständen unserer Erfahrungswelt zurechenbar wären. Gegenstände haben nicht an und für sich Zwecke, aber sie können im Hinblick auf Zwecke betrachtet und ggf. zur Verfolgung von Zwecken verwendet werden. Dies gilt insbesondere, aber nicht nur, für Gegenstände, die für einen zweckmäßigen Gebrauch hergestellt worden sind (Artefakte, insbesondere Werkzeuge und Maschinen). Weiterhin ist es auch erforderlich, Zwecke (als Vergegenwärtigungen eines reflektierenden Bewusstseins) von „Motiven“ zu unterscheiden, die man Akteuren zurechnet, um ihr Verhalten pseudo-kausal erklärbar zu machen.

genstände und Einrichtungen im Hinblick auf Zwecke, denen sie dienen können, *reflektiert* werden. Eine solche Reflexion setzt nicht voraus, dass es sich um Gegenstände oder Einrichtungen handelt, die im Hinblick auf Zwecke *hergestellt* worden sind. Somit kann ein *reflektierender Systembegriff* entstehen: Irgendetwas ist ein System, wenn und insoweit es in einer teleologischen Betrachtungsweise reflektierbar gemacht werden kann. Zur Unterscheidung spreche ich von einem *poietischen Systembegriff*,[4] wenn als Sinnvoraussetzung daran festgehalten wird, dass ein geordneter Zusammenhang *hergestellt* worden ist.[5]

Die Ausweitung des poietischen zu einem reflektierenden Systembegriff hat bereits in der griechischen Antike begonnen. Wie H. Busche (1998: 970) berichtet, wurde der Begriff 'Teleologie' jedoch erst 1728 durch den Philosophen Christian Wolff geprägt:

> „Für die Betrachtung der Dinge nach ihrem «Zweck» («finis») reserviert er [Wolff] einen «Teil der Naturphilosophie, der die Zwecke der Dinge erklärt. Noch immer hat er keinen Namen, obwohl er sehr weitreichend und nützlich ist. Man könnte ihn Teleologie nennen».“

Ein Zusammenhang mit dem Systembegriff wird explizit durch Lambert hergestellt, auf den bereits weiter oben zur Erläuterung des traditionellen Systembegriffs hingewiesen wurde. Lambert ist zugleich ein gutes Beispiel für den Übergang zu einem reflektierenden Systembegriff. Nachdem sich seine Überlegungen zunächst an poietischen Vorstellungen orientieren, sagt er, dass „Systeme immer als *Systeme von Mitteln und Absichten* betrachtet werden können, sooft die Entschließungen des Willens mit in Erwägung kommen“, und er fügt dann hinzu:

> „Sehen wir fürerst hierbei auf den göttlichen Willen, so kann überhaupt der ganze Weltbau als ein System von Mitteln und Absichten angesehen werden. Die Erde und alle sich auf, in und mit derselben ereignenden Veränderungen machen ebenfalls ein solches, wiewohl viel spezielleres System aus. Beide geben Stoff zu der Teleologie, einer Wissenschaft, welche sich mit der Betrachtung der Absichten der Natur und ihrer Teile beschäftigt.“ (Lambert 1988: 135)

Hier wird deutlich, wie sich der traditionelle Systembegriff von seiner ursprünglichen Anbindung an eine menschliche Praxis, in der Menschen geordnete Zusammenhänge herstellen, ablöst und auf beliebige Erscheinungen und Sachverhalte anwendbar erscheint. Die Voraussetzung, die den Sinnzusammenhang begründet, liegt natürlich in der Annahme, dass in jedem Fall eine teleologische Betrachtungsweise zu wissenswerten Erkenntnissen führen kann.

[4] Das Fremdwörterbuch (Duden) erläutert: „*poietisch* <aus gr. poiētós „zu machen, bildend“, zu poieĩn „zustande bringen, schaffen“> bildend, das Schaffen betreffend; *-e Philosophie*: bei Plato die dem Herstellen von etwas dienende Wissenschaft (z.B. Architektur).“

[5] Es sollte beachtet werden, dass sich das Adjektiv 'poietisch' (wie auch 'reflektierend' usw.) hier auf den System*begriff*, nicht auf Systeme bezieht. Es wird also nicht von „poietischen Systemen“ gesprochen.

5. Konsequenzen für unseren Sprachgebrauch. Es sollte deutlich geworden sein, dass der Systembegriff in unterschiedlichen Bedeutungen verwendet werden kann. Unproblematisch (weil inhaltsleer) ist der abstrakte Systembegriff, da mit seiner Hilfe alles Beliebige als ein System bezeichnet werden kann, ohne dass damit irgendwelche bestimmten Konsequenzen verbunden sind. Wenn also das Wort 'System' ohne weitere Hinweise verwendet wird, kann man den abstrakten Systembegriff unterstellen. Dagegen sollte explizit darauf hingewiesen werden, wenn mit dem Systembegriff bestimmte Ordnungsvorstellungen verbunden und/oder teleologische Betrachtungsweisen vorgeschlagen werden sollen.

3.2 Relationen und Graphen

In diesem Abschnitt werden einige elementare Begriffsbildungen zur formalen Konzeption, Darstellung und Analyse von Netzwerken besprochen.

1. Relationale Aussagen. Von Beziehungen bzw. Relationen wird in unterschiedlichen Bedeutungen geredet, einige inhaltlich wichtige Unterscheidungen werden im Abschnitt 3.3 besprochen. Hier soll zunächst angenommen werden, dass man ohne weiteres relationale Aussagen formulieren kann, zum Beispiel: Zwei Menschen kennen sich oder sind befreundet oder sind verheiratet; ein Mensch erzielt ein höheres Einkommen als ein anderer; zwei Schüler sind Mitglieder derselben Schulklasse; ein Mensch ist Angestellter eines bestimmten Unternehmens; ein Unternehmen bezieht von einem anderen Unternehmen Vorleistungen für seine Güterproduktion; zwei Computer sind durch ein Netzwerk verbunden, so dass Daten ausgetauscht werden können. Dies sind Beispiele für *relationale Aussagen*: Aussagen, die sich gleichzeitig auf zwei (oder mehr) Objekte beziehen. Zu unterscheiden sind relationale Aussagen und Aussageformen. Zum Beispiel ist 'Franz ist verheiratet mit Karin' eine *relationale Aussage*, die ihrer Intention nach einen Sachverhalt ausdrückt und infolgedessen wahr oder falsch sein kann. Dagegen ist 'ω ist verheiratet mit ω'' eine *relationale Aussageform*, wenn ω und ω' nicht Namen bestimmter Objekte, sondern *logische Variablen* sind. Relationale Aussagen, die wahr oder falsch sein können, entstehen erst dann, wenn man in die logischen Variablen (Leerstellen) bestimmte Namen einsetzt (z.B. Franz und Karin).

Im Folgenden soll das Symbol $\sim$ dazu dienen, um auf *relationale Ausdrücke* zu verweisen. Wenn man inhaltlich bestimmte Aussagen machen möchte, muss natürlich eine Bedeutung vereinbart werden. Zum Beispiel könnte vereinbart werden, dass das Symbol $\sim$ bis auf weiteres als Abkürzung für den relationalen Ausdruck 'ist verheiratet mit' verwendet werden soll. Unabhängig von der Vereinbarung einer bestimmten Bedeutung können jedoch mit dem Symbol $\sim$ relationale Aussageformen formuliert werden, die allgemein die Form $\omega \sim \omega'$ haben. In dieser Schreibweise handelt es sich also um eine Aussageform. Erst wenn man dem Symbol $\sim$

eine bestimmte Bedeutung gibt und anstelle von ω und ω' Namen für bestimmte Objekte einsetzt, entsteht eine relationale Aussage, die wahr oder falsch sein kann.

2. Ein expliziter Relationsbegriff. Offenbar muss überlegt werden, auf welche Arten von Objekten man sich beziehen kann, um aus relationalen Aussageformen relationale Aussagen zu machen. Die Umgangssprache orientiert sich an der Bedeutung der relationalen Ausdrücke. Ist zum Beispiel für das Symbol $\sim$ die Bedeutung 'ist verheiratet mit' vereinbart worden, ist klar, dass man nur dann zu sinnvollen Aussagen gelangt, wenn man für ω und ω' Namen von Menschen einsetzt. Für die weiteren Überlegungen soll angenommen werden, dass man sich jeweils auf eine explizit definierte Menge beziehen kann, deren Elemente als Objekte für relationale Aussagen verwendet werden können. Zur symbolischen Repräsentation dient wie bei der Definition statistischer Variablen die Schreibweise $\Omega := \{\omega_1, \ldots, \omega_n\}$. Wie in Abschnitt 2.1 erläutert wurde, sind $\omega_1, \ldots, \omega_n$ (fiktive) Namen für die Objekte, auf die man sich gedanklich beziehen möchte, und das Symbol Ω dient zum Verweis auf die Menge dieser Namen bzw. Objekte.

Nach diesen Vorüberlegungen kann der Begriff einer Relation, wie er im weiteren verwendet werden soll, explizit definiert werden. Eine *Relation* besteht aus drei Bestandteilen:

- Es muss ein relationaler Ausdruck $\sim$ eingeführt werden, mit dem relationale Aussageformen der Gestalt $\omega \sim \omega'$ gebildet werden können. (Sobald man nicht nur rein formale Betrachtungen anstellen möchte, muss natürlich auch die inhaltliche Bedeutung angegeben werden.)
- Es muss eine Objektmenge $\Omega := \{\omega_1, \ldots, \omega_n\}$ angegeben werden, deren Elemente als Namen verwendet werden können, um relationale Aussagen zu bilden.
- Schließlich muss angegeben werden, welche der insgesamt möglichen relationalen Aussagen wahr bzw. falsch sind.

Es wäre also eine verkürzte und potentiell irreführende Redeweise, das Symbol $\sim$ eine Relation zu nennen. Dieses Symbol bildet nur ein Hilfsmittel zur Formulierung relationaler Aussagen. Die Relation selbst besteht vielmehr in der Gesamtheit der zutreffenden relationalen Aussagen, die man mithilfe des relationalen Ausdrucks $\sim$ über alle möglichen Paare von Objekten in der Objektmenge Ω machen kann. Sobald man sich dies klargemacht hat, kann man natürlich von einer Relation $(\Omega, \sim)$ sprechen und auch abkürzend von einer Relation $\sim$, wenn der Bezug auf eine bestimmte Objektmenge durch den Kontext gegeben ist.

Ein einfaches Beispiel kann die Begriffsbildungen illustrieren. Die Objektmenge besteht aus 5 Personen: $\Omega := \{\omega_1, \ldots, \omega_5\}$, und es soll festgestellt werden, wer mit wem verheiratet ist. Die Bedeutung des Symbols $\sim$ wird also durch 'ist verheiratet mit' festgelegt. Mithilfe der Aussageform $\omega \sim \omega'$ können in diesem Beispiel auf insgesamt 25 unterschiedliche Weisen

relationale Aussagen gebildet werden. Einige davon sind richtig, die übrigen sind falsch. Angenommen, dass ω_1 und ω_3 und ω_2 und ω_4 verheiratet sind, gibt es folgende Aussagen:

Zutreffende Aussagen	Unzutreffende Aussagen		
$\omega_1 \sim \omega_3$	$\omega_1 \sim \omega_1$	$\omega_2 \sim \omega_5$	$\omega_4 \sim \omega_4$
$\omega_3 \sim \omega_1$	$\omega_1 \sim \omega_2$	$\omega_3 \sim \omega_2$	$\omega_4 \sim \omega_5$
$\omega_2 \sim \omega_4$	$\omega_1 \sim \omega_4$	$\omega_3 \sim \omega_3$	$\omega_5 \sim \omega_1$
$\omega_4 \sim \omega_2$	$\omega_1 \sim \omega_5$	$\omega_3 \sim \omega_4$	$\omega_5 \sim \omega_2$
	$\omega_2 \sim \omega_1$	$\omega_3 \sim \omega_5$	$\omega_5 \sim \omega_3$
	$\omega_2 \sim \omega_2$	$\omega_4 \sim \omega_1$	$\omega_5 \sim \omega_4$
	$\omega_2 \sim \omega_3$	$\omega_4 \sim \omega_3$	$\omega_5 \sim \omega_5$

Die Relation besteht in diesem Beispiel aus der Gesamtheit der 25 Aussagen, von denen 4 zutreffend, die übrigen 21 nicht zutreffend sind.

Das Beispiel zeigt, dass sich eine Relation auf alle möglichen Paare von Objekten bezieht, die man aus den Elementen einer Objektmenge bilden kann. Um diese Paare zu bilden, verwendet man in der Mengenlehre den Begriff eines kartesischen Produkts. Bezieht man sich allgemein auf zwei Mengen A und B, besteht ihr *kartesisches Produkt*, geschrieben $A \times B$, aus allen geordneten Paaren der Form (a, b), wobei a ein Element aus A und b ein Element aus B ist. Ist zum Beispiel $A = \{a, b, c\}$ und $B = \{e, f\}$, erhält man: $A \times B = \{(a, e), (a, f), (b, e), (b, f), (c, e), (c, f)\}$. Bei endlichen Mengen gilt offenbar $|A \times B| = |A| \cdot |B|$, wobei, wenn M irgendeine endliche Menge ist, mit dem Ausdruck $|M|$ auf die Anzahl ihrer Elemente verwiesen werden soll.

Eine Relation für eine Objektmenge Ω gibt nun offenbar für jedes Element $(\omega, \omega') \in \Omega \times \Omega$ an, ob die relationale Aussage $\omega \sim \omega'$ zutrifft oder nicht. Somit kann man auch jede Relation für eine Objektmenge Ω durch eine Teilmenge des kartesischen Produkts $\Omega \times \Omega$ festlegen, die genau diejenigen Paare (ω, ω') enthält, für die die relationale Aussage zutrifft. In unserem Beispiel: $R^* := \{(\omega_1, \omega_3), (\omega_3, \omega_1), (\omega_2, \omega_4), (\omega_4, \omega_2)\}$. Diese Methode wird *Definition einer Relation durch ein kartesisches Produkt (einer Objektmenge Ω mit sich selbst)* genannt. Offenbar entspricht jeder Teilmenge von $\Omega \times \Omega$ eine spezifische Relation für die Elemente von Ω.

3. Relationale Variablen. Eine andere Möglichkeit, um sich begrifflich auf Relationen für eine Objektmenge Ω zu beziehen, besteht in der Verwendung *relationaler Variablen*. Mit diesem Begriff sind zunächst (später wird die Definition verallgemeinert) Funktionen gemeint, die folgende Form haben: $R : \Omega \times \Omega \longrightarrow \{0, 1\}$. R ist der Name der Funktion (der relationalen Variablen), $\Omega \times \Omega$ ist ihr Definitionsbereich, und $\{0, 1\}$ ist ihr Wertebereich. Die Funktion (relationale Variable) R ordnet also jedem Element $(\omega, \omega') \in \Omega \times \Omega$ einen Wert $R(\omega, \omega') \in \{0, 1\}$ zu, wobei folgende Bedeutung vereinbart wird:

$$R(\omega,\omega') = \begin{cases} 1 & \text{wenn } \omega \sim \omega' \text{ zutrifft} \\ 0 & \text{wenn } \omega \sim \omega' \text{ nicht zutrifft} \end{cases}$$

Wie sich später zeigen wird, ist der Begriff einer relationalen Variablen sehr nützlich, weil er sich leicht verallgemeinern lässt, um in komplexerer Weise von Relationen zu sprechen. Außerdem gibt es eine gedanklich einfache Parallele zu statistischen Variablen, also zu Funktionen $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, die jedem Element einer Objektmenge Ω einen Merkmalswert in einem Merkmalsraum $\tilde{\mathcal{X}}$ zuordnen. Der Unterschied besteht nur darin, dass eine statistische Variable jedem einzelnen Objekt, eine relationale Variable dagegen jedem Paar von Objekten einen Merkmalswert zuordnet.

An dieser Parallele knüpft auch eine weitere Möglichkeit zur Darstellung von Relationen an. Beziehen wir uns zunächst auf eine statistische Variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$. Ihre Werte (die Daten) können in Form einer Datenmatrix dargestellt werden, die folgende Form hat:

ω_1	$X(\omega_1)$
$\vdots$	$\vdots$
ω_n	$X(\omega_n)$

Jede Zeile bezieht sich auf jeweils ein Objekt der Objektmenge Ω. Die erste Spalte enthält den Namen des Objekts, die zweite Spalte den Merkmalswert, der dem Objekt durch die Variable zugeordnet wird. Auf ähnliche Weise kann man die Werte einer relationalen Variablen durch ein zweidimensionales Schema darstellen, das allgemein folgende Form hat:

	ω_1	$\cdots$	ω_n
ω_1	$R(\omega_1,\omega_1)$	$\cdots$	$R(\omega_1,\omega_n)$
$\vdots$	$\vdots$		$\vdots$
ω_n	$R(\omega_n,\omega_1)$	$\cdots$	$R(\omega_n,\omega_n)$

Für das oben angeführte Beispiel erhält man folgende Darstellung:

	ω_1	ω_2	ω_3	ω_4	ω_5
ω_1	0	0	1	0	0
ω_2	0	0	0	1	0
ω_3	1	0	0	0	0
ω_4	0	1	0	0	0
ω_5	0	0	0	0	0

Wenn ein Schema dieser Art verwendet wird, um eine Relation darzustellen, spricht man von einer *Adjazenzmatrix*.

4. Reflexivität, Symmetrie, Transitivität. Zur formalen Charakterisierung von Relationen gibt es zahlreiche Begriffsbildungen. An dieser Stelle genügen die folgenden, zu deren Erläuterung angenommen wird, dass eine Relation $(\Omega,\sim)$ gegeben ist.

– Die Relation wird *reflexiv* genannt, wenn für alle $\omega \in \Omega$ gilt: $\omega \sim \omega$.
– Die Relation wird *symmetrisch* genannt, wenn für alle $\omega,\omega' \in \Omega$ gilt: $\omega \sim \omega' \Longrightarrow \omega' \sim \omega$.[6]
– Die Relation wird *transitiv* genannt, wenn für alle $\omega,\omega',\omega'' \in \Omega$ gilt: $\omega \sim \omega'$ und $\omega' \sim \omega'' \Longrightarrow \omega \sim \omega''$.

Die oben als Beispiel verwendete Relation ist offenbar symmetrisch, jedoch weder reflexiv noch transitiv. Wenn eine Relation alle drei Eigenschaften hat, spricht man auch von einer *Äquivalenzrelation*. Es sei auch daran erinnert, dass jede Äquivalenzrelation, die für eine Objektmenge Ω definiert ist, einer *Partition* von Ω entspricht, d.h. einer Zerlegung von Ω in disjunkte Teilmengen (die in diesem Fall *Äquivalenzklassen* genannt werden).[7]

5. Ungerichtete Graphen. Unter einem *Graphen* versteht man allgemein eine Menge von *Knoten*, die durch *Kanten* (Linien oder Pfeile) verbunden sein können. Die Knoten dienen zur Repräsentation der Objekte, auf die man sich beziehen möchte, die Kanten werden zur Darstellung von Beziehungen zwischen den Knoten (Objekten) verwendet. Zur Notation dient die Schreibweise $\mathcal{G} := (\Omega,\mathcal{K})$, wobei $\Omega := \{\omega_1,\ldots,\omega_n\}$ die *Knotenmenge* des Graphen und $\mathcal{K} := \{\kappa_1,\ldots,\kappa_m\}$ die *Kantenmenge* des Graphen ist.

Diese Erläuterung zeigt bereits, dass es einen engen Zusammenhang zwischen Relationen und Graphen gibt. Zunächst besprechen wir *ungerichtete Graphen*, die den symmetrischen Relationen entsprechen. Sei also $(\Omega,\sim)$ eine symmetrische Relation. Dann kann man Ω auch als Knotenmenge eines Graphen betrachten und festlegen, dass es zwischen zwei Knoten $\omega,\omega' \in \Omega$ genau dann eine Kante gibt, wenn die relationale Aussage $\omega \sim \omega'$ zutrifft. Die Kantenmenge wird also durch

$$\mathcal{K} := \{\, \{\omega,\omega'\} \,|\, \omega \sim \omega' \text{ ist zutreffend} \,\}$$

definiert. Anstelle von geordneten Paaren der Form (ω,ω') werden in die-

[6] Hier wird der *logische Regelpfeil* verwendet, durch den Aussageformen verknüpft werden können, allgemein: $F(x,y,z,\ldots) \Longrightarrow G(x,y,z,\ldots)$. Eine Regel dieser Art besagt: Wenn man durch Einsetzen von Namen $x^*,y^*,z^*,\ldots$ in die logischen Variablen $x,y,z,\ldots$ zu einer wahren Aussage $F(x^*,y^*,z^*,\ldots)$ gelangt, dann garantiert die Regel, dass auch $G(x^*,y^*,z^*,\ldots)$ eine wahre Aussage ist.

[7] Als Beispiel kann man daran denken, dass korrespondierend zu jeder statistischen Variablen $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ durch

$$\omega \sim_x \omega' \iff X(\omega) = X(\omega')$$

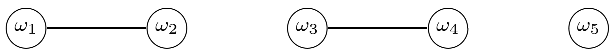
eine Äquivalenzrelation $\sim_x$ definiert werden kann. Ihre Äquivalenzklassen sind diejenigen Teilmengen von Ω, deren Mitglieder den gleichen Wert der Variablen X aufweisen. Wir nennen dies die *durch X induzierte Partition von Ω*.

sem Fall Mengen der Form $\{\omega, \omega'\}$ verwendet, da die Relation symmetrisch ist, so dass die Reihenfolge keine Rolle spielt.

Zur Illustration kann zunächst das bisher verwendete Beispiel dienen. In diesem Fall repräsentieren die Knoten die 5 Personen und die Kanten zeigen, welche der Personen miteinander verheiratet sind. In symbolischer Notation hat dieser Graph folgende Form:

$$\mathcal{G} := (\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}, \{\{\omega_1, \omega_3\}, \{\omega_2, \omega_4\}\})$$

Anhand dieses Beispiels kann auch die *graphische Darstellung* von Graphen erläutert werden. Jeder Knoten des Graphen wird durch einen Punkt (oder Kreis, Rechteck, ...) und jede Kante durch eine Verbindungslinie zwischen den zugehörigen Knoten dargestellt. In unserem Beispiel kann man folgende Darstellung verwenden:



Die Anordnung der Knoten kann beliebig erfolgen, denn sie hat keine Bedeutung.[8] Oft wählt man eine Anordnung, die möglichst keine oder nur wenige Überschneidungen der die Kanten repräsentierenden Linien erfordert. Ein Graph wird *planar* genannt, wenn man ihn vollständig ohne Überschneidungen darstellen kann.

Für ein weiteres Beispiel können Daten dienen, die in der ersten Stunde eines Seminars erhoben wurden, an dem 10 Personen teilgenommen haben. Das Ziel war herauszufinden, welche Teilnehmer „sich bereits kennen“. Um das zu präzisieren, wurde folgende Fragestellung gewählt: Haben jeweils zwei der Teilnehmer vor Beginn des Seminars schon mindestens 5 Minuten miteinander gesprochen? Um die Daten zu gewinnen, wurde zunächst eine Liste der Teilnehmer erstellt:

$$\Omega := \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10}\}$$

Dann wurde jeder Teilnehmer gefragt, mit welchen anderen Seminarteilnehmern er bereits vor Beginn des Seminars mindestens 5 Minuten gesprochen hat. Tabelle 3.2-1 zeigt das Ergebnis in Gestalt einer Adjazenzmatrix. Sie beschreibt einen Graphen, dessen Knoten aus den 10 Teilnehmern des Seminars bestehen. Die Einsen geben die Kanten des Graphen an und bedeuten, dass zwischen den jeweils beteiligten Knoten eine „Beziehung“ besteht, in diesem Beispiel dadurch definiert, dass bereits vor Beginn des

[8] Dies sollte betont werden: dass die räumliche Anordnung der Knoten bei der Darstellung eines Graphen keine Bedeutung hat. Dadurch unterscheidet sich nämlich diese Darstellung von Methoden zur Erzeugung räumlicher Bilder, bei denen ein vermeintlicher Erkenntnisgewinn durch die Suggestion räumlicher Interpretierbarkeit erzielt werden soll; exemplarisch sei auf Methoden der Korrespondenzanalyse und der multidimensionalen Skalierung hingewiesen.

Tabelle 3.2-1 Adjazenzmatrix der Seminardaten.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
ω_1	0	0	0	0	1	0	1	1	0	0
ω_2	0	0	0	0	0	0	0	1	0	0
ω_3	0	0	0	0	0	0	0	0	0	0
ω_4	0	0	0	0	1	1	0	0	0	0
ω_5	1	0	0	1	0	1	0	0	0	0
ω_6	0	0	0	1	1	0	0	1	1	0
ω_7	1	0	0	0	0	0	0	0	0	0
ω_8	1	1	0	0	0	1	0	0	0	0
ω_9	0	0	0	0	0	1	0	0	0	1
ω_{10}	0	0	0	0	0	0	0	0	1	0

Seminars eine Kommunikation stattgefunden hat. Da es sich um eine symmetrische Relation handelt, ist auch die Adjazenzmatrix symmetrisch und man kann zur Repräsentation einen ungerichteten Graphen verwenden.

6. Gerichtete Graphen. Oft sind Relationen nicht symmetrisch, dann werden *gerichtete Graphen* verwendet. Zur symbolischen Notation kann wie bei ungerichteten Graphen die Formulierung $\mathcal{G} := (\Omega, \mathcal{K})$ verwendet werden. Es ist jedoch zu berücksichtigen, dass bei gerichteten Graphen die Kantenmenge $\mathcal{K}$ aus *geordneten* Paaren von Knoten besteht, so dass bei zwei Knoten ω und ω' zwischen Kanten, die von ω zu ω' führen, und Kanten, die von ω' zu ω führen, unterschieden werden kann. Zur Unterscheidung wird von *gerichteten Kanten* gesprochen. In der graphischen Darstellung werden Pfeile (anstelle von Linien) verwendet.

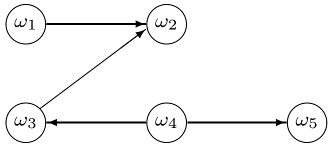
Als Beispiel betrachten wir eine Objektmenge Ω, die aus 5 Unternehmen besteht. Mit den relationalen Aussagen der Form $\omega \sim \omega'$ soll erfasst werden, ob das Unternehmen ω' Produkte des Unternehmens ω als Vorleistungen verwendet. Es werden folgende Beziehungen festgestellt:

$$\omega_1 \sim \omega_2,\ \omega_3 \sim \omega_2,\ \omega_4 \sim \omega_3,\ \omega_4 \sim \omega_5$$

so dass die Adjazenzmatrix folgendermaßen aussieht:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Offenbar ist diese Adjazenzmatrix und somit die Relation nicht symmetrisch. Zur Darstellung wird deshalb ein gerichteter Graph verwendet, dessen Kantenmenge durch $\mathcal{K} := \{(\omega_1, \omega_2), (\omega_3, \omega_2), (\omega_4, \omega_3), (\omega_4, \omega_5)\}$ definiert ist. Es handelt sich um gerichtete Kanten, und die graphische Darstellung sieht folgendermaßen aus:

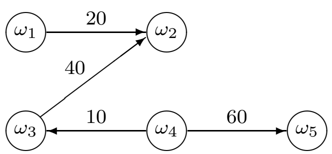


7. Bewertete Graphen. Bei einer Relation $(\Omega, \sim)$ wird nur festgestellt, ob für jeweils zwei Objekte $\omega, \omega' \in \Omega$ die relationale Aussage $\omega \sim \omega'$ zutrifft oder nicht. Zum Beispiel: Zwei Personen sind verheiratet oder nicht verheiratet. Oft ist es jedoch von Interesse, qualitative oder quantitative Unterschiede in der Art der Beziehung zu erfassen. Zum Beispiel könnte man bei persönlichen Beziehungen zwischen Bekanntschaften und Freundschaften unterscheiden; oder bei dem im vorangegangenen Paragraphen verwendeten Beispiel könnte man unterscheiden, in welchem Umfang Vorleistungen bezogen werden. Um solche Unterscheidungen berücksichtigen zu können, werden *bewertete Graphen* verwendet: Jeder (gerichteten oder ungerichteten) Kante des Graphen wird dann eine Zahl zugeordnet, die die durch die Kante repräsentierte Beziehung charakterisiert.

Als Beispiel verwenden wir wieder eine Objektmenge, die aus 5 Unternehmen besteht. In diesem Fall soll es sich jedoch um Aktiengesellschaften handeln, so dass man feststellen kann, wie viel Prozent des Aktienkapitals eines Unternehmens von einem anderen Unternehmen gehalten wird. Solche Daten können wiederum in Form einer Adjazenzmatrix dargestellt werden, wobei jetzt aber in den einzelnen Feldern der Matrix die Prozentanteile des Kapitalbesitzes eingetragen werden. In unserem Beispiel sieht die Adjazenzmatrix vielleicht folgendermaßen aus:

$$\begin{pmatrix} 0 & 20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 40 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 60 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Das Unternehmen ω_1 hält am Unternehmen ω_2 20 % der Kapitalanteile usw. Man erhält dann folgende graphische Darstellung:



Zur symbolischen Notation bewerteter Graphen wird in der Literatur oft die Formulierung $\mathcal{G} := (\Omega, \mathcal{K}, v)$ verwendet. Ω ist die Knotenmenge, $\mathcal{K}$ die Kantenmenge. Hinzu kommt eine Funktion $v : \mathcal{K} \longrightarrow \mathbf{R}$, die jeder Kante

$\kappa \in \mathcal{K}$ eine Zahl $v(\kappa) \in \mathbf{R}$ zuordnet und als *Bewertung der Kante* bezeichnet wird (wobei natürlich eine jeweils sinnvolle Bedeutung vereinbart werden muss).

8. Allgemeine relationale Variablen. Als einheitlicher begrifflicher Rahmen für Graphen aller Art eignen sich am besten relationale Variablen, die in folgender Form als Funktionen definiert sind: $R : \Omega \times \Omega \longrightarrow \tilde{\mathcal{R}}$. Hierbei ist Ω eine Objektmenge und $\tilde{\mathcal{R}}$ ein im Prinzip beliebig konzipierbarer Merkmalsraum. Die relationale Variable R ordnet jedem Element $(\omega, \omega') \in \Omega \times \Omega$ einen Merkmalswert $R(\omega, \omega') \in \tilde{\mathcal{R}}$ zu. Wie bereits besprochen wurde, genügt für unbewertete Graphen ein Merkmalsraum $\tilde{\mathcal{R}} := \{0, 1\}$, da nur unterschieden werden muss, ob zwischen zwei Objekten eine Beziehung besteht oder nicht. Wenn man differenziertere Merkmalsräume verwendet, können jedoch auch beliebige bewertete Graphen repräsentiert werden. Für das zuvor besprochene Beispiel kann man als Merkmalsraum zum Beispiel die Zahlen von 0 bis 100 verwenden, und $R(\omega, \omega')$ bedeutet dann den Prozentanteil des Kapitals des Unternehmens ω', den das Unternehmen ω besitzt. Relationale Variablen bieten also sehr flexible Formulierungsmöglichkeiten. Außerdem lassen sich viele Überlegungen und Unterscheidungen, die für statistische Variablen bereits eingeführt worden sind, unmittelbar übertragen.

Besonders wichtig ist, dass man analog zu mehrdimensionalen statistischen Variablen auch mehrdimensionale relationale Variablen betrachten kann. Allgemein haben sie folgende Form:

$$(R_1, \ldots, R_m) : \Omega \times \Omega \longrightarrow \tilde{\mathcal{R}}_1 \times \ldots \times \tilde{\mathcal{R}}_m$$

Es werden also für die Objektmenge Ω gleichzeitig m relationale Variablen $R_1, \ldots, R_m$ mit den zugehörigen Merkmalsräumen $\tilde{\mathcal{R}}_1, \ldots, \tilde{\mathcal{R}}_m$ definiert. Die korrespondierenden Graphen werden in der Literatur oft *Multigraphen* genannt. Als Beispiel kann man sich vorstellen, dass bei einer Menge von Unternehmen sowohl Kapitalverflechtungen als auch personelle Verflechtungen und Austausch von Gütern erfasst wird.

9. Ein allgemeiner Netzwerkbegriff. Offenbar ist der Begriff einer relationalen statistischen Variablen sehr allgemein und umfasst das Reden von Relationen und Graphen. Er kann deshalb auch verwendet werden, um in einer allgemeinen Weise von Netzwerken zu sprechen. Wir verwenden in diesem Text folgende Definition: Etwas ist ein *Netzwerk,* wenn bzw. insoweit es durch eine (ggf. mehrdimensionale und/oder multi-modale) relationale Variable repräsentiert werden kann.

Das unspezifische Reden von „etwas“ soll es erlauben, sich mit dem Netzwerkbegriff nicht nur auf materielle Aspekte unserer Erfahrungswelt zu beziehen, sondern auch auf abstrakte Mengen wie z.B. Zahlen oder Merkmalsräume. Für diesen allgemeinen Netzwerkbegriff werden somit keine besonderen Anforderungen an die Beschaffenheit der Objekte oder der sie verknüpfenden Beziehungen gestellt. Insofern für das Reden von Netzwerken eine bestimmte formale Repräsentation vorausgesetzt wird, ist der Begriff dennoch enger als der abstrakte Systembegriff.

3.3 Relationale Strukturbegriffe

Der Strukturbegriff wird hauptsächlich in zwei unterschiedlichen Bedeutungen verwendet: Einerseits in einer statistischen Bedeutung, in der sich das Wort auf eine oder mehrere Merkmalsverteilungen in einer statistischen Gesamtheit bezieht; damit haben wir uns in Abschnitt 2.2 beschäftigt. Andererseits wird das Wort verwendet, um in einer vergleichsweise unspezifischen Weise auf die Gliederung und den Aufbau irgendeines (realen oder fiktiven) Sachverhalts zu verweisen. In dieser zweiten Bedeutung sagte zum Beispiel Ludwig Wittgenstein in seinem *Tractatus* (2.032): „Die Art und Weise, wie die Gegenstände im Sachverhalt zusammenhängen, ist die Struktur des Sachverhalts.“ In einer ähnlichen Formulierung heißt es bei G. Hernes (1976: 518): „A structure is a configuration of parts, and a structural description is a characterization of the way the components in a set are interrelated.“ An dieses Wortverständnis knüpfen die relationalen (Sozial-) Strukturbegriffe an, mit denen wir uns in diesem Abschnitt beschäftigen.

1. Definition eines relationalen Strukturbegriffs. Zu einer allgemeinen Definition eines relationalen Strukturbegriffs gelangt man, wenn man sich auf Gesamtheiten von Objekten bezieht, die durch Beziehungen miteinander verbunden sind. In diesem Kontext bezieht sich das Wort darauf, wie die Objekte durch Beziehungen zusammenhängen, und wir sprechen dann von einem *relationalen Strukturbegriff.*

In dieser allgemeinen Bedeutung kann der relationale Strukturbegriff auf beliebige Systeme (im Sinne des abstrakten Systembegriffs) angewendet werden, insbesondere auf Netzwerke, so wie dieser Begriff in Abschnitt 3.2 (§ 9) definiert worden ist. Allerdings muss auf eine Ambivalenz in der Begriffsverwendung geachtet werden. Man betrachte dafür als Beispiel das Netzwerk aus Abschnitt 3.2 (§ 5), in dem Beziehungen des „Sich-bereits-Kennens“ zwischen 10 Teilnehmern eines Seminars dargestellt werden. Was ist die Struktur dieses Netzwerks? Wenn man den relationalen Strukturbegriff in seiner allgemeinen Bedeutung verwendet, gibt es zunächst keinen begrifflichen Unterschied zwischen dem Netzwerk und seiner Struktur; oder anders formuliert: Man beschreibt die Struktur eines Netzwerks, indem man das Netzwerk beschreibt.[9] Um diese Ambivalenz zu vermeiden, werden wir – in Übereinstimmung auch mit den anfangs angeführten Zitaten von Wittgenstein und Hernes – festlegen, dass sich der relationale

[9]So heißt es z.B. bei David Krackhardt (1987: 113): „The structure of any system is defined as a set of relational statements between all pairs of actors in the system.“

Strukturbegriff nur auf formale Eigenschaften eines Netzwerks (oder Systems) bezieht; in dem Beispiel etwa auf die Anzahl der Objekte und den Grad der Dichte ihrer Beziehungen, um nur zwei formale Aspekte zu nennen, nicht jedoch auf die Art der Objekte und die inhaltliche Bedeutung der Beziehungen.[10] Infolgedessen ist es auch möglich, dass unterschiedliche Netzwerke die gleiche relationale Struktur aufweisen.

2. Unterschiedliche Arten von Beziehungen. Um Netzwerke zu verstehen, ist nicht nur ihre Struktur von Bedeutung; zuerst muss man die Knoten und Beziehungen verstehen, auf die das Netzwerk Bezug nimmt. Darin unterscheiden sich auch zunächst die Netzwerke, die in der empirischen Sozialforschung betrachtet werden können. Allerdings ist es kaum möglich, eine vollständige Übersicht über alle Möglichkeiten zu gewinnen, in denen von Beziehungen gesprochen werden kann. Wir beschränken uns deshalb an dieser Stelle darauf, einige gelegentlich nützliche allgemeine Unterscheidungen anzudeuten.[11]

a) Eine Möglichkeit, von Beziehungen zwischen zwei oder mehr Objekten zu sprechen, beruht auf einem Vergleich von Eigenschaften, durch die man die Objekte zunächst jeweils separat charakterisieren kann; wir sprechen dann von *komparativen Beziehungen*. Zum Beispiel: ω ist größer als ω' oder ist älter als ω' oder ist gleichalt wie ω'. Insbesondere kann man solche Beziehungen bilden, wenn zunächst eine statistische Variable gegeben ist, indem man die Merkmalswerte von zwei oder mehr Objekten vergleicht.

b) Eine andere Möglichkeit zur Definition von Beziehungen zwischen zwei oder mehr Objekten besteht darin, auf eine Situation oder einen Kontext Bezug zu nehmen, dem die Objekte in einer bestimmten Weise angehören. So gelangt man zu *kontextabhängigen Beziehungen*. Zur Charakterisierung solcher Beziehungen kann man sowohl von Ereignissen als auch von Sachverhalten ausgehen.

   (α) Einerseits kann man von *ereignisförmigen Beziehungen* zwischen zwei oder mehr Objekten sprechen, wenn die Objekte irgendwie in ein die Beziehung konstituierendes Ereignis einbezogen sind. Zum Beispiel zwei Personen, die sich unterhalten, bei einem Verkehrsunfall zusam-

[10]Dieser Vorschlag zum Sprachgebrauch entspricht auch folgender Bemerkung von F. U. Pappi (1987: 15): „Netzwerke sind nach unserer Definition empirische Systeme. Sie lassen sich formal als Graphen darstellen.“ Etwas ausführlicher heißt es bei K.-D. Opp und H. J. Hummell (1973: 67): Es „soll im folgenden unter *Struktur* eine spezielle formale Eigenschaft von Netzwerken (Relationengebilden) verstanden werden. Beschreibt man zwei Relationengebilde durch Boolesche Matrizen [= Adjazenzmatrizen], dann sollen die *Netzwerke von gleicher Struktur* genau dann sein, wenn die zugeordneten Matrizen identisch bzw. durch Permutationsmatrizen ineinander transformierbar sind.“

[11]In der Literatur, die sich mit sozialen Netzwerken beschäftigt, beziehen sich zahlreiche Überlegungen auch auf inhaltliche Unterscheidungen; man vgl. die Diskussion bei F. U. Pappi (1987: 16ff.).

menstoßen oder an der gleichen Landtagswahl teilgenommen haben. Das zuletzt angeführte Beispiel zeigt, dass eine ereignisförmige Beziehung zwischen zwei Objekten nicht impliziert, dass es zwischen den Objekten auch einen direkten (oder allgemeiner: irgendwie kausal relevanten) Kontakt gibt. Wir sprechen deshalb im engeren Sinn von einer *Interaktionsbeziehung* (oder kurz: *Interaktion*), wenn in irgendeiner Form ein kommunikativer Austausch und/oder eine physische Wechselwirkung stattfindet.[12] Natürlich kann man zur Definition ereignisförmiger Beziehungen auch zeitliche Sequenzen mehrerer Ereignisse verwenden; man denke z.B. an Beschreibungen persönlicher Beziehungen, bei denen fast immer solche Bezugnahmen auf ihre Geschichte stattfinden. In jedem Fall, auch wenn nur auf ein Ereignis Bezug genommen wird, setzen empirische Feststellungen über ereignisförmige Beziehungen eine retrospektive Betrachtungsweise voraus.

(β) Andererseits kann man zur Definition von Beziehungen auch von Sachverhalten ausgehen, bei denen es sich nicht um Ereignisse handelt; zum Beispiel: Zwei Orte sind durch eine Straße miteinander verbunden. Es ist allerdings fraglich, ob der Unterscheidung zwischen Ereignissen und Sachverhalten auch eine relevante Unterscheidung zwischen Arten von Beziehungen entspricht. Denn in einer allgemeinen Sprechweise kann man auch dann von einem Sachverhalt sprechen, wenn ein Ereignis stattgefunden hat; und andererseits impliziert die gedankliche Bezugnahme auf einen Sachverhalt nicht, dass er während eines längeren Zeitraums (unverändert) existiert. Das mag der Fall sein (wie vermutlich bei der Straße zwischen den beiden Orten), aber es muss nicht der Fall sein (wie z.B. bei zwei Computern, die nur für einen kurzen Zeitraum durch ein Kabel miteinander verbunden werden, um Daten zu übertragen).

Anhand von Beispielen kann man sich verdeutlichen, dass es auch zwischen komparativen und kontextabhängigen Beziehungen keine vollständig scharfe Unterscheidung gibt. Ein wichtiges Beipiel ist räumliche Nähe, die man sowohl als eine komparative als auch als eine kontextabhängige Beziehung betrachten kann.

3. Faktische und modale Betrachtungsweisen. Wichtiger als allgemeine Unterscheidungen zwischen Arten von Beziehungen ist der Umstand, dass es in vielen Fällen bei ihrer Betrachtung und Darstellung eine wesentliche Ambivalenz gibt. Zum Beispiel: Zwei Computer sind durch ein Kabel für den Austausch von Daten miteinander verbunden. Das ist einerseits eine empirisch feststellbare Tatsache, die jedoch andererseits auf eine Möglichkeit, nämlich einen Austausch von Daten, verweist. Oder: Eine Person ist bei einem Unternehmen angestellt. Diese Feststellung verweist einerseits auf eine bestimmte Tatsache – dass irgendwann ein Arbeitsvertrag vereinbart wurde und immer noch besteht –, andererseits auf mögliche Verhaltensweisen, die infolge des Arbeitsvertrags realisiert werden sollten. Wie in diesen Beispielen kann man in vielen Fällen zwei Aspekte unterscheiden:

[12] Der in diesem Text verwendete Begriff einer Interaktion setzt also nicht unbedingt Akteure voraus; und auch dann, wenn er sich auf individuelle Akteure bezieht, muss es sich nicht um „soziales Handeln“ im Sinne Max Webers oder um „bewertende Austauschprozesse“ wie etwa bei George Homans (1961: 35) handeln.

a) Einerseits einen *faktischen* Aspekt, der sich auf empirisch fixierbare Sachverhalte oder Ereignisse (z.B. eine Kabelverbindung oder das Vorhandensein eines Arbeitsvertrags) bezieht, durch die eine Beziehung faktisch begründet wird; und

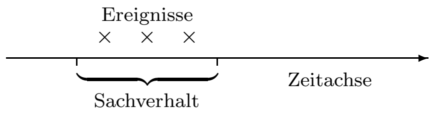
b) andererseits einen *modalen* Aspekt, der sich – je nach der Art der beteiligten Objekte oder Personen – auf mögliche Verhaltensweisen bezieht, die infolge der faktischen Beziehung möglich oder wahrscheinlich oder normativ gefordert werden.

Es handelt sich dabei nicht um unterschiedliche Arten von Beziehungen, sondern um unterschiedliche Betrachtungsweisen einer Beziehung. Um noch ein Beispiel anzuführen: Die Aussage, dass zwei Personen verheiratet sind, kann einerseits bedeuten, dass zu einem bestimmten Zeitpunkt ein Ereignis stattgefunden hat, durch das die beiden Personen verheiratet wurden. Andererseits können aber auch bestimmte Verhaltensweisen gemeint sein, die durch dieses Ereignis möglich und/oder normativ gefordert werden.

Zu betonen ist, dass sich die modale Betrachtungsweise auf Möglichkeiten bezieht. Zwar kann man bei allen Beziehungen, die eine modale Betrachtungsweise erlauben, auch eine *retrospektive* Betrachtungsweise einnehmen (deren Reichweite natürlich von der bisherigen Dauer der Beziehung abhängt). So kann man z.B. feststellen, in welchem Umfang Daten zwischen den beiden Computern ausgetauscht worden sind oder wie sich die beiden Personen während ihrer bisherigen Ehe zueinander verhalten haben. Bei der modalen Betrachtungsweise geht es jedoch nicht um eine retrospektive Feststellung von Interaktionen, sondern darum, *wie durch Beziehungen mögliche Verhaltensweisen der jeweils beteiligten Objekte konstituiert werden.* Dabei zielt das Erkenntnisinteresse nicht nur auf Unterschiede in den Verhaltensweisen selbst, sondern auch darauf, wie sie durch die Beziehung ermöglicht, wahrscheinlich gemacht oder normativ gefordert werden.

4. Beziehungen und mögliche Ereignisse. Denken wir noch einmal an die beiden Computer. Offenbar kann der Sachverhalt, dass sie durch ein Kabel miteinander verbunden sind, verwendet werden, um eine Beziehung zwischen ihnen festzustellen. Natürlich könnten auch zahlreiche andere Sachverhalte verwendet werden, um andere Beziehungen zwischen ihnen festzustellen; zum Beispiel, dass sie nebeneinander auf einem Tisch stehen oder sich im gleichen Raum befinden. Bei den Beziehungen, die auf diese Weise definiert werden können, handelt es sich ersichtlich nicht um ereignisförmige Beziehungen, insbesondere nicht um Interaktionen.

Ein Zusammenhang kann jedoch hergestellt werden, wenn die durch einen Sachverhalt definierte Beziehung eine modale Betrachtungsweise erlaubt. In unserem Beispiel ist das der Fall, wenn die beiden Computer durch ein Kabel verbunden sind, so dass es möglich wird, Daten auszutauschen. Wie das folgende Bild andeutet, ermöglicht dann der die Beziehung definierende Sachverhalt Ereignisse einer bestimmten Art:



Die Ereignisse (Austausch von Daten) werden durch den Sachverhalt (die Kabelverbindung) nicht verursacht, sondern ermöglicht. Insofern bilden sie keinen realen, sondern nur einen modalen Aspekt der durch den Sachverhalt definierten Beziehung zwischen den beiden Computern. Wenn jedoch solche Ereignisse stattfinden, kann man sich retrospektiv auf sie beziehen und dadurch auch eine ereignisförmige Beziehung definieren. Sie ist natürlich mit der ursprünglich durch den Sachverhalt definierten Beziehung nicht identisch.

Formal analog verhält es sich, wenn ereignisförmige Beziehungen eine modale Betrachtungsweise erlauben, also als Bedingungen für mögliche spätere Ereignisse betrachtet werden können. Das Basisereignis, durch das die Beziehung zunächst zustande kommt, kann dann nämlich als zeitlicher Beginn eines Sachverhalts aufgefasst werden, während dessen Vorhandensein wie im oben skizzierten Bild weitere Ereignisse stattfinden können. Zum Beispiel besteht das Basisereignis darin, dass ein Arbeitsvertrag abgeschlossen wird; und dadurch entsteht dann für eine gewisse Zeit ein bestimmter Sachverhalt, der seinerseits einen Rahmen für weitere Ereignisse (anderer Art) bildet.

5. Soziale Beziehungen. In der sozialwissenschaftlichen Literatur wird in unterschiedlichen Bedeutungen von „sozialen Beziehungen“ gesprochen. Soziologen orientieren sich oft an folgender Definition Max Webers:

> „Soziale „Beziehung“ soll ein seinem Sinngehalt nach aufeinander gegenseitig *eingestelltes* und dadurch orientiertes Sichverhalten mehrerer heißen. Die soziale Beziehung *besteht* also durchaus und ganz ausschließlich: in der *Chance*, daß in einer (sinnhaft) angebbaren Art sozial gehandelt wird, einerlei zunächst: worauf diese Chance beruht.“ (Weber 1921/1976: 13)

Diese Definition sozialer Beziehungen ist jedoch für die Sozialforschung, wie sie in diesem Text verstanden wird, ungeeignet.

– Zunächst deshalb, weil sie von vornherein nur personelle Beziehungen einbezieht und, noch enger, nur „soziales Handeln“ in Betracht zieht.

Somit werden viele wichtige Arten von Beziehungen, wie z.B. Beziehungen zwischen Organisationen oder indirekte Abhängigkeitsbeziehungen zwischen Personen, die nicht durch „sinnhafte Orientierungen“ erschlossen werden können, nicht erfasst.

– Ein zweites Problem betrifft das Verständnis faktischer und modaler Aspekte von Beziehungen. Webers Definition bezieht sich zunächst auf einen faktischen Aspekt („Sichverhalten“), wechselt dann aber durch ein Reden von „Chancen“ unvermittelt in eine scheinbar modale Betrachtungsweise.

– Schließlich liefert aber diese modale Betrachtungsweise gerade bei Beziehungen zwischen sozialen Akteuren in den meisten Fällen kein adäquates Verständnis. Denn Webers Chancenbegriff bezieht sich auf „statistische Wahrscheinlichkeiten“ und infolgedessen auf Häufigkeiten von Verhaltensweisen, die bei einer retrospektiven Betrachtung von Beziehungen festgestellt werden können.[13] Die wesentliche Differenz zwischen einer modalen und einer retrospektiven Betrachtungsweise von Beziehungen wird infolgedessen nicht nur verwischt, sondern verkehrt. Folgte man Webers Definition, ergäbe sich z.B. der Inhalt der Beziehung, die zwischen einer Person und einem Unternehmen durch den Abschluss eines Arbeitsvertrags zustande kommt, durch das nachfolgende Verhalten der Person und des Unternehmens (denn auf dieses Verhalten bezieht sich Webers Chancenbegriff). Eine modale Betrachtungsweise dieser Beziehung müsste dagegen auf die normativen Festlegungen des Arbeitsvertrags Bezug nehmen.

Besonders wichtig erscheint mir der zuletzt genannte Kritikpunkt. Denn dabei geht es nicht nur um ein richtiges Verständnis modaler Betrachtungsweisen von Beziehungen (insbesondere zwischen sozialen Akteuren) und ihre Unterscheidung von retrospektiven Betrachtungsweisen.[14] Vielmehr geht es auch um die Frage, ob und ggf. wie man zumindest in einigen Fällen Beziehungen als Bedingungen für das Verhalten beteiligter Akteure betrachten kann. Webers Ansatz verstellt schon den Zugang zu dieser Fragestellung, denn durch empirische Häufigkeiten definierte „Chancen“ für das je faktische Verhalten (im Kontext einer Beziehung) können nicht als dessen Bedingungen verstanden werden.

Es ist jedoch nicht erforderlich, Webers Definition durch eine andere

---

[13] Insbesondere bezieht sich Webers Chancenbegriff nicht auf Handlungschancen, die von Akteuren reflektiert und wahrgenommen werden können. Zur Unterscheidung zwischen Handlungschancen und „statistischen Chancen“ vgl. man die Ausführungen bei Rohwer und Pötter (2002b: 166ff.).

[14] Man kann sich das Problem an beliebig vielen Beispielen verdeutlichen. Angenommen, ω hat mit ω' einen Kaufvertrag abgeschlossen, die vereinbarte Ware aber nicht geliefert. Sollte man dann (unter Berufung auf Webers Definition) sagen, dass ω' die durch den Kaufvertrag begründete Beziehung zu ω missverstanden habe, da offenbar das „wahrscheinliche“ Verhalten von ω falsch eingeschätzt wurde?

zu ersetzen, denn es ist möglich, ganz ohne einen speziellen Begriff „sozialer“ Beziehungen auszukommen. Wichtig ist vielmehr, jeweils deutlich zu machen, welche Arten von Beziehungen betrachtet werden sollen und wie das geschehen soll, nämlich in einer faktischen (retrospektiven) oder modalen Betrachtungsweise.

6. Soziale Netzwerke und relationale Sozialstrukturbegriffe. Mit dem relationalen Strukturbegriff kann man sich auf beliebige Mengen von Objekten beziehen, zwischen denen es irgendwelche Beziehungen gibt, insbesondere auf beliebige Netzwerke. Einschränkend sprechen wir von *sozialen Netzwerken*, wenn sich die Knoten des Netzwerks auf *soziale Akteure* (Personen und/oder Organisationen, insbesondere Haushalte und Unternehmen) beziehen.[15]

Diese Definition impliziert, dass sich die Sozialforschung nicht nur mit sozialen Netzwerken beschäftigt. Denn erstens sind nicht alle für Fragen der Sozialforschung relevanten Netzwerke im Sinne der obigen Definition soziale Netzwerke, man denke z.B. an Verkehrsverbindungen zwischen Städten. Außerdem gibt es zahlreiche Aspekte gesellschaftlicher Verhältnisse, die nicht (im Sinne unserer formalen Definition) als Netzwerke beschrieben werden können, z.B. statistische Strukturen und Institutionen. Deshalb werden wir auch vermeiden, irgendeinen allgemeinen relationalen Sozialstrukturbegriff zu definieren, wie dies von einigen Autoren versucht worden ist.[16] Stattdessen beziehen wir relationale Strukturbegriffe immer nur auf diejenigen Aspekte gesellschaftlicher Verhältnisse, die durch Netzwerke explizit repräsentiert werden können.[17]

[15] Dies entspricht folgender Definition von S. Wasserman und K. Faust (1994: 20): „A *social network* consists of a finite set or sets of actors and the relation or relations defined on them.“

[16] In der Literatur wird oft auf einen Vortrag Alfred R. Radcliffe-Browns aus dem Jahr 1940 Bezug genommen: „For a preliminary definition of social phenomena it seems sufficiently clear that what we have to deal with are relations of association between individual organisms. [...] Let us consider what are the concrete, observable facts with which the social anthropologist is concerned. If we set out to study, for example, the aboriginal inhabitants of a part of Australia, we find a certain number of individual human beings in a certain natural environment. We can observe acts of behaviour of these individuals, including, of course, their acts of speech, and the material products of past actions. We do not observe a ‘culture’, since that word denotes, not any concrete reality, but an abstraction, and as it is commonly used a vague abstraction. But direct observation does reveal to us that these human beings are connected by a complex network of social relations. I use the term ‘social structure’ to denote this network of actually existing relations. It is this that I regard it as my business to study if I am working, not as an ethnologist or psychologist, but as a social anthropologist.“ (Radcliffe-Brown 1940: 189f.)

[17] Natürlich spricht nichts dagegen, diese Aspekte auch als „Aspekte einer Sozialstruktur“ zu bezeichnen; so kann man z.B. folgende Formulierung von F. U. Pappi (1987: 12) verstehen: „Für den Soziologen ist die Netzwerkanalyse eine Methode zur Untersuchung von sozialen Strukturen. Eine Sozialstruktur wird repräsentiert durch die Beziehungen zwischen sozialen Einheiten wie Personen, Gruppen, Organisationen usw.“

7. Sind relationale Strukturen zeitlich stabil? Einige der Fragen, die in Abschnitt 2.2 im Hinblick auf statistische Strukturen besprochen wurden, stellen sich gleichermaßen für relationale Strukturen. Das betrifft zunächst die Frage der zeitlichen Stabilität. Die in Abschnitt 2.2 (§ 2) zitierten Annahmen über zeitliche Stabilität beziehen sich fast immer auch auf relationale Strukturen. Dementsprechend definierte ein früher Vertreter der Netzwerkanalyse, Edward O. Laumann (1966: 3), „Sozialstruktur“ als ein „persistent system of social relationships among social positions“.[18]

Wie für statistische gilt jedoch auch für relationale Strukturen: dass ihr Begriff keinerlei Annahmen über ihre zeitliche Stabilität impliziert. Somit kann man auch stets fragen, wie sich relationale Strukturen im Zeitablauf verändern. In einer retrospektiven Betrachtung mag sich dann zeigen, dass sich einige Strukturen schneller, andere langsamer verändert haben; aber auch abgesehen davon, dass dies bestenfalls im Nachhinein festgestellt werden kann,[19] ergeben sich daraus keine Einschränkungen für ein Reden von Strukturen.

8. Wie entstehen relationale Strukturen? Auch die Frage, wie Strukturen entstehen, stellt sich für statistische und relationale Strukturen in analoger Weise. So wie in Abschnitt 2.2 bei statistischen Strukturen können auch bei relationalen Strukturen drei Aspekte unterschieden werden:

- Man kann zunächst an die *substantiellen Prozesse* denken, durch die in der sozialen Realität die jeweils thematisierten Beziehungen entstanden sind (oder, in einer modalen Betrachtungsweise, entstehen könnten).
- Man kann weiterhin an die *datenerzeugenden Prozesse* denken, durch die Informationen (Daten) über in der sozialen Realität als gegeben vorausgesetzte Beziehungen entstehen.
- Und man kann schließlich an die gedanklichen und rechnerischen Prozesse denken, durch die aus den Daten bestimmte Netzwerke, Charakterisierungen und Modelle konstruiert werden.[20]

Offenbar interessieren in erster Linie die substantiellen Prozesse; und man kann auch sogleich feststellen: Wie diese Prozesse aufzufassen und begrifflich zu konzipieren sind, hängt vor allem von der Art der Beziehungen ab, deren Entstehen überlegt werden soll. Zum Beispiel: Wie entstehen

[18] Diese Definition wurde auch von anderen Autoren übernommen, so etwa von P. V. Marsden und N. Lin (1982: 9).

[19] Deshalb sind Formulierungen der folgenden Art offenbar problematisch: „*Beziehungen* entstehen, sobald Menschen in relativ stabile, kontinuierliche Muster spezifischer Interaktionen und/oder gegenseitiger Abhängigkeit eintreten [...].“ (Joas 2001: 16) Oder: „Soziale Beziehungen sind beständige Interaktionsmuster zwischen zwei oder mehr Personen.“ (Weymann 2001: 104)

[20] Es erscheint durchaus angemessen, hier von einer *Konstruktion* zu sprechen; denn bei der Frage, welche Knoten in die Definition eines Netzwerks einbezogen und welche Beziehungen betrachtet werden sollen, sind mehr oder weniger willkürliche Entscheidungen kaum zu vermeiden.

Verkehrsverbindungen zwischen Städten? Wie entstehen Verkehrsunfälle, durch die zwei oder mehr Menschen in eine physische Interaktionsbeziehung geraten? Wie entstehen Arbeitsverträge, durch die Menschen zu Mitarbeitern eines Unternehmens werden? Wie entstehen Freundschaften? Wie werden zwei Personen zu Mitgliedern derselben Schulklasse oder zu Teilnehmern desselben Seminars?

Die Liste solcher Fragestellungen könnte fast beliebig fortgesetzt werden. Bemerkenswert ist vor allem, dass es keine allgemeine Prozesskonzeption gibt, die sich gleichermaßen für alle Fragestellungen eignet. In einigen Fällen erscheint es sinnvoll, an Handlungsprozesse zu denken, an denen zwei oder mehr Menschen beteiligt sind; aber eine solche Vorstellung passt nicht immer, denn eine Beziehung kann auch dadurch enstehen, dass zwei oder mehr Prozesse zunächst unabhängig voneinander ablaufen, bevor sie irgendwann tatsächlich zu einer Interaktion führen oder auch nur zu einer Situation, die zur Feststellung einer komparativen Beziehung verwendet werden kann. Weiterhin kann man auch an Prozesse denken, die gar nicht als Handlungsprozesse verstanden werden können, wie z.B. die Ausbreitung von Krankheiten durch eine Übertragung von Viren oder Bakterien.

Wir kommen also zunächst zu dem Ergebnis, dass eine allgemeine Antwort auf die Frage, wie relationale Strukturen entstehen, nicht gegeben werden kann.

9. Relationale Strukturen als Bedingungen? Schließlich stellt sich, wie für statistische Strukturen, auch für relationale Strukturen die Frage, ob und ggf. in welcher Weise sie *als Bedingungen* verstanden werden können. Das setzt offenbar voraus, dass die durch das Netzwerk erfassten Beziehungen einer modalen Betrachtungsweise zugänglich sind. Alle weiteren Überlegungen müssen dann aber darauf Bezug nehmen, wie und für wen bzw. was diese Beziehungen Bedingungen sein könnten; und darüber lässt sich in allgemeiner Weise kaum etwas aussagen. Wir werden uns mit dieser Frage deshalb nur bei der Diskussion bestimmter Beispiele beschäftigen.

Kapitel 4

Prozesse und Ablaufschemas

Eine wichtige Aufgabe der empirischen Sozialforschung besteht darin, Veränderungen gesellschaftlicher Verhältnisse zu ermitteln und darzustellen. In diesem Kapitel werden einige unterschiedliche Möglichkeiten zur Konzeptualisierung solcher Prozesse besprochen. Es gibt zwei Abschnitte. Im ersten Abschnitt werden einige Prozessbegriffe unterschieden, und es wird überlegt, wie man von historischen Prozessen und Ablaufschemas (bzw. wiederholbaren Prozessen) sprechen kann. In einem zweiten Abschnitt werden schematische Prozesskonzeptionen besprochen: Zeitreihen, insbesondere schematische Lebensverläufe, sowie statistische Prozesse, die aus zeitlichen Folgen statistischer Variablen bestehen.

4.1 Historische Prozesse und Ablaufschemas

1. Einige Varianten des Prozessbegriffs. In der Brockhaus-Enzyklopädie (Studienausgabe 2001, Bd. 17: 566) findet man als Erläuterung des Prozessbegriffs: „Verlauf, Ablauf, Hergang, Entwicklung“. Der Begriff kann somit vollständig allgemein verwendet werden und setzt weder bestimmte sachliche Bezüge voraus (*was* sich verändert), noch impliziert er spezifische Annahmen über das als Prozess bezeichnete Geschehen (*wie* es sich verändert).[1] Infolgedessen ist es auch kaum möglich, eine erschöpfende Gliederung unterschiedlicher Prozesskonzeptionen vorzunehmen. Wir unterscheiden deshalb zunächst nur einige Varianten.

[1] Dementsprechend heißt es in einer begriffsgeschichtlichen Studie von Kurt Röttgers (1983: 93) über den heutigen Sprachgebrauch: „Alles, was sich irgendwie verändert, ist

– Das Reden von Prozessen kann sich auf die Entwicklung von Zuständen identifizierbarer Objekte beziehen; zum Beispiel: die Entwicklung der Körpertemperatur bei einem bestimmten Menschen während eines bestimmten Zeitraums, oder die Entwicklung der Niederschlagsmenge in einem bestimmten räumlichen Gebiet.
– Man kann sich auf die Entwicklung von Zuständen bei einer Mehrzahl von Objekten beziehen. Verwendet man dafür statistische Begriffsbildungen, gelangt man zu statistischen Prozessen (i.e.S.), die aus zeitlichen Folgen statistischer Variablen bestehen.
– Zu andersartigen Vorstellungen über Prozesse gelangt man, wenn man von Ereignissen ausgeht. In einer allgemeinen Formulierung erscheinen dann Prozesse als zeitlich geordnete Folgen von Ereignissen. Damit ein solcher Prozessbegriff sinnvoll verwendbar wird, muss allerdings spezifiziert werden, wie von Ereignissen gesprochen werden soll.
– Zu einer spezielleren Variante der zuletzt genannten Konzeption gelangt man, wenn man an Handlungen denkt. Prozesse erscheinen dann als Handlungszusammenhänge, die aus einer zeitlichen Abfolge einer Mehrzahl von Tätigkeiten eines oder mehrerer Akteure bestehen. Wir nennen sie *Handlungsprozesse*.

2. Prozesse werden konstruiert. Beim Reden von Prozessen muss angegeben werden, welche Arten von Veränderungen betrachtet werden sollen (*was* sich verändert); außerdem muss überlegt werden, *wie* es sich verändert bzw. verändern kann. Die zweite Frage wird uns erst später beschäftigen. Hier soll zunächst darauf aufmerksam gemacht werden, dass eine Beantwortung der ersten Frage stets eine weitgehende Selektion von Aspekten eines realen oder vorstellbaren Geschehens erfordert. Man betrachtet zum Beispiel die Entwicklung der Körpertemperatur eines Patienten und abstrahiert zugleich von beliebig vielen anderen Aspekten, die ebenfalls betrachtet werden könnten; oder man bezieht sich auf Handlungsprozesse, ohne die materiellen Kontexte, in denen sich die Handlungen abspielen, explizit in der Prozesskonzeption zu berücksichtigen.

Prozessdefinitionen beruhen also stets auf spezifischen Abstraktionen, einer Selektion bestimmter Aspekte, über deren zeitliche Entwicklung man nachdenken möchte; und alle weiteren Überlegungen beziehen sich dann ausschließlich auf den zuvor definierten Prozess, d.h. auf die jeweils ausgewählten Aspekte. So wird auch die Verwendung von Adjektiven zur Charakterisierung von Prozessen verständlich, wenn etwa von physikalischen, chemischen oder demographischen Prozessen gesprochen wird. Solche Redeweisen zeigen, dass man sich mit dem Prozessbegriff nicht unmittelbar

ein Prozeß oder befindet sich in einem Prozeß, so daß Ausdrücke wie ‘Der Prozeß der Veränderung von X’ semantisch eine bloße Verdoppelung des Ausdrucks, pragmatisch das manchmal erwünschte bloße Hinauszögern der Artikulation eines Gedankens darstellt.“

auf Vorkommnisse und Abläufe in der menschlichen Erfahrungswelt bezieht, sondern auf Modelle, die zur Reflexion jeweils spezifischer Aspekte solcher Vorkommnisse und Abläufe und ihrer möglichen (zukünftigen) Entwicklung konstruiert werden.[2]

Diese Aussage gilt insbesondere für die an Lebensverläufen orientierte Sozialforschung. Bereits Konstruktionen individueller Biographien sind unvermeidlich sehr selektiv; und das gilt erst recht, wenn versucht wird, mithilfe statistischer Daten zu vergleichenden Aussagen über Lebensverläufe der Mitglieder einer Gesellschaft zu gelangen. Dann wird auch deutlich, dass man sich auf bestimmte Aspekte beschränken muss, aus denen Gesichtspunkte für einen Vergleich gewonnen werden können.

3. Historische Prozesse. Von *historischen Prozessen* soll in diesem Text gesprochen werden, wenn sich die Prozesskonstruktion auf einen empirisch identifizierbaren Ablauf in der menschlichen Erfahrungswelt bezieht.

– Das Reden von historischen Prozessen setzt also eine Bezugnahme auf eine menschliche Praxis voraus, durch die bzw. von der aus Prozesse identifiziert werden können.[3]
– Dem entspricht, dass zur Konzeption historischer Prozesse ein anthropozentrisches, der menschlichen Praxis gemäßes Zeitverständnis vorausgesetzt wird. Nicht nur wird vorausgesetzt, dass es zwischen Ereignissen zeitliche Beziehungen gibt, sondern außerdem eine fundamentale Unterscheidung zwischen zeitlichen Modalitäten: zwischen einer Vergangenheit, die bisher realisierte und insoweit nicht mehr veränderbare Sachverhalte umfasst, einer offenen Zukunft, die aus bisher nicht realisierten Möglichkeiten besteht, und schließlich einer flüchtigen Gegenwart, in der jeweils bestimmte Möglichkeiten realisiert und dadurch der Vergangenheit hinzugefügt werden. Offenbar verdankt sich dieses Zeitverständnis den Erfahrungen menschlicher Praxis.[4]

[2]Hierzu passt folgende Bemerkung von Rainer M. Lepsius (1976: 121): „Die Vorstellung, daß den systematischen Einzelwissenschaften jeweils abgrenzbare Teilbereiche der Erfahrung als Gegenstände ihrer Arbeit zugewiesen werden könnten, ist irrig, und insofern auch die Vorstellung, der Geschichtswissenschaft würde durch die Ausdifferenzierung der Sozialwissenschaften der Objektbereich verkleinert. Es gibt keine Erfahrungsbestände, die als solche soziologisch oder historisch sind. Erst die Umformulierung der Erfahrungsobjekte in Erkenntnisobjekte durch die Anwendung bestimmter Fragestellungen, kategorialer Bezugssysteme und Lösungswege formuliert konventionalisierte »Zuständigkeiten« von Wissenschaften.“

[3]Dem entspricht eine von A. C. Danto (1965/1980: 49) vorgeschlagene „minimale Charakterisierung der historiographischen Tätigkeit“: „daß das Unterfangen, dem Historiker sich letztendlich widmen, der Versuch ist, wahre Feststellungen über Ereignisse aus ihrer *eigenen* Vergangenheit zu treffen oder wahre Beschreibungen davon zu geben.“

[4]Ausführliche Überlegungen zu diesem anthropozentrischen Zeitverständnis findet man bei Michael Oakeshott (1983: 7ff.). Es unterscheidet sich von Zeitvorstellungen, wie sie oftmals für physikalische Modelle angenommen werden und im Kontext der „New Theory of Time“ (Oaklander und Smith, 1994) diskutiert werden.

– Auch für historische Prozesse gilt, dass sie konstruiert werden. Es gibt indessen keine bestimmten Anforderungen an die begrifflichen Hilfsmittel. Man kann historische Prozesse als Handlungszusammenhänge oder allgemeiner als zeitliche Folgen von Ereignissen konzipieren, man kann aber auch statistische Begriffsbildungen verwenden, um historische Prozesse (z.B. die Bevölkerungsentwicklung in Deutschland während eines bestimmten Zeitraums) darzustellen. Unser Begriff historischer Prozesse soll also insbesondere keine Festlegung auf die Idee einer „Ereignisgeschichte" beinhalten.[5]
– Unser Begriff historischer Prozesse soll auch keine Annahmen über die zeitliche Dauer voraussetzen. Christian Meier (1978: 56) hat wohl Recht: „Unter historischen Prozessen versteht man in der Regel längere, nämlich Jahrzehnte oder gar Jahrhunderte übergreifende Abläufe." Dem entsprechend wird bei zeitlich kürzeren Geschehnissen oft von „Ereignissen" gesprochen. Diese Unterscheidung soll hier jedoch ausdrücklich nicht gemeint sein. Gleichwohl wird zwischen Ereignissen und Prozessen unterschieden. Wenn von Ereignissen gesprochen wird, sind zeitlich datierbare Vorkommnisse in unserer Erfahrungswelt gemeint (z.B. ein bestimmter Verkehrsunfall[6]), der Prozessbegriff bezieht sich dagegen auf theoretische Konstruktionen (beliebiger zeitlicher Dauer). Somit ist es auch möglich, ein Ereignis als einen Prozess zu betrachten, z.B. sich eine explizite Vorstellung vom Ablauf eines Verkehrsunfalls zu machen.
– Insofern die empirische Identifizierbarkeit historischer Prozesse gefordert wird, sind sie stets räumlich und zeitlich beschränkt; sie haben einen Anfang und ein (ggf. vorläufiges) Ende und gehören somit zur

[5] Bei Historikern und Theoretikern der Geschichtswissenschaft findet man oftmals die Vorstellung, dass „Geschichte" aus einer Folge von Ereignissen, insbesondere aus menschlichen Handlungen besteht. So spricht z.B. Gordon Leff (1969: 4) von „history" „as the totality of human actions and endeavour"; und bei Christian Meier (1978: 11) heißt es: „Mit der Prozeß-Kategorie werden innerhalb der sozialen Welt bestimmte Handlungszusammenhänge wahrgenommen." Natürlich ist es zulässig, historische Prozesse so zu definieren; aber bereits innerhalb der Geschichtswissenschaft gibt es noch andere Prozesskonzeptionen, und die empirische Sozialforschung beschäftigt sich sogar überwiegend mit Prozessen, die weder unmittelbar als Handlungszusammenhänge noch ohne weiteres als „Folgen" menschlicher Tätigkeiten konzipiert werden können.

[6] Dem entspricht folgende Bemerkung von H.-R. Jauss (1973: 554): „Ereignis ist eine objektive, für das historische Geschehen selbst konstitutive Kategorie. Das Ereignis liegt dem Zugriff des Historikers immer schon voraus; es ist nicht [im Unterschied zu unserem Prozessbegriff, G.R.] ein subjektives Schema narrativer Aneignung, sondern dessen äußere Bedingung." Das angeführte Beispiel soll darauf hinweisen, dass es sich bei Ereignissen auch um durchaus triviale Vorkommnisse handeln kann; dies unterscheidet den allgemeinen von einem emphatischen Ereignisbegriff, den Christian Meier (a.a.O., S. 47) so erläutert: „Als Ereignis bezeichnen wir im Alltag primär ein besonderes, aus dem Üblichen herausragendes Geschehen. Die Historie gebraucht das Wort im gleichen Sinne und meint damit zumeist die bemerkenswerten, »denk- (und überlieferungs-) würdigen« Handlungen und Handlungszusammenhänge sowie anderswie bewirkten Einschnitte des politischen und militärischen Bühnengeschehens."

Vergangenheit (der Praxis, durch die sie identifiziert werden). Zur Verdeutlichung kann man an einen Spaziergang, den Entstehungsprozess eines Gebäudes oder an die Bevölkerungsentwicklung in einer bestimmten Region während eines bestimmten Zeitraums denken. Offenbar kann man beliebig viele historische Prozesse dieser Art konzipieren. Oftmals lassen sich auch Beziehungen zwischen mehreren Prozessen herstellen. Es ist jedoch fragwürdig, ob man sinnvoll von einer Gesamtheit aller historischen Prozesse, die in bestimmter Weise miteinander verbunden sind, sprechen kann.

4. Ablaufschemas. Insofern bei der Konzeption historischer Prozesse ein anthropozentrisches Zeitverständnis vorausgesetzt wird, sind sie „einmalig". Unabhängig von dieser Feststellung kann man jedoch von *wiederholbaren Prozessen* sprechen, d.h. von Prozessen, die in ähnlicher Form mehrmals ablaufen können. Wiederholbarkeit in diesem Sinn setzt nur voraus, dass mehrere Prozesse unter bestimmten Aspekten als vergleichbar betrachtet werden können.

Bei wiederholbaren Prozessen ist offenbar eine begriffliche Unterscheidung erforderlich: Einerseits kann man sich auf die Form des Prozessablaufs beziehen, wir nennen dies ein *Ablaufschema* (eines wiederholbaren Prozesses); andererseits kann man sich auf jeweils individuelle Prozessabläufe beziehen, die als Realisationen des Ablaufschemas betrachtet werden können. Sofern diese Realisationen nicht nur fiktiv vorgestellt werden, sondern tatsächlich stattfinden, handelt es sich um historische Prozesse. Somit gibt es auch keinen begrifflichen Gegensatz zwischen historischen und wiederholbaren Prozessen. Als Gegensatz zu historischen Prozessen kann man an fiktive Prozesse denken, die man sich nur vorstellt. Wenn man einen historischen (oder fiktiven) Prozess wiederholbar nennt, ist dagegen gemeint, dass man ihn als Realisation eines Ablaufschemas betrachten möchte und dass es noch andere Realisationen dieses Ablaufschemas gibt oder geben kann.[7] Einige Beispiele können das verdeutlichen.

– Eine wichtige Klasse von Beispielen liefern Computerprogramme (oder abstrakter: Algorithmen). Bei einem Computerprogramm muss man offenbar unterscheiden zwischen einerseits dem Programm, das in diesem Fall das Ablaufschema bildet und die möglichen Programmabläufe festlegt, und andererseits den Prozessrealisationen, also den Programmabläufen, die stattfinden, wenn das Programm gestartet wird. Jeder Programmablauf ist seinerseits ein historischer Prozess, der in einem bestimmten räumlichen und zeitlichen Kontext stattfindet.[8]
– Weitere leicht durchschaubare Beispiele liefern Gesellschaftsspiele wie

[7] *Wiederholbar zu sein* ist also keine Eigenschaft, die einem Prozess „an und für sich" zukommt, sie resultiert vielmehr aus einer jeweils bestimmten Betrachtungsweise eines historischen oder fiktiven Prozesses.

[8] Man vgl. dazu auch die Überlegungen von B. C. Smith (1996: 32ff.).

z.B. Schach oder Skat. Einerseits gibt es Spielregeln, durch die festgelegt wird, wie Spiele ablaufen können; andererseits gibt es jeweils bestimmte Spielabläufe, die man sich vorstellen oder als historische Prozesse realisieren kann.

– Weiterhin kann man an viele andere Handlungsprozesse denken, die mehr oder weniger detailliert durch Regeln bestimmt werden und für die es insofern ein Ablaufschema gibt, zum Beispiel die Zubereitung von Speisen (nach einem Rezept) oder die Durchführung einer medizinischen Diagnose (nach den dafür gültigen ärztlichen Regeln).

– Schließlich kann man auch menschliche Lebensverläufe als wiederholbare Prozesse auffassen. Zwar kann niemand das eigene Leben wiederholen, aber die Idee der Wiederholbarkeit bezieht sich nur darauf, dass man die Lebensverläufe mehrerer Menschen (aller Mitglieder einer Gesellschaft) vergleichend betrachten und zu diesem Zweck Ablaufschemas konstruieren kann. Eine Möglichkeit liefern sogenannte Biographieschemas, die im nächsten Abschnitt besprochen werden.

Die Beispiele zeigen, dass wiederholbare Prozesse von ganz unterschiedlicher Art sein können. Während Computerprogramme Beispiele für mechanische Prozesse sind, die durch Akteure nur initialisiert werden,[9] handelt es sich bei Gesellschaftsspielen um Handlungsprozesse, bei denen auch der Ablauf durch Akteure beeinflusst wird. Schließlich können Lebensverläufe weder als rein mechanische Prozesse noch ausschließlich als Handlungsprozesse begriffen werden.

Weiterhin zeigen die Beispiele, dass es bei Ablaufschemas unterschiedliche *Spielräume für Prozessrealisationen* geben kann. Einen Extremfall bilden Algorithmen, die einen Prozess eindeutig festlegen. Die meisten Ablaufschemas lassen dagegen in mehr oder weniger weiten Grenzen unterschiedliche Prozessabläufe zu.[10]

[9] Wir sprechen in diesem Text von *mechanischen* Prozessen, wenn Akteure gar nicht beteiligt sind oder nur als Auslöser des Prozesses, also ohne Einfluss darauf zu nehmen, wie der Prozess abläuft. (Die obige Aussage bezieht sich somit nicht auf sogenannte interaktive Programme, bei denen ein Akteur in den Programmablauf eingreifen kann.) In dieser Bedeutung bezieht sich das Adjektiv ‘mechanisch’ also nicht auf die Mechanik als Teilgebiet der Physik, sondern dient der Charakterisierung einer bestimmten Art von Prozessen. In einer ganz ähnlichen Bedeutung bemerkte z.B. die Wissenschaftshistorikerin Lorraine Daston (1998: 34): „Eine hervorstechende Form der Objektivität – nennen wir sie >mechanische Objektivität< – zielt auf die Ausschaltung aller Formen des menschlichen Eingriffs in die Natur ab [. . .].“

[10] Der Begriff eines Ablaufschemas soll also nicht beinhalten, dass nur ein in irgendeinem Sinn „kleiner“ Spielraum für dem Schema entsprechende Prozesse besteht. Deshalb kann auch nicht bereits durch den Verweis auf ein Ablaufschema von „typischen“ Prozessabläufen gesprochen werden; obwohl – wie etwa von Karl-Georg Faber (1971: 95) ausgeführt worden ist – der Typenbegriff sich ebenfalls bloß in abstrakter Weise auf reales bzw. realisierbares Geschehen bezieht.

5. Durch Regeln bestimmte Prozesse. In vielen Fällen können Ablaufschemas für wiederholbare Prozesse durch Regeln beschrieben werden. Man kann dann davon sprechen, dass die Prozessrealisationen „durch Regeln bestimmt“ werden. Allerdings muss darauf geachtet werden, was mit dieser Formulierung ausgesagt werden kann.

Zunächst ist klar: Dass ein Prozess durch Regeln bestimmt wird, impliziert nicht, dass der Prozessablauf determiniert ist. Gilbert Ryle (1949/1982: 98ff.) hat das am Beispiel des Schachspiels verdeutlicht: obwohl durch Regeln bestimmt, ist jeder einzelne Spielverlauf nicht (jedenfalls nicht durch die Spielregeln) determiniert. Das Beispiel zeigt auch, dass Regeln, die einen Prozess bestimmen, nicht als Ursachen verstanden werden können, die gewissermaßen bewirken, dass sich der Prozessablauf an die Regeln hält (und zwar gilt diese Feststellung ganz unabhängig von der Größe des durch die Regeln gegebenen Spielraums für Prozessrealisationen): Die Regeln, die es für Schachspiele gibt, können weder bewirken noch garantieren, dass sich die Spieler an die Regeln halten.

Dies gilt aber auch für mechanische Prozesse, an deren Ablauf keine Akteure beteiligt sind. Zum Beispiel kann auch ein Computerprogramm weder bewirken noch garantieren, dass bei seiner Aktivierung ein dem festgelegten Ablaufschema entsprechender Prozess stattfindet. Wie der historische Prozess abläuft, der bei der Programmaktivierung beginnt, hängt vielmehr von zahlreichen Bedingungen ab, die durch das Computerprogramm überhaupt nicht in Betracht gezogen werden (z.B. von der Stromversorgung des Prozessors).

Allerdings ist es durchaus von Bedeutung, ob Akteure einen Prozess nur initialisieren oder ob es auch von ihrem Verhalten abhängt, wie der Prozess abläuft (wie dies inbesondere, aber nicht nur bei Handlungsprozessen der Fall ist). Denn wenn letzteres der Fall ist, kann man in zwei unterschiedlichen Bedeutungen von der Existenz von Regeln sprechen.

– Zunächst kann gemeint sein, dass sich die Regeln auf ein Modell beziehen, dass *aus einer Beobachterperspektive* konstruiert wird, um wiederholbare Prozessabläufe vorstellbar und reflektierbar zu machen. Bei mechanischen Prozessen ist offenbar nur dieses Verständnis von Regeln möglich.

– Wenn es sich jedoch um Prozesse handelt, an deren Ablauf Akteure beteiligt sind, wird es möglich und ist oftmals der Fall, dass auch diese Akteure über ein Regelwissen verfügen. Es gibt dann ein Regelwissen *aus der Akteursperspektive*, und wie ein Prozess abläuft hängt auch davon ab, wie dieses Regelwissen beschaffen ist und wie die beteiligten Akteure es in ihrem Verhalten verwenden.

Oftmals kann aus beiden Perspektiven das gleiche (oder zumindest ein teilweise gleiches) Regelwissen unterstellt werden, z.B. bei Gesellschaftsspielen, bei denen Spieler und Zuschauer die Regeln gleichermaßen kennen. Anders verhält es sich jedoch in vielen Fällen, in denen Wissenschaftler

Modelle für Prozesse konstruieren, an denen Menschen beteiligt sind. Dies geschieht normalerweise in einer Beobachterperspektive, und es kann meistens nicht angenommen werden, dass das wissenschaftlich konstruierte Regelwissen auch bei den Akteuren der Prozesse vorhanden ist (z.B. bei einem Modell zur Erklärung des Zustandekommens von Staus auf Autobahnen). Dementsprechend muss auch zwischen Verhaltensregelmäßigkeiten, die aus einer Beobachterperspektive festgestellt werden können, und Regeln, an denen sich Akteure in ihrem Verhalten orientieren, begrifflich unterschieden werden.

4.2 Zeitreihen und statistische Prozesse

Wenn man sich im Rahmen der empirischen Sozialforschung mit Prozessen beschäftigt, werden diese fast immer als Realisationen theoretisch konzipierter Ablaufschemas betrachtet. Möglichkeiten zur Konstruktion von Ablaufschemas hängen in erster Linie von der Konzeption der Prozesse ab. In diesem Abschnitt beziehen wir uns auf Zeitreihen und statistische Prozesse.

1. Zeitachsen. Offenbar benötigt man zur Konzeption von Ablaufschemas einen zeitlichen Rahmen. Meistens wird eine Zeitachse verwendet. Es gibt hauptsächlich zwei Varianten:

– Man kann sich die Zeit als eine *Folge von Zeitstellen* (z.B. Sekunden, Stunden, Tage, Monate, Jahre) vorstellen. Zur Repräsentation der Zeitstellen werden die natürlichen oder ganzen Zahlen verwendet, und man spricht von einer *diskreten Zeitachse*.

– Man kann versuchen, sich die Zeit als ein linear geordnetes Kontinuum von Zeitpunkten vorzustellen. Zur Repräsentation werden in diesem Fall die reellen Zahlen verwendet, und man spricht von einer *stetigen* oder *kontinuierlichen Zeitachse*.

Unabhängig von dieser Unterscheidung, die die begriffliche Repräsentation von Zeit(stellen) betrifft, kann man Verwendungskontexte unterscheiden. Zunächst kann man an eine *historische Zeitachse* denken, die zur Repräsentation der historischen Zeit dient, in der sich das Leben der Menschen tatsächlich abspielt. Gedankliche Bezugnahmen auf diese Zeitachse erfolgen mithilfe von Kalendern und Uhren.[11] Davon zu unterscheiden sind *Modell-Zeitachsen* (auch *Prozesszeitachsen* genannt), die zur Konstruktion von Modellen verwendet werden, die einer modalen Reflexion von Prozessabläufen dienen sollen.

Will man eine explizite Repräsentation zeitlicher Bezüge vornehmen, muss man sich für eine diskrete oder eine stetige Darstellung entscheiden.

[11] Darüber, wie sich solche Orientierungsmittel historisch entwickelt und verändert haben, gibt es eine umfangreiche Literatur, man vgl. z.B. E. G. Richards (1998).

Wir werden in diesem Text in den meisten Fällen eine diskrete Zeitachse zugrunde legen und dafür die Notation $\mathcal{T} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ verwenden, so dass $\mathcal{T}$ formal der Menge der ganzen Zahlen entspricht und die Ordnungsrelation zwischen diesen Zahlen ($\leq$ und ge) als zeitliche Relation zwischen den durch sie bezeichneten Zeitstellen verstanden werden kann. Festlegungen über die Art der Zeitstellen und Verknüpfungen mit einer historischen Zeitachse können bei Bedarf erfolgen. Natürlich benötigt man zur Repräsentation von Daten oft nur einen Teil der Zeitachse; wir verwenden dann die Notation $\mathcal{T}^*$, womit stets eine zusammenhängende und meistens (wenn nicht ausdrücklich anders angegeben) auch eine endliche Teilmenge von $\mathcal{T}$ gemeint sein soll.

2. Zeitreihen. Sobald man über den Begriff einer Zeitachse verfügt, kann man in sehr allgemeiner Weise von *Zeitreihen* sprechen. Zur formalen Vergegenwärtigung kann ein *Zeitreihenschema*

$$X : \mathcal{T}^* \longrightarrow \tilde{\mathcal{X}} \tag{4.1}$$

verwendet werden, also eine Funktion, durch die jeder Zeitstelle t einer Zeitachse $\mathcal{T}^*$ ein Wert $X(t)$ in einem Wertebereich $\tilde{\mathcal{X}}$ zugeordnet wird. Je nachdem ob es sich bei $\mathcal{T}^*$ um eine kontinuierliche oder diskrete Zeitachse handelt, kann man kontinuierliche (stetige) und diskrete Zeitreihen unterscheiden. Hinsichtlich des Wertebereichs $\tilde{\mathcal{X}}$ kann man außerdem folgende Unterscheidungen treffen:

– *Einfache Zeitreihen.* In diesem Fall wird jeder Zeitstelle t ein einfacher Wert $X(t)$ zugeordnet, so dass $\tilde{\mathcal{X}}$ numerisch durch reelle Zahlen repräsentiert werden kann; zum Beispiel: die Entwicklung der Körpertemperatur eines Patienten oder der Bevölkerungszahl eines Landes während eines gewissen Zeitraums.

– *Vektorielle Zeitreihen.* In diesem Fall wird jeder Zeitstelle t ein Vektor $X(t) = (X_1(t), \ldots, X_m(t))$ zugeordnet. Als Beispiel kann man daran denken, dass bei neugeborenen Kindern für einen gewissen Zeitraum sowohl die Körpergröße als auch das Körpergewicht erfasst wird.

– *Funktionale Zeitreihen.* In diesem Fall wird jeder Zeitstelle t eine Funktion zugeordnet; $X(t)$ ist dann keine Zahl, sondern eine Funktion (und wir verwenden dann meistens die Schreibweise X_t anstelle von $X(t)$). Bei diesen Funktionen kann es sich insbesondere um statistische Variablen handeln, so dass statistische Prozesse entstehen (das wird weiter unten genauer besprochen).

Bei einfachen und vektoriellen Zeitreihen kann der Wertebereich $\tilde{\mathcal{X}}$ dem Merkmalsraum einer ein- bzw. mehrdimensionalen statistischen Variablen entsprechen (man vgl. die Ausführungen in Abschnitt 2.1). Auch dann ist jedoch die Analogie zwischen dem Zeitreihenschema (4.1) und dem für statistische Variablen verwendeten Schema $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ rein formal, denn eine Zeitachse $\mathcal{T}^*$ ist keine Objektmenge. Tatsächlich wird in der allgemeinen Definition von Zeitreihen überhaupt kein bestimmter Objektbezug vorgenommen; insofern handelt es sich um eine rein formale Definition, deren inhaltliche Bedeutung sich nur aus einem Anwendungskontext ergeben kann.

3. Schematische Lebensverläufe. Die Methodenliteratur befasst sich überwiegend mit Zeitreihen, deren Wertebereiche quantitativ, meistens auch metrisch sind. In der empirischen Sozialforschung sind auch qualitative Zeitreihen wichtig, da sie sich zur Konzeptualisierung schematischer Lebensverläufe eignen. Dabei kann man grundsätzlich an beliebige Objekte, insbesondere auch an Menschen denken. Die allgemeine Vorstellung besteht darin, dass sich ein Objekt im Zeitablauf in unterschiedlichen Zuständen befinden kann. Um dies zu erfassen, eignet sich offenbar das allgemeine Zeitreihenschema $Y : \mathcal{T}^* \longrightarrow \tilde{\mathcal{Y}}$, wobei jetzt $\tilde{\mathcal{Y}}$ ein *Zustandsraum* ist, dessen Elemente Zustände sind, in denen sich Objekte einer bestimmten Art während der durch $\mathcal{T}^*$ gegebenen Zeitstellen befinden können. Eine Zeitreihe $\{Y(t) \,|\, t \in \mathcal{T}^*\}$ beschreibt dann aspekthaft einen Ausschnitt eines Lebensverlaufs. Die Formulierung ist allgemein genug, um sich sowohl auf zeitliche Ausschnitte eines Lebensverlaufs als auch auf die gesamte Lebensdauer zu beziehen.

Wiederum muss unterschieden werden zwischen spezifischen Lebensverläufen (Zeitreihen), die für ein jeweils spezifisches Objekt erfasst werden, und dem zugrundeliegenden Zeitreihenschema, das ein Ablaufschema für *mögliche* Lebensverläufe liefert. In diesem Zusammenhang wird das Ablaufschema auch als ein *Biographieschema* bezeichnet (Rohwer und Pötter 2001: 186), dessen Definition zweierlei erfordert:

- Zunächst muss ein Zustandsraum $\tilde{\mathcal{Y}}$ festgelegt werden, so dass dem Objekt in jeder Zeitstelle $t \in \mathcal{T}^*$ genau ein Zustand zukommt (natürlich kann auch ein mehrdimensionaler Zustandsraum verwendet werden, der eine simultane Bezugnahme auf mehrere Zustände erlaubt);
- außerdem muss für jeden Zustand festgelegt werden, in welche Folgezustände ein Wechsel stattfinden kann.

Um die Begriffsbildung zu illustrieren, zeigt Abbildung 4.2-2 ein Biographieschema für die Bildung und Auflösung von Lebensgemeinschaften. Es gibt folgende Zustände: (A) Anfangszustand, in dem sich eine Person vor der ersten Lebensgemeinschaft befindet, (U) nicht-eheliche Lebensgemeinschaft, (H) eheliche Lebensgemeinschaft (verheiratet), (S) nach einer Trennung bzw. Scheidung, (E) Endzustand (Tod). Die Pfeile deuten die möglichen Zustandsveränderungen an. Zum Beispiel gibt es keinen Pfeil von A nach S, weil eine Trennung oder Scheidung das Bestehen einer Lebensgemeinschaft voraussetzt.

4. Ein spezieller Ereignisbegriff. Mit dem allgemeinen Zeitreihenschema (4.1) ist ein spezieller Ereignisbegriff verbunden, durch den Ereignisse *als*

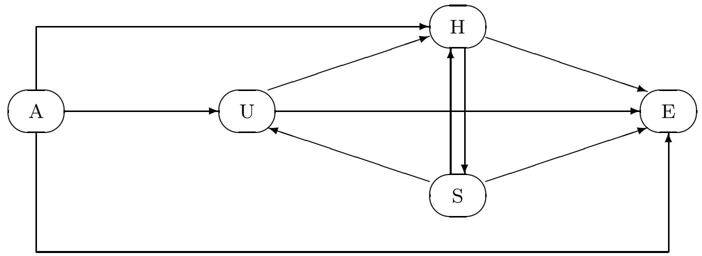


Abb. 4.2-2 Ein Biographieschema für die Bildung und Auflösung von Lebensgemeinschaften (Rohwer und Pötter 2001: 186).

Zustandswechsel aufgefasst werden. Wenn $\mathcal{T}^*$ eine diskrete Zeitachse ist, kann man mit diesem Begriff von einem Ereignis im Übergang von der Zeitstelle t zur nächsten Zeitstelle $t+1$ sprechen, wenn $X(t+1) \neq X(t)$ ist. Ein so definiertes Ereignis kann also durch die Angabe einer Zeitstelle t und der beiden Zustände $X(t)$ und $X(t+1)$ vollständig charakterisiert werden (wobei angenommen wird, dass die ontologischen Bezüge durch Erläuterungen des vorausgesetzten Zeitreihenschemas gegeben sind).

Dieser Ereignisbegriff wird insbesondere in der Literatur, die sich mit statistischen Methoden zur Modellierung qualitativer Zeitreihen beschäftigt, oft verwendet.[12] Es handelt sich jedoch um einen durchaus speziellen Ereignisbegriff.

- Tatsächlich können die meisten Ereignisse nicht hinreichend durch Zustandswechsel beteiligter Objekte beschrieben werden. Vielmehr verweist der Ereignisbegriff in seiner normalen Verwendung zunächst auf ein Geschehen, an dem oft Tätigkeiten von Akteuren beteiligt sind, die sich nicht durch einen Zustandswechsel charakterisieren lassen (man denke z.B. an einen Verkehrsunfall).
- Hier muss auch daran erinnert werden, dass das Zeitreihenschema, auf das sich der spezielle Ereignisbegriff bezieht, selbst bereits eine durchaus spezifische Konzeption von Prozessen als *zeitliche Folgen von Zuständen* impliziert. Insbesondere können Handlungsprozesse durch dieses Schema nicht erfasst werden.
- Schließlich ist bemerkenswert, dass mit dem speziellen Ereignisbegriff der direkte Realitätsbezug verschwindet, der mit dem normalen Reden von Ereignissen verbunden ist. Wie bereits erwähnt wurde (Anm. 6 auf S. 58) bezieht man sich mit dem Ereignisbegriff normalerweise auf Geschehnisse in der menschlichen Erfahrungswelt. Der spezielle Ereignisbegriff bezieht sich dagegen auf Prozesse, die als Modelle eines realen

[12] Dementsprechend wird auch von „Techniques of Event History Modeling" (Blossfeld und Rohwer 2002) gesprochen.

oder vorstellbaren Geschehens konzipiert werden. Dann ist z.B. mit einer Heirat nicht das reale Ereignis gemeint, das in einer bestimmten Weise stattgefunden hat, sondern ein daraus ableitbarer Zustandswechsel bei einer Person, die hinterher verheiratet ist.

5. Zeitliche Folgen statistischer Variablen. Das allgemeine Zeitreihenschema (4.1) umfasst insbesondere *statistische Prozesse*, worunter in diesem Text zeitliche Folgen statistischer Variablen verstanden werden sollen. Wir unterscheiden drei Varianten, wobei der Zeitindex t jeweils Werte in einer diskreten Zeitachse $\mathcal{T}^*$ annehmen kann.

a) Wir sprechen von einem *synchron aggregierten* statistischen Prozess, wenn es für jede Zeitstelle t eine Objektmenge Ω_t gibt, so dass der Prozess aus einer Folge von Variablen besteht:

$$X_t : \Omega_t \longrightarrow \tilde{\mathcal{X}} \tag{4.2}$$

b) Wir sprechen von einem *diachron aggregierten* statistischen Prozess, wenn es nur eine Objektmenge Ω gibt, deren Elementen jedoch für alle Zeitstellen bestimmte Merkmalswerte zugeordnet werden können.[13] Der Prozess besteht somit aus einer Folge von Variablen

$$X_t : \Omega \longrightarrow \tilde{\mathcal{X}} \tag{4.3}$$

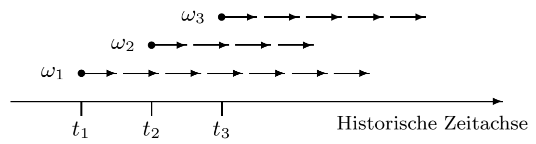
c) Schließlich gibt es die Möglichkeit, dass sich die Zugehörigkeit von Elementen zur Objektmenge im Zeitablauf verändern kann und Informationen darüber bei der Prozesskonstruktion berücksichtigt werden können (so dass anders als im Fall (a) Elemente während ihrer Mitgliedschaft in der Objektmenge identifizierbar bleiben). Wir sprechen dann von *transitorisch aggregierten* statistischen Prozesse. Ein wichtiges Beispiel bilden demographische Prozesse, bei denen sich eine Gesellschaft durch Geburten, Sterbefälle und Migrationen verändert (Definitionen erfolgen in Abschnitt 5.1 in Teil I).

6. Längsschnittgesamtheiten und Prozesszeitachsen. Offenbar können sich bei synchron aggregierten statistischen Prozessen die Objektmengen Ω_t in jeder Zeitstelle verändern. Dagegen geht man bei diachron aggregierten statistischen Prozessen von einer *Längsschnittgesamtheit* Ω aus, die für die Prozessdefinition als unveränderlich angenommen wird. Man kann sich vorstellen, dass jedem Element von Ω zunächst ein individueller Prozess zugeordnet ist und dass der statistische Prozess aus einer Aggregation dieser individuellen Prozesse entsteht.

Natürlich verlaufen die individuellen Prozesse nicht unbedingt zeitlich

[13]Diese Formulierung soll so verstanden werden, dass ggf. auch eine Kennzeichnung verwendet werden kann, die besagt, dass ein Objekt in einer Zeitstelle noch nicht oder nicht mehr existiert.

parallel, und sie haben auch meistens unterschiedliche zeitliche Dauern, wie durch folgendes Bild illustriert wird:



In diesem Bild wird angenommen, dass die individuellen Prozesse bei drei Objekten in drei verschiedenen Zeitstellen einer historischen Zeitachse beginnen. Bezieht sich die Prozessdefinition z.B. auf Ehedauern, kann man sich vorstellen, dass es sich um drei Personen mit einem unterschiedlichen Heiratsdatum handelt. Das Bild macht auch deutlich, dass es zur Bildung einer Längsschnittgesamtheit drei Möglichkeiten gibt:

a) Man kann Objekte zusammenfassen, bei denen der Prozess in der gleichen Zeitstelle beginnt (z.B. Personen, die im gleichen Jahr geheiratet haben, wenn man als Zeitstellen Jahre verwendet);

b) man kann Objekte zusammenfassen, bei denen der Prozess in der gleichen Zeitstelle aufhört (z.B. Personen, die im gleichen Jahr geschieden worden sind); und

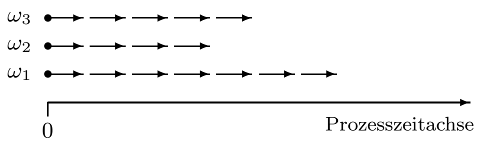
c) man kann Objekte zusammenfassen, bei denen der Prozess irgendwann während eines längeren Zeitraums begonnen (oder aufgehört) hat.

In der Literatur wird meistens die erste Variante empfohlen; man spricht dann von einem *Kohortenansatz*, wobei unter einer *Kohorte* eine Gesamtheit von Menschen verstanden wird, bei denen ein Prozess (einer bestimmten Art) in der gleichen historischen Zeitstelle begonnen hat. Allerdings gibt es keine vollständig scharfe Abgrenzung zum Fall (c), denn man kann auch von längeren Zeitperioden ausgehen, in denen das eine Kohorte definierende Ereignis stattgefunden hat.[14] Immerhin gibt es einen deutlichen Unterschied zum Fall (b), insbesondere wenn man von individuellen Prozessen ausgeht, die sich in ihrer zeitlichen Dauer erheblich unterscheiden können.

Ein Kohortenansatz wird in der an Lebensverläufen orientierten Sozialforschung in erster Linie als ein Hilfsmittel zur Darstellung historischer Veränderungen in der Entwicklung von Lebensverläufen verwendet. Unabhängig davon, wie eine Längsschnittgesamtheit gebildet wird, geht es jedoch stets um einen statistischen Vergleich der für die Elemente der Gesamtheit erfassten individuellen Prozesse. Dafür wird in jedem Fall eine *Prozesszeitachse* verwendet, so dass man sich vorstellen kann, dass alle

[14]Dem entspricht z.B. folgende Definition von N. D. Glenn (1977: 8): „a cohort is defined as those people within a geographically or otherwise delineated population who experienced the same significant life event within a given period of time.“

individuellen Prozesse in der gleichen Prozesszeitstelle beginnen und somit während ihrer Dauer synchron ablaufen. In unserem Beispiel kann der Übergang von einer historischen zu einer Prozesszeitachse folgendermaßen veranschaulicht werden:



Es wird also angenommen, dass bei jedem Objekt der Prozess in einer Zeitstellt $t = 0$ beginnt. Dementsprechend wird in diesem Text zur numerischen Repräsentation einer diskreten Prozesszeitachse die Notation $\mathcal{T}_0 := \{0, 1, 2, 3, \ldots\}$ verwendet.

7. Individuelle und aggregierte Prozesse. Zur Erläuterung dieser Unterscheidung beziehen wir uns zunächst auf das Schema (4.3) für diachron aggregierte statistische Prozesse.

- Einerseits kann man sich dann bei jedem Objekt $\omega \in \Omega$ auf einen diesem Objekt individuell zurechenbaren Prozess $\{X_t(\omega) \,|\, t \in \mathcal{T}^*\}$ beziehen.[15] Offenbar kann ein solcher Prozess auch als eine Zeitreihe bzw. als Realisierung eines Zeitreihenschemas aufgefasst werden.
- Andererseits kann man sich auf einen *aggregierten Prozess* beziehen, der aus der statistischen Aggregation der individuellen Prozesse entsteht und durch die gemeinsame Verteilung der Variablen X_t oder daraus ableitbare Zeitreihen charakterisiert werden kann.

Daraus ergibt sich zugleich ein wichtiger Unterschied zu synchron aggregierten statistischen Prozessen, bei denen man sich nur auf eine zeitliche Folge von Verteilungen der Variablen X_t, nicht jedoch auf ihre gemeinsame Verteilung beziehen kann.[16] Ein weiterer bemerkenswerter Unterschied besteht darin, dass es in diesem Fall nicht möglich ist, sich in gedanklich bestimmter Weise auf korrespondierende individuelle Prozesse zu beziehen. Denn entweder, wie im Ansatz des Schemas (4.2), wird bereits bei der Prozessdefinition gar nicht von individuellen Prozessen ausgegangen; oder es wird zwar von individuellen Prozessen ausgegangen, dann jedoch von der diachronen Identität der diese individuellen Prozesse konstituie-

[15] Bei dieser Schreibweise wird vorausgesetzt, dass es für alle Zeitstellen zumindest formal zurechenbare Zustände gibt. Natürlich ist es möglich, die Definition jedes individuellen Prozesses auf diejenigen Zeitstellen einzuschränken, in denen das betreffende Objekt existiert.

[16] Es sei angemerkt, dass solche zeitlichen Folgen von Verteilungen bei Rohwer und Pötter (2001: 192) missverständlich als „diachrone Zustandsverteilungen" bezeichnet werden. Besser sollte von zeitlichen Folgen synchroner (Zustands-)Verteilungen gesprochen werden.

renden Objekte abstrahiert.

Man kann sich leicht klarmachen, dass mit dieser Abstraktion ein erheblicher Informationsverlust verbunden sein kann. Zur Illustration kann bereits eine Gesamtheit dienen, die nur aus zwei Personen (ω_1 und ω_2) besteht, die sich in zwei verschiedenenen Zuständen (etwa 0 = erwerbstätig und 1 = arbeitslos) befinden können. Dann können zwei Prozessvarianten etwa folgendermaßen aussehen:

t	0	1	2	3	4	5	6
ω_1	1	0	1	0	1	0	1
ω_2	0	1	0	1	0	1	0

t	0	1	2	3	4	5	6
ω_1	0	0	0	0	0	0	0
ω_2	1	1	1	1	1	1	1

Offenbar unterscheiden sich die beiden Varianten, denn im ersten Fall sind beide Personen abwechselnd erwerbstätig und arbeitslos, wohingegen im zweiten Fall eine Person nie, die andere Person immer arbeitslos ist. Dennoch würde bei einer synchronen Aggregation in beiden Fällen der gleiche statistische Prozess enstehen.

Kapitel 5

Demographische Prozesse

Zu den elementaren Aufgaben der empirischen Sozialforschung gehören Untersuchungen demographischer Prozesse, zunächst zur zeitlichen Entwicklung der Anzahl der Menschen, aus denen Gesellschaften bestehen, und ihrer Gliederungen nach dem Alter und dem Geschlecht. In diesem Kapitel besprechen wir im ersten Abschnitt einige Definitionen zur Konzeption und Charakterisierung demographischer Prozesse, dann werden Daten zur Bevölkerungsentwicklung in Deutschland dargestellt. Weitere Überlegungen folgen in späteren Kapiteln.

5.1 Einige Begriffe der Demographie

1. Der begriffliche Rahmen. Um demographische Prozesse explizit definieren zu können, ist ein zeitlicher und räumlicher Kontext erforderlich. Als zeitlichen Kontext setzen wir in diesem Kapitel eine diskrete Zeitachse $\mathcal{T}^*$ voraus, deren Elemente auf Zeitstellen (z.B. Tage, Monate oder Jahre) verweisen; und als räumlichen Kontext verwenden wir einen topographischen Raum $\mathcal{R}$, der aus Raumstellen besteht, in denen Menschen leben können.[1] Wie bei den Zeitstellen ist auch bei diesen Raumstellen zunächst keine

[1]Ausführliche Erläuterungen zu den Begriffsbildungen und Notationen für Zeitachsen und Zeitreihen findet man im Abschnitt 4.2.

genaue Definition erforderlich. Natürlich ist es zur Betrachtung von Migrationsvorgängen erforderlich, dass es mindestens zwei Raumstellen gibt, zwischen denen Wanderungsbewegungen stattfinden können; solange von solchen Wanderungen abgesehen wird, genügt aber bereits die Vorstellung, dass $\mathcal{R}$ nur eine Raumstelle enthält.

Schließlich soll Ω_t die Gesamtheit der Menschen bezeichnen, die während der Zeitstelle t innerhalb des Raums $\mathcal{R}$ leben. Elemente von Ω_t sind also Namen, die auf jeweils bestimmte Menschen verweisen. Der Zeitindex t ist erforderlich, weil sich nicht nur der Umfang, sondern auch die Zusammensetzung dieser Bevölkerungsmengen im Zeitablauf verändern kann. Dabei gehen wir von folgenden Konventionen aus: Wenn ein Kind in einer Zeitstelle t geboren wird, ist es Mitglied von Ω_t, aber nicht von irgendeiner früheren Bevölkerungsmenge; und wenn ein Mensch in einer Zeitstelle t stirbt, ist er Mitglied von Ω_t, aber nicht von irgendeiner späteren Bevölkerungsmenge. Ω_t umfasst also alle Menschen, deren Lebenslinien sich mit der Zeitstelle t ganz oder partiell überschneiden, d.h. die vor dem Ende von t geboren wurden, aber nicht bereits vor dem Beginn von t gestorben sind.

Eine analoge Definition wird verwendet, wenn sich Ω_t nur auf einen Teilraum $\mathcal{R}^* \subset \mathcal{R}$ bezieht, so das Zu- und Abwanderungen möglich sind. Ω_t umfasst dann alle Menschen, die während einer Zeitspanne, die sich mit der Zeitstelle t überschneidet, innerhalb des Teilraums $\mathcal{R}^*$ leben.[2]

Zur Veranschaulichung der Begriffsbildungen kann man sich eine kleine Insel vorstellen, die nur von wenigen Menschen bewohnt wird. Gelegentlich wird ein Kind geboren oder ein Mensch stirbt, und gelegentlich verlässt ein Bewohner die Insel oder jemand kommt als ein neuer Bewohner auf die Insel. Man kann sich vorstellen, dass ein Chronist über diese Vorgänge Buch führt. Wir nehmen an, dass nur jährliche Angaben erfolgen (so dass es sich bei den Zeitstellen der vorausgesetzten Zeitachse um Jahre handelt) und dass die Chronik im Jahr 1960 beginnt und bis zum Jahr 1985 fortgesetzt wird. Im ersten Jahr macht der Chronist eine Bestandsaufnahme und stellt fest, dass die Insel von 10 Personen bewohnt wird. Die folgende Tabelle zeigt ihre Namen, ihr Alter (in vollendeten Lebensjahren) und ihr Geschlecht (0 = männlich, 1 = weiblich):

Name	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
Alter	40	38	4	16	63	70	25	8	63	11
Geschlecht	0	1	1	0	1	0	1	0	1	0

Diese Tabelle bildet den Anfang der Chronik. In den folgenden Jahren nimmt der Chronist immer dann neue Einträge vor, wenn ein demographisch relevantes Ereignis (eine Geburt, ein Todesfall oder eine Zu- oder

[2]Für empirische Bevölkerungszählungen kann es natürlich sinnvoll sein, weitere Präzisierungen und Einschränkungen vorzunehmen und sich z.B. nur auf die „Wohnbevölkerung“ zu beziehen.

Tabelle 5.1-1 Chronik der fiktiven Insel 1960 – 1985.

Jahr	Name	Alter	Geschlecht	Art des Ereignisses
1961	ω_4	17	0	verlässt die Insel
1963	ω_6	73	0	stirbt
1964	ω_{11}	30	0	wird neuer Bewohner
1966	ω_{12}	0	1	wird geboren
1970	ω_{13}	0	0	wird geboren
1971	ω_9	74	1	stirbt
1975	ω_8	23	0	verlässt die Insel
1975	ω_{14}	26	1	wird neuer Bewohner
1980	ω_{15}	0	0	wird geboren
1982	ω_{16}	0	1	wird geboren
1985	ω_5	88	1	stirbt

Abwanderung) stattfindet. Tabelle 5.1-1 zeigt die Einträge bis zum Ende der Chronik im Jahr 1985.

2. Definitionen demographischer Prozesse. Der in § 1 eingeführte begriffliche Rahmen erlaubt einfache Definitionen demographischer Prozesse, indem man sich auf zeitliche Folgen der Bevölkerungsmengen Ω_t bezieht, wobei der Zeitindex t Werte in einer Zeitachse $\mathcal{T}^*$ annehmen kann. Wir unterscheiden zwei Varianten:

– Eine zeitliche Folge von Bevölkerungsmengen Ω_t wird ein *demographischer Prozess ohne externe Migration* genannt, wenn Ω_t stets alle Menschen umfasst, die innerhalb eines als Kontext vorausgesetzten topographischen Raums $\mathcal{R}$ leben.

– Eine zeitliche Folge von Bevölkerungsmengen Ω_t wird ein *demographischer Prozess mit externer Migration* genannt, wenn sich Ω_t auf einen Teilraum $\mathcal{R}^* \subset \mathcal{R}$ bezieht, so dass auch Migrationsvorgänge zwischen diesem Teilraum und seiner räumlichen Umgebung $\mathcal{R} \setminus \mathcal{R}^*$ stattfinden können.[3]

Es sei angemerkt, dass demographische Prozesse nicht unmittelbar auch statistische Prozesse sind, worunter wir zeitliche Folgen statistischer Variablen verstehen (vgl. Abschnitt 4.2). Statistische Prozesse entstehen jedoch, sobald man die Bevölkerungsmengen Ω_t als Referenzmengen statistischer Variablen verwendet.

3. Definitionen der Bevölkerungszahl. Der Begriff einer Bevölkerungszahl bezieht sich auf die Anzahl der Menschen, die während einer Zeitstelle

[3] Mit der Schreibweise $\mathcal{R} \setminus \mathcal{R}^*$ ist die Menge aller Elemente von $\mathcal{R}$ gemeint, die nicht zur Menge $\mathcal{R}^*$ gehören.

in einem bestimmten Gebiet leben. Dabei gibt es hauptsächlich zwei Unschärfen. Die erste betrifft die Formulierung „in einem bestimmten Gebiet leben“. Offenbar sind Unterscheidungen möglich. In der Bevölkerungsstatistik wird insbesondere zwischen der ortsanwesenden und der Wohnbevölkerung unterschieden.

Problematischer ist die Bezugnahme auf Zeitstellen, denn Zeitstellen haben stets eine mehr oder weniger große zeitliche Ausdehnung, so dass sich die Bevölkerungszahl während einer Zeitstelle verändern kann. Für dieses Problem gibt es keine einfache Lösung. Wir verwenden je nach Anwendungskontext eine von drei unterschiedlichen Definitionen.

a) Die erste Definition geht von den in § 1 eingeführten Bevölkerungsmengen aus: $n_t := |\Omega_t|$.[4] Bei dieser Definition ist n_t die Anzahl aller Menschen, die während der Zeitstelle t in dem Gebiet, auf das Bezug genommen wird, gelebt haben. Dabei kann es sich auch um eine kurze Zeitspanne innerhalb der Zeitstelle t handeln, so dass durch n_t auch Menschen erfasst werden, die während t geboren werden, sterben, oder zu- oder abwandern.

b) Bei Zeitstellen, die eine größere zeitliche Ausdehnung aufweisen (beispielsweise Jahre), kann es sinnvoll sein, sich auf die Bevölkerungszahl zum Beginn oder zum Ende der Zeitstelle zu beziehen. Wir verwenden dann die Notationen:

$$n_t^{\vdash} := \text{Anzahl der Menschen zu Beginn der Zeitstelle } t$$
$$n_t^{\dashv} := \text{Anzahl der Menschen zum Ende der Zeitstelle } t$$

und legen die definitorische Gleichsetzung $n_t^{\dashv} = n_{t+1}^{\vdash}$ zugrunde. Werden Zeitstellen als Jahre aufgefasst, kann man sich unter $n_t^{\vdash}$ und $n_t^{\dashv}$ die Bevölkerungszahlen am 1. Januar bzw. am 31. Dezember vorstellen.

c) Schließlich kann man versuchen, die durchschnittliche Bevölkerungszahl während einer Zeitstelle zu definieren. Wir verwenden dafür die Notation $\bar{n}_t$, ohne damit irgendeine bestimmte Definition zu implizieren. Den wichtigsten Anwendungsfall bilden die jahresdurchschnittlichen Bevölkerungszahlen der amtlichen Statistik, bei deren Berechnung von unterschiedlichen Definitionen ausgegangen wird.[5]

[4] In diesem Text werden die Zeichen ‘=’ und ‘:=’ unterschieden. Ein Gleichheitszeichen mit vorangestelltem Doppelpunkt wird verwendet, um anzudeuten, dass eine definitorische Gleichsetzung vorgenommen wird, d.h. der Ausdruck auf der linken Seite wird durch den Ausdruck auf der rechten Seite definiert. Dagegen setzt ein einfaches Gleichheitszeichen voraus, dass beide Seiten schon definiert sind.

[5] Im definitorischen Apparat der STATIS-Datenbank findet man folgende Hinweise: „Der Bevölkerungsstand gibt die Zahl der Personen an, die zur Bevölkerung gehören, nachgewiesen zu verschiedenen Zeitpunkten. Der Bevölkerungsstand im Jahresdurchschnitt insgesamt ist das arithmetische Mittel aus zwölf Monatswerten, die wiederum Durchschnitte aus dem Bevölkerungsstand am Anfang und Ende jeden Monats sind.

4. Buchführungsgleichungen. Weiterhin verwenden wir folgende Bezeichnungen für Geburten und Sterbefälle:

$$b_t := \text{Anzahl der Kinder, die in der Zeitstelle } t \text{ geboren werden}$$

$$d_t := \text{Anzahl der Menschen, die in der Zeitstelle } t \text{ sterben}$$

Somit kann für einen demographischen Prozess ohne externe Migration zunächst folgende *Buchführungsgleichung* formuliert werden:

$$n_{t+1} = n_t + b_{t+1} - d_t \tag{5.1}$$

Für einen demographischen Prozess mit externer Migration werden außerdem die Bezeichnungen

$$m_t^i := \text{Anzahl der Menschen, die in } t \text{ nach } \mathcal{R}^* \text{ einwandern}$$

$$m_t^o := \text{Anzahl der Menschen, die in } t \text{ aus } \mathcal{R}^* \text{ auswandern}$$

verwendet, so dass die Buchführungsgleichung folgende Form annimmt:

$$n_{t+1} = n_t + b_{t+1} - d_t + m_{t+1}^i - m_t^o \tag{5.2}$$

Zuwanderungen werden wie Geburten und Abwanderungen werden wie Sterbefälle verbucht. Dementsprechend werden Personen, die in der gleichen Zeitstelle t zuwandern und wieder auswandern (oder sterben), als Mitglieder der Bevölkerungsmenge Ω_t betrachtet.[6]

Anstatt sich mit n_t auf die Bevölkerungsmengen Ω_t zu beziehen, kann man auch die Bevölkerungszahlen zum Beginn und Ende der Zeitstellen verwenden. Die Buchführungsgleichung bekommt dann die Form

$$n_{t+1}^{\vdash} = n_t^{\dashv} = n_t^{\vdash} + b_t - d_t + m_t^i - m_t^o \tag{5.3}$$

Zur Berechnung des durchschnittlichen Bevölkerungsstandes nach Altersjahren und Geschlecht wird ein vereinfachtes Verfahren angewendet: Es werden lediglich die arithmetischen Durchschnittswerte aus dem Bevölkerungsstand jeder Gruppe zum Jahresanfang und -ende gebildet und mit einem Korrekturfaktor multipliziert. Dieser Korrekturfaktor ist der Quotient aus dem durchschnittlichen Bevölkerungsstand insgesamt und der Summe aller vereinfacht berechneten Durchschnittswerte des Bevölkerungsstandes in den einzelnen Altersjahren.“

Die Definitionen und Berechnungsmethoden haben sich im Laufe der Zeit verändert: „In den Jahren 1961, 1970 und 1987 wurden keine Durchschnittwerte gebildet, sondern die Ergebnisse der jeweiligen Volks- und Berufszählungen nachgewiesen.“ „Bis 1953 und von 1956 bis 1960 wurde zur Berechnung des Bevölkerungsstandes im Durchschnitt insgesamt das arithmetische Mittel aus jeweils vier Vierteljahreswerten gebildet; dagegen wurde der Bevölkerungsstand von 1953 bis 1955, von 1962 bis 1969 und wird seit 1971 – wie oben beschrieben – als Durchschnitt aus Monatswerten berechnet.“

Für die ehemalige DDR wird mitgeteilt: „Als Bevölkerungsdurchschnittszahl für ein Kalenderjahr galt bis einschließlich Berichtsjahr 1988 jeweils die zum 30.6. fortgeschriebene Einwohnerzahl.“ (Fachserie 1, Reihe 1, 1999: 15)

[6] Personen, die innerhalb der gleichen Zeitstelle mehrfach zuwandern, sollten natürlich nur einmal gezählt werden.

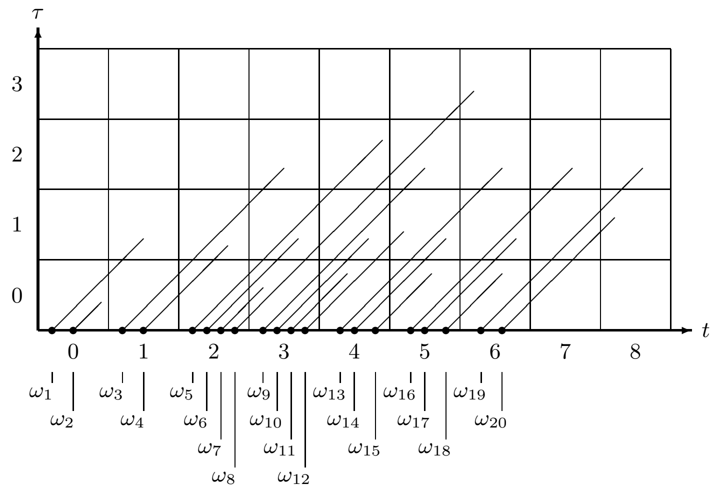


Abb. 5.1-1 Ein Lexis-Diagramm mit 20 Lebenslinien.

Es ist bemerkenswert, dass Buchführungsgleichungen nicht ohne weiteres auch mit durchschnittlichen Bevölkerungszahlen formuliert werden können.

5. Überlegungen anhand eines Lexis-Diagramms. Zur Verdeutlichung der bisherigen Begriffsbildungen eignet sich ein *Lexis-Diagramm.*[7] Abbildung 5.1-1 illustriert die Konstruktion. Die horizontale Achse repräsentiert die historische Zeit (t), die vertikale Achse das Lebensalter (τ). Beide werden als kontinuierliche Zeitachsen konzipiert, so dass Zeitstellen als Zeitintervalle aufgefasst werden können. Wir nehmen an, dass es sich um links geschlossene und rechts offene Zeitintervalle handelt, also:

$$\text{Zeitintervall } 0 \equiv [0,1[\,, \text{ Zeitintervall } 1 \equiv [1,2[\,, \text{ usw.}$$

Als eine allgemeine Formulierung verwenden wir

$$\text{Zeitstelle } t \equiv [t^{\vdash}, t^{\dashv}[$$

wobei $t^{\vdash}$ den Anfangs- und $t^{\dashv}$ den Endzeitpunkt der Zeitstelle t bezeichnen soll. (Dabei muss es sich nicht um ganzzahlige Werte handeln.)

Für jede Person ω wird nun angenommen, dass es einen genauen Geburtszeitpunkt $x(\omega)$ und einen genauen Sterbezeitpunkt $y(\omega)$ gibt.[8] Für

[7] Benannt nach dem Demographen Wilhelm Lexis (1837–1914).

[8] Es handelt sich offenbar um eine fiktive Idealisierung, da alle realen Ereignisse eine zeitliche Ausdehnung aufweisen; nicht einmal Anfangs- und Endzeitpunkte können auf einer kontinuierlichen Zeitachse genau fixiert werden.

jede Person kann somit eine Lebenslinie in das Lexis-Diagramm eingetragen werden: eine diagonale Linie, die bei $x(\omega)$ beginnt und bei $y(\omega)$ endet. Abbildung 5.1-1 zeigt solche Lebenslinien für 20 Personen.

Anhand dieses Lexis-Diagramms können zunächst unsere bisherigen Buchführungskonventionen noch einmal verdeutlicht werden. Offenbar gibt es folgende Bevölkerungsmengen:

$$\Omega_0 = \{\omega_1, \omega_2\}, \;\; \Omega_1 = \{\omega_1, \omega_3, \omega_4\}, \;\; \text{usw.}$$

Das entspricht der in § 1 eingeführten allgemeinen Definition, die jetzt auch in der Form

$$\Omega_t = \{\omega \,|\, x(\omega) < t^{\dashv}, y(\omega) \geq t^{\vdash}\}$$

angegeben werden kann. Anhand des Lexis-Diagramms in Abbildung 5.1-1 kann man auch die Buchführungsgleichung $n_{t+1} = n_t + b_{t+1} - d_t$ nachvollziehen, wie die folgende Tabelle zeigt:

t	n_t	d_t	b_t
0	2	1	2
1	3	1	2
2	6	1	4
3	9	3	4
4	9	3	3
5	9	4	3
6	7	4	2
7	3	1	0
8	2	2	0

Es sei angemerkt, dass sich auch Lebenslinien von Zuwanderern in ein Lexis-Diagramm eintragen lassen. Grundsätzlich beginnen auch sie auf der Ordinate im Alter 0; für die Buchführung sind jedoch nur diejenigen Abschnitte dieser Lebenslinien zu berücksichtigen, während der sich eine Person in dem Gebiet aufhält, für das die Buchführungsgleichung aufgestellt werden soll.

6. Gliederungen nach Geschlecht und Alter. Für demographische Überlegungen sind in erster Linie Unterscheidungen nach dem Geschlecht und dem Alter von Bedeutung. Für Unterscheidungen nach dem Geschlecht verwenden wir die Indizes m (männlich) und f (weiblich). So sind z.B. Ω_t^m die männlichen und Ω_t^f die weiblichen Personen, die in der Zeitstelle t leben, und n_t^m bzw. n_t^f sind die entsprechenden Anzahlen.

Um Personen nach ihrem Alter zu unterscheiden, verwenden wir je nach Anwendungskontext unterschiedliche Altersbegriffe:

– Der Begriff eines *exakten Alters* geht davon aus, dass man sich sowohl für die Geburt einer Person als auch für ihre Altersbestimmung auf genaue Zeitpunkte beziehen kann. Das exakte Alter ist dann als Differenz zwischen dem Erfassungs- und dem Geburtszeitpunkt definiert.

– Das *gewöhnliche Alter* ist das in vollendeten Zeitstellen erfasste exakte Alter. Geht man von Jahren aus, ist das gewöhnliche Alter einer Person bis zum 1. Geburtstag 0 Jahre, vom 1. bis zum 2. Geburtstag 1 Jahr usw. Anstelle von Jahren kann man offenbar beliebige Zeitstellen verwenden. Wir setzen im Folgenden voraus, dass zur Erfassung des gewöhnlichen Alters die Zeitstellen der zugrundeliegenden Zeitachse, durch die der demographische Prozess definiert wird, verwendet werden.

– Das *demographische Alter* einer Person in einer Zeitstelle t ist $t - t'$, wobei t' die Zeitstelle ist, in der die Person geboren wurde.

Im Unterschied zum exakten und zum gewöhnlichen Alter einer Person verändert sich das demographische Alter während einer Zeitstelle (z.B. während eines Jahres) nicht. In jeder Zeitstelle gibt es für jede Person genau ein demographisches Alter. Zur Erfassung des exakten oder gewöhnlichen Alters einer Person muss man sich jedoch auf bestimmte Zeitpunkte (innerhalb von Zeitstellen) beziehen.

Gliederungen einer Gesellschaft nach dem Alter können bei allen drei Bevölkerungsbegriffen ansetzen, die in § 3 unterschieden wurden. Allerdings gibt es dabei unterschiedliche Möglichkeiten und Restriktionen.

a) Setzt man bei den Bevölkerungsmengen Ω_t an, wobei sich t auf Zeitstellen einer Zeitachse $\mathcal{T}^*$ bezieht, möchte man sich auf Teilmengen beziehen, die Menschen gleichen Alters zusammenfassen. Wir definieren: $\Omega_{t,\tau}$ umfasst alle Personen, die in der Zeitstelle $t - \tau$ geboren wurden; insbesondere besteht $\Omega_{t,0}$ aus den in der Zeitstelle t geborenen Kindern. Offenbar kann man auch sagen, dass $\Omega_{t,\tau}$ alle Personen umfasst, die in der Zeitstelle t das demographische Alter τ haben.[9] Ihre Anzahl wird durch $n_{t,\tau} := |\Omega_{t,\tau}|$ bezeichnet.

b) Jeweils zum Ende einer Zeitstelle stimmen das gewöhnliche und das demographische Alter überein. Unmittelbar deutlich ist das bei der Verwendung von Jahren: Menschen, die zum Ende eines Jahres t im gewöhnlichen Alter τ sind, wurden im Jahr $t - \tau$ geboren. Somit lässt sich eine Gliederung nach dem Alter besonders leicht vornehmen, wenn sich die Bevölkerungszahlen auf das Ende (oder den Anfang) von Zeitstellen beziehen. Wir verwenden $n_{t,\tau}^{\vdash}$ bzw. $n_{t,\tau}^{\dashv}$ für die Anzahl der Menschen, die am Anfang bzw. zum Ende der Zeitstelle t im Alter τ sind.

c) Schließlich kann eine Gliederung nach dem Alter auch bei durchschnitt-

[9] Geht man dagegen vom gewöhnlichen Altersbegriff aus, ist zu beachten, dass die meisten Personen in $\Omega_{t,\tau}$ nicht während der gesamten Zeitstelle t im Alter τ sind, sondern erst zum Ende dieser Zeitstelle. Dies gilt natürlich nur für Personen, die bis zum Ende der Zeitstelle überleben. In Abbildung 5.1-1 ist zum Beispiel $\omega_7 \in \Omega_{3,1}$, ω_7 stirbt jedoch vor dem Erreichen des Alters 1.

lichen Bevölkerungszahlen ansetzen. Wir verwenden dann die Bezeichnung $\bar{n}_{t,\tau}$ für die Anzahl der Menschen, die im Durchschnitt der Zeitstelle t im Alter τ sind.[10]

Gliederungen nach dem Alter sind auch bei Sterbefällen und bei Zu- und Abwanderungen sinnvoll. Dabei muss wiederum zwischen dem gewöhnlichen und dem demographischen Alter unterschieden werden. Wir verwenden folgende Notationen:

- $d_{t,\tau}$ Anzahl Menschen, die in der Zeitstelle t im demographischen Alter τ sterben.
- $m^i_{t,\tau}$ Anzahl Menschen, die in der Zeitstelle t im demographischen Alter τ nach $\mathcal{R}^*$ einwandern.
- $m^o_{t,\tau}$ Anzahl Menschen, die in der Zeitstelle t im demographischen Alter τ aus $\mathcal{R}^*$ auswandern.

Bei diesen Definitionen ist stets das demographische Alter gemeint, das sich auf das Geburtsjahr $t-\tau$ bezieht. Wenn dagegen das gewöhnliche Alter gemeint ist, verwenden wir zur zusätzlichen Kennzeichnung einen Querstrich. So ist zum Beispiel $\bar{d}_{t,\tau}$ die Anzahl der Menschen, die in der Zeitstelle t im gewöhnlichen Alter τ sterben.[11]

7. Altersspezifische Buchführungsgleichungen. Die im vorangegangenen Paragraphen eingeführte Definition altersspezifischer Bevölkerungsmengen liefert nicht nur konsistente Partitionierungen der Bevölkerungsmengen Ω_t bzgl. des Alters, sie erlaubt auch einfache Formen altersspezifischer Buchführungsgleichungen. Für einen demographischen Prozess ohne externe Migration gilt $n_{t+1,0} = b_{t+1}$ und allgemein:

$$n_{t+1,\tau+1} \;=\; n_{t,\tau} - d_{t,\tau} \tag{5.4}$$

Und für einen demographischen Prozess mit externer Migration findet man $n_{t+1,0} = b_{t+1} + m^i_{t+1,0}$ und allgemein:

$$n_{t+1,\tau+1} \;=\; n_{t,\tau} - d_{t,\tau} + m^i_{t+1,\tau+1} - m^o_{t,\tau} \tag{5.5}$$

Entsprechende Gleichungen können formuliert werden, wenn man sich auf die Bevölkerungszahl zum Anfang bzw. Ende von Zeitstellen bezieht, für einen demographischen Prozess mit externer Migration:

$$n^{\vdash}_{t+1,\tau+1} \;=\; n^{\dashv}_{t,\tau} \;=\; n^{\vdash}_{t,\tau} - d_{t,\tau} + m^i_{t,\tau} - m^o_{t,\tau} \tag{5.6}$$

[10] Diese Definition ist offenbar unklar. Zum Verständnis der Vorgehensweise des Statistischen Bundesamts, das jahresdurchschnittliche Bevölkerungszahlen nach dem Alter gliedert, vgl. man die Anmerkung 5.

[11] Bezieht man sich auf das Lexis-Diagramm in Abbildung 5.1-1, findet man zum Beispiel $d_{3,1} = 2$ und $\bar{d}_{3,1} = 1$.

Es ist bemerkenswert, dass man für die Sterbefälle und die Zu- und Abwanderungen auch in diesem Fall das demographische Alter verwenden muss.[12]

8. Charakterisierungen demographischer Prozesse. In der Literatur sind sehr viele Konzepte und Maßzahlen zur Charakterisierung demographischer Prozesse vorgeschlagen worden.[13] Hier genügen einige elementare Definitionen:

- Man kann *Anzahlen* verwenden, und zwar sowohl zur Charakterisierung von Bevölkerungsmengen als auch zum Zählen demographischer Ereignisse. Dem entsprechen die in den vorangegangenen Paragraphen eingeführten Bezeichnungen.
- Ausgehend von Anzahlen können *Veränderungsraten* berechnet werden.[14] Wir verwenden die Notation $\rho_t := (x_{t+1} - x_t)/x_t$, wobei x irgendeine für die Zeitstellen definierte Größe ist.[15]
- Zur Erfassung der Häufigkeit von Sterbefällen kann zunächst eine *allgemeine Sterbeziffer* verwendet werden, die als Quotient d_t/n_t oder $d_t/\bar{n}_t$ definiert ist und meistens *pro 1000* angegeben wird.[16] Für analytische Zwecke sind altersspezifische, außerdem nach dem Geschlecht differenzierte Sterbeziffern besser geeignet, deren Definition wir in Abschnitt 6.2 besprechen werden.
- Analog zu einer allgemeinen Sterbeziffer kann auch eine *allgemeine Geburtenziffer* durch b_t/n_t oder $b_t/\bar{n}_t$ definiert werden; auch sie wird meistens *pro 1000* angegeben. Wiederum sind für analytische Zwecke altersspezifische Geburtenziffern, mit denen wir uns in Kapitel 7 beschäftigen werden, besser geeignet.

Es sei angemerkt, dass anstelle von Geburten- und Sterbeziffern auch von Geburten- bzw. Sterbe*raten* gesprochen wird.

[12] Das Statistische Bundesamt, dessen Bevölkerungsfortschreibungen sich auf den Bevölkerungsstand zum Jahresende beziehen, geht z.B. bei der Altersgliederung der Zu- und Abwanderungen so vor: „Die Bestimmung des Alters der wandernden Personen geschieht mittels Auszählung nach Geburtsjahren. Dabei werden die Personen eines bestimmten Geburtsjahrganges jeweils dem Altersjahr zugeordnet, dem sie am Jahresende angehören (Beispiel für 1999 Geburtsjahr: 1999 = Altersjahr 0 bis unter 1; Geburtsjahr 1998 = Altersjahr 1 bis unter 2 usw.).“ (Fachserie 1, Reihe 1, 1999: 14)

[13] Man vgl. z.B. die Übersichten bei U. Mueller (1993 und 2000).

[14] Da sie auch negative Werte annehmen können, ist der Ausdruck ‘Veränderungsrate’ besser als der ebenfalls oft verwendete Ausdruck ‘Wachstumsrate’.

[15] Für Entwicklungen, die mehrere Perioden umfassen, werden auch *durchschnittliche Veränderungsraten* berechnet. Betrachtet man zum Beispiel die Entwicklung von x_t zu $x_{t'}$, wobei $t' > t$ ist, ist die hier durch $\bar{\rho}$ bezeichnete durchschnittliche Veränderungsrate durch $x_{t'} = x_t\,(1+\bar{\rho})^{t'-t}$ definiert.

[16] Das Statistische Bundesamt verwendet die zweite Variante. Im Jahr 1999 hatten diese Sterbeziffern den Wert 9.8 für Männer und 10.8 für Frauen (Fachserie 1, Reihe 1, 1999: 55).

5.2 Daten zur Bevölkerungsentwicklung

In diesem Abschnitt besprechen wir einige Daten der amtlichen Statistik zur Bevölkerungsentwicklung in Deutschland.

1. Datenquellen der amtlichen Statistik. Wir beginnen mit einigen Bemerkungen zu den Datenquellen.[17] Die meisten elementaren demographischen Daten stammen von der amtlichen Statistik, in der BRD vom *Statistischen Bundesamt* (`www.destatis.de`). Deren hauptsächliche Datenquellen sind einerseits Volkszählungen, die die Anzahl (und einige Merkmale) der Menschen an einem Stichtag erfassen,[18] und andererseits Bevölkerungsregister, durch die Geburten, Todesfälle und Zu- und Abwanderungen erfasst werden:

– Bevölkerungsregister für Geburten, Todesfälle und Heiraten bzw. Ehelösungen werden von den Standesämtern geführt.[19]

– Bevölkerungsregister zum Wohnort werden von den Einwohnermeldeämtern geführt. Außerdem gibt es ein Ausländerzentralregister für Personen ohne deutsche Staatsangehörigkeit. Daten aus diesen Registern werden vom Statistischen Bundesamt für die Statistiken zur internen und externen Migration verwendet.[20]

Über die Geschichte der Volkszählungen in Deutschland heißt es in einer Publikation des Statistischen Bundesamts:

„Nach der territorialen Neuordnung der Nachfolgestaaten des Heiligen Römischen Reichs Deutscher Nation auf dem Wiener Kongreß wurde 1816 erstmals in Preußen innerhalb der neuen Grenzen eine Volkszählung durchgeführt. Die anderen Länder des Deutschen Bundes führten in der Folgezeit Volkszählungen durch, deren Ergebnisse jedoch wegen der unterschiedlichen Erhebungszeitpunkte und der unterschiedlichen Abgrenzung der Merkmale kaum untereinander vergleichbar sind. Erst mit der Schaffung des Norddeutschen Zollvereins 1834 wurde im größten Teil des späteren Deutschen Reichs eine größere Einheitlichkeit des Vorgehens erreicht. Von da an fand bis 1867 alle drei Jahre Anfang Dezember eine Volkszählung in den Mitgliedsländern des Zollvereins statt. Die übrigen deutschen Länder schlossen sich diesem Verfahren erst 1867 an, so daß am 3. Dezember dieses Jahres erstmals in allen deutschen Ländern zum gleichen Zeitpunkt gezählt wurde. Die nächste Volkszählung erfolgte dann nach der Reichsgründung, am 1. Dezember 1871. Vom 1. Dezember 1875 an wurden Volkszählungen im Fünf-Jahres-Turnus durchgeführt. Die letzte Zählung vor dem Ersten Weltkrieg war am 1. Dezember 1910. Danach vergingen fast 15 Jahre, bis am 16. Juni 1925 wieder eine das gesamte damalige Reichsgebiet umfassende Volkszählung stattfinden konnte. Eine vorher – im Oktober 1919 – durchgeführte Zählung hatte, da die Verhältnisse noch nicht wieder konsolidiert waren, nur behelfsmäßigen Charakter. Der mit der Zählung 1925 wieder angestrebte Fünf-Jahres-Rhythmus konnte infolge der Weltwirtschaftskrise nicht eingehalten werden. So fand die nächste Zählung erst acht Jahre später am 16. Juni 1933 statt, der im Abstand von sechs Jahren am 19. Mai 1939 die letzte Zählung vor dem Ausbruch des Zweiten Weltkrieges folgte. Die nächste Volkszählung, die am 29. Oktober 1946 auf Anordnung der Besatzungsmächte durchgeführt wurde, konnte aus den gleichen Gründen wie die von 1919 die normalerweise geforderten Ansprüche nicht erfüllen, war aber für die Bewältigung der damaligen Notsituation von großer Bedeutung. Es war die letzte Zählung, die mit einem einheitlichen Erhebungsprogramm in den vier Besatzungszonen gleichzeitig stattfand. Ihr folgte am 13. September 1950 die erste Volkszählung im Bundesgebiet. Weitere Volkszählungen im Abstand von etwa zehn Jahren fanden am 6. Juni 1961 und am 27. Mai 1970 statt.“ (Statistisches Bundesamt 1972: 89)

Eine weitere, die bisher letzte Volkszählung fand in der BRD am 25. Mai 1987 statt. In der ehemaligen DDR wurden Volkszählungen in den Jahren 1950 (31.8.), 1964 (31.12.) und 1981 (31.12.) durchgeführt.

Zu den jeweils verwendeten Bevölkerungsbegriffen findet man in derselben Quelle folgende Erläuterungen:

„Die Zählungen vor dem 3. Dezember 1867 hatten nicht immer einen einheitlichen Bevölkerungsbegriff. In den durch Zollverträge miteinander verbundenen Ländern wurde zwischen 1834 und 1867 die sog. *Zollabrechnungsbevölkerung* festgestellt. Es handelt sich hierbei im wesentlichen um die dauerhaft wohnhafte Bevölkerung. Dieser Bevölkerungsbegriff wurde 1863 dahingehend präzisiert, daß Personen, die länger als ein Jahr abwesend waren, nicht zur Zollabrechnungsbevölkerung gezählt wurden. Bei der Zählung 1867 wurde daneben erstmals auch die *ortsanwesende Bevölkerung* festgestellt, d.h. alle Personen, die sich zum Stichtag der Zählung im Zählungsgebiet aufhielten. Dieser Bevölkerungsbegriff stand in der Folgezeit im Vordergrund. Im Kaiserreich wurde die ortsanwesende Bevölkerung allein als maßgeblich nachgewiesen. Bei der Zählung 1925 wurde erstmals der Begriff der *Wohnbevölkerung* verwendet, der in etwa an den Bevölkerungsbegriff zwischen 1834 und 1867 anschließt. Zur Wohnbevölkerung zählten alle Personen, die am Zählungsstichtag im Zählungsgebiet ihren ständigen Wohnsitz hatten, einschl. der vorübergehend Abwesenden sowie ausschließlich der vorübergehend Anwesenden. Personen mit mehreren Wohnsitzen wurden an dem Ort zur Bevölkerung gezählt, an dem sie sich am Stichtag der Zählung befanden. Davon abweichend wurden Untermieter (einschl. Hausangestellte, Schüler und Studierende mit zweitem Wohnsitz) stets an ihrem Arbeits- bzw. Studienort zur Wohnbevölkerung gerechnet. Dieser Bevölkerungsbegriff liegt, mit nur unwesentlichen Abweichungen, allen seitherigen Volkszählungen sowie der Bevölkerungsentwicklung zugrunde [. . .].“ (Statistisches Bundesamt 1972: 89)

[17] Eine umfassende Einführung in die Datenquellen und ihre institutionellen Grundlagen gibt H. Rinne (1996). Ausführliche Informationen zu demgraphischen Daten findet man bei C. Schmid (2000).

[18] Informationen über den Fragebogen, der bei der letzten Volkszählung in der BRD im Jahr 1987 verwendet wurde, findet man bei Würzberger, Störtzbach and Stürmer (1986). Eine Übersicht über die in Deutschland durchgeführten Volks-, Berufs- und Betriebszählungen gibt H. Rinne (1996: 60).

[19] Diese Register wurden im Jahr 1875 eingeführt. Informationen zur Geschichte findet man bei Schütz (1977). Darstellungen der für die Registratur verwendeten Formblätter wurden vom Statistischen Bundesamt in der Fachserie 1, Reihe 1, 1990: 312–323, veröffentlicht.

[20] Vgl. Fachserie 1, Reihe 1, 1999: 13 –14.

Einige Veränderungen in den Definitionen wurden nach der Einführung eines neuen Meldegesetzes im Jahr 1983 vorgenommen.[21]

2. Die Entwicklung der Bevölkerungszahl. Daten der amtlichen Statistik beginnen mit der ersten Volkszählung in Preußen im Jahr 1816. Eine Schwierigkeit liegt offenbar darin, dass sich die Grenzen Deutschlands seither häufig verändert haben. Beschränkt man sich auf die Entwicklung nach dem 2. Weltkrieg, genügt es jedoch, zwischen dem Gebiet der früheren BRD (einschließlich West-Berlin),[22] und dem Gebiet der früheren DDR (einschließlich Ost-Berlin) zu unterscheiden.

Um einen Eindruck von der Entwicklung seit der ersten Volkszählung in Preußen zu gewinnen, können die Daten der Tabelle 5.2-1 verwendet werden, die sich auf das Gebiet der früheren BRD beziehen.[23] Aus der graphischen Darstellung in Abbildung 5.2-1 erkennt man, dass das langfri-

[21] Dazu heißt es im Statistischen Jahrbuch 2001 (S. 41): „Die Einwohnerzahlen basierten bis zum Frühjahr 1983 auf dem Wohnbevölkerungsbegriff. Danach gehörten Personen mit nur einer Wohnung zur Wohnbevölkerung der Gemeinde, in der sich diese Wohnung befand. Personen mit mehr als einer Wohnung oder Unterkunft im früheren Bundesgebiet wurden der Wohnbevölkerung derjenigen Gemeinde zugeordnet, von der aus sie zur Arbeit oder Ausbildung gingen. Bei Personen, die weder berufstätig waren, noch sich in der Ausbildung befanden, war die Wohnung oder Unterkunft maßgebend, in der sie sich überwiegend aufhielten.
Mit der Einführung neuer Meldegesetze in allen Bundesländern haben die statistischen Ämter die Fortschreibung der Einwohnerzahlen auf den neuen Begriff der Bevölkerung am Ort der alleinigen bzw. Hauptwohnung umgestellt.“ (Es folgen Erläuterungen zum Begriff der Hauptwohnung.)

[22] Das Saarland wurde erst 1957 ein Teil dieses Gebiets; in vielen Zeitreihen der amtlichen Statistik wird es jedoch bereits für den Zeitraum 1950 – 56 einbezogen.

[23] Im Statistischen Jahrbuch 2001 (S. 44), dem die Daten entnommen wurden, findet man folgende Hinweise zu den Quellen:

a) Die Angaben für 1961, 1970, und 1987 beruhen auf Volkszählungen und beziehen sich auf die deren Stichtage (6. Juni 1961, 27 Mai 1970 und 25. Mai 1987). Bei den restlichen Angaben für den Zeitraum seit 1946 handelt es sich um Schätzwerte der jahresdurchschnittlichen Bevölkerungszahl. Dazu heißt es (S. 41): „Bei den [. . .] für die Jahre 1950 bis 1970 nachgewiesenen Fortschreibungszahlen handelt es sich um rückgerechnete Einwohnerzahlen aufgrund der Ergebnisse der Wohnungsstatistik vom 25.9.1956 (1950 bis 1955), der Volkszählung vom 6.6.1961 (1957 bis 1960) und der Volkszählung vom 27.5.1970 (1962 bis 1969). Die für die Jahre ab 1970 bis einschl. 1986 nachgewiesenen Bevölkerungszahlen sind Fortschreibungsdaten, die von den Ergebnissen der Volkszählung 1970 ausgehen. Die ab 30.6.1987 nachgewiesenen Bevölkerungszahlen beruhen auf den Ergebnissen der Volkszählung 1987.“

b) Die Quellen der Angaben für frühere Perioden werden nicht explizit dokumentiert. Man kann annehmen, dass seit 1871 zunächst Daten der Volkszählungen verwendet wurden, die in den Jahren 1871, 1880, 1890, 1900, 1910, 1925, 1933, and 1939 stattfanden, und dass es sich für die Jahre zwischen den Volkszählungen um Schätzwerte handelt.

c) Schließlich kann angenommen werden, dass auch die Angaben für den Zeitraum vor 1871 aus den Daten von Volkszählungen abgeleitet worden sind, die in 3-Jahres-Intervallen in Preußen seit 1816 und später auch in anderen Ländern des damaligen Zollvereins durchgeführt wurden. Es ist jedoch unklar, wie die Umrechnung auf das Gebiet der früheren BRD erfolgte.

Tabelle 5.2-1 Bevölkerungszahlen (in 1000) im Gebiet der früheren BRD. Quelle: Statistisches Jahrbuch 2001: 44.

t	$\bar{n}_t$	t	$\bar{n}_t$	t	$\bar{n}_t$	t	$\bar{n}_t$
1816	13720	1925	39017	1954	51180	1977	61419
1819	14150	1926	39351	1955	52382	1978	61350
1822	14580	1927	39592	1956	53008	1979	61382
1825	15130	1928	39861	1957	53656	1980	61538
1828	15270	1929	40107	1958	54292	1981	61663
1831	15860	1930	40334	1959	54876	1982	61596
1834	16170	1931	40527	1960	55433	1983	61383
1837	16570	1932	40737	1961	56175	1984	61126
1840	17010	1933	40956	1962	56837	1985	60975
1843	17440	1934	41168	1963	57389	1986	61010
1846	17780	1935	41457	1964	57971	1987	61077
1849	17970	1936	41781	1965	58619	1988	61450
1852	18230	1937	42118	1966	59148	1989	62063
1855	18230	1938	42576	1967	59286	1990	63254
1858	18600	1939	43008	1968	59500	1991	64074
1861	19050	1946	46190	1969	60067	1992	64865
1864	19600	1947	46992	1970	60651	1993	65534
1867	19950	1948	48251	1971	61280	1994	65858
1871	20410	1949	49198	1972	61697	1995	66156
1880	22820	1950	49989	1973	61987	1996	66444
1890	25433	1951	50528	1974	62071	1997	66647
1900	29838	1952	50859	1975	61847	1998	66697
1910	35590	1953	51350	1976	61574	1999	66834

Tabelle 5.2-2 Bevölkerungszahlen (in 1000) im Gebiet der früheren BRD ($\bar{n}_t^a$) und der früheren DDR ($\bar{n}_t^b$). Quelle: Statistisches Jahrbuch 2001: 44.

t	$\bar{n}_t^a$	$\bar{n}_t^b$	t	$\bar{n}_t^a$	$\bar{n}_t^b$	t	$\bar{n}_t^a$	$\bar{n}_t^b$
1950	49989	18388	1967	59286	17082	1984	61126	16671
1951	50528		1968	59500	17084	1985	60975	16644
1952	50859		1969	60067	17076	1986	61010	16624
1953	51350	18178	1970	60651	17058	1987	61077	16641
1954	51180	18059	1971	61280	17061	1988	61450	16666
1955	52382	17944	1972	61697	17043	1989	62063	16614
1956	53008	17716	1973	61987	16980	1990	63254	16111
1957	53656	17517	1974	62071	16925	1991	64074	15910
1958	54292	17355	1975	61847	16850	1992	64865	15730
1959	54876	17298	1976	61574	16786	1993	65534	15645
1960	55433	17241	1977	61419	16765	1994	65858	15564
1961	56175	17125	1978	61350	16756	1995	66156	15505
1962	56837	17102	1979	61382	16745	1996	66444	15451
1963	57389	17155	1980	61538	16737	1997	66647	15405
1964	57971	16992	1981	61663	16736	1998	66697	15332
1965	58619	17028	1982	61596	16697	1999	66834	15253
1966	59148	17066	1983	61383	16699			

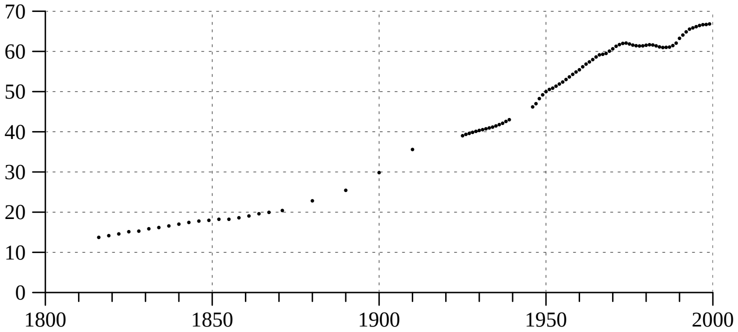


Abb. 5.2-1 Graphische Darstellung der Daten aus Tabelle 5.2-1. Ordinate: Anzahl in Millionen.

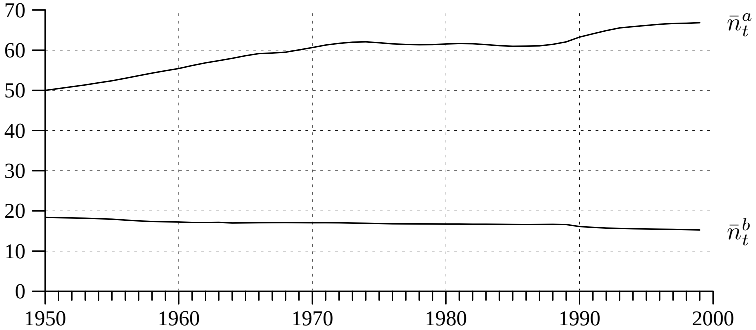


Abb. 5.2-2 Graphische Darstellung der Daten aus Tabelle 5.2-2. Ordinate: Anzahl in Millionen.

stige Bevölkerungswachstum erst Anfang der 1970er Jahre aufgehört hat; erst durch die Migrationsprozesse nach der Wiedervereinigung kam es dann erneut zu einer Bevölkerungszunahme. Dagegen gab es im Gebiet der ehemaligen DDR in der gesamten Nachkriegszeit eine geringfügige Bevölkerungsabnahme, wie der in Abbildung 5.2-2 dargestellte Vergleich zeigt.

3. Geburten- und Sterbefälle. Die Bevölkerungszahl verändert sich durch Geburten, Sterbefälle und Zu- und Abwanderungen. Wir betrachten zunächst nur Geburten und Sterbefälle. Tabelle 5.2-3 zeigt die Entwicklung in den Gebieten der früheren BRD und DDR seit 1950. (Die symbolischen Notationen entsprechen den in Abschnitt 5.1 (§ 3) erläuterten Definitionen, die zusätzlichen Indizes a und b dienen zur Unterscheidung der Gebiete.) Abbildung 5.2-3 veranschaulicht die Entwicklung.

Ein Vergleich der Abbildungen 5.2-2 und 5.2-3 zeigt, dass es eine Korrespondenz zwischen Perioden unterschiedlicher Bevölkerungsentwicklungen und Perioden, in denen die Anzahl der Geburten größer oder kleiner als

Tabelle 5.2-3 Anzahlen (in 1000) der Geburten (b_t) und Todesfälle (d_t) im Gebiet der früheren BRD (linke Hälfte) und der früheren DDR (rechte Hälfte). Quelle: Fachserie 1. Reihe 1, 1999: 43–44.

t	b_t^a	d_t^a	$b_t^a - d_t^a$	t	b_t^b	d_t^b	$b_t^b - d_t^b$
1950	812.8	528.7	284.1	1950	303.9	219.6	84.3
1951	795.6	543.9	251.7	1951	310.8	208.8	102.0
1952	799.1	546.0	253.1	1952	306.0	221.7	84.3
1953	796.1	578.0	218.1	1953	298.9	212.6	86.3
1954	816.0	555.5	260.6	1954	293.7	219.8	73.9
1955	820.1	581.9	238.3	1955	293.3	214.1	79.2
1956	855.9	599.4	256.5	1956	281.3	212.7	68.6
1957	892.2	615.0	277.2	1957	273.3	225.2	48.1
1958	904.5	597.3	307.2	1958	271.4	221.1	50.3
1959	951.9	605.5	346.4	1959	292.0	229.9	62.1
1960	968.6	643.0	325.7	1960	293.0	233.8	59.2
1961	1012.7	627.6	385.1	1961	300.8	222.7	78.1
1962	1018.6	644.8	373.7	1962	298.0	234.0	64.0
1963	1054.1	673.1	381.1	1963	301.5	222.0	79.5
1964	1065.4	644.1	421.3	1964	291.9	226.2	65.7
1965	1044.3	677.6	366.7	1965	281.1	230.3	50.8
1966	1050.3	686.3	364.0	1966	268.0	225.7	42.3
1967	1019.5	687.3	332.1	1967	252.8	227.1	25.7
1968	969.8	734.0	235.8	1968	245.1	242.5	2.7
1969	903.5	744.4	159.1	1969	238.9	243.7	-4.8
1970	810.8	734.8	76.0	1970	236.9	240.8	-3.9
1971	778.5	730.7	47.9	1971	234.9	235.0	-0.1
1972	701.2	731.3	-30.0	1972	200.4	234.4	-34.0
1973	635.6	731.0	-95.4	1973	180.3	232.0	-51.6
1974	626.4	727.5	-101.1	1974	179.1	229.1	-49.9
1975	600.5	749.3	-148.7	1975	181.8	240.4	-58.6
1976	602.9	733.1	-130.3	1976	195.5	233.7	-38.2
1977	582.3	704.9	-122.6	1977	223.2	226.2	-3.1
1978	576.5	723.2	-146.8	1978	232.2	232.3	-0.2
1979	582.0	711.7	-129.7	1979	235.2	232.7	2.5
1980	620.7	714.1	-93.5	1980	245.1	238.3	6.9
1981	624.6	722.2	-97.6	1981	237.5	232.2	5.3
1982	621.2	715.9	-94.7	1982	240.1	228.0	12.1
1983	594.2	718.3	-124.2	1983	233.8	222.7	11.1
1984	584.2	696.1	-112.0	1984	228.1	221.2	7.0
1985	586.2	704.3	-118.1	1985	227.6	225.4	2.3
1986	626.0	701.9	-75.9	1986	222.3	223.5	-1.3
1987	642.0	687.4	-45.4	1987	226.0	213.9	12.1
1988	677.3	687.5	-10.3	1988	215.7	213.1	2.6
1989	681.5	697.7	-16.2	1989	198.9	205.7	-6.8
1990	727.2	713.3	13.9	1990	178.5	208.1	-29.6
1991	722.2	708.8	13.4	1991	107.8	202.4	-94.7
1992	720.8	695.3	25.5	1992	88.3	190.2	-101.9
1993	717.9	711.6	6.3	1993	80.5	185.6	-105.1
1994	690.9	703.3	-12.4	1994	78.7	181.4	-102.7
1995	681.4	706.5	-25.1	1995	83.8	178.1	-94.2
1996	702.7	708.3	-5.6	1996	93.3	174.5	-81.2
1997	711.9	692.8	19.1	1997	100.3	167.5	-67.3
1998	682.2	688.1	-5.9	1998	102.9	164.3	-61.4
1999	664.0	685.0	-21.0	1999	106.7	161.3	-54.6

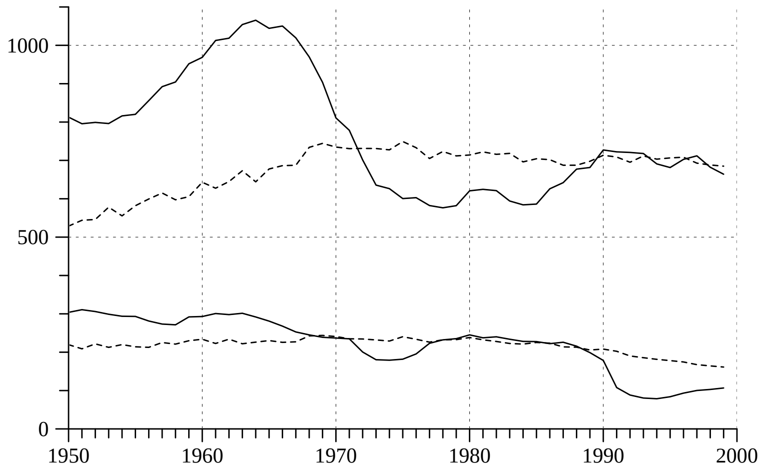


Abb. 5.2-3 Anzahlen (in 1000) der Geburten (durchgezogene Linien) und Todesfälle (gestrichelte Linien) im Gebiet der früheren BRD (obere Hälfte) und früheren DDR (untere Hälfte). Die Daten entsprechen den Angaben in Tabelle 5.2-3.

die Anzahl der Todesfälle ist, gibt. Besonders bemerkenswert ist die erhebliche Variation in der Anzahl der Geburten. Im Gebiet der früheren BRD folgte auf einen „Baby-Boom“ in den 1960er Jahren ein deutlicher Rückgang. Im Gebiet der früheren DDR fand nach der Wiedervereinigung ein außerordentlich großer Geburtenrückgang statt.

Diese Entwicklungen sollten jedoch in einem längerfristigen historischen Kontext betrachtet werden. Informativ sind bereits die allgemeinen Geburten- und Sterbeziffern, die vom Statistischen Bundesamt für die meisten Jahre seit 1841 publiziert wurden. Wir verwenden die in Abschnitt 5.1 (§ 8) erläuterten Definitionen, also $b_t/\bar{n}_t$ für die allgemeine Geburtenziffer und $d_t/\bar{n}_t$ für die allgemeine Sterbeziffer (jeweils multipliziert mit 1000). Tabelle 5.2-4 zeigt die Daten.[24] Die graphische Veranschaulichung in Abbildung 5.2-4 zeigt eindrucksvoll den langfristigen Rückgang sowohl der Sterbeziffer als auch der Geburtenziffer. Man erkennt auch, dass bis etwa 1970 die Geburtenziffern größer als die Sterbeziffern waren; eine Ausnahme bildeten nur die Jahre nach dem ersten Weltkrieg.

[24] Für den Zeitraum 1841 – 1943 wird in der Quelle der Ausdruck ‘Reichsgebiet’ verwendet, für die Jahre 1938 bis 1943 mit dem zusätzlichen Hinweis „Gebietsstand 31.12.1937“ (Statistisches Bundesamt 1972: 103). Für die Jahre 1871–1918 (und vermutlich auch für frühere Jahre, vgl. Statistisches Jahrbuch für das Deutsche Reich 1919: 2), beziehen sich die Daten auf das Gebiet des Deutschen Reichs. In anderen Quellen findet man für den Zeitraum vor 1871 auch andere Angaben, die sich auf das Gebiet des Deutschen Zollvereins beziehen.

Tabelle 5.2-4 Allgemeine Geburtenziffer (CBR) und allgemeine Sterbeziffer (CDR) in Deutschland. Angaben für den Zeitraum 1841 – 1943 beziehen sich auf das Gebiet des Deutschen Reichs in wechselnden Grenzen (s. Fußnote 24); für den Zeitraum 1946 – 1999 beziehen sie sich auf das Gebiet der früheren BRD. Quellen: Stat. Bundesamt, Bevölkerung und Wirtschaft 1872 – 1972: 101-103, und Fachserie 1, Reihe 1, 1999: 50.

t	CBR	CDR	t	CBR	CDR	t	CBR	CDR	t	CBR	CDR
1841	36.4	26.2	1881	37.0	25.5	1921	25.3	13.9	1961	18.0	11.2
1842	37.6	27.1	1882	37.2	25.7	1922	23.0	14.4	1962	17.9	11.3
1843	36.0	26.9	1883	36.6	25.9	1923	21.1	13.9	1963	18.3	11.7
1844	35.9	24.5	1884	37.2	26.0	1924	20.5	12.3	1964	18.2	11.0
1845	37.3	25.3	1885	37.0	25.7	1925	20.7	11.9	1965	17.7	11.5
1846	36.0	27.1	1886	37.1	26.2	1926	19.5	11.7	1966	17.6	11.5
1847	33.3	28.3	1887	36.9	24.2	1927	18.4	12.0	1967	17.0	11.5
1848	33.3	29.0	1888	36.6	23.7	1928	18.6	11.6	1968	16.1	12.2
1849	38.1	27.1	1889	36.4	23.7	1929	17.9	12.6	1969	14.8	12.2
1850	37.2	25.6	1890	35.7	24.4	1930	17.5	11.1	1970	13.4	12.1
1851	36.7	25.0	1891	37.0	23.4	1931	16.0	11.2	1971	12.7	11.9
1852	35.5	28.4	1892	35.7	24.1	1932	15.1	10.8	1972	11.3	11.8
1853	34.6	27.2	1893	36.8	24.6	1933	14.7	11.2	1973	10.3	11.8
1854	34.0	27.0	1894	35.9	22.3	1934	18.0	10.9	1974	10.1	11.7
1855	32.2	28.1	1895	36.1	22.1	1935	18.9	11.8	1975	9.7	12.1
1856	33.3	25.2	1896	36.3	20.8	1936	19.0	11.8	1976	9.8	11.9
1857	36.0	27.2	1897	36.1	21.3	1937	18.8	11.7	1977	9.5	11.5
1858	36.8	26.8	1898	36.1	20.5	1938	19.6	11.6	1978	9.4	11.8
1859	37.5	25.7	1899	35.9	21.5	1939	20.4	12.3	1979	9.5	11.6
1860	36.3	23.2	1900	35.6	22.1	1940	20.0	12.7	1980	10.1	11.6
1861	35.7	25.6	1901	35.7	20.7	1941	18.6	12.0	1981	10.1	11.7
1862	35.4	24.6	1902	35.1	19.4	1942	14.9	12.0	1982	10.1	11.6
1863	37.5	25.7	1903	33.8	20.0	1943	16.0	12.1	1983	9.7	11.7
1864	37.8	26.2	1904	34.0	19.6	1944			1984	9.5	11.4
1865	37.6	27.6	1905	33.0	19.8	1945			1985	9.6	11.6
1866	37.8	30.6	1906	33.1	18.2	1946	16.1	13.0	1986	10.3	11.5
1867	36.8	26.1	1907	32.3	18.0	1947	16.4	12.1	1987	10.5	11.3
1868	36.8	27.6	1908	32.1	18.1	1948	16.5	10.5	1988	11.0	11.2
1869	37.8	26.9	1909	31.0	17.2	1949	16.8	10.4	1989	11.0	11.2
1870	38.5	27.4	1910	29.8	16.2	1950	16.2	10.5	1990	11.5	11.3
1871	34.5	24.6	1911	28.6	17.3	1951	15.7	10.8	1991	11.3	11.1
1872	39.5	29.0	1912	28.3	15.6	1952	15.7	10.7	1992	11.1	10.7
1873	39.7	28.3	1913	27.5	15.0	1953	15.5	11.3	1993	11.0	10.9
1874	40.1	26.7	1914	26.8	19.0	1954	15.7	10.7	1994	10.5	10.7
1875	40.6	27.6	1915	20.4	21.4	1955	15.7	11.1	1995	10.3	10.7
1876	40.9	26.3	1916	15.2	19.2	1956	16.1	11.3	1996	10.5	10.6
1877	40.0	26.4	1917	13.9	20.6	1957	16.6	11.5	1997	10.7	10.4
1878	38.9	26.2	1918	14.3	24.8	1958	16.7	11.0	1998	10.2	10.3
1879	38.9	25.6	1919	20.0	15.6	1959	17.3	11.0	1999	9.9	10.3
1880	37.6	26.0	1920	25.9	15.1	1960	17.4	11.6			

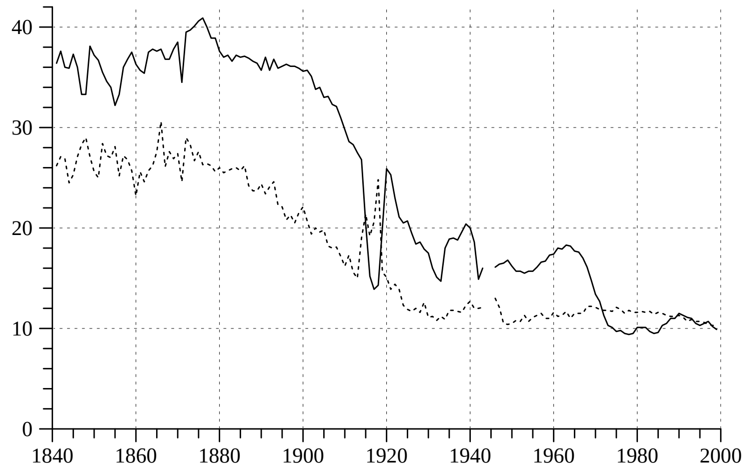


Abb. 5.2-4 Allgemeine Geburtenziffer (durchgezogene Linie) und Sterbeziffer (gestrichelte Linie) für den Zeitraum 1841–1999 in Deutschland entsprechend den Daten der Tabelle 5.2-4.

4. Buchführungsgleichungen und Migrationsvorgänge. Um zu verstehen, wie die Bevölkerungsentwicklung nicht nur von Geburten und Todesfällen, sondern auch von Migrationsvorgängen abhängt, eignen sich die in Abschnitt 5.1 (§ 3) besprochenen Buchführungsgleichungen. Da das Statistische Bundesamt auch Bevölkerungszahlen publiziert, die sich auf das Ende eines Kalenderjahrs beziehen, verwenden wir die Variante

$$n_{t+1}^{\vdash} = n_t^{\dashv} = n_t^{\vdash} + b_t - d_t + m_t^i - m_t^o$$

Hierbei bezieht sich $n_t^{\vdash}$ auf die Bevölkerungszahl zu Beginn und $n_t^{\dashv}$ auf die Bevölkerungszahl am Ende des Kalenderjahrs t. Die übrigen Größen wurden bereits definiert: Geburten (b_t), Todesfälle (d_t), Zuwanderung (m_t^i) und Abwanderung (m_t^o), jeweils Stromgrößen für das Kalenderjahr t.

Werte für $n_t^{\vdash}$ findet man in den Tabellen 5.2-5 und 5.2-6, die sich auf die Gebiete der früheren BRD bzw. der früheren DDR beziehen. Außerdem findet man in den Tabellen die Anzahl der Geburten (b_t) und der Todesfälle (d_t), die mit den Angaben in der Tabelle 5.2-3 übereinstimmen. Daraus können mit der aus der Buchführungsgleichung (5.3) ableitbaren Gleichung

$$(m_t^i - m_t^o) = (n_t^{\dashv} - n_t^{\vdash}) - (b_t - d_t)$$

die Migrationssalden berechnet werden, deren Werte ebenfalls in den Tabellen 5.2-5 und 5.2-6 angegeben werden.

Abbildung 5.2-5 zeigt, wie sich diese Migrationssalden entwickelt haben. Im Gebiet der früheren BRD war in den meisten Jahren die Zuwanderung größer als die Abwanderung. Bis zum Jahr 1961, in dem die DDR

Tabelle 5.2-5 Komponenten der Bevölkerungsveränderung (in 1000) im Gebiet der früheren BRD. Quelle: Fachserie 1, Reihe 1, 1999: 30 und 43.

t	$n_t^{\vdash}$	b_t	d_t	$b_t - d_t$	$m_t^i - m_t^o$
1951	50336.1	795.6	543.9	251.7	138.2
1952	50726.0	799.1	546.0	253.1	72.8
1953	51051.9	796.1	578.0	218.1	369.6
1954	51639.6	816.0	555.5	260.6	226.6
1955	52126.8	820.1	581.9	238.3	333.2
1956	52698.3	855.9	599.4	256.5	364.0
1957	53318.8	892.2	615.0	277.2	397.8
1958	53993.8	904.5	597.3	307.2	305.0
1959	54606.0	951.9	605.5	346.4	171.0
1960	55123.4	968.6	643.0	325.7	335.7
1961	55784.8	1012.7	627.6	385.1	419.2
1962	56589.1	1018.6	644.8	373.7	284.4
1963	57247.2	1054.1	673.1	381.1	236.2
1964	57864.5	1065.4	644.1	421.3	301.7
1965	58587.5	1044.3	677.6	366.7	342.4
1966	59296.6	1050.3	686.3	364.0	132.3
1967	59792.9	1019.5	687.3	332.1	-176.5
1968	59948.5	969.8	734.0	235.8	278.7
1969	60463.0	903.5	744.4	159.1	572.5
1970	61194.6	810.8	734.8	76.0	-269.4
1971	61001.2	778.5	730.7	47.9	453.4
1972	61502.5	701.2	731.3	-30.0	336.9
1973	61809.4	635.6	731.0	-95.4	387.4
1974	62101.4	626.4	727.5	-101.1	-8.8
1975	61991.5	600.5	749.3	-148.7	-198.2
1976	61644.6	602.9	733.1	-130.3	-72.3
1977	61442.0	582.3	704.9	-122.6	33.3
1978	61352.7	576.5	723.2	-146.8	115.8
1979	61321.7	582.0	711.7	-129.7	247.3
1980	61439.3	620.7	714.1	-93.5	312.1
1981	61657.9	624.6	722.2	-97.6	152.4
1982	61712.7	621.2	715.9	-94.7	-71.9
1983	61546.1	594.2	718.3	-124.2	-115.2
1984	61306.7	584.2	696.1	-112.0	-145.4
1985	61049.3	586.2	704.3	-118.1	89.3
1986	61020.5	626.0	701.9	-75.9	195.9
1987	61140.5	642.0	687.4	-45.4	143.0
1988	61238.1	677.3	687.5	-10.3	487.3
1989	61715.1	681.5	697.7	-16.2	980.1
1990	62679.0	727.2	713.3	13.9	1032.8
1991	63725.7	722.2	708.8	13.4	745.7
1992	64484.8	720.8	695.3	25.5	778.9
1993	65289.2	717.9	711.6	6.3	444.2
1994	65739.7	690.9	703.3	-12.4	279.9
1995	66007.2	681.4	706.5	-25.1	359.9
1996	66342.0	702.7	708.3	-5.6	247.0
1997	66583.4	711.9	692.8	19.1	85.5
1998	66688.0	682.2	688.1	-5.9	65.2
1999	66747.3	664.0	685.0	-21.0	219.9

Tabelle 5.2-6 Komponenten der Bevölkerungsveränderung (in 1000) im Gebiet der früheren DDR. Quelle: Fachserie 1, Reihe 1, 1999: 30 und 44.

t	$n_t^{\vdash}$	b_t	d_t	$b_t - d_t$	$m_t^i - m_t^o$
1951	18388.2	310.8	208.8	102.0	-140.1
1952	18350.1	306.0	221.7	84.3	-134.3
1953	18300.1	298.9	212.6	86.3	-274.3
1954	18112.1	293.7	219.8	73.9	-184.5
1955	18001.5	293.3	214.1	79.2	-248.5
1956	17832.2	281.3	212.7	68.6	-297.2
1957	17603.6	273.3	225.2	48.1	-241.0
1958	17410.7	271.4	221.1	50.3	-149.3
1959	17311.7	292.0	229.9	62.1	-87.9
1960	17285.9	293.0	233.8	59.2	-156.6
1961	17188.5	300.8	222.7	78.1	-187.3
1962	17079.3	298.0	234.0	64.0	-7.4
1963	17135.9	301.5	222.0	79.5	-34.3
1964	17181.1	291.9	226.2	65.7	-243.2
1965	17003.6	281.1	230.3	50.8	-14.7
1966	17039.7	268.0	225.7	42.3	-10.6
1967	17071.4	252.8	227.1	25.7	-7.2
1968	17089.9	245.1	242.5	2.7	-5.4
1969	17087.2	238.9	243.7	-4.8	-7.9
1970	17074.5	236.9	240.8	-3.9	-2.3
1971	17068.3	234.9	235.0	-0.1	-14.5
1972	17053.7	200.4	234.4	-34.0	-8.4
1973	17011.3	180.3	232.0	-51.6	-8.4
1974	16951.3	179.1	229.1	-49.9	-10.6
1975	16890.8	181.8	240.4	-58.6	-12.0
1976	16820.2	195.5	233.7	-38.2	-14.9
1977	16767.0	223.2	226.2	-3.1	-6.0
1978	16757.9	232.2	232.3	-0.2	-6.3
1979	16751.4	235.2	232.7	2.5	-13.6
1980	16740.3	245.1	238.3	6.9	-7.7
1981	16739.5	237.5	232.2	5.3	-39.2
1982	16705.6	240.1	228.0	12.1	-15.4
1983	16702.3	233.8	222.7	11.1	-11.9
1984	16701.5	228.1	221.2	7.0	-48.5
1985	16660.0	227.6	225.4	2.3	-22.2
1986	16640.1	222.3	223.5	-1.3	1.1
1987	16639.9	226.0	213.9	12.1	9.4
1988	16661.4	215.7	213.1	2.6	10.6
1989	16674.6	198.9	205.7	-6.8	-234.0
1990	16433.8	178.5	208.1	-29.6	-376.6
1991	16027.6	107.8	202.4	-94.7	-143.1
1992	15789.8	88.3	190.2	-101.9	-2.5
1993	15685.4	80.5	185.6	-105.1	18.1
1994	15598.4	78.7	181.4	-102.7	35.7
1995	15531.4	83.8	178.1	-94.2	38.3
1996	15475.5	93.3	174.5	-81.2	34.4
1997	15428.7	100.3	167.5	-67.3	8.0
1998	15369.4	102.9	164.3	-61.4	-18.3
1999	15289.7	106.7	161.3	-54.6	-17.8

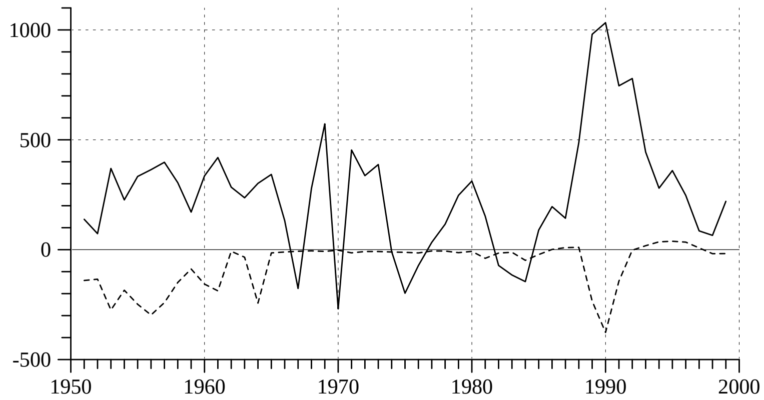


Abb. 5.2-5 Saldo der Migration (Zuwanderung minus Abwanderung) im Gebiet der früheren BRD (durchgezogene Linie) und der früheren DDR (gestrichelte Linie). Ordinate in 1000. Die Daten entsprechen den Angaben in den Tabellen 5.2-5 und 5.2-6.

ihre Grenzen schloss, kam die Zuwanderung großenteils aus deren Gebiet. Eine weitere Wanderungsbewegung von Ost- nach Westdeutschland fand in den Jahren unmittelbar nach der Wiedervereinigung statt. (Man beachte, dass die Wanderungen zwischen Ost- und Westdeutschland auch noch nach der Wiedervereinigung in den Migrationssalden enthalten sind.)

5. Gliederungen nach Geschlecht und Alter. Demographische Charakterisierungen von Gesellschaften gehen fast immer von Gliederungen der Bevölkerung nach dem Geschlecht und dem Alter aus. Der begriffliche Ansatz beruht auf der Konzeption statistischer Variablen

$$(S_t, A_t) : \Omega_t \longrightarrow \tilde{\mathcal{S}} \times \tilde{\mathcal{A}}$$

Die Referenzmenge Ω_t repräsentiert die Bevölkerung im Jahr (allgemein: in der Zeitstelle) t. S_t mit dem Merkmalsraum $\tilde{\mathcal{S}} := \{0, 1\}$ erfasst das Geschlecht;[25] und A_t mit dem Merkmalsraum $\tilde{\mathcal{A}} := \{0, 1, 2, 3, \ldots\}$ erfasst das Alter. Ist also zum Beispiel $(S_t, A_t)(\omega) = (1, 40)$, bedeutet dies, dass ω der Name einer 40jährigen Frau ist.

Hier muss natürlich darauf geachtet werden, wie das Alter definiert ist. Verwendet man das demographische Alter (Abschnitt 5.1, § 6), gibt es einfache Zusammenhänge mit den bisherigen Notationen, beispielsweise $n_t = |\Omega_t|$ und $n_{t,\tau} = |\{\omega \in \Omega_t \,|\, A_t(\omega) = \tau\}|$. Das demographische Alter stimmt jedoch nur dann mit dem gewöhnlichen Alter (in vollendeten Jahren) überein, wenn man sich auf das Jahresende bezieht. Werden jahresdurchschnittliche Bevölkerungszahlen nach dem Alter gegliedert, wie in

[25] Als Konvention verwenden wir in diesem Text: 0 = männlich, 1 = weiblich.

Tabelle 5.2-7 Anzahl Männer ($\bar{n}^m_\tau$) und Frauen ($\bar{n}^f_\tau$) im Alter τ im Jahresdurchschnitt 1999 in Deutschland; 95* umfasst alle Altersjahre $\tau \geq 95$. Quelle: Segment 685 der STATIS Datenbank des Statistischen Bundesamts.

τ	$\bar{n}^m_\tau$	$\bar{n}^f_\tau$	τ	$\bar{n}^m_\tau$	$\bar{n}^f_\tau$	τ	$\bar{n}^m_\tau$	$\bar{n}^f_\tau$
0	399633	378251	32	732651	686810	64	472336	511058
1	410782	389437	33	748125	697478	65	409837	445060
2	413836	391872	34	758218	705906	66	361225	399667
3	403107	381972	35	761731	711736	67	359314	405178
4	398813	377741	36	746559	700436	68	365748	423645
5	409761	387869	37	727433	687469	69	362427	431229
6	422128	400905	38	711239	675052	70	347956	425406
7	435168	413392	39	690941	656171	71	319299	416670
8	466447	442107	40	663561	630073	72	283797	410392
9	485976	461036	41	641087	608883	73	258704	416006
10	490076	464833	42	627205	597762	74	228253	407353
11	492537	465361	43	610005	584657	75	202335	390714
12	482637	456705	44	594357	575662	76	196282	387353
13	469636	445784	45	578806	566937	77	193182	395011
14	461216	437317	46	570259	561714	78	180053	390951
15	463159	438441	47	566062	558353	79	144258	328396
16	472798	446710	48	563942	556038	80	96563	221190
17	479914	453245	49	558611	548502	81	68960	161774
18	481413	457174	50	529900	517217	82	65068	156443
19	473334	451301	51	494176	482985	83	70399	177521
20	462189	441633	52	450618	442241	84	79463	213228
21	460967	441747	53	397251	392979	85	78125	219593
22	460272	441953	54	434885	432973	86	67852	197433
23	456346	436715	55	502194	498201	87	55469	168730
24	458828	438622	56	500889	496426	88	44330	141995
25	469223	449325	57	545671	544075	89	35770	120032
26	500605	477112	58	612495	615985	90	27775	97204
27	557051	528212	59	621650	630333	91	20817	75582
28	603444	568797	60	593275	606556	92	15772	58386
29	644108	604417	61	551237	569910	93	11724	42726
30	686013	642576	62	522738	546881	94	8658	30973
31	712393	668147	63	502833	534670	95*	21807	69931

Tabelle 5.2-7 für das Jahr 1999, handelt es sich nicht um das demographische Alter.[26] Sie können aber als Näherungswerte auch für eine Gliederung nach dem demographischen Alter verwendet werden.[27]

Bildet man in Tabelle 5.2-7 die Summen der Einträge für Männer bzw. Frauen, findet man

$$\bar{n}^m_{1999} = 40047972 \quad \text{und} \quad \bar{n}^f_{1999} = 42038610$$

[26] Vgl. zur Berechnung die Ausführungen in Anmerkung 5.

[27] Hinter den scheinbar exakten Angaben in der Tabelle verbirgt sich natürlich nur eine rechnerische, keine empirische Genauigkeit.

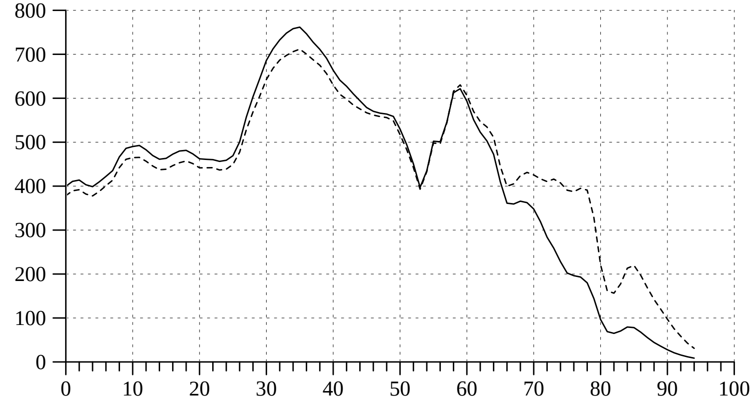


Abb. 5.2-6 Graphische Darstellung der absoluten Häufigkeiten (in 1000) aus Tabelle 5.2-7 bis zum Alter von 94 Jahren für Männer (durchgezogene Linie) und Frauen (gestrichelte Linie).

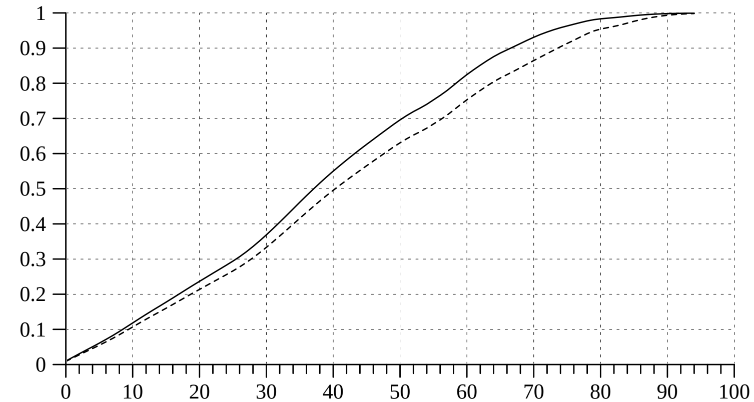


Abb. 5.2-7 Darstellung der Altersverteilungen für Männer (durchgezogen) und Frauen (gestrichelt) im Jahr 1999 durch Verteilungsfunktionen. Daten aus Tabelle 5.2-7.

so dass sich auch sogleich relative Häufigkeiten ausrechnen lassen. Mit der Einschränkung, dass die Altersjahre ab 95 in einer nach oben offenen Altersklasse zusammengefasst sind, liefert die Tabelle 5.2-7 eine vollständige Darstellung der Verteilung der Variablen (S_t, A_t) für $t = 1999$.

Allerdings erkennt man auch, dass diese tabellarische Darstellung zwar als Ausgangspunkt für weitere Rechnungen verwendet werden kann, aber nicht ohne weiteres eine reflektierbare Anschauung der Verteilungen erlaubt. Diesem Zweck dienen graphische Darstellungen. Zur Illustration zeigt Abb. 5.2-6 eine graphische Darstellung der absoluten Häufigkeiten aus der Tabelle. Offenbar erlaubt diese Darstellung anschauliche Vergleiche

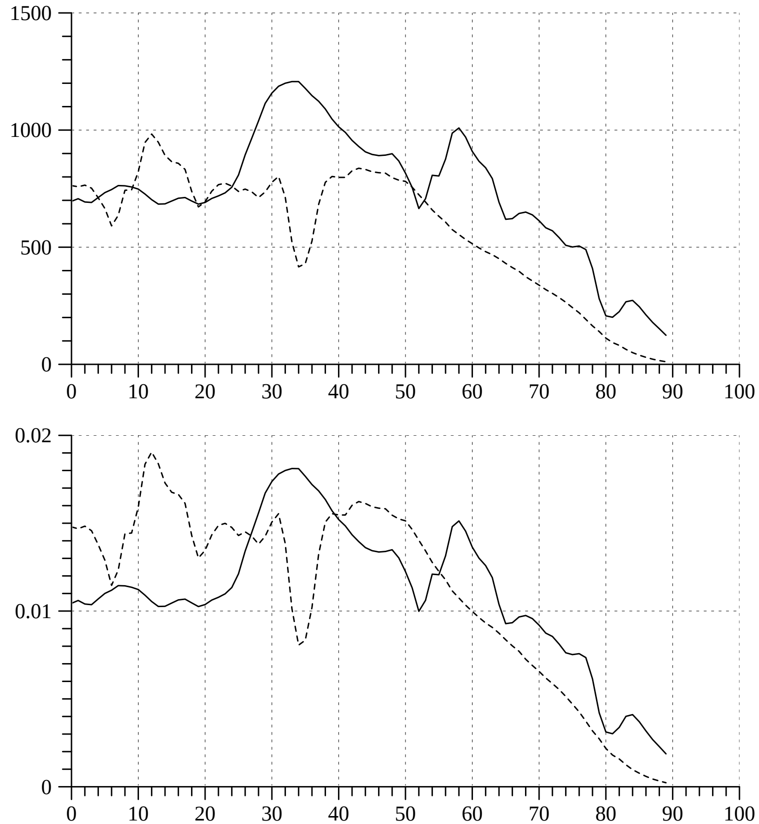


Abb. 5.2-8 Altersverteilungen im Gebiet der früheren BRD im Jahr 1998 (durchgezogene Linie) und im Jahr 1952 (gestrichelte Linie); oben in absoluten Häufigkeiten (in 1000), unten in relativen Häufigkeiten.

der Altersverteilungen von Männern und Frauen. Außerdem sind jeweils deutliche Unregelmäßigkeiten erkennbar (z.B. der Einschnitt im Altersbereich von 50 bis 55 Jahren), die sich mit Annahmen über einen historischen Prozess verknüpfen lassen, durch den die Verteilungen entstanden sind.

Informativ ist auch eine Darstellung der Altersverteilungen durch Verteilungsfunktionen. Abbildung 5.2-7 zeigt dies für die Daten aus Tabelle 5.2-7. Offenbar kann man unmittelbar für jedes Alter ablesen, wie groß die Anteile der jeweils jüngeren bzw. älteren Bevölkerung sind.

6. Veränderungen in der Altersstruktur. Um Einsichten in die Veränderung der Altersstruktur zu gewinnen, verwenden wir einen Datensatz des Statistischen Bundesamts, der sich auf das Gebiet der früheren BRD für

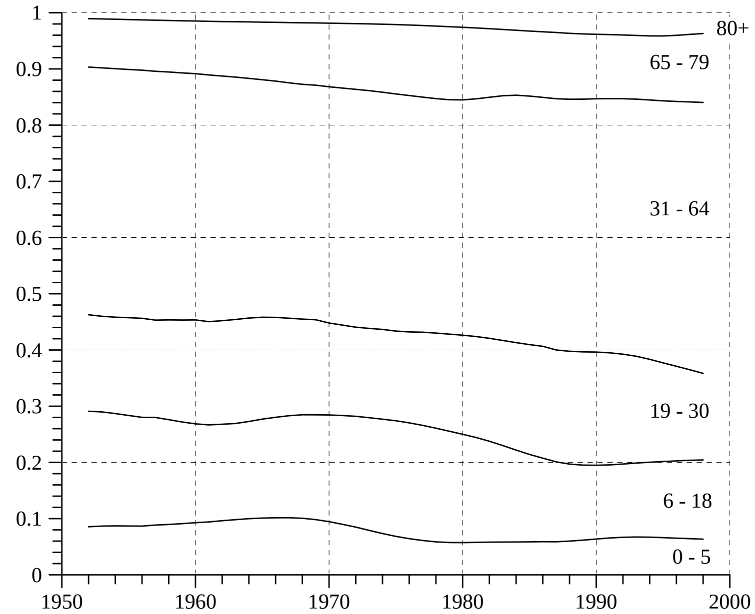


Abb. 5.2-9 Entwicklung der Altersverteilung im Gebiet der früheren BRD. Die Abbildung zeigt relative Häufigkeiten für sechs Altersklassen, deren Abgrenzung auf der rechten Seite angegeben wird.

den Zeitraum 1952 bis 1998 bezieht.[28] Fasst man die Daten für Männer und Frauen zusammen, sieht der Datensatz folgendermaßen aus:

τ	1952	$\cdots$	1998
0	$\bar{n}_{1952,0}$	$\cdots$	$\bar{n}_{1998,0}$
1	$\bar{n}_{1952,1}$	$\cdots$	$\bar{n}_{1998,1}$
2	$\bar{n}_{1952,2}$	$\cdots$	$\bar{n}_{1998,2}$
$\vdots$	$\vdots$		$\vdots$
90^*	$n_{1952,90^*}$	$\cdots$	$n_{1998,90^*}$

Für jedes Jahr kann also eine Altersverteilung bis zu einer nach oben offenen Altersklasse 90^* ermittelt werden.

In Abbildung 5.2-8 werden die Altersverteilungen der Jahre 1952 und 1998 verglichen. Sowohl die absoluten als auch die relativen Häufigkeitsverteilungen zeigen, dass eine erhebliche Umschichtung von jüngeren zu älteren Personen stattgefunden hat. Bemerkenswert sind auch die irregulären

[28] Segment 36 der STATIS-Datenbank nach einem Update im Juni 2000. Wie in Tabelle 5.2-7 handelt es sich um nach dem Alter und Geschlecht gegliederte jahresdurchschnittliche Bevölkerungszahlen.

Verteilungsformen, die als Folge der demographischen Entwicklung seit dem 1. Weltkrieg entstanden sind.

Zu überlegen bleibt, wie sich der Veränderungsprozess seit 1952 charakterisieren lässt. Eine einfache Möglichkeit besteht darin, das Durchschnittsalter zu berechnen. Man findet, dass es sich kontinuierlich von 38 Jahren in 1952 auf 44 Jahre in 1998 erhöht hat.[29] Um auch einen Einblick in Veränderungen der Altersverteilung zu gewinnen, können relative Häufigkeiten für Altersklassen betrachtet werden. Wie Abbildung 5.2-9 zeigt, können Veränderungen solcher Häufigkeiten auch graphisch dargestellt werden.

[29]Für diese Berechnung haben wir für die nach oben offene Altersklasse 90* ein Alter von 90 Jahren angenommen; natürlich sind auch andere Annahmen möglich, die dann zu einem höheren Durchschnittsalter führen würden.

Kapitel 6

Lebensdauern und Sterbetafeln

Demographische Prozesse werden in erster Linie durch Geburten und Todesfälle, bei räumlicher Eingrenzung außerdem durch Migrationsvorgänge bestimmt. In diesem Kapitel beschäftigen wir uns mit Unterschieden und historischen Veränderungen in den Lebensdauern. Wir beginnen mit einem Abschnitt, der sich in etwas allgemeinerer Weise mit Verweildauervariablen und Übergangsraten beschäftigt. Im zweiten Abschnitt werden Kohorten- und Perioden-Sterbetafeln erklärt, und es wird eine Perioden-Sterbetafel für die BRD 1999 konstruiert. Schließlich werden im dritten Abschnitt anhand von Sterbetafeln der amtlichen Statistik langfristige Veränderungen der Lebensdauerverteilungen in Deutschland seit Ende des 19. Jahrhunderts besprochen.

6.1 Verweildauern und Übergangsraten

In diesem Abschnitt werden Verweildauervariablen und einige an diesen Begriff anschließende Definitionen besprochen. Dabei orientieren wir uns an den Ausführungen bei Rohwer und Pötter (2001, Teil III), wo ebenfalls von einer diskreten Zeitachse ausgegangen wird.[1]. Methoden zur Schätzung

[1]Entsprechende Begriffsbildungen und Modellansätze für eine stetige Zeitachse werden bei Blossfeld und Rohwer (2002) behandelt.

von Verweildauerverteilungen mit unvollständigen Daten werden erst in späteren Kapiteln besprochen, wenn sie benötigt werden (in Abschnitt 7.3). Die folgenden Begriffsbildungen knüpfen an die Ausführungen in den Abschnitten 2.1 und 4.2 an.

1. Summarische und sequentielle Prozessdarstellungen. Prozesse können auf unterschiedliche Weisen dargestellt werden. Teilweise folgt dies bereits aus der Konzeptualisierung der Prozesse, etwa als Handlungsprozesse, bei denen oft eine narrative Darstellung sinnvoll ist, oder als statistische Prozesse, zu deren Darstellung statistische Begriffe verwendet werden. Weitgehend unabhängig von unterschiedlichen Prozesskonzeptionen kann man summarische und sequentielle Prozessdarstellungen unterscheiden:

– Wir sprechen von einer *summarischen* Prozessdarstellung, wenn sich Aussagen unmittelbar auf den gesamten Prozess beziehen. Bei Lebensverläufen handelt es sich z.B. um Aussagen über die Lebensdauer oder den erreichten Schulabschluß oder die Anzahl von Eheschließungen.

– Dagegen sprechen wir von einer *sequentiellen* Prozessdarstellung, wenn der Prozess als eine zeitliche Abfolge von Teilprozessen oder Ereignissen dargestellt wird. Typische Beispiele liefern narrative Darstellungen von Handlungsprozessen; bestimmte Varianten sequentieller Darstellungen sind jedoch auch bei statistischen Prozessen möglich.

2. Episoden und Verweildauervariablen. Die einfachste summarische Darstellung bezieht sich auf die Prozessdauer. Von besonderem Interesse ist dies beim Vergleich mehrerer Realisationen eines wiederholbaren Prozesses. Bezieht man sich auf schematische Lebensverläufe, werden Zeitdauern von *Episoden* verglichen, die als Zeitspannen zwischen zwei Ereignissen im Sinne von Zustandswechseln definiert sind; zum Beispiel Ehe-Episoden, die mit einer Eheschließung beginnen und mit dem Tod eines Ehepartners oder einer Scheidung enden. Zur Definition kann man auch von dem Prozessschema $X_t : \Omega \longrightarrow \tilde{\mathcal{X}}$ ausgehen.[2] Für jedes Objekt $\omega \in \Omega$ gibt es dann einen bestimmten individuellen Prozess, der aus einer zeitlichen Folge von Zuständen $X_t(\omega)$ besteht und der sich somit auch als eine Folge von Episoden auffassen lässt, während der sich der Zustand nicht verändert. (Zur Illustration kann das Biographieschema in Abb. 4.2-2 verwendet werden.)

Zur vollständigen Charakterisierung einer Episode innerhalb eines Prozesses sind im allgemeinen vier Angaben erforderlich: Eine Angabe des Zustands, der während der Episode eingenommen wird (auch Anfangs- oder Ausgangszustand der Episode genannt); eine Angabe der Zeitstelle, in der dieser Zustand zum ersten Mal eingenommen wird; eine Angabe der Zeitstelle, in der zum ersten Mal ein neuer Zustand eingenommen wird; und eine Angabe dieses neuen Zustands, durch den die Episode beendet wird. Wenn man sich nicht für die genaue zeitliche Lagerung von Episoden innerhalb eines sie umfassenden Prozesses interessiert, sondern nur für ihre zeitliche Dauer und die Art ihrer Beendigung, können *Verweildauervariablen* verwendet werden, die sich allgemein durch folgendes Schema definieren lassen:

$$(T, D) : \Omega \longrightarrow \mathcal{T}_0 \times \tilde{\mathcal{D}} \tag{6.1}$$

In dieser zweidimensionalen statistischen Variablen bezieht sich die Komponente T auf die Episodendauer, zu deren Erfassung eine Prozesszeitachse $\mathcal{T}_0 := \{0, 1, 2, 3, \ldots\}$ verwendet wird, deren Elemente sich auf Zeiteinheiten beziehen (z.B. Tage, Monate oder Jahre).[3] Die zweite Komponente D erfasst den Folgezustand (bzw. das Ereignis), durch dessen Eintreten die Episode beendet wird. Die möglichen Folgezustände werden durch den Zustandsraum $\tilde{\mathcal{D}}$ erfasst.

Als Beispiel kann man sich vorstellen, dass es sich bei Ω um eine Menge verheirateter Personen handelt und für jede Person $\omega \in \Omega$ durch $T(\omega)$ die Dauer der Ehe und durch $D(\omega)$ erfasst wird, ob die Ehe durch den Tod eines Ehepartners oder eine Scheidung beendet wird.

3. Lebensdauern gestorbener Personen. Es ist offensichtlich, dass Verweildauervariablen zunächst nur summarische Darstellungen liefern: man erhält Informationen nicht über den Ablauf, sondern nur über die Dauer von Episoden und über die Art des Ereignisses, durch das sie beendet worden sind. Natürlich kann man sich auch auf eine Erfassung der Episodendauer beschränken, ohne Arten von Ereignissen, durch die Episoden beendet werden können, zu unterscheiden. Zur Definition von Verweildauervariablen genügt dann das einfache Schema $T : \Omega \longrightarrow \mathcal{T}_0$. Jedem Objekt $\omega \in \Omega$ wird dann nur ein Wert $T(\omega)$ zugeordnet, der angibt, wie lange bei diesem Objekt die Episode gedauert hat.

Als Beispiel beziehen wir uns auf Lebensdauern von Menschen in Deutschland. Ω ist in diesem Fall eine Gesamtheit von Menschen, und die Verweildauervariable T ordnet jeder Person ω ihre Lebensdauer $T(\omega)$ zu. Allerdings ist es in diesem Fall schwierig, von einem Kohortenansatz auszugehen, etwa von einer Gesamtheit von Menschen, die in einem bestimmten Jahr in Deutschland geboren wurden.[4] Denn bezieht man sich auf Geburtsjahre innerhalb der letzten 100 Jahre, würde es bei vielen oder sogar den meisten Mitgliedern solcher Geburtskohorten noch keine bestimmten Werte für ihre Lebensdauer geben. Außerdem bezieht sich die amtliche Statistik nicht auf Geburtskohorten, sondern auf die jeweils während eines Jahres gestorbenen Personen. Wir folgen deshalb zunächst diesem Ansatz und definieren Ω, um uns auf ein konkretes Beispiel beziehen zu können, als die Menge der Personen, die 1999 in Deutschland gestorben sind. Dann gibt es für jede Person $\omega \in \Omega$ eine Lebensdauer $T(\omega)$.

[2] Man vgl. Abschnitt 4.2 (§ 5), wo dieses Schema eingeführt wird.

[3] Jede zeitliche Dauer wird also durch eine Anzahl *vollendeter* Zeitstellen erfasst; inso-

Tabelle 6.1-1 Anzahl Männer ($\bar{d}_\tau^m$) und Frauen ($\bar{d}_\tau^f$), die 1999 in Deutschland im Alter von τ Jahren gestorben sind; 95* umfasst alle Altersjahre $\tau \geq 95$. Quelle: Segmente 1124–26 der STATIS Datenbank des Statistischen Bundesamts. Auch berechenbar aus den Angaben der Fachserie 1, Reihe 1, 1999: 234ff.

τ	$\bar{d}_\tau^m$	$\bar{d}_\tau^f$	τ	$\bar{d}_\tau^m$	$\bar{d}_\tau^f$	τ	$\bar{d}_\tau^m$	$\bar{d}_\tau^f$
0	1979	1517	32	666	308	64	8835	4451
1	173	137	33	740	345	65	8207	4173
2	126	83	34	809	365	66	7990	4074
3	90	63	35	872	431	67	8984	4766
4	79	54	36	1066	433	68	10288	5572
5	52	41	37	1076	539	69	11122	6279
6	70	43	38	1126	559	70	11690	6858
7	66	52	39	1300	629	71	11439	7489
8	76	44	40	1334	667	72	10934	8366
9	56	54	41	1418	717	73	11017	9404
10	67	42	42	1572	786	74	10847	10147
11	71	42	43	1719	858	75	10404	11172
12	76	54	44	1788	892	76	11027	12516
13	79	50	45	1983	980	77	12292	14616
14	114	65	46	2055	1097	78	12610	16287
15	145	84	47	2223	1136	79	12452	17041
16	194	108	48	2422	1222	80	7484	10678
17	314	148	49	2555	1320	81	6478	9951
18	487	159	50	2745	1354	82	6763	11125
19	454	168	51	2691	1414	83	7840	13397
20	435	137	52	2873	1421	84	10704	20226
21	468	156	53	2432	1287	85	11001	22124
22	412	125	54	3182	1618	86	10565	22470
23	401	118	55	4012	2023	87	9517	21892
24	441	135	56	4244	2010	88	8191	20816
25	377	155	57	5189	2493	89	7471	19758
26	443	143	58	5906	2831	90	6466	18053
27	473	169	59	6970	3345	91	5174	15700
28	533	200	60	7303	3514	92	4072	13384
29	528	214	61	7359	3497	93	3170	10696
30	614	235	62	7819	3692	94	2446	8332
31	627	263	63	8422	3985	95*	4871	20949

Tabelle 6.1-1 zeigt die vom Statistischen Bundesamt über diese Lebensdauern verfügbaren Daten in Form einer nach dem Geschlecht differenzierten Häufigkeitsverteilung. Es handelt sich um absolute Häufigkeiten; z.B. erkennt man, dass 1999 im Alter von 70 Jahren 11690 Männer und 6858 Frauen gestorben sind. Das Alter ist in vollendeten Lebensjahren zum Zeitpunkt des Todes angegeben; die Angabe 95* umfasst alle Lebensjahre, die größer oder gleich 95 sind.[5]

fern kann die Verweildauervariable auch den Wert 0 annehmen.

[4] Man vgl. hierzu die Ausführungen in Abschnitt 4.2, § 6.

[5] Das Statistische Bundesamt publiziert auch Tabellen, in denen zusätzlich das Geburtsjahr angegeben wird; für 1999 in der Fachserie 1, Reihe 1, 1999: 232ff. Daraus

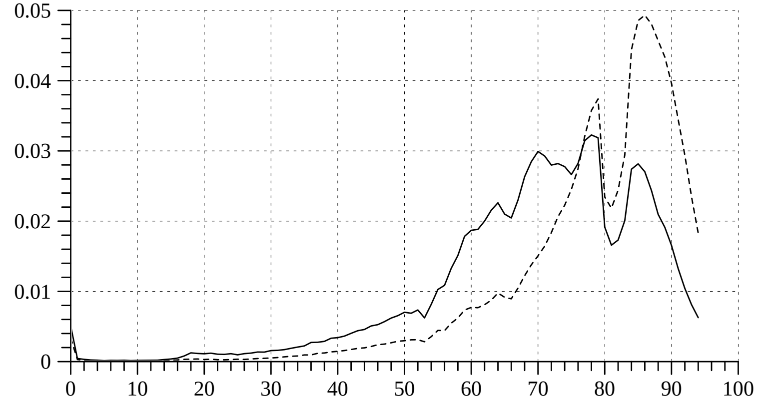


Abb. 6.1-1 Darstellung der Verteilung der Lebensdauern der 1999 in Deutschland gestorbenen Männer (durchgezogene Linie) und Frauen (gestrichelte Linie) durch Häufigkeitsfunktionen. Berechnet aus den Daten in Tabelle 6.1-1.

Allerdings gewinnt man aus der Tabelle nicht ohne weiteres ein anschauliches Bild der Verteilung der Lebensdauern. Dafür eignen sich Schaubilder, in denen z.B. Häufigkeitsverteilungen dargestellt werden können. Unter Verwendung der im Abschnitt 2.1 eingeführten Bezeichnungen handelt es sich bei der *Häufigkeitsfunktion* einer Verweildauervariablen T um eine Funktion P[T],[6] die jedem möglichen Wert dieser Variablen, also in unserem Beispiel jedem Alter t, eine relative Häufigkeit

$$\mathrm{P}(t) = \frac{1}{|\Omega|}\,|\{\omega \in \Omega \,|\, T(\omega) = t\}|$$

zuordnet.[7] Da nach den Angaben von Tabelle 6.1-1 insgesamt 390742 Männer und 455588 Frauen gestorben sind, findet man z.B. für $t = 70$ die relativen Häufigkeiten 0.03 bei Männern und 0.015 bei Frauen. Abbildung 6.1-1 zeigt die Häufigkeitsverteilungen für Männer und Frauen. Offenbar handelt es sich um Altersverteilungen, die in Form einer summarischen Darstellung zeigen, wie alt die 1999 in Deutschland gestorbenen Männer bzw. Frauen geworden sind.

kann man z.B. ersehen, dass von den 11690 Männern, die 1999 im Alter von 70 Jahren gestoren sind, 5810 in 1929 und 5880 in 1928 geboren wurden.

[6] P, oder ausführlich P[T], ist der Funktionsname. Erst durch Anhängen eines Arguments in runden Klammern, zum Beispiel P(t), erhält man einen bestimmten Wert der Funktion. Anstelle von P(t) findet man auch die Schreibweise P($T = t$), wenn auf eine Variable T Bezug genommen wird.

[7] In diesem Text wird zum Verweis auf Altersangaben wahlweise das Symbol t oder das Symbol τ verwendet.

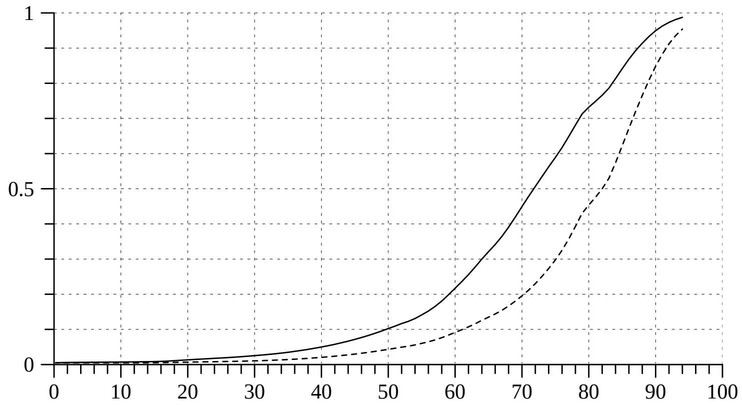


Abb. 6.1-2 Darstellung der Verteilung der Lebensdauern der 1999 in Deutschland gestorbenen Männer (durchgezogene Linie) und Frauen (gestrichelte Linie) durch Verteilungsfunktionen. Berechnet aus den Daten in Tabelle 6.1-1.

Weiterhin können für graphische Darstellungen auch Verteilungsfunktionen verwendet werden. Die *Verteilungsfunktion* einer Verweildauervariablen T ist eine Funktion F, die jedem möglichen Wert von t eine kumulierte Häufigkeit

$$F(t) := \frac{1}{|\Omega|}\,|\{\omega \in \Omega \,|\, T(\omega) \leq t\}| \;=\; \sum_{\tau \leq t} \mathrm{P}(\tau)$$

zuordnet. In unserem Anwendungsfall ist also $F(t)$ der Anteil der bis zu einem Alter von t Jahren gestorbenen Männer bzw. Frauen. Aus der Darstellung in Abbildung 6.1-2 erkennt man z.B., dass etwa 50 % der Männer älter als 72, etwa 50 % der Frauen älter als 82 Jahre geworden sind.

4. Survivor- und Ratenfunktionen. Offenbar erhält man in unserem Beispiel eine summarische Darstellung von Lebensdauern, d.h. eines Aspekts der Lebensverläufe der 1999 in Deutschland gestorbenen Männer bzw. Frauen. Allerdings ist die Definition der in diesem Beispiel verwendeten Gesamtheit problematisch, da sie Personen umfasst, die in ganz unterschiedlichen historischen Perioden gelebt haben. Infolgedessen vermischen sich in den Lebensdauerverteilungen, die im vorangegangenen Paragraphen betrachtet wurden, altersspezifische Mortalitätsbedingungen mit historischen Veränderungen in den Geburtenhäufigkeiten. So sind z.B. die niedrigen Häufigkeitswerte im Altersbereich von 80 bis 83 Jahren eine Folge der niedrigen Geburtenraten infolge des 1. Weltkriegs. Wir besprechen deshalb im nächsten Abschnitt zwei Möglichkeiten, um zu besseren Darstellungen von Mortalitätsbedingungen zu gelangen.

Zuvor definieren wir zwei Begriffe, die zur Darstellung und Analyse von

Verweildauervariablen oft nützlich sind. Dafür beziehen wir uns auf eine beliebige Verweildauervariable T. Als *Survivorfunktion* (dieser Verweildauervariablen) wird eine Funktion G bezeichnet, die jedem möglichen Wert t der Variablen eine Häufigkeit

$$G(t) := \frac{1}{|\Omega|}\,|\{\omega \in \Omega \,|\, T(\omega) \geq t\}| \;=\; \sum_{\tau \geq t} \mathrm{P}(\tau)$$

zuordnet, also den Anteil derjenigen Personen oder Objekte, bei denen die Verweildauer mindestens den Wert t annimmt. Bezieht sich T zum Beispiel auf die Lebensdauer, wäre $G(t)$ der Anteil derjenigen Personen, die mindestens t Jahre alt geworden sind. (Dieser demographische Anwendungskontext motiviert auch den Ausdruck 'Survivorfunktion'.) Offenbar kann man die Survivorfunktion als eine Art Komplement zur Verteilungsfunktion auffassen, denn es gilt: $G(t) = 1 - F(t) + \mathrm{P}(t)$.[8]

Weiterhin wird als *Ratenfunktion* einer Verweildauervariablen T eine Funktion r bezeichnet, die jedem möglichen Wert t der Variablen eine bedingte Häufigkeit

$$r(t) := \frac{\mathrm{P}(t)}{G(t)} = \frac{|\{\omega \in \Omega \,|\, T(\omega) = t\}|}{|\{\omega \in \Omega \,|\, T(\omega) \geq t\}|}$$

zuordnet.[9] Es handelt sich um eine bedingte Häufigkeit, die auch in der Form $r(t) = \mathrm{P}[T|T \geq t](t)$ geschrieben werden kann. Erfasst zum Beispiel T die Lebensdauer in Jahren, wäre $r(80)$ die Häufigkeit einer Lebensdauer von 80 Jahren bei denjenigen Personen, die mindestens 80 Jahre alt geworden sind.

Wichtig ist, dass Charakterisierungen der Verteilung einer Verweildauervariablen durch Häufigkeits-, Verteilungs-, Survivor- und Ratenfunktionen äquivalent sind, d.h. aus jeweils einer dieser vier Funktionen können alle anderen formal abgeleitet werden. Insbesondere kann mit folgender Formel die Survivorfunktion aus einer Ratenfunktion abgeleitet werden:

$$G(t) = \prod_{j=0}^{t-1} (1 - r(j)) \tag{6.2}$$

Sie bildet eine Grundlage zahlreicher statistischer Methoden und wird insbesondere zur Konstruktion von Sterbetafeln verwendet.

5. Übergangsraten bei mehreren Folgezuständen. Ratenfunktionen können auch für allgemeine Verweildauervariablen (T, D) definiert werden, bei denen nicht nur eine Verweildauer T, sondern außerdem ein Folgezustand

[8] Es sei erwähnt, dass man in der Literatur auch Definitionen der Survivorfunktion durch $1 - F$ findet; bei diskreten Zeitachsen ist jedoch die oben angegebene Definition vorteilhafter.

[9] Erläuterungen zum Begriff einer bedingten Verteilung findet man im Abschnitt 15.1.

D erfasst wird (vgl. § 2). Dann gibt es für jeden möglichen Folgezustand $d \in \tilde{\mathcal{D}}$ eine *zustandsspezifische Ratenfunktion*

$$r_d(t) := \frac{|\{\omega \in \Omega \,|\, T(\omega) = t, D(\omega) = d\}|}{|\{\omega \in \Omega \,|\, T(\omega) \geq t\}|}$$

$r_d(t)$ wird auch als *Übergangsrate in den Folgezustand* d (in der Zeitstelle t) bezeichnet. Im Zähler steht die Anzahl der Objekte, bei denen in der Zeitstelle t ein Übergang in den Folgezustand d stattfindet; und im Nenner steht die Anzahl der Objekte, die sich in der Zeitstelle t noch im Ausgangszustand befinden und bei denen infolgedessen in dieser Zeitstelle ein Übergang in den Folgezustand d erfolgen könnte. Offenbar sind diese Funktionen additiv:

$$r(t) = \sum_{d \in \tilde{\mathcal{D}}} r_d(t)$$

so dass man aus der Addition der zustandsspezifischen Übergangsraten die Rate für das Verlassen des Ausgangszustands erhält.

6.2 Kohorten- und Perioden-Sterbetafeln

1. Notationen für Geburtskohorten. Wir beziehen uns auf den in Abschnitt 4.2 besprochenen Kohortenansatz und verwenden zur Notation:

$\mathcal{C}_t$ bezeichnet eine Menge von Menschen, die im Kalenderjahr t geboren wurden; dabei wird irgendeine bestimmte räumliche Abgrenzung vorausgesetzt. Solche Mengen werden auch *Geburtskohorten* genannt (wenn in diesem Text ohne Zusatz von *Kohorten* gesprochen wird, sind stets Geburtskohorten gemeint).

Weiterhin werden Indizes m und f verwendet, um nach dem Geschlecht zu unterscheiden. So ist zum Beispiel $\mathcal{C}_t^f$ eine Menge von Frauen, die im Jahr t geboren wurden.

2. Altersspezifische Sterbeziffern. Zur Berechnung von Sterbetafeln werden altersspezifische Sterbeziffern (Mortalitätsraten) verwendet. Da sich die Sterblichkeit von Männern und Frauen, besonders in höheren Lebensjahren, erheblich unterscheidet, wird bei der Definition auch nach dem Geschlecht differenziert. Wir verwenden meistens die Definitionen:

$$\bar{\delta}^m_{t,\tau} := \bar{d}^m_{t,\tau} / \bar{n}^m_{t,\tau} \text{ für Männer und } \bar{\delta}^f_{t,\tau} := \bar{d}^f_{t,\tau} / \bar{n}^f_{t,\tau} \text{ für Frauen}$$

Im Zähler steht die Anzahl der in der Zeitstelle t im Alter τ gestorbenen Männer bzw. Frauen, im Nenner steht die durchschnittliche Anzahl der Männer bzw. Frauen des Alters τ in der Zeitstelle t. Offenbar können mit den in Abschnitt 5.1 besprochenen Begriffsbildungen auch noch andere Definitionen vorgenommen werden. In der hier angegebenen Definition

können altersspezifische Sterbeziffern jedoch besonders leicht aus publizierten Datenquellen, für das Jahr 1999 aus den Tabellen 6.1-1 und 5.2-7, berechnet werden. Zum Beispiel findet man:

$$\bar{\delta}^m_{1999,70} = \frac{11690}{347956} = 0.0336 \quad \text{und} \quad \bar{\delta}^f_{1999,70} = \frac{6858}{425406} = 0.0161$$

In diesem Alter ist also die Sterblichkeit der Frauen nur etwa halb so groß wie die der Männer.

3. Kohorten- und Perioden-Sterbetafeln. Man unterscheidet Kohorten- und Perioden-Sterbetafeln.[10] Eine *Kohorten-Sterbetafel* bezieht sich auf die Lebensdauerverteilung einer Kohorte. Ausgangspunkt ist eine für eine Kohorte definierte Variable $T_t : \mathcal{C}_t \longrightarrow \mathcal{T}_0 = \{0, 1, 2, 3, \ldots\}$, die die Lebensdauern der Mitglieder von $\mathcal{C}_t$ erfasst. (Wir nehmen an, dass diese Lebensdauern in vollendeten Lebensjahren erfasst werden.) Eine Kohorten-Sterbetafel besteht dann in einer Darstellung der Suvivorfunktion (und ggf. auch der Ratenfunktion) der Variablen T_t.

Offenbar ist eine Kohorten-Sterbetafel erst dann empirisch bestimmt, wenn alle Mitglieder der Kohorte gestorben sind. Vollständige Kohorten-Sterbetafeln können also nur für Kohorten berechnet werden, deren Mitglieder vor mindestens 100 Jahren geboren wurden.[11] Meistens, insbesondere von der amtlichen Statistik, werden deshalb keine Kohorten-, sondern Perioden-Sterbetafeln berechnet, die von den Sterbeziffern eines Kalenderjahrs oder einer kurzen Kalenderzeitperiode ausgehen. Zur Erläuterung beziehen wir uns auf ein Kalenderjahr t, für das die altersspezifischen Sterbeziffern (Mortalitätsraten) $\delta_{t,\tau}$ ermittelt worden sind (dabei wird auf Männer oder Frauen Bezug genommen). Dann kann in formaler Analogie zur Formel (6.2) durch

$$G_t(\tau) := \prod_{j=0}^{\tau-1} (1 - \delta_{t,j}) \qquad (6.3)$$

eine Funktion G_t definiert werden, die sich als Survivorfunktion der Lebensdauerverteilung einer fiktiven Gesamtheit interpretieren lässt, deren Absterbeprozess den Mortalitätsraten $\delta_{t,0}, \delta_{t,1}, \delta_{t,2}, \delta_{t,3}, \ldots$ folgt. Eine Tabellierung dieser Funktion (und ggf. der Mortalitätsraten, aus denen sie berechnet worden ist) wird als *Perioden-Sterbetafel* (für die Periode t) bezeichnet.[12] Sie bezieht sich nicht auf den realen Absterbeprozess irgend-

[10] Wenn ohne Zusatz von *Sterbetafeln* gesprochen wird, sind meistens (auch in diesem Text) Perioden-Sterbetafeln gemeint.

[11] Da auch meistens die erforderlichen Daten nicht oder nur unvollständig vorhanden sind, sind in der Literatur verschiedene Möglichkeiten diskutiert worden, um Kohorten-Sterbetafeln wenigstens näherungsweise zu schätzen. Man vgl. R. H. Dinkel (1984, 1992).

[12] Dabei ist es üblich, von einer fiktiven Gesamtheit von 100000 Personen auszugehen, so dass Rundungsprobleme vernachlässigt werden können und es ausreicht, ganze Zahlen zu tabellieren.

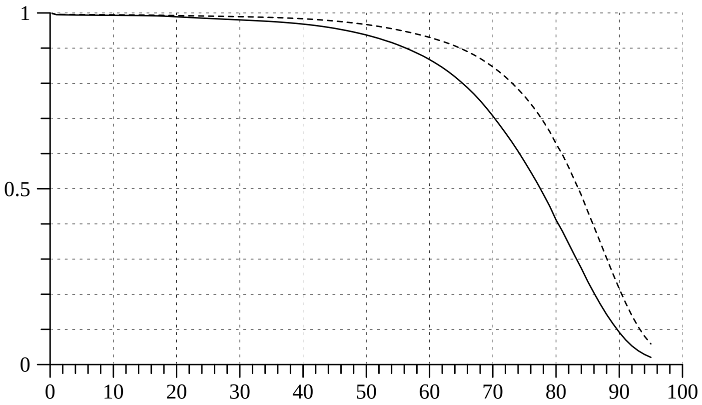


Abb. 6.2-1 Darstellung der Survivorfunktionen G^m_{1999} (durchgezogene Linie) und G^f_{1999} (gestrichelte Linie), die aus den altersspezifischen Sterbeziffern des Jahres 1999 in der BRD berechnet worden sind.

einer Geburtskohorte, sondern erfasst die Mortalitätsbedingungen derjenigen Periode, auf die sich die für ihre Konstruktion verwendeten Mortalitätsraten beziehen.

4. Eine Perioden-Sterbetafel für die BRD 1999. Zur Illustration verwenden wir die altersspezifischen Sterbeziffern für 1999 in Deutschland, die aus den Daten in den Tabellen 5.2-7 und 6.1-1 berechnet werden können. Mithilfe der Formel (6.3) können daraus Perioden-Sterbetafeln (Survivorfunktionen) für das Jahr 1999 berechnet werden: G^m_{1999} für Männer und G^f_{1999} für Frauen. Abbildung 6.2-1 zeigt den Verlauf dieser Survivorfunktionen.

Offenbar unterscheiden sich diese Survivorfunktionen von den in Abbildung 6.1-2 dargestellten Verteilungsfunktionen (bzw. ihnen korrespondierenden Survivorfunktionen), die sich auf die Lebensdauern der 1999 gestorbenen Männer und Frauen beziehen. Die Unterschiede werden zum Beispiel durch einen Vergleich der Medianwerte der Lebensdauern deutlich. Aus den Survivorfunktionen der Perioden-Sterbetafeln für 1999 findet man etwa 77.5 Jahre für Männer und 83.5 Jahre für Frauen. Für die 1999 gestorbenen Männer und Frauen sind die Medianwerte jedoch deutlich niedriger, nämlich etwa 72 bzw. 82 Jahre.

5. Berechnung von Lebenserwartungen. Sowohl Kohorten- als auch Perioden-Sterbetafeln können zur Berechnung von sogenannten Lebenserwartungen verwendet werden.[13] Zur Definition kann auf eine beliebige Ver-

[13]Wie man sehen wird, wird bei der Definition auf realisierte Daten Bezug genommen, so dass das Reden von „Erwartungen“ eigentlich falsch ist.

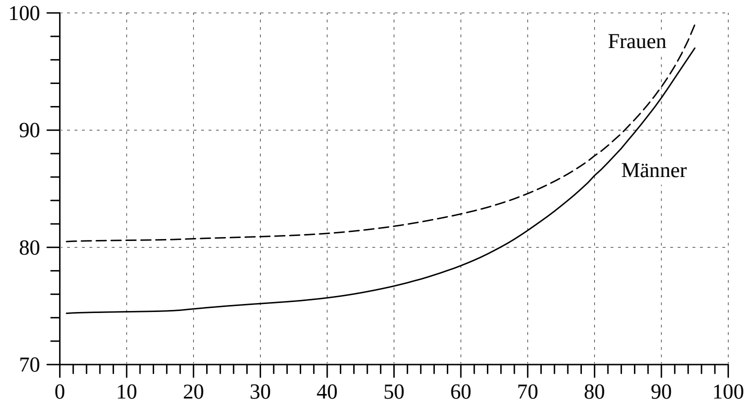


Abb. 6.2-2 Lebenserwartungen 1999 in Deutschland, berechnet aus den Survivorfunktionen G^m_{1999} und G^f_{1999}.

weildauervariable T Bezug genommen werden. Zunächst kann durch

$$\mathrm{M}(T) := \sum_{j=0}^{\infty} j\,\mathrm{P}(j)$$

der *Mittelwert* von T definiert werden.[14] Bezieht sich T auf Lebensdauern, kann $\mathrm{M}(T)$ als durchschnittliche Lebensdauer interpretiert werden. Sie erfasst die Lebenserwartung bei der Geburt. Man kann aber auch fragen, wie lange diejenigen Personen noch leben werden, die bereits ein bestimmtes Alter erreicht haben. Dann ist ein *bedingter Mittelwert* zu berechnen, der allgemein folgendermaßen definiert ist:

$$\mathrm{M}(T|T \geq t) = \frac{\sum_{j=t}^{\infty} j\,\mathrm{P}(j)}{\sum_{j=t}^{\infty} \mathrm{P}(j)}$$

Es handelt sich um den Mittelwert der Variablen T unter der Bedingung $T \geq t$, d.h. um den Mittelwert in der Teilgesamtheit $\{\omega \in \Omega \,|\, T(\omega) \geq t\}$. Bezieht sich T auf Lebensdauern, wird $\mathrm{M}(T|T \geq t)$ die *Lebenserwartung im Alter* t genannt; gemeint ist also die durchschnittliche Lebensdauer derjenigen Personen, die mindestens t Jahre alt geworden sind. Dementsprechend erhält man durch $\mathrm{M}(T|T \geq t) - t$ die *fernere Lebenserwartung im Alter* t, d.h. die durchschnittliche restliche Lebensdauer derjenigen Personen, die mindestens t Jahre alt geworden sind.

Abbildung 6.2-2 zeigt Lebenserwartungen, die ausgehend von den in § 4 besprochenen Perioden-Sterbetafeln für die BRD 1999 berechnet worden sind. Dabei wurde für die nach oben offene Altersklasse 95* angenommen,

[14]Wenn ohne Zusatz vom Mittelwert einer Variablen gesprochen wird, ist stets dieser arithmetische Mittelwert gemeint.

dass Männer durchschnittlich 97 und Frauen durchschnittlich 99 Jahre alt werden.

6.3 Veränderungen der Lebensdauern

In diesem Abschnitt wird dargestellt, wie sich die Lebensdauern bzw. Mortalitätsbedingungen in Deutschland verändert haben. Da wir dafür Perioden-Sterbetafeln der amtlichen Statistik verwenden, beginnen wir mit einigen Bemerkungen zu deren Konstruktion.

1. Sterbetafeln der amtlichen Statistik. Im vorangegangenen Abschnitt wurden altersspezifische Sterbeziffern für ein einzelnes Jahr (1999) zur Konstruktion einer Perioden-Sterbetafel verwendet. Diese sehr einfache Methode wird oft modifiziert. Das Statistische Bundesamt berechnet mit etwas unterschiedlichen Methoden zwei Arten von Perioden-Sterbetafeln:

– *Allgemeine Sterbetafeln* beziehen sich auf (dreijährliche) Perioden, in deren Mitte eine Volkszählung stattgefunden hat. Die jünste allgemeine Sterbetafel bezieht sich auf die Volkszählung 1987 und umfasst die Periode 1986 – 88. Die Methoden zur Berechnung allgemeiner Sterbetafeln haben sich im Laufe der Geschichte der amtlichen Statistik oft verändert.[15] Für die letzten beiden allgemeinen Sterbetafeln (1970 – 72 und 1986 – 88) beginnen die Berechnungen mit altersspezifischen Sterbeziffern, die in unserer Notation durch $\bar{\delta}_{t,\tau} = \bar{d}_{t,\tau}/\bar{n}_{t,\tau}$ definiert sind, wobei sich τ auf das Alter in vollendeten Lebensjahren und t auf das Kalenderjahr bezieht. Diese Raten werden dann in folgender Weise modifiziert:

$$q_{t,\tau} := \frac{\bar{d}_{t,\tau}}{\bar{n}_{t,\tau} + \frac{\bar{d}_{t,\tau}}{2}}$$

Die Überlegung besteht darin, dass etwa die Hälfte der Personen, die im Alter τ sterben, durch $\bar{n}_{t,\tau}$ nicht erfasst wird und dass es deshalb sinnvoll ist, den Umfang der Risikomenge im Nenner um $\bar{d}_{t,\tau}/2$ zu erhöhen.[16] Die derart modifizierten Mortalitätsraten werden dann auf folgende Weise für einen Dreijahreszeitraum zusammengefasst:

$$q_{(t),\tau} := \frac{\bar{d}_{t-1,\tau} + \bar{d}_{t,\tau} + \bar{d}_{t+1,\tau}}{\bar{n}_{t-1,\tau} + \bar{n}_{t,\tau} + \bar{n}_{t+1,\tau} + \frac{\bar{d}_{t-1,\tau}+\bar{d}_{t,\tau}+\bar{d}_{t+1,\tau}}{2}}$$

Dabei bezieht sich t auf das mittlere Jahr, also z.B. $t = 1987$ bei der Sterbetafel für die Periode 1986 – 88.

[15] Detaillierte Erklärungen findet man in der Fachserie 1, Reihe 1-S.2, Allgemeine Sterbetafel für die Bundesrepublik Deutschland 1986/88.

[16] Dieses Verfahren wird auch als „Sterbeziffernmethode nach Farr" bezeichnet. Es sei auch erwähnt, dass für das Lebensalter 0 ein anderes Verfahren verwendet wird.

– *Abgekürzte Sterbetafeln* werden ebenfalls für Dreijahresperioden berechnet.[17] Der wichtigste Unterschied besteht darin, dass die oben definierten Mortalitätsraten $q_{(t),\tau}$ bei den abgekürzten Sterbetafeln unmittelbar verwendet werden, bei den allgemeinen Sterbetafeln jedoch zunächst mit einem aufwendigen Rechenverfahren geglättet werden (vgl. Fachserie 1, Reihe 1-S.2: 13).

2. Allgemeine Sterbetafeln 1871 – 1988. Allgemeine Sterbetafeln wurden in Deutschland für folgende Perioden veröffentlicht:

Periode	Publikation
1871 – 1880	Statistik des Deutschen Reichs, Bd. 246: 14*-17*.
1881 – 1890	Statistik des Deutschen Reichs, Bd. 246: 14*-17*.
1891 – 1900	Statistik des Deutschen Reichs, Bd. 246: 14*-17*.
1901 – 1910	Statistik des Deutschen Reichs, Bd. 246: 14*-17*.
1910 – 1911	Statistik des Deutschen Reichs, Bd. 275. Statistisches Jahrbuch für das Deutsche Reich 1919: 50-51.
1924 – 1926	Statistik des Deutschen Reichs, Bd. 360 and 401. Statistisches Jahrbuch für das Deutsche Reich 1928: 38-39.
1932 – 1934	Statistik des Deutschen Reichs, Bd. 495: 86-87. Statistisches Jahrbuch für das Deutsche Reich 1936: 45-46.
1949 – 1951	Statistik der Bundesrepublik Deutschland, Bd. 75 und 173. Statistisches Jahrbuch für die Bundesrepublik Deutschland 1954: 62-63.
1960 – 1962	Statistisches Jahrbuch für die Bundesrepublik Deutschland 1965: 67-68. Weitere Informationen bei Schwarz (1964).
1970 – 1972	Fachserie 1, Reihe 2, Sonderheft 1. Allgemeine Sterbetafel für die Bundesrepublik Deutschland 1970/72. Weitere Informationen bei Meyer and Rückert (1974).
1986 – 1988	Fachserie 1, Reihe 1, Sonderheft 2. Allgemeine Sterbetafel für die Bundesrepublik Deutschland 1986/88. Weitere Informationen bei Meyer and Paul (1991).

In allen Fällen handelt es sich um Perioden-Sterbetafeln. Bis 1932 – 34 beziehen sie sich auf das Gebiet des früheren Deutschen Reichs; alle anderen Sterbefafeln beziehen sich auf das Gebiet der früheren BRD. Die Tabellen 6.3-1 und 6.3-2 stellen die Daten der Sterbetafeln in Gestalt von Survivorfunktionen (×100000) dar. Jeweils beginnend mit $l_0^m = l_0^f = 100000$, zeigen l_τ^m und l_τ^f wieviele Männer bzw. Frauen im Alter τ noch leben.[18]

[17] Sie werden vom Statistischen Bundesamt jährlich seit 1957 berechnet und in der Fachserie 1 (Reihe 1) publiziert.

[18] Das Symbol l wird also als Abkürzung für das 100000 fache der Survivorfunktion

Tabelle 6.3-1 Survivorfunktionen der Männer in den allgemeinen Sterbetafeln für Deutschland. Quellen: s. Text.

	1871/ 1881	1881/ 1890	1891/ 1900	1901/ 1910	1910/ 1911	1924/ 1926	1932/ 1934	1949/ 1951	1960/ 1962	1970/ 1972	1986/ 1988
τ	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m
0	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000
1	74727	75831	76614	79766	81855	88462	91465	93823	96467	97400	99075
2	69876	70998	72631	76585	79211	87030	90618	93433	96244	97249	99005
3	67557	68729	70999	75442	78255	86477	90211	93203	96109	97152	98956
4	65997	67212	69945	74727	77662	86127	89901	93022	96013	97067	98921
5	64871	66127	69194	74211	77213	85855	89654	92880	95929	96989	98891
6	64028	65330	68641	73820	76873	85647	89446	92768	95852	96918	98862
7	63369	64711	68214	73506	76596	85477	89255	92673	95782	96854	98835
8	62849	64221	67874	73244	76361	85330	89081	92586	95721	96795	98809
9	62431	63836	67599	73023	76161	85197	88927	92513	95667	96741	98786
10	62089	63526	67369	72827	75984	85070	88793	92444	95620	96692	98764
11	61800	63265	67167	72650	75818	84950	88675	92379	95577	96647	98744
12	61547	63036	66983	72487	75662	84837	88567	92315	95536	96604	98724
13	61320	62830	66811	72334	75517	84726	88464	92250	95493	96561	98704
14	61108	62636	66641	72179	75365	84607	88360	92178	95445	96515	98681
15	60892	62441	66462	72007	75189	84469	88244	92097	95388	96459	98652
16	60657	62226	66259	71808	74986	84306	88105	92001	95316	96383	98612
17	60383	61972	66017	71573	74746	84110	87939	91892	95225	96273	98557
18	60063	61675	65731	71300	74470	83874	87746	91767	95112	96118	98483
19	59696	61340	65405	70989	74165	83592	87531	91625	94973	95927	98389
20	59287	60970	65049	70647	73832	83268	87298	91466	94812	95732	98284
21	58843	60572	64674	70291	73488	82912	87051	91294	94637	95541	98175
22	58369	60156	64292	69935	73143	82539	86795	91113	94457	95357	98068
23	57871	59734	63912	69582	72800	82162	86539	90924	94280	95182	97964
24	57378	59315	63539	69232	72466	81792	86285	90730	94110	95016	97862
25	56892	58897	63168	68881	72130	81429	86032	90531	93948	94858	97763
26	56410	58474	62796	68528	71789	81072	85777	90329	93789	94705	97664
27	55927	58047	62420	68173	71446	80721	85516	90125	93633	94555	97567
28	55442	57613	62043	67817	71105	80380	85251	89922	93478	94405	97468
29	54951	57169	61663	67458	70768	80049	84984	89720	93323	94253	97367
30	54454	56713	61274	67092	70425	79726	84715	89518	93166	94097	97262
31	53949	56243	60873	66719	70070	79404	84440	89314	93008	93937	97153
32	53434	55755	60459	66338	69705	79080	84157	89104	92846	93773	97039
33	52908	55245	60030	65946	69332	78758	83863	88887	92679	93604	96920
34	52369	54715	59581	65536	68948	78436	83555	88662	92505	93429	96794
35	51815	54168	59111	65104	68545	78111	83234	88428	92322	93245	96661
36	51244	53599	58618	64650	68125	77779	82905	88184	92129	93049	96519
37	50656	53009	58099	64175	67693	77433	82571	87930	91924	92838	96367
38	50049	52406	57557	63676	67233	77073	82224	87666	91705	92610	96203
39	49422	51788	56992	63149	66741	76701	81860	87391	91470	92361	96026
40	48775	51148	56402	62598	66227	76313	81481	87102	91218	92089	95834
41	48110	50486	55785	62021	65682	75905	81088	86795	90949	91794	95624
42	47428	49806	55142	61413	65113	75473	80676	86468	90662	91475	95394
43	46729	49112	54470	60773	64518	75016	80240	86120	90354	91131	95141
44	46010	48402	53768	60105	63894	74536	79776	85746	90021	90761	94863
45	45272	47668	53037	59405	63238	74032	79285	85342	89659	90363	94555
46	44511	46910	52282	58666	62542	73496	78763	84902	89262	89934	94216
47	43728	46135	51507	57892	61810	72927	78207	84417	88825	89468	93841
48	42919	45347	50708	57084	61036	72326	77617	83883	88344	88958	93428
49	42086	44534	49875	56233	60215	71688	76990	83294	87814	88398	92973
50	41228	43684	49002	55340	59349	71006	76322	82648	87230	87781	92471

Tabelle 6.3-1 (Forts.) Survivorfunktionen der Männer in den allgemeinen Sterbetafeln für Deutschland. Quellen: s. Text.

	1871/ 1881	1881/ 1890	1891/ 1900	1901/ 1910	1910/ 1911	1924/ 1926	1932/ 1934	1949/ 1951	1960/ 1962	1970/ 1972	1986/ 1988
τ	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m	l_τ^m
51	40343	42800	48092	54403	58435	70274	75605	81945	86585	87104	91917
52	39433	41890	47150	53419	57473	69497	74834	81186	85871	86369	91305
53	38497	40956	46179	52388	56457	68670	74004	80371	85078	85574	90630
54	37534	39990	45176	51312	55395	67780	73109	79497	84197	84717	89887
55	36544	38989	44133	50186	54290	66818	72147	78562	83221	83789	89071
56	35524	37949	43047	49003	53114	65784	71124	77560	82142	82779	88177
57	34474	36872	41922	47772	51869	64678	70043	76490	80952	81673	87204
58	33392	35774	40760	46500	50563	63495	68889	75352	79644	80460	86146
59	32276	34643	39558	45180	49177	62232	67640	74141	78212	79130	85002
60	31124	33456	38308	43807	47736	60883	66293	72852	76652	77675	83767
61	29935	32221	37008	42379	46246	59444	64853	71474	74963	76087	82439
62	28708	30954	35657	40892	44663	57914	63321	70003	73144	74357	81014
63	27442	29658	34255	39343	43013	56285	61695	68437	71198	72477	79486
64	26139	28322	32799	37737	41312	54553	59962	66772	69128	70440	77851
65	24802	26940	31294	36079	39527	52715	58106	64999	66941	68242	76106
66	23433	25520	29743	34381	37695	50769	56128	63110	64643	65882	74245
67	22037	24076	28155	32637	35842	48705	54033	61104	62240	63361	72262
68	20620	22622	26531	30838	33933	46527	51822	58985	59739	60685	70150
69	19189	21154	24877	28998	31946	44256	49495	56751	57145	57864	67901
70	17750	19665	23195	27136	29905	41906	47059	54394	54461	54909	65508
71	16310	18160	21494	25254	27850	39472	44517	51903	51691	51838	62966
72	14880	16649	19784	23345	25741	36948	41872	49278	48835	48673	60270
73	13468	15145	18080	21416	23587	34348	39138	46529	45894	45438	57419
74	12085	13655	16391	19490	21450	31697	36341	43666	42873	42161	54417
75	10743	12188	14730	17586	19328	28998	33479	40700	39784	38872	51273
76	9454	10761	13109	15715	17216	26275	30553	37644	36647	35601	48000
77	8228	9404	11543	13902	15184	23589	27609	34524	33487	32373	44620
78	7077	8130	10049	12169	13278	20989	24703	31372	30334	29212	41157
79	6010	6934	8640	10525	11440	18479	21863	28222	27215	26137	37645
80	5035	5833	7330	8987	9711	16066	19122	25106	24156	23167	34119
81	4156	4837	6129	7568	8152	13785	16509	22059	21186	20321	30618
82	3378	3944	5044	6275	6708	11664	14038	19118	18337	17619	27183
83	2700	3158	4075	5116	5396	9712	11725	16324	15644	15083	23856
84	2120	2481	3225	4094	4253	7941	9607	13715	13142	12735	20678
85	1635	1909	2497	3212	3297	6371	7732	11321	10861	10595	17687
86	1236	1437	1893	2468	2519	5015	6126	9168	8819	8678	14914
87	917	1057	1405	1856	1882	3872	4765	7274	7026	6990	12385
88	666	758	1018	1364	1374	2930	3623	5655	5479	5529	10119
89	474	530	718	978	982	2182	2698	4294	4171	4287	8126
90	330	360	492	683	679	1599	1966	3175	3092	3251	6406
91	225	238	327	464	457	1144	1400	2278	2229	2407	4952
92	150	152	211	307	299	801	974	1589	1565	1735	3750
93	97	94	132	197	190	549	662	1082	1070	1215	2778
94	61	57	80	123	117	368	438	719	713	824	2011
95	38	33	46	74	70	241	283	466	463	539	1421
96	23	18	27	44	40	154	178	294	293	339	979
97	13	10	14	25	23	97	109	181	181	204	656
98	7	5	7	14	12	59	65	108	110	117	428
99	4	3	4	7	6	35	37	63	65	64	271
100	2	1	2	4	3	20	21	36	38	33	167

Tabelle 6.3-2 Survivorfunktionen der Frauen in den allgemeinen Sterbetafeln für Deutschland. Quellen: s. Text.

	1871/ 1881	1881/ 1890	1891/ 1900	1901/ 1910	1910/ 1911	1924/ 1926	1932/ 1934	1949/ 1951	1960/ 1962	1970/ 1972	1986/ 1988
τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ
0	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000
1	78260	79311	80138	82952	84695	90608	93161	95091	97222	98016	99298
2	73280	74404	76137	79761	82070	89255	92394	94749	97027	97888	99241
3	70892	72073	74482	78594	81126	88743	92026	94545	96922	97810	99201
4	69295	70514	73406	77867	80523	88422	91761	94390	96845	97745	99174
5	68126	69377	72623	77334	80077	88169	91535	94270	96782	97690	99153
6	67249	68537	72038	76924	79730	87975	91338	94177	96728	97641	99136
7	66572	67881	71577	76587	79445	87817	91160	94100	96682	97597	99119
8	66035	67358	71206	76301	79206	87683	91003	94041	96643	97558	99103
9	65599	66942	70903	76058	79001	87563	90870	93986	96609	97523	99088
10	65237	66601	70646	75845	78816	87452	90753	93937	96579	97492	99073
11	64926	66309	70420	75651	78642	87347	90650	93893	96552	97465	99058
12	64649	66049	70210	75467	78476	87243	90557	93850	96525	97439	99044
13	64390	65801	70003	75285	78311	87134	90467	93805	96498	97413	99029
14	64136	65555	69789	75094	78131	87013	90373	93756	96468	97384	99013
15	63878	65306	69562	74887	77930	86877	90270	93701	96434	97349	98995
16	63609	65045	69319	74661	77710	86719	90152	93637	96395	97305	98974
17	63322	64764	69060	74411	77470	86534	90016	93564	96351	97251	98947
18	63013	64468	68787	74143	77216	86319	89858	93484	96301	97189	98916
19	62681	64160	68500	73861	76945	86075	89680	93394	96246	97124	98881
20	62324	63838	68201	73564	76659	85808	89490	93295	96188	97059	98843
21	61941	63500	67888	73254	76362	85523	89287	93188	96128	96996	98806
22	61534	63142	67559	72929	76052	85226	89072	93073	96068	96934	98768
23	61102	62762	67212	72586	75730	84920	88849	92955	96008	96874	98731
24	60648	62360	66848	72225	75397	84602	88622	92834	95948	96815	98694
25	60174	61937	66467	71849	75043	84275	88390	92711	95884	96755	98657
26	59680	61497	66072	71463	74668	83943	88151	92586	95814	96694	98619
27	59170	61042	65666	71070	74283	83610	87904	92457	95739	96632	98579
28	58647	60570	65249	70669	73896	83274	87653	92324	95660	96567	98538
29	58111	60082	64822	70261	73513	82937	87397	92185	95575	96499	98493
30	57566	59584	64385	69848	73115	82597	87139	92039	95485	96429	98446
31	57010	59076	63937	69432	72703	82254	86876	91887	95390	96355	98395
32	56445	58554	63479	69008	72291	81909	86607	91729	95290	96276	98340
33	55869	58018	63010	68575	71876	81559	86329	91565	95184	96190	98280
34	55282	57473	62533	68132	71457	81205	86044	91396	95071	96098	98216
35	54685	56921	62047	67679	71020	80847	85754	91221	94949	95997	98146
36	54078	56360	61549	67215	70554	80482	85455	91039	94818	95886	98071
37	53462	55789	61041	66744	70080	80105	85145	90850	94676	95764	97988
38	52837	55215	60524	66266	69610	79720	84819	90651	94524	95632	97896
39	52207	54638	59998	65779	69139	79324	84481	90443	94360	95488	97796
40	51576	54054	59467	65283	68659	78917	84135	90225	94184	95331	97685
41	50946	53467	58931	64779	68172	78498	83779	89995	93995	95161	97564
42	50320	52880	58391	64269	67689	78068	83410	89749	93792	94975	97431
43	49701	52297	57848	63754	67194	77627	83027	89486	93573	94773	97286
44	49090	51720	57302	63238	66692	77175	82630	89204	93337	94551	97127
45	48481	51146	56751	62717	66187	76704	82211	88901	93081	94308	96954
46	47870	50569	56195	62181	65661	76210	81763	88574	92803	94042	96766
47	47248	49983	55628	61628	65105	75688	81282	88221	92500	93750	96562
48	46605	49385	55040	61053	64510	75136	80767	87841	92173	93427	96341
49	45939	48765	54423	60449	63883	74557	80213	87432	91821	93072	96102
50	45245	48110	53768	59812	63231	73943	79620	86991	91442	92683	95842

Tabelle 6.3-2 (Forts.) Survivorfunktionen der Frauen in den allgemeinen Sterbetafeln für Deutschland. Quellen: s. Text.

	1871/ 1881	1881/ 1890	1891/ 1900	1901/ 1910	1910/ 1911	1924/ 1926	1932/ 1934	1949/ 1951	1960/ 1962	1970/ 1972	1986/ 1988
τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ	l^f_τ
51	44521	47418	53078	59138	62547	73289	78990	86516	91035	92260	95559
52	43767	46692	52354	58418	61827	72592	78322	86003	90597	91806	95252
53	42981	45934	51594	57648	61048	71854	77613	85451	90125	91323	94918
54	42162	45136	50791	56837	60219	71071	76855	84860	89615	90813	94553
55	41308	44293	49938	55984	59350	70236	76038	84225	89063	90272	94156
56	40414	43396	49032	55077	58441	69342	75162	83540	88464	89696	93723
57	39472	42448	48072	54106	57468	68383	74225	82796	87814	89078	93252
58	38476	41462	47054	53067	56398	67357	73221	81989	87105	88411	92738
59	37418	40415	45971	51959	55245	66257	72142	81115	86331	87689	92179
60	36293	39287	44814	50780	54016	65076	70984	80166	85484	86903	91569
61	35101	38087	43582	49524	52713	63809	69745	79131	84556	86044	90903
62	33843	36823	42272	48176	51320	62448	68409	77994	83538	85101	90178
63	32521	35497	40880	46725	49816	60973	66960	76744	82420	84062	89387
64	31140	34102	39398	45178	48199	59377	65396	75374	81191	82915	88526
65	29703	32628	37828	43540	46484	57671	63712	73875	79839	81647	87587
66	28217	31088	36179	41816	44693	55852	61895	72232	78352	80250	86565
67	26686	29506	34460	40007	42782	53901	59933	70428	76720	78713	85451
68	25118	27897	32675	38111	40773	51813	57822	68455	74932	77027	84236
69	23521	26252	30826	36129	38663	49597	55568	66312	72976	75179	82909
70	21901	24546	28917	34078	36448	47255	53184	63994	70840	73157	81459
71	20265	22786	26956	31963	34191	44799	50652	61491	68513	70948	79869
72	18617	21000	24957	29777	31830	42248	47951	58794	65981	68539	78124
73	16960	19204	22938	27535	29379	39609	45118	55905	63235	65920	76206
74	15307	17416	20914	25273	26933	36869	42182	52837	60267	63084	74096
75	13677	15645	18900	23006	24517	34024	39132	49605	57076	60033	71775
76	12090	13892	16919	20745	22106	31126	35989	46226	53674	56774	69230
77	10569	12219	15000	18526	19673	28217	32820	42721	50082	53323	66447
78	9131	10661	13163	16372	17336	25335	29670	39118	46331	49702	63419
79	7795	9192	11417	14299	15112	22487	26559	35457	42458	45934	60148
80	6570	7815	9773	12348	12981	19711	23500	31787	38507	42046	56640
81	5464	6550	8252	10539	11016	17075	20527	28163	34529	38076	52912
82	4479	5408	6869	8864	9184	14624	17691	24642	30579	34071	48992
83	3614	4394	5626	7329	7499	12353	15026	21282	26717	30091	44916
84	2867	3511	4524	5955	6030	10262	12561	18132	23004	26204	40734
85	2232	2756	3568	4752	4794	8372	10323	15225	19500	22478	36501
86	1705	2124	2764	3719	3746	6712	8324	12582	16258	18974	32282
87	1276	1605	2104	2850	2856	5290	6567	10213	13319	15744	28146
88	935	1189	1571	2138	2140	4101	5075	8132	10705	12826	24160
89	671	862	1149	1571	1574	3128	3857	6335	8147	10245	20393
90	471	612	821	1131	1126	2356	2868	4815	6480	8016	16903
91	323	424	573	797	786	1736	2083	3567	4872	6139	13738
92	217	288	390	549	534	1256	1476	2571	3580	4597	10935
93	142	191	260	370	354	891	1019	1814	2571	3362	8511
94	90	123	169	244	228	620	683	1253	1805	2409	6468
95	56	78	107	157	142	423	445	846	1240	1671	4792
96	34	48	66	99	87	283	281	559	834	1134	3457
97	20	29	40	61	51	185	172	361	550	750	2425
98	11	17	24	38	29	119	101	227	356	483	1651
99	6	10	14	22	16	74	58	140	227	303	1090
100	3	6	8	13	9	45	31	84	142	185	697

3. Darstellung der Survivorfunktionen. Mithilfe der in den Tabellen angeführten Daten können die Veränderungen in den Mortalitätsbedingungen bzw. Lebensdauern untersucht werden.[19] Einen ersten Eindruck erhält man durch das Schaubild 6.3-1, in dem die Survivorfunktionen der Männer und Frauen dargestellt werden. Jeweils von unten nach oben folgen die Funktionen der zeitlichen Ordnung, so dass das Ausmaß der Veränderungen gut sichtbar wird. Zum Beispiel starben in der Periode 1871 – 81 ungefähr 41 % der Männer und 38 % der Frauen vor ihrem 20sten Lebensjahr; diese Anteile sanken dann kontinuierlich bis auf ungefähr 1–2 % in der Periode 1986 – 88.

Besonders groß war der Rückgang der Sterblichkeit bei den neugeborenen Kindern. Die folgende Tabelle zeigt die Anteile (in %) der Neugeborenen, die während ihres ersten Lebensjahrs starben:

Periode	männlich	weiblich
1871 – 1881	25.3	21.7
1881 – 1890	24.2	20.7
1891 – 1900	23.4	19.9
1901 – 1910	20.2	17.0
1910 – 1911	18.1	15.3
1924 – 1926	11.5	9.4
1932 – 1934	8.5	6.8
1949 – 1951	6.2	4.9
1960 – 1962	3.5	2.8
1970 – 1972	2.6	2.0
1986 – 1988	0.9	0.7

4. Veränderungen der Lebenserwartung. Zur Darstellung der Veränderungen können auch die in Abschnitt 6.2 (§ 5) definierten Lebenserwartungen verwendet werden. Die Berechnung erfolgt (jeweils separat für Männer und Frauen) mit der Formel

$$\mathrm{M}(T|T \geq \tau) := \frac{\sum_{j=\tau}^{100} j d_j}{\sum_{j=\tau}^{100} d_j}$$

wobei $d_j := l_j - l_{j+1}$ die Anzahl der im Alter j gestorbenen Personen bezeichnet.[20] Abb. 6.3-2 zeigt, wie sich die Funktionen $\tau \longrightarrow \mathrm{M}(T|T \geq \tau)$ bei Männer und Frauen seit 1871–81 verändert haben; die Kurven beginnen

verwendet. Außerdem wird zur Vereinfachung der Notation der Bezug auf die jeweilige Periode nicht explizit angeführt.

[19] Eine ausführliche Analyse dieser Daten (mit Ausnahme der letzten Sterbetafel für 1986–88) findet man bei H. Proebsting (1984).

[20] Man könnte ein halbes Jahr hinzufügen, um zu berücksichtigen, dass das Alter in vollendeten Lebensjahren erfasst wird.

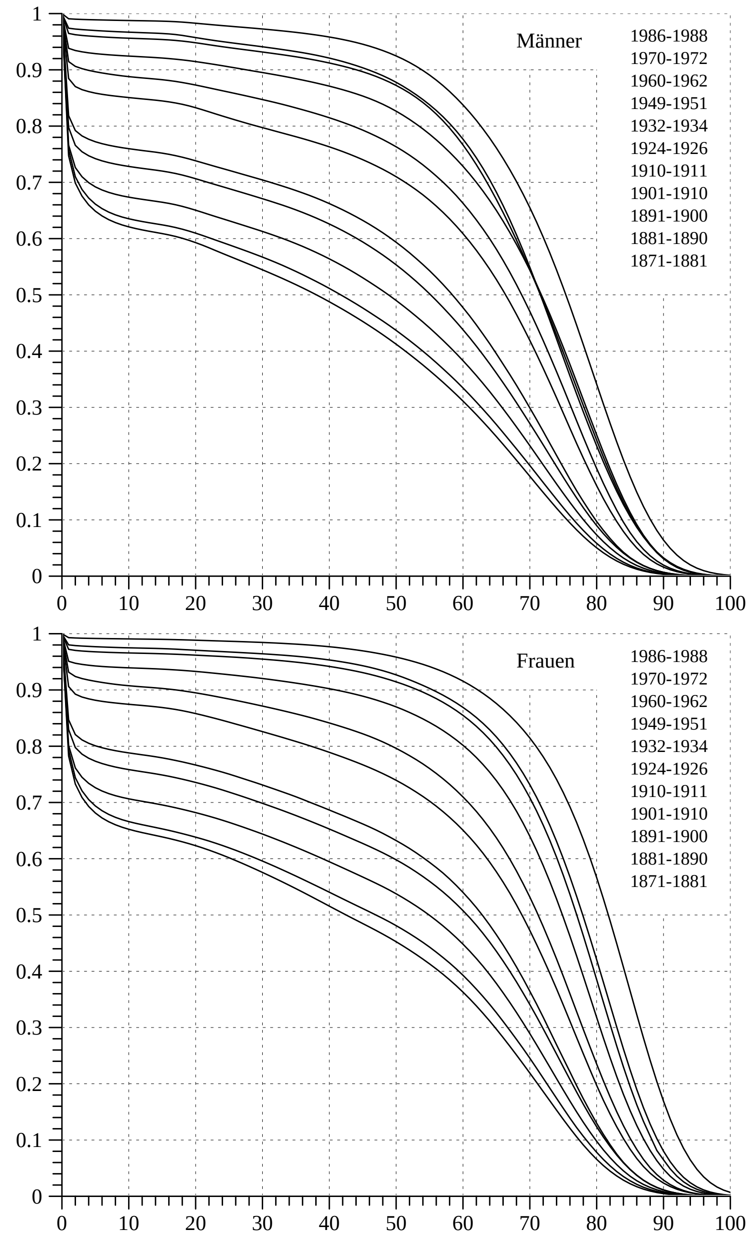


Abb. 6.3-1 Survivorfunktionen der Männer und Frauen in Deutschland, 1871 – 1988. Chronologische Ordnung im Alter von 10 Jahren.

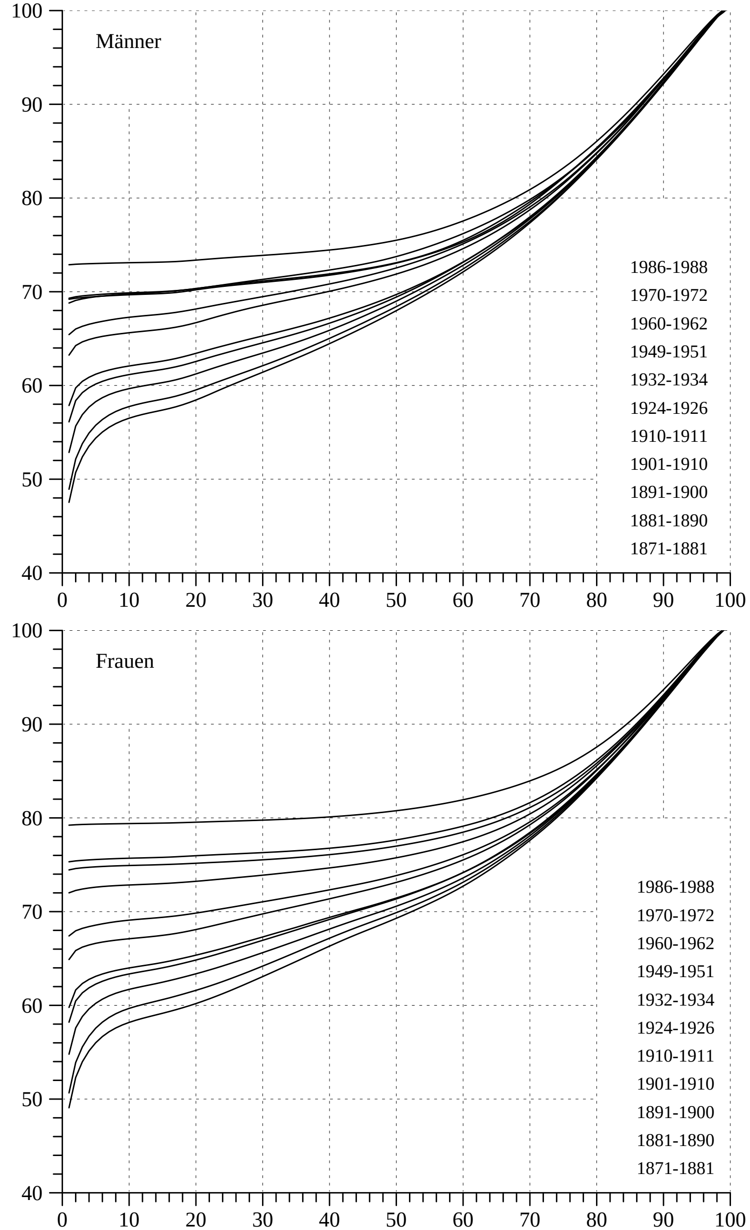


Abb. 6.3-2 Veränderungen der Lebenserwartungen ($\mathrm{M}(T|T \geq \tau)$ als Funktion des Alters τ auf der Abszisse) in Deutschland 1871 – 1988.

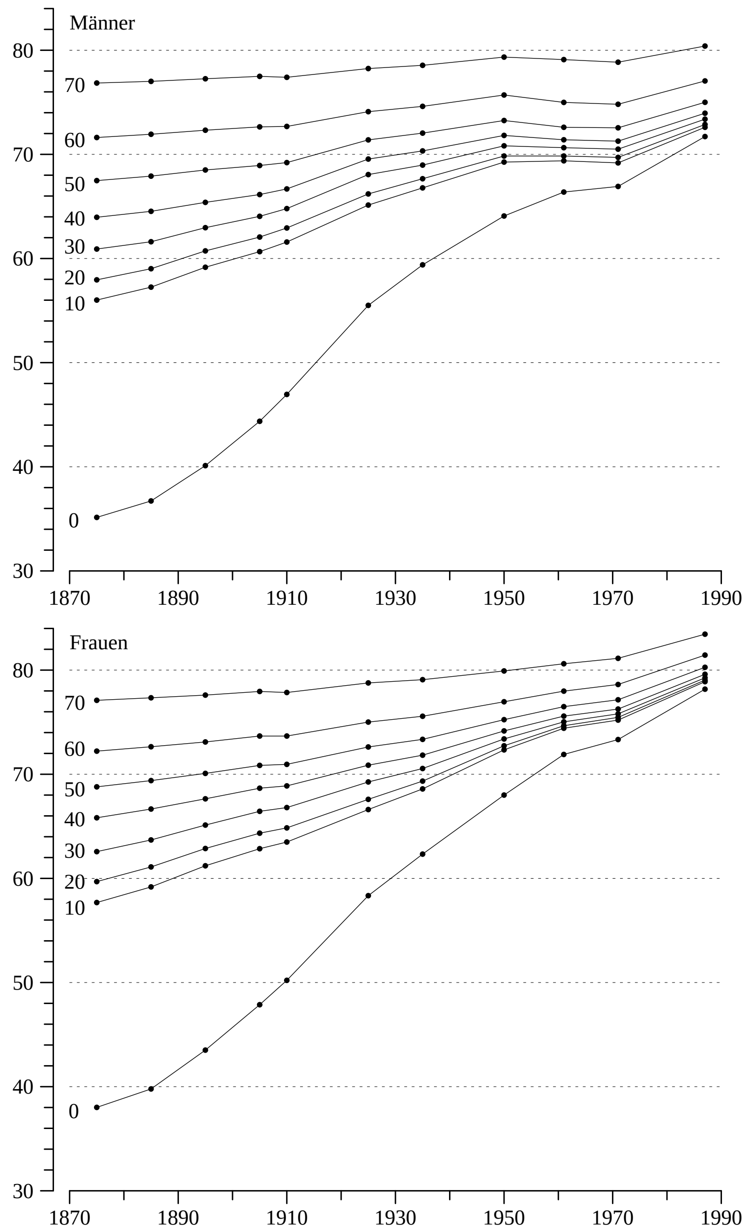


Abb. 6.3-3 Veränderungen der Lebenserwartungen $\mathrm{M}(T|T \geq \tau)$, berechnet für $\tau = 0, 10, 20, 30, 40, 50, 60, 70$, in Deutschland 1871 – 1988.

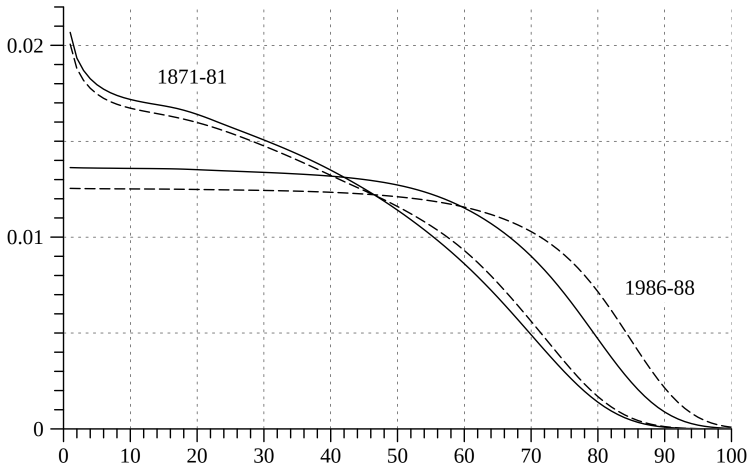


Abb. 6.3-4 Aus den Sterbetafeln 1871 – 81 und 1986 – 88 abgeleitete hypothetische Altersverteilungen der Männer (durchgezogene Linien) und Frauen (gestrichelte Linien).

im Alter $\tau = 1$ und verlaufen wieder in chronologischer Ordnung von unten nach oben. Eine ergänzende Berechnung der Lebenserwartungen Neugeborener ($\tau = 0$) zeigt, dass sie sich im Vergleich der Perioden 1871–81 und 1986–88 bei den Männern von 36 auf 72 und bei den Frauen von 38 auf 79 Jahre erhöht haben.

Abbildung 6.3-3 stellt die Veränderungen der Lebenserwartung auf einer historischen Zeitachse dar. Für einige ausgewählte Altersjahre ($\tau = 0, 10, 20, 30, 40, 50, 60, 70$) wird sichtbar gemacht, wie sich $\mathrm{M}(T|T \geq \tau)$ im Laufe der durch die Sterbetafeln erfassten Perioden erhöht hat. Offensichtlich ist die Zunahme der Lebenserwartung bei den Neugeborenen am größten und nimmt mit dem Alter ab.

5. Implikationen für die Altersverteilung. Im Allgemeinen hängt die Altersverteilung einer Bevölkerung nicht nur von den Mortalitätsraten ab, sondern auch von der Entwicklung der Geburten und von Migrationsvorgängen. Es ist deshalb nicht ohne weiteres möglich, den Einfluss von Veränderungen der Mortalität auf die Altersverteilung zu isolieren. Man kann jedoch eine hypothetische Überlegung anstellen, indem man annimmt, dass während eines längeren Zeitraums von etwa 100 Jahren jedes Jahr die gleiche Anzahl von Kindern geboren wird und entsprechend einer gegebenen, sich nicht ändernden Sterbetafel (Survivorfunktion) überlebt bzw. stirbt; außerdem sollen während dieses Zeitraums keine Zu- oder Abwanderungen vorkommen. Dann entwickelt sich schließlich eine stabile Altersverteilung, die ausschließlich durch die in der Sterbetafel erfassten

Mortalitätsbedingungen bestimmt wird.

Zur Verdeutlichung beziehen wir uns auf eine Sterbetafel l_τ und nehmen an, dass jedes Jahr 100000 Kinder geboren werden. Offenbar ist dann die Anzahl der Personen des Alters τ, die in jedem Jahr leben, gerade gleich l_τ, und ihr Anteil an allen jeweils lebenden Personen ist $l_\tau / \sum_{j=0}^{100} l_j$. Diese Anteile liefern die durch eine Sterbetafel implizierte Altersverteilung. Abbildung 6.3-4 zeigt sie, getrennt für Männer und Frauen, für die Sterbetafeln der Perioden 1871 – 81 und 1986 – 88. Man erkennt, wie der korrespondierende Rückgang in der Mortalität zu einer Veränderung der Altersverteilung geführt hat, nämlich zu einer relativen Abnahme jüngerer und Zunahme älterer Personen.

Kapitel 7

Statistik der Geburten

In diesem Kapitel beschäftigen wir uns mit der statistischen Erfassung von Geburten. Im ersten Abschnitt werden einige elementare Definitionen vorgestellt und mit Daten der amtlichen Statistik für Deutschland illustriert. Dann wird zur Untersuchung von Veränderungen des generativen Verhaltens ein Kohortenansatz besprochen. Schließlich wird im dritten Abschnitt zunächst auf einige Grenzen der amtlichen Geburtenstatistik hingewiesen, die es erforderlich machen, auch nicht-amtliche Surveys heranzuziehen; dann werden Daten aus den Lebensverlaufsstudien des Max-Planck-Instituts für Bildungsforschung verwendet, die sich auf die Geburtskohorten von 1920 bis 1960 beziehen.

7.1 Entwicklung der Geburtenziffern

1. Die Entwicklung seit 1950. Wir verwenden die in Abschnitt 5.1 eingeführten Definitionen und Notationen. Bezugnehmend auf irgendein Gebiet bezeichnet b_t die im Jahr t geborenen Kinder. Tabelle 7.1-1 zeigt Werte dieser Größe für die Gebiete der ehemaligen BRD (b_t^a) und der ehemaligen DDR (b_t^b), Abbildung 7.1-1 veranschaulicht ihre Entwicklung. Man

Tabelle 7.1-1 Anzahl Geburten (Lebendgeborene in 1000) im Gebiet der früheren BRD (b_t^a) und der früheren DDR (b_t^b). Quelle: Fachserie 1. Reihe 1, 1999: 43–44.

t	b_t^a	b_t^b	t	b_t^a	b_t^b	t	b_t^a	b_t^b
1950	812.8	303.9	1967	1019.5	252.8	1984	584.2	228.1
1951	795.6	310.8	1968	969.8	245.1	1985	586.2	227.6
1952	799.1	306.0	1969	903.5	238.9	1986	626.0	222.3
1953	796.1	298.9	1970	810.8	236.9	1987	642.0	226.0
1954	816.0	293.7	1971	778.5	234.9	1988	677.3	215.7
1955	820.1	293.3	1972	701.2	200.4	1989	681.5	198.9
1956	855.9	281.3	1973	635.6	180.3	1990	727.2	178.5
1957	892.2	273.3	1974	626.4	179.1	1991	722.2	107.8
1958	904.5	271.4	1975	600.5	181.8	1992	720.8	88.3
1959	951.9	292.0	1976	602.9	195.5	1993	717.9	80.5
1960	968.6	293.0	1977	582.3	223.2	1994	690.9	78.7
1961	1012.7	300.8	1978	576.5	232.2	1995	681.4	83.8
1962	1018.6	298.0	1979	582.0	235.2	1996	702.7	93.3
1963	1054.1	301.5	1980	620.7	245.1	1997	711.9	100.3
1964	1065.4	291.9	1981	624.6	237.5	1998	682.2	102.9
1965	1044.3	281.1	1982	621.2	240.1	1999	664.0	106.7
1966	1050.3	268.0	1983	594.2	233.8			

erkennt deutlich einen „Baby-Boom“ im Zeitraum von etwa 1955–65 in der früheren BRD.

Offenbar hängt die Anzahl der Geburten auch von der Größe der Bevölkerung und insbesondere von der Anzahl der Frauen im gebärfähigen Alter ab. In der Demographie werden deshalb spezielle Geburtenziffern bzw. -raten verwendet:

– Die *allgemeine Geburtenziffer*, bei der die Anzahl der Geburten eines Jahres auf den jahresdurchschnittlichen Bevölkerungsstand bezogen wird: $b_t/\bar{n}_t$.[1]

– Die *allgemeine Geburtenrate*, bei der die Anzahl der Geburten eines Jahres auf die jahresdurchschnittliche Anzahl der Frauen im gebärfähigen Alter bezogen wird.[2] Für die Abgrenzung des gebärfähigen Alters gibt es keine festen Bestimmungen. Das Statistische Bundesamt verwendet unterschiedliche Abgrenzungen, oft 15–45, 15–49 oder 15–50 Jahre. Wir verwenden in allgemeinen Definitionen und Formeln die Symbole τ_a und τ_b für den Beginn bzw. das Ende der reproduktiven Phase.

Beide Größen werden üblicherweise *pro 1000* angegeben. Allgemeine Geburtenziffern können unmittelbar aus den Angaben in den Tabellen 7.1-1 und 5.2-1 berechnet werden. Zum Beispiel findet man für das Jahr 1999

[1] In der englischsprachigen Literatur wird diese allgemeine Geburtenziffer auch als „crude birth rate“ oder „crude fertility rate“ bezeichnet.

[2] In der älteren deutschen Literatur wird sie auch als „allgemeine Fruchtbarkeitsziffer“ bezeichnet, in der englischsprachigen Literatur meistens als „general birth rate“ oder „general fertility rate“.

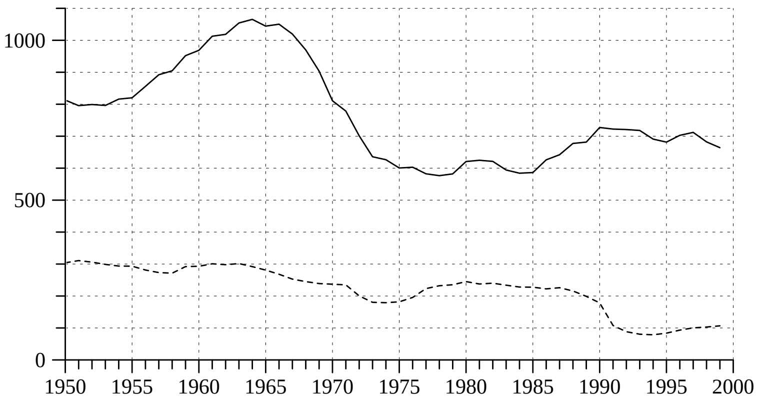


Abb. 7.1-1 Anzahl Geburten (Lebendgeborene in 1000) im Gebiet der früheren BRD (durchgezogene Linie) und der früheren DDR (gestrichelte Linie). Werte aus Tabelle 7.1-1.

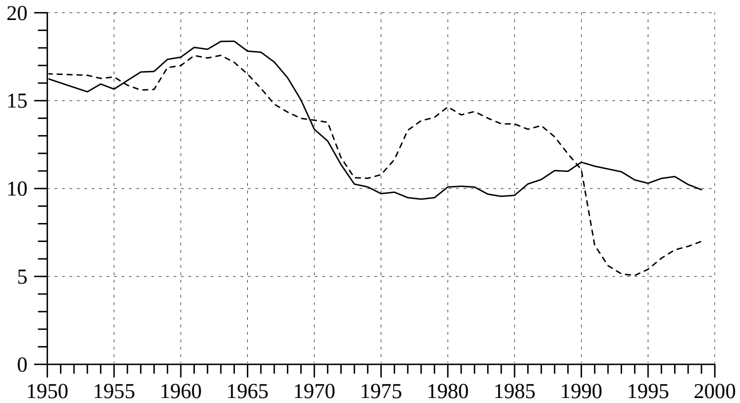


Abb. 7.1-2 Allgemeine Geburtenziffer im Gebiet der früheren BRD (durchgezogene Linie) und der früheren DDR (gestrichelte Linie). Berechnet aus Tab. 7.1-1 und 5.2-1.

im Gebiet der ehemaligen BRD den Wert: 727.2 / 63254 ($\times$1000) = 11.5, d.h. 11.5 Geburten pro 1000 der Bevölkerung. Abbildung 7.1-2 zeigt, wie sich diese Geburtenziffern seit 1950 entwickelt haben. Offenbar hat es auch in der ehemaligen DDR Ende der 1950er, Anfang der 1960er Jahre einen „Baby-Boom“ gegeben, der aus der Abbildung 7.1-1 nicht unmittelbar erkennbar ist.

2. Die langfristige Entwicklung. Die Entwicklung seit Ende des 2. Weltkriegs kann als Teil eines langfristigen Rückgangs der Geburten betrachtet

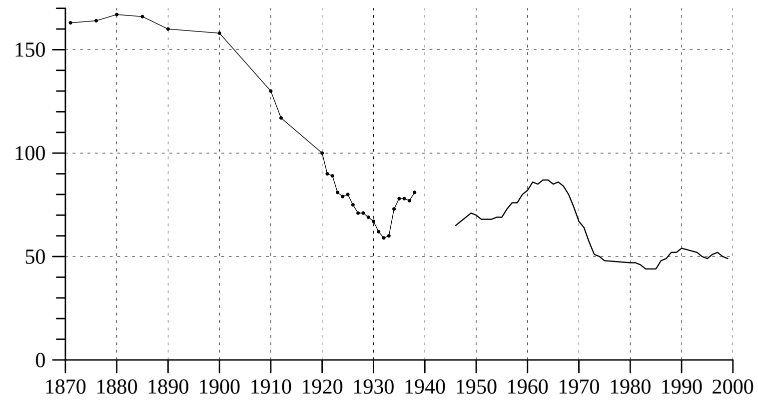


Abb. 7.1-3 Entwicklung der allgemeinen Geburtenrate in Deutschland seit 1871. Die bis 1938 verfügbaren Daten beziehen sich auf das Reichsgebiet und sind durch Punkte markiert. Ab 1946 jährliche Daten, die sich auf das Gebiet der früheren BRD beziehen. Quelle: Statistisches Bundesamt, Bevölkerung und Wirtschaft 1872–1972: 109, und Fachserie 1, Reihe 1.

werden. Abbildung 7.1-3 zeigt das anhand der Enwicklung der allgemeinen Geburtenrate seit 1871.

3. Altersspezifische Geburtenziffern. Grundlegend für differenziertere Betrachtungen sind altersspezifische Geburtenziffern. Wie bei den altersspezifischen Sterbeziffern gibt es auch in diesem Fall unterschiedliche Definitionsmöglichkeiten. Das Statistische Bundesamt verwendet folgende Definition:

$$\tilde{\beta}_{t,\tau} := \tilde{b}_{t,\tau} \,/\, \tilde{n}^{f}_{t,\tau}$$

Im Zähler steht die Anzahl der Kinder, die im Jahr t von Frauen des Geburtsjahrgangs $t-\tau$, also im demographischen Alter τ, geboren wurden; und im Nenner steht die für das Jahr t jahresdurchschnittliche Anzahl der Frauen, die im Jahr $t-\tau$ geboren wurden.

Tabelle 7.1-2 zeigt Daten des Statistischen Bundesamts, mit denen diese Geburtenziffern für 1999 berechnet werden können.[3] Die graphische Darstellung in Abbildung 7.1-4 zeigt deutlich, wie die Geburtenziffern vom Alter abhängen.

4. Berechnung von Reproduktionsraten. Die allgemeine Geburtenrate kann als ein gewichteter Mittelwert der altersspezifischen Geburtenziffern ver-

[3] Die für den Nenner erforderlichen Angaben stammen aus der internen Tabelle B13c-1999. Ich danke Hans-Peter Bosse (Statistisches Bundesamt), der diese Angaben zur Verfügung gestellt hat.

Tabelle 7.1-2 Daten zur Berechnung altersspezifischer Geburtenraten für Deutschland 1999.

$\tilde{b}_{1999,\tau}$ Anzahl Kinder, die 1999 von Frauen des Geburtsjahrgangs $1999-\tau$ geboren wurden (Fachserie 1, Reihe 1, 1999: 215).

$\tilde{n}^f_{1999,\tau}$ Jahresdurchschnittliche Anzahl von Frauen des Geburtsjahrgangs $1999-\tau$ (Tab. B13c-1999 des Statistischen Bundesamts).

$\tilde{\beta}_{1999,\tau}$ aus $\tilde{b}_{1999,\tau}$ und $\tilde{n}^f_{1999,\tau}$ berechnete altersspezifische Geburtenziffern pro 1000.

τ	$\tilde{b}_{1999,\tau}$	$\tilde{n}^f_{1999,\tau}$	$\tilde{\beta}_{1999,\tau}$
≤14	80		
15	341	436782	0.78
16	1234	441006	2.80
17	3085	452610	6.82
18	6332	454730	13.92
19	11158	460706	24.22
20	15558	442599	35.15
21	19693	440781	44.68
22	24009	443065	54.19
23	27326	440361	62.05
24	30436	432779	70.33
25	35493	444718	79.81
26	39850	454341	87.71
27	45348	500610	90.59
28	52632	555333	94.78
29	56566	582220	97.16
30	60007	626937	95.71
31	60093	657849	91.35
32	56767	677296	83.81
33	50623	696136	72.72
34	43428	699210	62.11
35	36185	713016	50.75
36	28680	710250	40.38
37	21055	690981	30.47
38	15398	684141	22.51
39	11165	666236	16.76
40	7540	646050	11.67
41	4627	614752	7.53
42	2963	603257	4.91
43	1619	592163	2.73
44	789	577973	1.37
45	342	573468	0.60
46	163	560591	0.29
47	58	563369	0.10
48	48	553593	0.09
49	25	558612	0.04
50	12	538511	0.02
≥51	16		

standen werden. Somit hängt die allgemeine Geburtenrate von den altersspezifischen Geburtenziffern und der Altersverteilung der Frauen ab.

Oft findet man auch Angaben zu einer *zusammengefassten Geburtenziffer*, die folgendermaßen definiert ist:

$$\mathrm{TFR}_t := \sum_{\tau=\tau_a}^{\tau_b} \tilde{\beta}_{t,\tau}$$

Die Abkürzung TFR verdankt sich der englischen Bezeichnung 'total fertility rate'. Diese zusammengefasste Geburtenziffer wird meistens *pro 1000* angegeben, für Deutschland im Jahr 1999 hat sie den Wert 1360.9.[4] Zur Interpretation kann man sich vorstellen, dass unter der Voraussetzung der zugrunde liegenden altersspezifischen Geburtenziffern eine Gesamtheit von 1000 Frauen durchschnittlich 1360.9 Kinder zur Welt bringen würde.

Diese Überlegung führt auch sogleich zum Begriff einer *Bruttorepro-*

[4] Fachserie 1. Reihe 1, 1999: 49. Man erhält den gleichen Wert, wenn man die Einträge für $\tilde{\beta}_{1999,\tau}$ in Tabelle 7.1-2 für den Altersbereich 15–49 addiert.

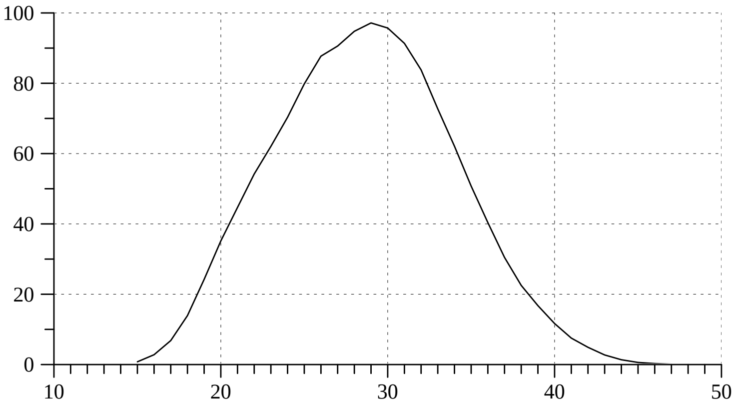


Abb. 7.1-4 Darstellung der altersspezifischen Geburtenziffern (pro 1000) aus Tabelle 7.1-2.

duktionsrate:

$$\mathrm{BRR}_t := \sigma_{t,f} \sum_{\tau=\tau_a}^{\tau_b} \tilde{\beta}_{t,\tau}$$

wobei $\sigma_{t,f}$ der Anteil der Mädchen an den Geburten im Jahr t ist.[5] Die Bruttoreproduktionsrate gibt also die Anzahl der Mädchen an, die von einer Gesamtheit von 1000 Frauen im Durchschnitt geboren werden.

Allerdings wird bei dieser Überlegung davon abstrahiert, dass einige Frauen bereits vor dem Ende der reproduktiven Phase (im Alter τ_b) sterben. Die Berücksichtigung dieser Mortalität führt zur Definition einer Nettoreproduktionsrate. Dafür werden die Werte einer Survivorfunktion $G^f_{t,\tau} := \prod_{j=0}^{\tau-1}(1-\delta^f_{t,j})$ verwendet, die den Anteil der Frauen erfasst, die im Alter τ noch leben. Somit gibt die Größe $\sum_{\tau=\tau_a}^{\tau_b} \tilde{\beta}_{t,\tau} G^f_{t,\tau}$ an, wieviele Kinder unter den angenommenen Mortalitätsbedingungen im Durchschnitt von einer Frau bis zum Ende der reproduktiven Phase geboren werden. Beschränkt man sich auf die Geburten von Mädchen, gelangt man zur Definition der *Nettoreproduktionsrate*:

$$\mathrm{NRR}_t := \sigma_{t,f} \sum_{\tau=\tau_a}^{\tau_b} \tilde{\beta}_{t,\tau} G^f_{t,\tau}$$

Offenbar liegt auch dieser Definition eine Periodenbetrachtung (im Unterschied zu einer Kohortenbetrachtung) zugrunde. Das Statistische Bundesamt verwendet bei seinen Berechnungen Werte für die Survivorfunktion $G^f_{t,\tau}$, die aus einer Perioden-Sterbetafel gewonnen werden. Unter Verwendung der Sterbetafel 1986–88 und einer reproduktiven Phase von 15 bis

[5] Die Berechnung stützt sich auf die Anzahl der Geburten. Zum Beispiel wurden 1999 in Deutschland 374448 Mädchen und 396296 Jungen geboren (Fachserie 1, Reihe 1, 1999: 42), somit ist $\sigma_{1999,f} = 0.486$ und $\sigma_{1999,m} = 0.514$.

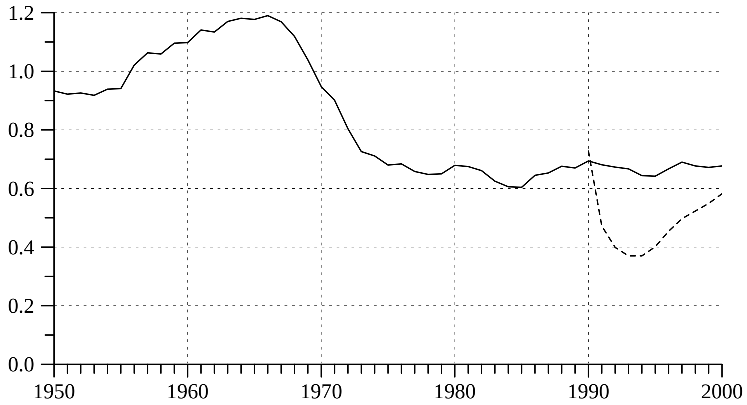


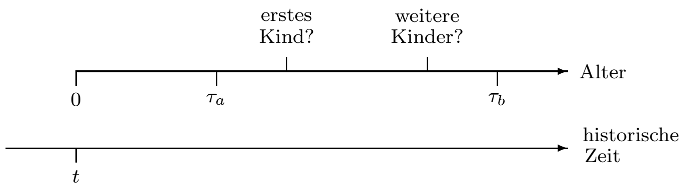
Abb. 7.1-5 Nettoreproduktionsraten im Gebiet der ehemaligen BRD (durchgezogene Linie) und in den neuen Bundesländern (gestrichelte Linie). Quelle: Fachserie 1, Reihe 1, Jg. 1996, 1999, 2000-2002.

50 Jahren wurde z.B. für Deutschland im Jahr 1999 ein Wert von 0.651 berechnet (Fachserie 1. Reihe 1, 1999: 53). Da die Mortalität der Frauen bis zum Ende der reproduktiven Phase in Deutschland sehr niedrig ist, ist dieser Wert kaum kleiner als derjenige der Bruttoreproduktionsrate.

Abbildung 7.1-5 zeigt, wie sich die Nettoreproduktionsraten im Gebiet der ehemaligen BRD und ab 1990 in den neuen Bundesländern entwickelt hat. Ersichtlich lag sie ab etwa 1970 unterhalb des Wertes 1, der einer einfachen Reproduktion in einer Situation ohne Migration entsprechen würde.

7.2 Geburtenziffern im Kohortenvergleich

1. Geburten in einer Kohortenbetrachtung. Im vorangegangenen Abschnitt wurde von einer Periodenbetrachtung ausgegangen, jetzt beziehen wir uns auf Kohorten. Zur Bezeichnung verwenden wir: $\mathcal{C}_t^f$:= Gesamtheit von Frauen, die im Jahr t geboren wurden. Wie die folgende Graphik verdeutlicht, verlaufen die Lebensläufe der Mitglieder einer solchen Geburtskohorte von Frauen auf einer historischen Zeitachse parallel:



Alle Mitglieder von $\mathcal{C}_t^f$ beginnen ihren Lebenslauf im Jahr t im Alter $\tau = 0$, und sie können somit im Hinblick darauf, ob und wann sie Kinder bekommen, unmittelbar verglichen werden.

Um ihre Lebensdauern und ihre Geburten zu erfassen, kann folgende Variable verwendet werden:

$$(T_t, K_{t,\tau_a}, \ldots, K_{t,\tau_b}) : \mathcal{C}_t^f \longrightarrow \mathcal{T}^* \times \mathbf{N}^{\tau_b - \tau_a + 1}$$

Hierbei erfasst T_t die Lebensdauer, so dass $T_t(\omega)$ das Alter ist, in dem ω stirbt; und für jedes Alter τ während der reproduktiven Phase erfasst $K_{t,\tau}(\omega)$ die Anzahl der Kinder, die von ω in diesem Alter geboren werden.[6] Das Alter kann wahlweise als gewöhnliches oder als demographisches Alter bestimmt werden. Wir werden zunächst vom demographischen Alter ausgehen, weil sich dann einfache Parallelen zu den altersspezifischen Geburtenziffern der amtlichen Statistik herstellen lassen.

Alle weiteren Größen können von der eben definierten Variablen abgeleitet werden. Zum Beispiel erhält man durch

$$\bar{K}_{t,\tau}(\omega) := \sum_{j=\tau_a}^{\tau} K_{t,j}(\omega)$$

die Gesamtzahl der von ω bis zum Alter τ geborenen Kinder; und durch $\sum_{\omega \in \mathcal{C}_t^f} \bar{K}_{t,\tau}(\omega)$ erhält man die Gesamtzahl der von Frauen der Geburtskohorte t bis zum Alter τ geborenen Kinder.

Weiterhin können *altersspezifische Kohorten-Geburtenziffern* definiert werden.[7] Wir verwenden die Definition

$$\gamma_{t,\tau} := \frac{\sum_{\omega \in \mathcal{C}_t^f} K_{t,\tau}(\omega)}{|\{\omega \in \mathcal{C}_t^f \mid T_t(\omega) \geq \tau\}|}$$

Im Nenner steht die Anzahl der Frauen, die bis zum Alter τ überleben, im Zähler steht die Anzahl der Kinder, die von diesen Frauen im Alter τ geboren werden. Schließlich können auch *kumulierte Kohorten-Geburtenziffern*

$$\bar{\gamma}_{t,\tau} := \sum_{j=\tau_a}^{\tau} \gamma_{t,j}$$

definiert werden. Wird bis zum Ende der reproduktiven Phase kumuliert, erhält man die *zusammengefasste Kohorten-Geburtenziffer*: $\bar{\gamma}_{t,\tau_b}$.

2. Daten zur Simulation einer Kohortenbetrachtung. Eine Kohortenbetrachtung im strengen Sinn kann mit den Daten der amtlichen Geburtenstatistik nicht vorgenommen werden, da die altersspezifischen Geburtenziffern $\gamma_{t,\tau}$ nicht ermittelt werden können. Man kann jedoch versuchen,

[6] $\mathbf{N}$ wird in diesem Text zur Bezeichnung der natürlichen Zahlen (einschließlich Null) verwendet. Als kombinierter Merkmalsraum für die $K_{t,\tau}$-Variablen kann also das aus $\tau_b - \tau_a + 1$ Komponenten bestehende kartesische Produkt $\mathbf{N} \times \cdots \times \mathbf{N}$ verwendet werden.

[7] Die Bezeichnung soll auf den Unterschied zu den in Abschnitt 7.1 (§ 3) definierten Geburtenziffern hinweisen, die zur Unterscheidung auch altersspezifische Perioden-Geburtenziffern genannt werden.

Tabelle 7.2-1 Altersspezifische Geburtenziffern (pro 1000) von Frauen der Geburtskohorten 1930, ..., 1970. Quelle: Fachserie 1, Reihe 1, 1999: 198–200.

	Geburtsjahr								
Alter	1930	1935	1940	1945	1950	1955	1960	1965	1970
15	0.3	0.2	0.4	0.8	0.9	1.2	1.0	0.7	0.6
16	2.1	2.2	2.3	5.0	5.5	7.8	5.0	3.1	2.2
17	10.0	9.8	10.7	18.9	21.8	26.8	13.8	8.1	6.5
18	28.9	26.8	28.0	46.6	53.8	43.7	26.0	14.4	14.2
19	52.7	52.2	56.9	82.4	90.5	58.6	40.1	23.6	25.7
20	74.6	77.3	85.9	113.1	109.8	67.1	55.9	32.4	37.7
21	96.6	104.2	120.0	141.0	115.5	78.9	67.1	43.0	47.8
22	114.2	130.1	143.3	159.8	109.9	86.1	77.3	55.1	55.8
23	125.3	145.8	163.3	155.9	105.9	93.6	83.5	68.1	61.9
24	134.9	161.6	173.2	138.6	110.3	99.5	89.2	79.6	67.6
25	139.4	167.5	171.7	125.3	110.3	111.1	97.4	94.9	75.0
26	145.9	170.0	169.0	118.9	110.9	112.9	109.0	101.2	86.9
27	149.1	161.7	156.0	102.5	105.0	110.0	112.8	104.3	95.7
28	141.8	155.1	138.0	88.5	98.0	101.2	114.7	107.4	96.8
29	136.5	143.2	116.9	80.9	91.3	93.5	108.0	103.5	99.3
30	123.9	127.6	94.1	72.8	85.8	86.4	104.1	99.7	
31	113.6	112.6	78.2	63.3	74.8	81.7	91.8	97.1	
32	98.9	95.6	61.0	53.1	63.3	72.7	80.4	91.3	
33	89.5	78.7	46.8	45.1	50.8	63.6	68.5	78.7	
34	78.7	65.3	38.8	37.6	41.5	52.6	56.5	68.1	
35	65.6	50.6	30.5	32.6	35.1	45.8	47.7		
36	56.4	40.4	24.2	26.0	29.0	35.6	40.3		
37	45.0	29.8	18.4	19.9	23.3	27.5	33.1		
38	36.1	21.2	13.5	14.6	18.4	20.4	24.9		
39	27.6	15.5	10.2	10.6	12.9	15.1	18.8		
40	19.7	10.7	7.5	7.6	10.2	10.6			
41	14.3	7.3	5.2	5.2	6.9	7.4			
42	8.5	4.4	3.3	3.7	4.3	5.0			
43	5.1	2.6	1.9	2.2	2.6	2.8			
44	2.7	1.3	1.0	1.3	1.4	1.5			
45	1.3	0.8	0.6	0.8	0.7				
46	0.6	0.4	0.3	0.3	0.3				
47	0.3	0.2	0.2	0.2	0.2				
48	0.1	0.1	0.1	0.1	0.1				
49	0.1	0.0	0.1	0.0	0.1				

durch geeignet datierte Periodengrößen Näherungswerte zu finden:

$$\gamma_{t,\tau} \approx \gamma^*_{t,\tau} := \tilde{\beta}_{t+\tau,\tau}$$

Offenbar wären bei einem demographischen Prozess ohne externe Migration $\gamma_{t,\tau}$ und $\gamma^*_{t,\tau}$ identisch.

Tabelle 7.2-1 zeigt altersspezifische Geburtenziffern der amtlichen Statistik in der Form der Größen $\gamma^*_{t,\tau}$ (pro 1000), so dass sie unmittelbar

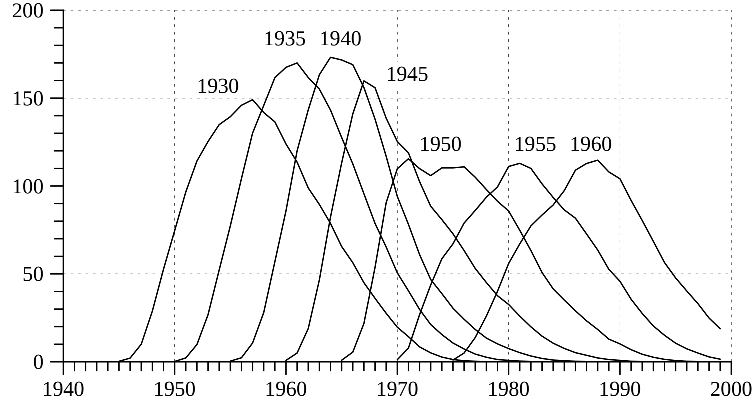


Abb. 7.2-1 Altersspezifische Geburtenziffern (pro 1000) von Frauen der Geburtskohorten 1930, ..., 1970. Daten aus Tabelle 7.2-1.

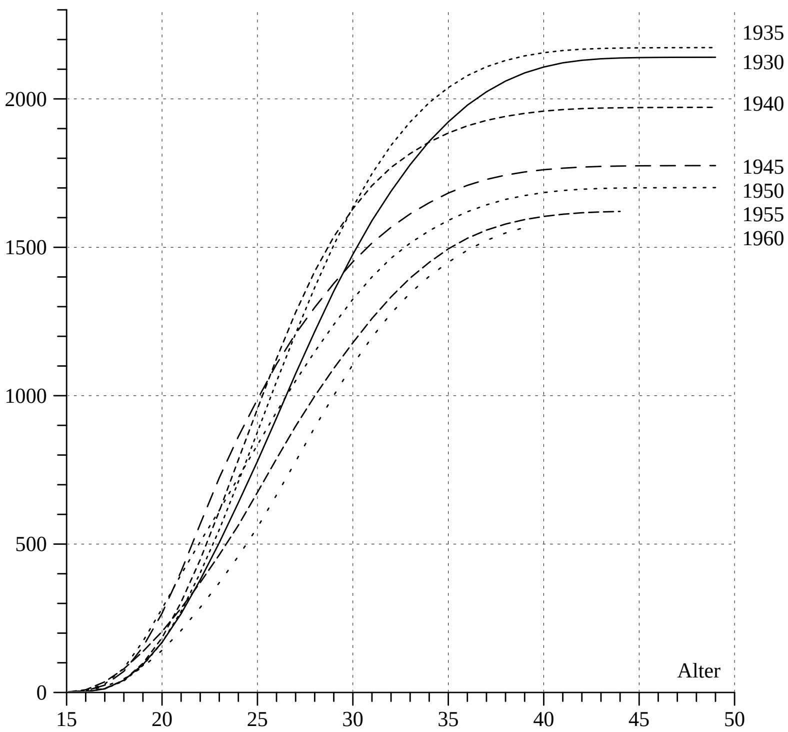


Abb. 7.2-2 Kumulierte altersspezifische Geburtenziffern $\bar{\gamma}^*_{t,\tau}$ (pro 1000) der Geburtskohorten $t = 1930, \ldots, 1970$. Daten aus Tabelle 7.2-1.

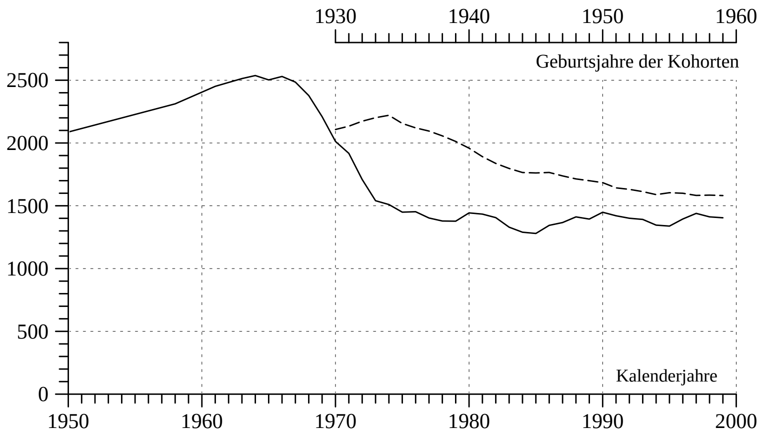


Abb. 7.2-3 Vergleich zusammengefasster Geburtenziffern (durchgezogene Linie) und kumulierter Geburtenziffern für Kohorten 1930–1959 (gestrichelt), jeweils berechnet für eine reproduktive Phase von 15 bis 40 Jahren.

auf unterschiedliche Geburtskohorten bezogen werden können. Eine graphische Darstellung auf einer historischen Zeitachse erfolgt in Abbildung 7.2-1. Man erkennt, wie die Geburten der Kohorten aufeinander folgen und sich überschneiden, und es wird auch deutlich, dass bei den jüngeren Kohorten insgesamt weniger Geburten auftreten.

Um genauer sichtbar zu machen, wie sich die Verteilungen der Geburtenziffern in den aufeinanderfolgenden Kohorten verändert haben, ist es nützlich, die kumulierten Geburtenziffern der Kohorten zu betrachten, mit den hier verfügbaren Daten also die Funktionen

$$\tau \longrightarrow \bar{\gamma}^*_{t,\tau} := \sum_{j=\tau_a}^{\tau} \gamma^*_{t,j} = \sum_{j=\tau_a}^{\tau} \tilde{\beta}_{t+j,j}$$

Sie werden in Abbildung 7.2-2 dargestellt. Man erkennt nicht nur deutlicher die Unterschiede in den zusammengefassten Geburtenziffern, sondern auch Unterschiede in der zeitlichen Verteilung der Geburten während der reproduktiven Phase. Vergleicht man zum Beispiel die Geburtskohorten 1945 und 1955, kann man eine deutliche Verschiebung der Geburten in ein höheres Alter feststellen.

3. Kohorten- und Perioden-Geburtenziffern. Abbildung 7.2-2 deutet an, dass die zusammengefassten Kohorten-Geburtenziffern etwa beginnend mit der Geburtskohorte 1935 fortgesetzt kleiner wurden. Das erkennt man auch anhand der gestrichelten Linie in der Abbildung 7.2-3, die $\bar{\gamma}^*_{t,40}$ für $t = 1930, \ldots, 1959$ zeigt.[8]

[8] Das Alter 40 wurde gewählt, da die altersspezifischen Geburtenziffern in Tabelle 7.2-1 nur bis zum Jahr 1999 verfügbar sind.

Tabelle 7.2-2 Durchschnittsalter bei der Geburt von Kindern ($\bar{\tau}_t$) und bis zum Alter 40 kumulierte Kohorten-Geburtenziffern ($\bar{\gamma}^*_{t,40}$, pro 1000) für die Geburtskohorten $t = 1930, \ldots, 1959$. Berechnet aus altersspezifischen Geburtenziffern in Fachserie 1, Reihe 1, 1999: 198–200.

t	$\bar{\tau}_t$	$\bar{\gamma}^*_{t,40}$	t	$\bar{\tau}_t$	$\bar{\gamma}^*_{t,40}$	t	$\bar{\tau}_t$	$\bar{\gamma}^*_{t,40}$
1930	27.7	2107.3	1940	26.1	1958.8	1950	26.1	1684.5
1931	27.7	2133.7	1941	26.0	1891.2	1951	26.2	1642.6
1932	27.6	2173.4	1942	25.9	1837.5	1952	26.4	1630.8
1933	27.4	2201.4	1943	25.7	1797.2	1953	26.6	1612.9
1934	27.1	2220.7	1944	25.6	1765.1	1954	26.8	1589.0
1935	27.1	2155.7	1945	25.5	1761.4	1955	26.9	1604.0
1936	26.9	2120.3	1946	25.5	1765.3	1956	27.1	1599.7
1937	26.7	2095.1	1947	25.6	1738.0	1957	27.3	1582.5
1938	26.5	2056.6	1948	25.7	1714.2	1958	27.5	1585.2
1939	26.3	2012.1	1949	25.9	1700.0	1959	27.6	1581.4

Die Abbildung zeigt auch (als durchgezogene Linie) die Entwicklung der zusammengefassten Perioden-Geburtenziffer, die für diesen Vergleich als $\sum_{\tau=15}^{40} \tilde{\beta}_{t,\tau}$ (für die Kalenderjahre $t = 1950, \ldots, 1999$) berechnet wurde. Zwar sind beide Versionen einer zusammengefassten Geburtenziffer nicht unmittelbar vergleichbar. Die Abbildung deutet jedoch an, dass die zusammengefassten Kohorten-Geburtenziffern weniger starke Schwankungen aufweisen als die Perioden-Geburtenziffern. Man kann vermuten, dass der hauptsächliche Grund dafür ist, dass Schwankungen der zusammengefassten Perioden-Geburtenziffern auch durch Veränderungen in der zeitlichen Lagerung von Geburten in den Lebensverläufen von Frauen entstehen können. Für die zusammengefassten Kohorten-Geburtenziffern sind dagegen solche Veränderungen folgenlos.

4. Kinderzahlen und Durchschnittsalter bei der Geburt. Es ist bemerkenswert, dass es keinen einfachen Zusammenhang zwischen dem durchschnittlichen Alter, in dem die Frauen einer Geburtskohorte ihre Kinder bekommen, und der Anzahl der schließlich geborenen Kinder gibt. Um das zu zeigen, verwenden wir als Näherungswert für die Anzahl der Kinder die bis zum Alter 40 kumulierte Kohorten-Geburtenziffer $\bar{\gamma}^*_{t,40}$; und als Indikator für das durchschnittliche Alter, in dem die Frauen einer Geburtskohorte t ihre Kinder bekommen, verwenden wir den Mittelwert

$$\bar{\tau}_t := \frac{\sum_{\tau=15}^{40} \tau\, \gamma^*_{t,\tau}}{\sum_{\tau=15}^{40} \gamma^*_{t,\tau}} = \frac{\sum_{\tau=15}^{40} \tau\, \beta_{t+\tau,\tau}}{\sum_{\tau=15}^{40} \beta_{t+\tau,\tau}}$$

Beide Größen, $\bar{\gamma}^*_{t,40}$ und $\bar{\tau}_t$, können aus den altersspezifischen Geburtenziffern der amtlichen Geburtenstatistik berechnet werden; Tabelle 7.2-2 zeigt ihre Werte für die Geburtskohorten $t = 1930, \ldots, 1959$. Die graphische Darstellung in Abbildung 7.2-4 zeigt nicht nur, dass das durchschnittliche Alter der Frauen bei der Geburt von Kindern zunächst bis zur Kohorte

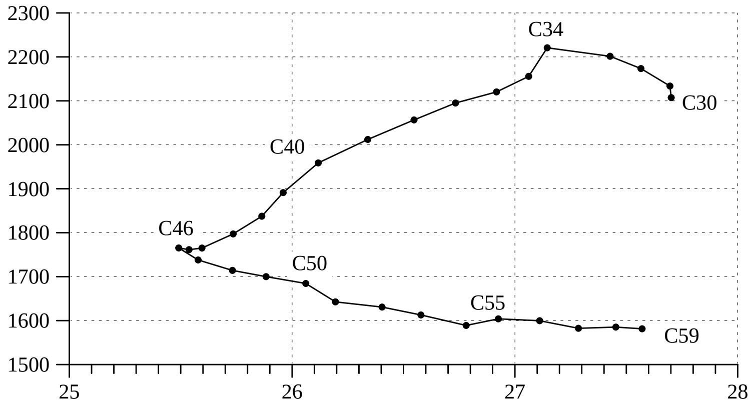


Abb. 7.2-4 Darstellung der Daten aus Tabelle 7.2-2. Die Abszisse bezieht sich auf das Durchschnittsalter bei der Geburt von Kindern, die Ordinate erfasst die bis zum Alter 40 kumulierte Geburtenziffer (pro 1000).

1946 gesunken, dann jedoch wieder angestiegen ist; sie zeigt auch, dass der Zusammenhang mit der Anzahl der Kinder jeweils unterschiedlich ist.

7.3 Daten aus retrospektiven Surveys

1. Bemerkungen zur amtlichen Geburtenstatistik. Einige wichtige Fragen können mit den Daten der amtlichen Geburtenstatistik nicht beantwortet werden, insbesondere:

a) Wie sieht die Verteilung des Alters bei der Geburt des ersten Kindes aus und wie hat sie sich verändert?

b) Wie groß ist der Anteil der Frauen, die kinderlos bleiben, und wie hat sich dieser Anteil verändert?

c) Wie verteilen sich die Kinderzahlen? Wieviel Prozent der Frauen haben ein, zwei oder mehr Kinder? Wie haben sich diese Verteilungen verändert?

Hauptsächlich liegt dies daran, dass sich die amtliche Geburtenstatistik vielfach nur auf ehelich geborene Kinder oder auf Kinder verheirateter Mütter bezieht. Zum Beispiel wird nur das durchschnittliche Alter der Mütter bei der Geburt ihrer „ehelich lebendgeborenen ersten Kinder“ erfasst (Fachserie 1, Reihe 1, 1999: 54).

Die Probleme beginnen bereits bei der statistischen Erfassung der Geburten in den Standesämtern (eine Dokumentation der Erhebungsunterlagen findet man in der Fachserie 1, Reihe 1, 1990: 313). Nur bei ehelich geborenen Kindern wird erfasst, um das „wievielte Kind dieser Ehe“ es sich

Tabelle 7.3-1 Anteile nicht-ehelicher Geburten (pro 100 Geburten insgesamt) in Deutschland und den Gebieten der früheren BRD und DDR. Quellen: Statistisches Bundesamt, Bevölkerung und Wirtschaft 1872–1972: 107–108, und Fachserie 1, Reihe 1, 1999: 50–51.

Deutschland (Reichsgebiet)				Gebiete der früheren					
					BRD	DDR		BRD	DDR
Jahr		Jahr		Jahr			Jahr		
1872	8.78	1908	8.77	1946	16.38	19.25	1973	6.27	15.64
1873	9.13	1909	8.92	1947	11.85	15.11	1974	6.27	16.29
1874	8.57	1910	8.96	1948	10.23	12.69	1975	6.12	16.14
1875	8.56	1911	9.08	1949	9.31	11.89	1976	6.35	16.21
1876		1912	9 44	1950	9 73	12 79	1977	6 47	15 77
1877	8.58	1913	9.60	1951	9.64	13.15	1978	6.96	17.34
1878	8.57	1914	9.69	1952	9.03	13.00	1979	7.13	19.59
1879	8.76	1915	11.07	1953	8.67	13.03	1980	7.56	22.84
1880		1916	10 95	1954	8 42	13 25	1981	7 90	25 58
1881	8.97	1917	11.41	1955	7.86	13.00	1982	8.49	29.29
1882	9.19	1918	12.96	1956	7.47	13.19	1983	8.83	32.04
1883	9.13	1919	11.03	1957	7.19	13.18	1984	9.07	33.55
1884	9.42	1920	11.22	1958	6.85	12.37	1985	9.40	33.81
1885	9.36	1921	10.56	1959	6.69	12.01	1986	9.55	34.43
1886	9.38	1922	10.63	1960	6.33	11.60	1987	9.71	32.80
1887	9.34	1923	10.31	1961	5.95	11.13	1988	10.03	33.44
		1924	10 41	1962	5 56	10 08	1989	10 22	33 64
1893	9.05	1925	11.82	1963	5.23	9.34	1990	10.49	34.99
1894	9.27	1926	12.37	1964	4.99	9.42	1991	11.11	41.72
1895	8.98	1927	12.28	1965	4.69	9.81	1992	11.59	41.82
1896	9.27	1928	12.21	1966	4.56	9.99	1993	11.87	41.09
1897	9.13	1929	12.07	1967	4.61	10.70	1994	12.43	41.44
1898	9.03	1930	12.00	1968	4.76	11.49	1995	12.89	41.77
1899	8.88	1931	11.75	1969	5.04	12.41	1996	13.68	42.39
1900	8.63	1932	11.63	1970	5.46	13.30	1997	14.27	44.10
1901	8.48	1933	10.67	1971	5.81	15.12	1998	15.92	47.15
1902	8.39	1934	8.53	1972	6.05	16.20	1999	17.67	49.94
1903	8.24	1935	7.77						
1904	8.31	1936	7.70						
1905	8.43	1937	7.66						
1906	8.41	1938	7.60						
1907	8.60								

handelt. Somit erfolgt eine Zählung der Parität nur bei ehelich geborenen (einschließlich später legitimierten) Kindern, und sie beginnt überdies bei mehrfachen Ehen stets bei Null.

2. Entwicklung der nicht-ehelichen Geburten. Um deutlich zu machen, wie problematisch die Fokussierung auf eheliche Geburten ist, werfen wir einen kurzen Blick auf die Entwicklung der nicht-ehelichen Geburten. Tabelle 7.3-1 zeigt, wie sich die Anteile nicht-ehelicher Geburten seit 1872 entwickelt haben, Abbildung 7.3-1 zeigt sie in graphischer Darstellung. Man erkennt, dass bis etwa 1933 die Anteile bei etwa 10 Prozent lagen.[9] Ein er-

[9]Tatsächlich war der Anteil in früheren Zeiten teilweise noch erheblich größer. Zum Beispiel dokumentiert F. Lindner (1900: 217) einen Anteil von etwa 20 % nicht-ehelicher

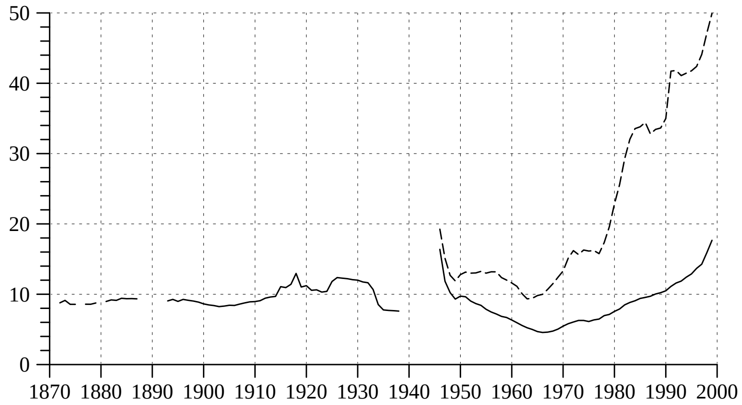


Abb. 7.3-1 Graphische Darstellung der Daten über die Anteile nicht-ehelicher Geburten aus Tabelle 7.3-1.

heblicher Anstieg nicht-ehelicher Geburten begann dann Mitte der 1960er Jahre, besonders im Gebiet der früheren DDR, wo inzwischen ein Anteil von etwa 50 % erreicht ist.[10] Zumindest für die statistische Erfassung und für sozialstrukturelle Analysen wird somit eine Unterscheidung zwischen ehelichen und nicht-ehelichen Geburten zunehmend obsolet.

3. Daten aus den Lebensverlaufsstudien. Die Grenzen der amtlichen Geburtenstatistik machen es erforderlich, auf Daten aus nicht-amtlichen Surveys zurückzugreifen. In den folgenden Paragraphen verwenden wir Daten aus Lebensverlaufsstudien, die am Max-Planck-Institut für Bildungsforschung unter der Leitung von Karl Ulrich Mayer durchgeführt wurden. Es handelt sich um eine Serie retrospektiver Surveys, in denen Mitglieder ausgewählter Geburtskohorten über zentrale Ereignisse und Bedingungen ihrer bisherigen Lebensverläufe befragt wurden. Die Daten sind gut dokumentiert und für die Forschung verfügbar.[11] Im Folgenden verwenden wir nur Daten aus den älteren westdeutschen Teilstudien:[12]

- Die Daten des ersten Surveys (LV I) wurden während der Jahre 1981–83 erhoben und beziehen sich auf 2171 Personen der Geburtskohorten 1929–31, 1939–41, und 1949–51.
- Die Daten eines zweiten Surveys (LV II) beziehen sich auf Personen der Geburtskohorten 1919–21. Zuerst wurden im Zeitraum 1985–86 407 Personen befragt (LV IIA), dann im Zeitraum 1987–88 weitere 1005 Personen (LV IIT).
- Daten eines dritten Surveys (LV III) wurden 1989 erhoben und beziehen sich auf 2008 Personen der Geburtskohorten 1954–56 und 1959–61.

Die Daten wurden im Gebiet der ehemaligen BRD erhoben und umfassen Personen mit einer deutschen Staatsangehörigkeit. Für die folgenden Berechnungen beziehen wir uns auf die in den Surveys LV I, LV IIT, und LV III (nur Geburtsjahre 1959–61) erfassten Frauen. Folgende Tabelle zeigt unsere Notation für die Geburtskohorten und die Fallzahlen:[13]

Geburtskohorte	Geburtsjahre	Männer	Frauen	Interviews
C20	1919 – 21	373	632	1987 – 88
C30	1929 – 31	349	359	1981 – 83
C40	1939 – 41	375	355	1981 – 83
C50	1949 – 51	365	368	1981 – 83
C60	1959 – 61	512	489	1989

Es soll untersucht werden, wie sich das Alter bei der Geburt des ersten Kindes und die Häufigkeitsverteilungen der Kinder in der historischen Abfolge der Kohorten verändert haben.[14]

4. Alter bei der Geburt des ersten Kindes. Wir beginnen mit Daten zum Alter bei der Geburt des ersten Kindes. Zur formalen Repräsentation dieser Daten kann für jede Geburtskohorte eine zweidimensionale statistische Variable

$$(T_c, D_c) : \Omega_c \longrightarrow \mathcal{T}_0 \times \tilde{\mathcal{D}}$$

verwendet werden. Ω_c ist die Referenzmenge der Frauen der Geburtskohorte c (C20, ..., C60). D_c mit dem Merkmalsraum $\tilde{\mathcal{D}} := \{0, 1\}$ erfasst, ob eine Frau mindestens ein Kind geboren hat ($D_c = 1$) oder ob das bisher nicht der Fall gewesen ist ($D_c = 0$);[15] und T_c mit dem Merkmalsraum $\mathcal{T}_0 := \{0, 1, 2, \ldots\}$ erfasst das Alter bei der Geburt des ersten Kindes (wenn $D = 1$ ist) oder das Alter zum Interviewzeitpunkt (wenn $D = 0$ ist).

Geburten für das Königreich Bayern während des Zeitraums 1825–1868. Eine Diskussion der Veränderungen, die während des 19. Jahrhunderts stattfanden, findet man bei P. Kottmann (1987).

[10] Man vgl. dazu auch die Ausführungen von J. Huinink (1998).

[11] Sie sind beim Zentralarchiv für empirische Sozialforschung in Köln erhältlich.

[12] Einen Überblick gibt M. Wagner (1996).

[13] Drei der 632 Frauen der Geburtskohorte C20 haben keine gültigen Geburtsjahre für ihre Kinder angegeben und werden aus den weiteren Berechnungen ausgeschlossen.

[14] Es sei erwähnt, dass die Lebensverlaufsdaten bereits in zahlreichen früheren Arbeiten zur Untersuchung dieser und ähnlicher Fragen verwendet worden sind; man vgl. etwa J. Huinink (1987, 1988, 1989), H.-P. Blossfeld und J. Huinink (1989), N. B. Tuma und J. Huinink (1990).

[15] Die Lebensverlaufsdaten erlauben, zwischen eigenen, adoptierten und Stiefkindern zu unterscheiden. In den folgenden Untersuchungen beziehen wir uns nur auf eigene Kinder der Frauen.

Tabelle 7.3-2 Daten aus den Lebensverlaufsstudien zum Alter bei der Geburt des ersten Kindes ($d = 1$) bzw. zum Interviewzeitpunkt ($d = 0$).

	C20		C30		C40		C50		C60	
τ	$d=1$	$d=0$	$d=1$	$d=0$	$d=1$	$d=0$	$d=1$	$d=0$	$d=1$	$d=0$
15									1	
16	1						3		3	
17	2		1		5		4		6	
18	11		5		15		10		10	
19	21		16		21		30		16	
20	28		23		21		34		14	
21	37		23		25		36		22	
22	60		29		36		26		24	
23	60		22		40		22		17	
24	68		21		37		16		21	
25	41		32		21		13		27	
26	35		32		23		20		26	
27	28		31		13		22		35	12
28	28		16		21		14		27	85
29	22		17		6		12		8	80
30	15		16		10		13	21	1	54
31	15		7		5		3	22		
32	7		11		8		3	35		
33	10		5		3		1	8		
34	7		8							
35	5		3		2					
36	7		1		2					
37	5				2					
38	3		1							
39	1									
40	3					7				
41			1			11				
42						13				
43						8				
50				9						
51				16						
52				7						
53				6						
66		4								
67		47								
68		34								
69		24								
Insg.	520	109	321	38	316	39	282	86	258	231

Tabelle 7.3-2 zeigt die Verteilungen dieser Variablen in Form absoluter Häufigkeiten. Man erkennt zum Beispiel, dass 68 Frauen der Geburtskohorte C20 bei der Geburt ihres ersten Kindes 24 Jahre alt waren, 41 Frauen waren 25 Jahre alt usw. Insgesamt hatten in dieser Geburtskohorte 520 Frauen mindestens ein Kind und 109 Frauen blieben bis zum Interviewzeitpunkt kinderlos.

Bei den älteren Geburtskohorten kann man annehmen, dass die reproduktive Phase zum Interviewzeitpunkt bereits abgeschlossen ist, so dass sich auch Anteile der kinderlos gebliebenen Frauen unmittelbar schätzen lassen: 17 % in der Kohorte C20 und 11 % in der Kohorte C30. Bei den jüngeren Kohorten muss jedoch berücksichtigt werden, dass die Daten unvollständig sind, weil die reproduktive Phase zum Interviewzeitpunkt noch nicht abgeschlossen ist.

5. Schätzungen mit rechts zensierten Daten. Unvollständige Daten dieser Art treten sehr oft auf, wenn man Episodendauern (wie z.B. Dauern von Arbeitslosigkeitsepisoden, Ehedauern oder Zeitdauern bis zur Geburt des erstens Kindes) mithilfe retrospektiver Surveys ermitteln möchte; stets dann, wenn bei einer Person ein Ereignis, durch das die Episode beendet wird, bis zum Interviewzeitpunkt noch nicht eingetreten ist, aber später noch eintreten kann. Man spricht dann von *rechts zensierten Daten*.

Zu überlegen ist, ob bzw. wie man auch mit teilweise rechts zensierten Daten zu informativen Einsichten gelangen kann. Die Antwort hängt auch von der zeitlichen Lagerung der zensierten Daten ab. Als Beispiel betrachten wir die Angaben zur Kohorte C60 in Tabelle 7.3-2. Offenbar kann man die Verteilung des Alters bei der Geburt des ersten Kindes bis zu einem Alter von 26 Jahren problemlos berechnen. Schwierig wird es erst ab dem Alter 27, denn die 12 Frauen, die zur Zeit des Interviews 27 Jahre alt und kinderlos sind, könnten später noch ihr erstes Kind bekommen.

In vielen Fällen ist es dennoch sinnvoll möglich, die Berechnungen bis zu demjenigen Alter (oder allgemein bis zu derjenigen Episodendauer) auszudehnen, für das die letzte vollständige Beobachtung vorliegt. In unserem Beispiel wäre dies das Alter 30. Darüber hinaus stehen natürlich keine Informationen mehr zur Verfügung, so dass sich auch mit den Daten der Tabelle 7.3-2 nicht schätzen lässt, welcher Anteil der Frauen aus der Kohorte C60 am Ende der reproduktiven Phase kinderlos bleibt.

Verwendet wird das sogenannte *Kaplan-Meier-Verfahren* (zur Schätzung von Verweildauerverteilungen mit teilweise rechts zensierten Daten). Dieses Verfahren gibt es in unterschiedlichen Varianten. Wir besprechen zunächst die einfachste Variante und gehen erst im nächsten Paragraphen auf einige Komplikationen ein, die bei unserem gegenwärtigen Anwendungsfall auftreten.

Bei der einfachsten Variante des Verfahrens wird angenommen, dass die Episoden, für deren Dauer man sich interessiert, für alle Mitglieder einer Gesamtheit Ω definiert sind und einen bestimmten (obwohl möglicherweise nicht bekannten) Wert haben. Somit genügt zur theoretischen Erfassung der Episodendauern eine einfache Verweildauervariable

$$\hat{T} : \Omega \longrightarrow \mathcal{T}_0 := \{0, 1, 2, \ldots\}$$

Würde man für jedes Mitglied $\omega \in \Omega$ die Verweildauer $\hat{T}(\omega)$ kennen, könnte man offenbar die Verteilung von $\hat{T}$ sogleich berechnen. Jetzt nehmen wir jedoch an, dass die verfügbaren Daten teilweise rechts zensiert sind. Die Daten sind somit durch eine zweidimensionale Variable

$$(T, D) : \Omega \longrightarrow \mathcal{T}_0 \times \tilde{\mathcal{D}}$$

gegeben, wobei die Komponente T die bisherige Verweildauer erfasst und die Komponente D angibt, ob es sich um eine vollständige oder um eine rechts zensierte Beobachtung handelt. Somit ergibt sich folgender Zusammenhang zur eigentlich interessierenden Variablen $\hat{T}$:

– Wenn $D(\omega) = 1$ ist, liegt eine vollständige Beobachtung vor und es ist $T(\omega) = \hat{T}(\omega)$.
– Wenn $D(\omega) = 0$ ist, liegt eine rechts zensierte Beobachtung vor und es ist $T(\omega) \leq \hat{T}(\omega)$.[16]

Die Frage ist, was sich mit den durch (T, D) gegebenen Daten über die Verteilung von $\hat{T}$ aussagen lässt. Das Kaplan-Meier-Verfahren geht von einem Zusammenhang zwischen der Survivorfunktion $G[\hat{T}]$ und der Ratenfunktion $r[\hat{T}]$ aus, der bereits in Abschnitt 6.1 (§ 4) besprochen wurde:

$$G[\hat{T}](t) \;=\; \prod_{j=0}^{t-1} (1 - r[\hat{T}](j)) \tag{7.1}$$

Die Idee besteht darin, Werte der Ratenfunktion $r[\hat{T}]$ mithilfe der durch (T, D) gegebenen Daten zu schätzen, wobei angenommen wird, dass folgender Zusammenhang näherungsweise zutreffend ist:

$$\begin{aligned} r[\hat{T}](t) \;&=\; \frac{|\{\omega \in \Omega \,|\, \hat{T}(\omega) = t\}|}{|\{\omega \in \Omega \,|\, \hat{T}(\omega) \geq t\}|} \\ &\approx\; r^*[\hat{T}](t) \;:=\; \frac{|\{\omega \in \Omega \,|\, T(\omega) = t, D(\omega) = 1\}|}{|\{\omega \in \Omega \,|\, T(\omega) \geq t\}|} \end{aligned}$$

Geht man von dieser Näherung aus, erhält man durch

$$G^*[\hat{T}](t) \;:=\; \prod_{j=0}^{t-1} (1 - r^*[\hat{T}](j))$$

eine Schätzung für die Survivorfunktion $G[\hat{T}]$ und somit auch für alle anderen Charakterisierungen der Verteilung von $\hat{T}$.

6. Anwendung des Kaplan-Meier-Verfahrens. Bei dem einfachen Kaplan-Meier-Verfahren, das im vorangegangenen Paragraphen besprochen wurde, wird angenommen, dass es für alle Mitglieder einer Referenzgesamtheit Ω eine bestimmte Episodendauer gibt, die durch eine einfache Verweildauervariable $\hat{T}$ repräsentiert werden kann. Unser Anwendungsfall ist jedoch komplizierter, da nicht alle Frauen mindestens ein Kind zur Welt bringen und einige Frauen bereits vor dem Ende der reproduktiven Phase sterben.

[16] Geht man von einer diskreten Zeitachse mit relativ groben Zeitstellen aus, wie z.B. Monaten oder Jahren, sollte auch die Möglichkeit $T(\omega) = \hat{T}(\omega)$ zugelassen werden.

Tabelle 7.3-3 Anwendung des Kaplan-Meier-Verfahrens zur Berechnung einer Survivorfunktion für das Alter bei der Geburt des ersten Kindes. Die Daten beziehen sich auf die Kohorte C50 in der Tabelle 7.3-2.

t	noch kinderlos	Anzahl Geburten	zensierte Fälle	$r^*_{[\hat{T}]}(t)$	$1 - r^*_{[\hat{T}]}(t)$	$G^*_{[\hat{T}]}(t)$
16	368	3	0	0.0082	0.9918	1.0000
17	365	4	0	0.0110	0.9890	0.9918
18	361	10	0	0.0277	0.9723	0.9809
19	351	30	0	0.0855	0.9145	0.9537
20	321	34	0	0.1059	0.8941	0.8722
21	287	36	0	0.1254	0.8746	0.7798
22	251	26	0	0.1036	0.8964	0.6820
23	225	22	0	0.0978	0.9022	0.6114
24	203	16	0	0.0788	0.9212	0.5516
25	187	13	0	0.0695	0.9305	0.5081
26	174	20	0	0.1149	0.8851	0.4728
27	154	22	0	0.1429	0.8571	0.4185
28	132	14	0	0.1061	0.8939	0.3587
29	118	12	0	0.1017	0.8983	0.3206
30	106	13	21	0.1226	0.8774	0.2880
31	72	3	22	0.0417	0.9583	0.2527
32	47	3	35	0.0638	0.9362	0.2422
33	9	1	8	0.1111	0.8889	0.2267
34						0.2015

Tatsächlich ist für das Schätzproblem nur die Sterblichkeit vor dem Ende der reproduktiven Phase von Bedeutung. Denn wenn man annimmt, dass alle Mitglieder der Referenzmenge Ω bis zum Alter τ_b (dem Ende der reproduktiven Phase) überleben, genügt es, eine einfache Verweildauervariable $\hat{T}$ zu betrachten, die folgendermaßen definiert ist: Wenn $\hat{T}(\omega) \leq \tau_b$ ist, ist dies das Alter bei der Geburt des ersten Kindes, andernfalls die Lebensdauer von ω. Bei dieser Definition liefert der Wert der Verteilungsfunktion $F[\hat{T}](\tau_b)$ den Anteil der Frauen mit mindestens einem Kind, und durch $1 - F[\hat{T}](\tau_b)$ erhält man den Anteil der Frauen, die kinderlos bleiben. Zwar können diese Anteile nicht korrekt geschätzt werden, wenn bereits vor dem Ende der reproduktiven Phase eine größere Anzahl rechts zensierter Beobachtungen vorkommt. Dennoch kann bis zum höchsten Alter, das in den Daten für die Geburt eines Kindes auftritt, die Verteilungsfunktion $F[\hat{T}]$ geschätzt werden; denn sie ist durch die Gleichung

$$F[\hat{T}](t) \;=\; 1 - G[\hat{T}](t + 1)$$

mit der Survivorfunktion $G[\hat{T}]$ verknüpft, die sich mithilfe des einfachen Kaplan-Meier-Verfahrens schätzen lässt.

Um diesen Gedankengang zu verfolgen, nehmen wir an, dass unsere Daten aus einer Referenzmenge stammen, deren Mitglieder bis zum Ende der reproduktiven Phase (oder auch nur bis zum höchsten beobachteten

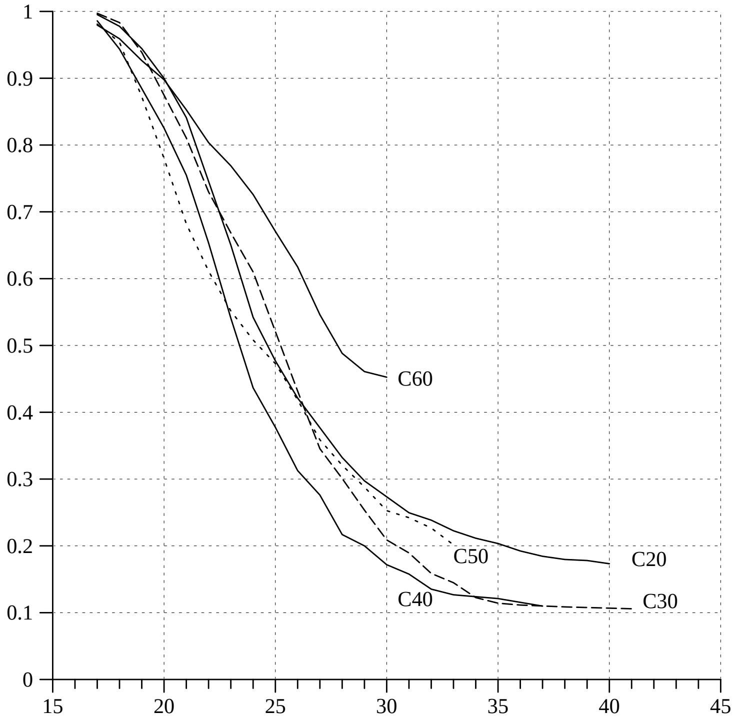


Abb. 7.3-2 Survivorfunktionen für das Alter bei der Geburt des ersten Kindes, berechnet mit den Daten aus Tabelle 7.3-2.

Alter bei der Geburt eines ersten Kindes) überleben.[17] Tabelle 7.3-3 illustriert die Berechnungen, wobei auf Daten für die Kohorte C50 aus der Tabelle 7.3-2 Bezug genommen wird.

Die Werte der geschätzten Survivorfunktion können unmittelbar interpretiert werden. Zum Beispiel bedeutet $G^*[\hat{T}](30) = 0.2880$, dass etwa 29 % der Frauen der Geburtskohorte C50 im Alter 30 noch kinderlos sind; umgekehrt haben bereits etwa 71 % mindestens ein Kind zur Welt gebracht.

7. Veränderungen des Alters bei der ersten Geburt. Die im vorangegangenen Paragraphen erläuterten Berechnungen können für alle in Tabelle 7.3-2 unterschiedenen Kohorten durchgeführt werden, so dass man Hinweise auf Veränderungen erhält. Abbildung 7.3-2 zeigt die berechneten Survivorfunktionen, also die Anteile der Frauen, die bis zum jeweiligen

[17] Da die Sterblichkeit von Frauen bis zum Ende der reproduktiven Phase in Deutschland sehr gering ist, erscheint diese Annahme relativ unproblematisch. In anderen Anwendungsfällen, wie z.B. bei der Schätzung von Ehedauern, müssen jedoch Mortalitätsprozesse explizit berücksichtigt werden (vgl. Rohwer 2006).

Tabelle 7.3-4 Anzahl der Kinder in den Lebensverlaufsdaten, differenziert nach Geburtskohorten und Alter (τ) der Mütter.

τ	C20	C30	C40	C50	C60
15					1
16	1			3	3
17	2	1	5	4	6
18	11	5	16	11	11
19	25	16	22	30	17
20	34	26	34	40	18
21	44	26	33	48	26
22	72	44	48	40	33
23	80	37	60	44	27
24	105	44	67	32	34
25	77	49	64	37	56
26	68	53	56	42	44
27	76	66	56	50	67
28	67	48	54	41	46
29	54	62	35	35	13
30	57	63	39	37	2
31	64	35	23	15	
32	41	43	23	9	
33	41	35	22	1	
34	42	33	13		
35	36	26	14		
36	39	22	11		
37	29	17	5		
38	19	15	1		
39	18	7	4		
40	10	2	2		
41	10	7	2		
42	3	5			
43	2	1			
44	3	1			
45	1				
46	1				
Insg.	1132	789	709	519	404

Alter noch kinderlos sind. Folgende Punkte erscheinen bemerkenswert:

- Bis zu einem Alter von etwa 27 Jahren sind die Survivorfunktionen bei den Kohorten C20 und C30 sehr ähnlich. Ab diesem Alter, also etwa Ende der 1950er Jahre, entsteht bei der Kohorte C30 ein vergleichsweise deutlich höherer Anteil von Frauen mit mindestens einem Kind. Schließlich ist der Anteil kinderloser Frauen in der Kohorte C30 deutlich kleiner als in der Kohorte C20.
- Verglichen mit C30, beginnen die Frauen der Kohorte C40 in einem jüngeren Alter mit der Geburt von Kindern. In beiden Kohorten gibt es aber schließlich einen ähnlich großen Anteil, etwa 90 %, von Frauen, die mindestens ein Kind zur Welt gebracht haben.
- Ähnlich wie in der Kohorte C40 beginnen auch die Frauen der Kohorte C50 mit der Geburt von Kindern in vergleichsweise jungen Jahren.

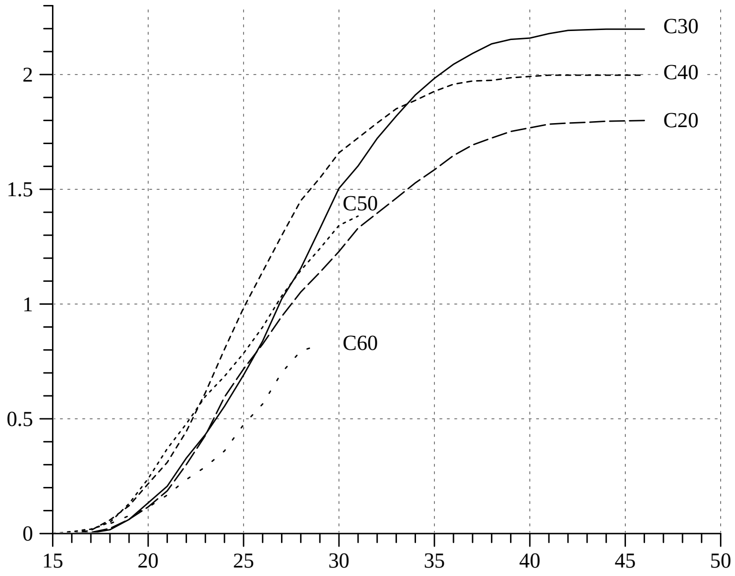


Abb. 7.3-3 Kumulierte Kohorten-Geburtenziffern in Abhängigkeit vom Alter (Abszisse), berechnet mit den Daten aus Tabelle 7.3-4.

Beginnend etwa in der Mitte der 1960er Jahre sinken jedoch die Geburtenraten, und man kann annehmen, dass (im Vergleich zu den beiden vorangegangenen Kohorten) ein erheblich größerer Anteil der Frauen der Geburtskohorte C50 schließlich kinderlos bleibt.

– Schließlich beginnen die Frauen der Kohorte C60 noch später mit der Geburt von Kindern, und man kann wiederum vermuten (aber mit den hier verfügbaren Daten natürlich nicht beweisen), dass ein noch größerer Anteil schließlich kinderlos bleibt.

8. Kumulierte Kohorten-Geburtenziffern. Jetzt verwenden wir die Lebensverlaufsdaten zur Berechnung kumulierter Kohorten-Geburtenziffern und vergleichen sie mit entsprechenden Angaben aus der amtlichen Statistik. Tabelle 7.3-4 zeigt die Daten. Wenn man von der (mit Retrospektivdaten grundsätzlich nicht erfassbaren) Sterblichkeit bis zum Interviewzeitpunkt absieht, genügt es zur Berechnung kumulierter Geburtenziffern $\bar{\gamma}_{c,\tau}$ für eine Geburtskohorte c, die Werte in der entsprechenden Spalte von Tabelle 7.3-4 bis zum Alter τ zu kumulieren und durch die Gesamtzahl der Frauen der betreffenden Kohorte zu dividieren. Abbildung 7.3-3 veranschaulicht den Verlauf der Funktionen $\tau \longrightarrow \bar{\gamma}_{c,\tau}$.

Folgende Tabelle vergleicht bis zum Alter τ (in der zweiten Spalte) zusammengefasste Kohorten-Geburtenziffern mit entsprechenden Größen

Tabelle 7.3-5 Anzahl von Frauen mit 0, 1, 2, 3, 4 und 5 oder mehr Kindern, berechnet aus den Lebensverlaufsdaten für die angegebenen Kohorten. Prozentangaben ohne Klammern beziehen sich auf alle Frauen der jeweiligen Kohorte mit mindestens einem Kind; Prozentangaben in Klammern geben den geschätzten Anteil der kinderlos bleibenden Frauen an.

	C20		C30		C40		C50		C60	
Kinder	N	%	N	%	N	%	N	%	N	%
0	109	(17)	38	(11)	39	(11)	86		231	
1	185	35.6	75	23.4	78	24.7	106	37.6	145	56.2
2	168	32.3	126	39.3	139	44.0	134	47.5	86	33.3
3	104	20.0	61	19.0	64	20.3	30	10.6	21	8.1
4	40	7.7	36	11.2	23	7.3	7	2.5	6	2.3
≥ 5	23	4.4	23	7.2	12	3.8	5	1.8		

der amtlichen Statistik:

Kohorte	τ	Lebensverlaufsdaten	Amtliche Statistik
C20	45	1.80	
C30	43	2.19	2.15
C40	40	1.99	1.96
C50	31	1.38	1.39
C60	29	0.82	0.99

Die Angaben in der Spalte ‘Amtliche Statistik’ wurden so berechnet, wie in Abschnitt 7.2 (§ 2) besprochen wurde. Es handelt sich um einfache Mittelwerte für die an den Geburtskohorten der Lebensverlaufsdaten beteiligten Geburtsjahre.[18] Offenbar liefern die Lebensverlaufsdaten und die amtliche Statistik sehr ähnliche Ergebnisse. Beide bestätigen die Annahme, dass die kumulierten Kohorten-Geburtenziffern zunächst, bei der Kohorte C30, gestiegen, dann jedoch zunehmend kleiner geworden sind.

9. Verteilungen für die Anzahl der Kinder. Kumulierte Kohorten-Geburtenziffern liefern Informationen über die von den Frauen einer Geburtskohorte insgesamt geborenen Kinder, nicht jedoch über die Verteilung der Kinderzahlen. Zu diesem Zweck muss die Anzahl von Kindern *pro Frau* betrachtet werden. Tabelle 7.3-5 zeigt, welche Informationen darüber mit den Lebensverlaufsdaten gewonnen werden können. Da für die Frauen der Kohorten C50 und C60 das Ende der reproduktiven Phase zum Zeitpunkt der Interviews noch nicht erreicht worden ist, sollte sich die Interpretation auf die Kohorten C20, C30 und C40 beschränken.

[18] Dafür wurden wiederum die in der Fachserie 1, Reihe 1, 1999: 198-200, für die Geburtsjahrgänge ab 1930 publizierten altersspezifischen Geburtenziffern verwendet. Für die Kohorte C20 stehen keine Daten zur Verfügung.

– Verglichen mit C20 hat ein größerer Anteil der Frauen der Kohorte C30 mindestens ein Kind zur Welt gebracht. Zugleich sank der Anteil der Frauen mit nur einem Kind, was zu einem Anstieg der durchschnittlichen Kinderzahl (bezogen auf alle Frauen mit mindestens einem Kind) von 2.2 in der Kohorte C20 auf 2.5 in der Kohorte C30 führte.
– Der Anteil der kinderlosen Frauen ist in der Kohorte C40 etwa gleich groß wie in der Kohorte C30. Es kann jedoch eine gewisse Tendenz zur Verringerung der Kinderzahlen pro Frau festgestellt werden. Insbesondere ist der Anteil der Frauen mit vier und mehr Kindern kleiner, dagegen der Anteil mit zwei Kindern größer geworden. Als Ergebnis ist die durchschnittliche Anzahl von Kindern von 2.5 bei der Kohorte C30 auf 2.3 bei der Kohorte C40 gesunken.

Kapitel 8

Demographische Projektionen

In diesem Kapitel besprechen wir einen Modellansatz, bei dem mithilfe von Annahmen über Geburten- und Sterberaten sowie über Zu- und Abwanderungen Regeln für die Entwicklung einer nach dem Alter und dem Geschlecht gegliederten Bevölkerung konstruiert werden. Wir beginnen im ersten Abschnitt mit einem Modell, das nur Geburten und Sterbefälle berücksichtigt. Dann folgt ein Abschnitt, in dem einige mathematische Eigenschaften des Modells besprochen werden. Schließlich wird in einem dritten Abschnitt das Modell erweitert, so dass auch Zu- und Abwanderungen berücksichtigt werden können.

8.1 Ein Makro-Modell ohne Migration

1. Notationen für die Modellkonstruktion. Wir beziehen uns zunächst auf einen demographischen Prozess $(\mathcal{R}, \mathcal{T}^*, \Omega_t)$ ohne externe Migration. $\mathcal{R}$ ist der nicht weiter differenzierte räumliche Kontext, $\mathcal{T}^*$ ist eine diskrete Zeitachse, und für jedes $t \in \mathcal{T}^*$ ist Ω_t die Menge der in dieser Zeitstelle lebenden Menschen.

Für Modelle zur Bevölkerungsprojektion wird von Buchführungsgleichungen ausgegangen, wie sie in Abschnitt 5.1 besprochen wurden. Wie dort ausgeführt wurde, gibt es unterschiedliche Varianten; für die Modellbildung eignen sich hauptsächlich zwei Varianten. Man kann sich auf den

Bevölkerungsstand zum Beginn bzw. Ende von Zeitstellen (fast immer Kalenderjahre) beziehen[1] oder auf die Gesamtzahl der Menschen, die während einer Zeitstelle leben. Wir gehen im Folgenden von der zweiten Variante aus, also (unter Verwendung der Notationen aus Abschnitt 5.1) von den altersspezifischen Buchführungsgleichungen

$$n_{t+1,0} = b_{t+1} \quad \text{und} \quad n_{t+1,\tau+1} = n_{t,\tau} - d_{t,\tau}$$

Man beachte, dass bei diesen Gleichungen das demographische Alter verwendet wird: $n_{t,\tau}$ erfasst die Anzahl der Personen, die in der Zeitstelle $t-\tau$ geboren wurden. Zur Vereinfachung der Notationen wird angenommen, dass es ein maximales Alter τ_m gibt.

Um das Modell zu formulieren, ist es zweckmäßig, einige Notationen aus der Matrizenrechnung zu verwenden.[2] Die nach dem Alter gegliederten männlichen und weiblichen Bevölkerungen werden durch Vektoren

$$\mathbf{n}_t^m := \begin{bmatrix} n_{t,1}^m \\ \vdots \\ n_{t,\tau_m}^m \end{bmatrix} \quad \text{und} \quad \mathbf{n}_t^f := \begin{bmatrix} n_{t,1}^f \\ \vdots \\ n_{t,\tau_m}^f \end{bmatrix}$$

repräsentiert. Für die Gesamtbevölkerung kann somit der Vektor $\mathbf{n}_t := \mathbf{n}_t^m + \mathbf{n}_t^f$ verwendet werden. Man beachte, dass die Zählung der Vektorelemente mit 1, nicht mit 0, beginnt; es werden also nur Personen erfasst, die mindestens das demographische Alter 1 erreichen.

Hiervon ausgehend kann der Zweck der Modellkonstruktion so formuliert werden: Es soll ein begrifflicher Rahmen entwickelt werden, um über mögliche Bevölkerungsentwicklungen

$$\mathbf{n}_0 \longrightarrow \mathbf{n}_1 \longrightarrow \mathbf{n}_2 \longrightarrow \cdots$$

die in einer beliebigen Zeitstelle $t=0$ mit einer anfänglichen Bevölkerung $\mathbf{n}_0$ beginnen, nachdenken zu können. Also müssen Regeln bestimmt werden, die es erlauben, von $\mathbf{n}_0$ zu $\mathbf{n}_1$, dann zu $\mathbf{n}_2$, $\mathbf{n}_3$ usw. zu gelangen. Da wir uns zunächst auf einen demographischen Prozess ohne externe Migration beziehen, genügt es, Geburten und Sterbefälle zu berücksichtigen. Da nur Frauen Kinder gebären können, muss der Prozess allerdings auf folgende Weise betrachtet werden:

$$\begin{tikzcd}
\mathbf{n}_0^m \arrow[r] & \mathbf{n}_1^m \arrow[r] & \mathbf{n}_2^m \arrow[r] & \cdots \\
\mathbf{n}_0^f \arrow[r] \arrow[ur] & \mathbf{n}_1^f \arrow[r] \arrow[ur] & \mathbf{n}_2^f \arrow[r] \arrow[ur] & \cdots
\end{tikzcd}$$

[1] Dieser Ansatz liegt z.B. den sog. koordinierten Bevölkerungsvorausberechnungen des Statistischen Bundesamts zugrunde. Eine ausführliche Darlegung der Berechnungsmethoden findet man bei M. Bretz (2000).

[2] Eine kurze Erläuterung dieser Notationen findet man bei Rohwer und Pötter (2002a, Anhang A). Wir übernehmen hier insbesondere die Konvention, Matrizen durch fettgedruckte Großbuchstaben und Vektoren durch fettgedruckte Kleinbuchstaben kenntlich zu machen.

Diese Darstellung soll andeuten, dass es möglich und zweckmäßig ist, zunächst ein Modell für die Entwicklung der weiblichen Bevölkerung zu entwickeln, woraus dann in einem zweiten Schritt auch Überlegungen zur Entwicklung der männlichen Bevölkerung gewonnen werden können.

2. Altersspezifische Geburten- und Sterbeziffern. Um Annahmen über Geburten und Sterbefälle zu formulieren, werden altersspezifische Geburten- und Sterbeziffern verwendet. Zur Notation altersspezifischer Sterbeziffern verwenden wir

$$\delta_{t,\tau}^m := d_{t,\tau}^m / n_{t,\tau}^m \quad \text{und} \quad \delta_{t,\tau}^f := d_{t,\tau}^f / n_{t,\tau}^f$$

Jeweils im Zähler steht die Anzahl der Männer bzw. Frauen, die in der Zeitstelle t im demographischen Alter τ gestorben sind. Zur Notation altersspezifischer Geburtenziffern verwenden wir

$$\beta_{t,\tau} := b_{t,\tau} / n_{t,\tau}^f$$

Im Nenner steht die Anzahl der Frauen des demographischen Alters $t-\tau$ in der Zeitstelle t, und im Zähler steht die Zahl der von ihnen in dieser Zeitstelle geborenen Kinder. Da bei der Modellbildung zwischen der Geburt von Jungen und Mädchen unterschieden wird und nur Kinder berücksichtigt werden, die mindestens das demographische Alter 1 erreichen, verwenden wir auch die Notationen

$$\beta_{t,\tau}^m := \sigma_m \beta_{t,\tau} (1 - \delta_{t,0}^m) \quad \text{und} \quad \beta_{t,\tau}^f := \sigma_f \beta_{t,\tau} (1 - \delta_{t,0}^f)$$

Hierbei bezeichnet σ_m den Anteil der männlichen und σ_f den Anteil der weiblichen Geburten ($\sigma_m + \sigma_f = 1$). Somit ist $\beta_{t,\tau}^m n_{t,\tau}^f$ die Anzahl der Jungen und $\beta_{t,\tau}^f n_{t,\tau}^f$ die Anzahl der Mädchen, die in der Zeitstelle t von Frauen des demographischen Alters τ geboren werden und in der Zeitstelle $t+1$ noch leben.

3. Ableitung der Modellgleichungen. Mit Hilfe der altersspezifischen Geburten- und Sterbeziffern können nun einfache Regeln formuliert werden, die es erlauben, die Bevölkerungsvektoren $\mathbf{n}_t^f$ und $\mathbf{n}_t^m$ im Zeitablauf fortzuschreiben. Zunächst zeigen die Gleichungen

$$n_{t+1,1}^f = \sum\nolimits_{\tau=1}^{\tau_m} \beta_{t,\tau}^f n_{t,\tau}^f \quad \text{und} \quad n_{t+1,1}^m = \sum\nolimits_{\tau=1}^{\tau_m} \beta_{t,\tau}^m n_{t,\tau}^f$$

wie aus den Geburten, die in der Zeitstelle t stattfinden, die Anzahlen der Kinder des Alters 1 in der Zeitstelle $t+1$ entstehen. Für die übrigen Altersklassen findet man die Anzahlen aus den Überlebenden der jeweils vorangegangenen Altersklasse, also durch

$$n_{t+1,\tau+1}^f = (1 - \delta_{t,\tau}^f)\, n_{t,\tau}^f \quad \text{und} \quad n_{t+1,\tau+1}^m = (1 - \delta_{t,\tau}^m)\, n_{t,\tau}^m$$

Diese Gleichungen können verwendet werden, um die Bevölkerungsvektoren $\mathbf{n}_t^f$ und $\mathbf{n}_t^m$ für alle $t > 0$ aus den Anfangsbevölkerungen $\mathbf{n}_0^f$ und $\mathbf{n}_0^m$

zu berechnen. Mit Hilfe von Matrizen kann das übersichtlich dargestellt werden. Zunächst werden zwei (τ_m, τ_m)-Matizen $\mathbf{B}_t^f$ und $\mathbf{B}_t^m$ definiert, die die altersspezifischen Geburtenziffern für Mädchen bzw. Jungen zusammenfassen:

$$\mathbf{B}_t^f := \begin{bmatrix} \beta_{t,1}^f & \cdots & \beta_{t,\tau_m}^f \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \quad \text{und} \quad \mathbf{B}_t^m := \begin{bmatrix} \beta_{t,1}^m & \cdots & \beta_{t,\tau_m}^m \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

Somit erhält man die Anzahl der Mädchen bzw. Jungen, die in der Zeitstelle $t+1$ das Alter 1 erreichen, durch das jeweils erste Elemente in den Vektoren $\mathbf{B}_t^f\,\mathbf{n}_t^f$ bzw. $\mathbf{B}_t^m\,\mathbf{n}_t^f$. Weiterhin werden zwei (τ_m, τ_m)-Matrizen $\mathbf{D}_{f,t}$ und $\mathbf{D}_{m,t}$ definiert, die die altersspezifischen Sterbeziffern für Frauen bzw. Männer zusammenfassen:

$$\mathbf{D}_t^f := \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1-\delta_{t,1}^f & 0 & \cdots & 0 & 0 \\ 0 & 1-\delta_{t,2}^f & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-\delta_{t,\tau_m-1}^f & 0 \end{bmatrix}$$

ist die Matrix für Frauen, und ganz analog wird die Matrix $\mathbf{D}_t^m$ für Männer definiert. Unter Verwendung dieser Matrizen kann man schließlich folgende Gleichungen für die Fortschreibung der Bevölkerungsvektoren bilden:

$$\mathbf{n}_{t+1}^f = \mathbf{D}_t^f\,\mathbf{n}_t^f + \mathbf{B}_t^f\,\mathbf{n}_t^f \quad \text{und} \quad \mathbf{n}_{t+1}^m = \mathbf{D}_t^m\,\mathbf{n}_t^m + \mathbf{B}_t^m\,\mathbf{n}_t^f$$

Die erste Gleichung für die Entwicklung der weiblichen Bevölkerung lässt sich noch weiter vereinfachen, indem man eine zusammengefasste Matrix $\mathbf{F}_t := \mathbf{B}_t^f + \mathbf{D}_t^f$ verwendet. Mittels dieser *Leslie-Matrix*[3] kann dann die Basisgleichung des *Leslie-Modells* kurz in der Form

$$\mathbf{n}_{t+1}^f = \mathbf{F}_t\,\mathbf{n}_t^f \tag{8.1}$$

geschrieben werden. Sie zeigt in kompakter Form, wie man aus dem Vektor für die weibliche Bevölkerung einer Zeitstelle t durch Multiplikation mit einer Leslie-Matrix den entsprechenden Vektor für die nächste Zeitstelle berechnen kann.

4. Implikationen konstanter Geburten- und Sterbeziffern. Das Modell (8.1) liefert einen allgemeinen Rahmen, um über mögliche Entwicklungen einer weiblichen Bevölkerung in Abhängigkeit von Annahmen über Geburten-

[3] Um an P. H. Leslie (1945) zu erinnern, der das hier beschriebene Modell zum erstenmal ausführlich analysiert und die besondere Form der Matrix $\mathbf{F}_t$ vorgeschlagen hat.

und Sterbeziffern nachzudenken. Insbesondere kann man die Frage stellen, welcher Entwicklungsprozess resultieren würde, wenn die Geburten- und Sterbeziffern langfristig konstant blieben. Dann kann eine zeitunabhängige Leslie-Matrix $\mathbf{F}$ angenommen werden, und das Basismodell wird zu

$$\mathbf{n}_{t+1}^f = \mathbf{F}\,\mathbf{n}_t^f \tag{8.2}$$

Daraus gewinnt man auch sogleich die Gleichung $\mathbf{n}_t^f = \mathbf{F}^t\,\mathbf{n}_0^f$.

Diese Gleichung hat nun eine bemerkenswerte Implikation: Langfristig konvergiert die Bevölkerungsentwicklung gegen einen Entwicklungspfad, bei dem folgende Beziehung gilt: $\mathbf{n}_{t+1}^f = (1+\rho^*)\,\mathbf{n}_t^f$. Das heisst, in allen Altersstufen wächst oder schrumpft die weibliche Bevölkerung mit der gleichen Rate ρ^*, so dass sich die Altersstruktur (die Verteilung der weiblichen Bevölkerung auf die unterschiedlichen Altersgruppen) nicht mehr verändert. Sie wird als *stabile Altersverteilung* und ρ^* wird als *intrinsische Wachstumsrate* der Leslie-Matrix $\mathbf{F}$ bezeichnet. (Mit einigen mathematischen Details dieser Implikationen des Modells beschäftigen wir uns im nächsten Abschnitt.)

5. Weibliche und männliche Bevölkerungen. Der eben dargestellte Gedankengang bezieht sich zunächst nur auf die Entwicklung der weiblichen Bevölkerung. Unter der Annahme, dass auch die Geburtenziffern $\beta_{t,\tau}^m$ und die Sterbeziffern $\delta_{t,\tau}^m$ langfristig konstant sind, gelangt man jedoch zu entsprechenden Aussagen für die männliche Bevölkerung. Man erkennt das, wenn man von der in § 3 abgeleiteten Gleichung

$$n_{t+1,1}^m = \textstyle\sum_{\tau=1}^{\tau_m} \beta_{t,\tau}^m n_{t,\tau}^f$$

ausgeht. Sie zeigt nämlich, dass sich eine Entwicklung der weiblichen Bevölkerung mit der langfristig konstanten Rate ρ^* (bei ebenfalls langfristig konstanter Altersverteilung) auf die Entwicklung der männliche Bevölkerung in der Altersklasse $\tau = 1$ überträgt. Wird schließlich ein Gleichgewichtszustand erreicht, ist $n_{t+1,1}^m = (1+\rho^*)\,n_{t,1}^m$. Bei langfristig konstanten Sterbeziffern δ_τ^m gilt jedoch

$$n_{t+\tau,\tau}^m = n_{t+1,1}^m \textstyle\prod_{j=1}^{\tau-1}(1-\delta_j^m)$$

und wenn sich $n_{t+1,1}^m$ mit einer konstanten Rate ρ^* verändert, gilt dies somit für die männliche Bevölkerung in allen Altersklassen, also auch für die männliche Bevölkerung insgesamt. Bei langfristig konstanten Geburten- und Sterbeziffern wächst oder schrumpft sie schließlich mit der gleichen Rate wie die weibliche Bevölkerung.

Ebenfalls resultiert schließlich eine stabile männliche Altersverteilung, die natürlich bei unterschiedlichen Sterbeziffern nicht mit derjenigen der weiblichen Bevölkerung übereinstimmt. Beide Altersverteilungen können auch leicht berechnet werden, sobald man die intrinsische Wachstumsrate

ρ^* kennt. Seien nämlich v^f_τ und v^m_τ die Besetzungen der Altersklassen in der weiblichen bzw. männlichen Bevölkerung, können sie ausgehend von zunächst beliebigen Anfangswerten (etwa $v^f_1 = v^m_1 = 1$) durch

$$v^f_\tau \;=\; \frac{1-\delta^f_{\tau-1}}{1+\rho^*}\, v^f_{\tau-1} \quad \text{bzw.} \quad v^m_\tau \;=\; \frac{1-\delta^m_{\tau-1}}{1+\rho^*}\, v^m_{\tau-1}$$

für $\tau = 2,\ldots,\tau_m$ rekursiv berechnet werden.

6. Eine Illustration mit Daten für die BRD. Zur Illustration des Modells verwenden wir die in der Tabelle 8.1-1 angeführten Daten:

- $\tilde{n}^m_{1999,\tau}$ und $\tilde{n}^f_{1999,\tau}$ sind für 1999 jahresdurchschnittliche Anzahlen von Männern bzw. Frauen, die im Jahr $1999-\tau$ geboren wurden;[4]
- $\tilde{d}^m_{1999,\tau}$ und $\tilde{d}^f_{1999,\tau}$ sind die Anzahlen der Männer bzw. Frauen, die 1999 im demographischen Alter τ gestorben sind;[5]
- $\tilde{b}_{1999,\tau}$ erfasst die Anzahl der Kinder, die 1999 von Frauen des Geburtsjahrgangs $1999-\tau$ geboren wurden.[6]

Die Altersklasse 95 umfasst alle Personen, die 1904 oder früher geboren wurden, und wird für die Modellberechnung nicht verwendet.

Aus den Angaben in dieser Tabelle haben wir die altersspezifischen Sterbeziffern durch

$$\tilde{d}^f_{1999,\tau}/\tilde{n}^f_{1999,\tau} \quad \text{bzw.} \quad \tilde{d}^m_{1999,\tau}/\tilde{n}^m_{1999,\tau}$$

berechnet. Die Berechnung der Geburtenziffern für Mädchen bzw. Jungen erfolgte nach der in § 2 besprochenen Vorgehensweise, wobei $\sigma_f = 0.486$ und $\sigma_m = 0.514$ angenommen wurden.[7] Beispielsweise für $\tau = 25$ findet man zunächst die altersspezifische Geburtenziffer $35493/444718 \approx 0.07981$; und durch Multiplikation mit

$$\sigma_f\,(1-\delta^f_{1999,0}) \;=\; 0.486\,(1-1268/187446) \;\approx\; 0.483$$

[4]Die Zahlen stammen aus der internen Tabelle B13c-1999 des Statistischen Bundesamts. Ich danke Hans-Peter Bosse, der diese Angaben zur Verfügung gestellt hat.

[5]Diese Zahlen wurden aus den Angaben in der Fachserie 1, Reihe 1, 1999: 232-234, berechnet. Bei einem Vergleich mit Tabelle 6.1-1 erkennt man die unterschiedliche Vorgehensweise. Zum Beispiel hatten von den 1517 Mädchen, die 1999 im Alter 0 starben, 1268 das Geburtsjahr 1999 und 249 das Geburtsjahr 1998, und von den 137 Mädchen, die 1999 im Alter 1 starben, hatten 67 das Geburtsjahr 1998 und 70 das Geburtsjahr 1997. Somit starben 316 Mädchen des Geburtsjahrs 1998. Offenbar erfasst die Zahl 1268 nicht alle 1999 geborenen Mädchen, die im Alter 0 gestorben sind.

[6]Diese Zahlen entsprechen denjenigen in der Tabelle 7.1-2. Bei der Verwendung der Zahlen für die Modellberechnung wird angenommen, dass die 80 Geburten der jüngsten Frauen im demographischen Alter 14 und die 16 Geburten der ältesten Frauen im demographischen Alter 51 stattfanden.

[7]Diese Werte ergeben sich daraus, dass 1999 insgesamt 396296 Jungen und 374448 Mädchen geboren wurden (Fachserie 1, Reihe 1, 1999: 42).

Tabelle 8.1-1 Für das Projektionsmodell verwendete Daten.

τ	Geburtsjahr	$\tilde{n}^m_{1999,\tau}$	$\tilde{n}^f_{1999,\tau}$	$\tilde{d}^m_{1999,\tau}$	$\tilde{d}^f_{1999,\tau}$	$\tilde{b}_{1999,\tau}$
0	1999	198967	187446	1653	1268	0
1	1998	404008	383127	418	316	0
2	1997	416773	395481	146	124	0
3	1996	409977	387882	103	57	0
4	1995	396310	376271	86	66	0
5	1994	401746	380029	65	44	0
6	1993	417330	395733	66	43	0
7	1992	426782	406028	66	50	0
8	1991	443616	421230	68	44	0
9	1990	488642	462841	69	51	0
10	1989	483080	459229	59	46	0
11	1988	496868	470524	70	38	0
12	1987	487395	460293	80	52	0
13	1986	477696	453391	71	53	0
14	1985	461993	438451	94	59	80
15	1984	461057	436782	119	66	341
16	1983	465441	441006	184	99	1234
17	1982	479905	452610	230	136	3085
18	1981	479347	454730	440	146	6332
19	1980	483538	460706	460	173	11158
20	1979	463719	442599	434	148	15558
21	1978	461012	440781	483	146	19693
22	1977	461198	443065	399	147	24009
23	1976	458894	440361	422	120	27326
24	1975	453389	432779	430	119	30436
25	1974	463793	444718	410	144	35493
26	1973	475000	454341	390	147	39850
27	1972	526608	500610	474	167	45348
28	1971	586626	555333	497	184	52632
29	1970	619612	582220	534	197	56566
30	1969	668147	626937	564	222	60007
31	1968	702583	657849	618	264	60093
32	1967	721025	677296	651	286	56767
33	1966	743011	696136	723	308	50623
34	1965	752728	699210	795	347	43428
35	1964	763581	713016	776	402	36185
36	1963	758935	710250	1014	436	28680
37	1962	734137	690981	1054	477	21055
38	1961	720491	684141	1109	574	15398
39	1960	701824	666236	1224	597	11165
40	1959	679685	646050	1302	636	7540
41	1958	647293	614752	1354	702	4627
42	1957	634515	603257	1499	744	2963
43	1956	619622	592163	1674	813	1619
44	1955	600493	577973	1745	887	789
45	1954	587969	573468	1913	959	342
46	1953	569320	560591	1965	1029	163
47	1952	571078	563369	2191	1123	58

Tabelle 8.1-1 (Forts.) Für das Projektionsmodell verwendete Daten.

τ	Geburtsjahr	$\tilde{n}^m_{1999,\tau}$	$\tilde{n}^f_{1999,\tau}$	$\tilde{d}^m_{1999,\tau}$	$\tilde{d}^f_{1999,\tau}$	$\tilde{b}_{1999,\tau}$
48	1951	561031	553593	2273	1160	48
49	1950	566366	558612	2556	1292	25
50	1949	550294	538511	2689	1318	12
51	1948	508496	495454	2634	1387	16
52	1947	479305	470252	2843	1422	0
53	1946	420860	413697	2651	1392	0
54	1945	376302	374914	2606	1326	0
55	1944	494656	492367	3781	1911	0
56	1943	506606	501819	4172	2009	0
57	1942	496592	492713	4331	2107	0
58	1941	594411	595754	5717	2799	0
59	1940	627760	634640	6559	3105	0
60	1939	613981	625152	7186	3468	0
61	1938	571615	587238	7320	3481	0
62	1937	532135	551748	7520	3552	0
63	1936	513671	541657	8188	3865	0
64	1935	493232	525302	8722	4271	0
65	1934	454851	490059	8723	4355	0
66	1933	365643	401128	7759	3908	0
67	1932	357615	401076	8454	4468	0
68	1931	358809	411323	9504	5153	0
69	1930	370874	436577	10878	6011	0
70	1929	353045	425943	11370	6484	0
71	1928	341756	424459	11871	7199	0
72	1927	299007	406401	11025	7809	0
73	1926	270961	412841	10866	8884	0
74	1925	248439	415761	11253	9900	0
75	1924	209879	396724	10366	10609	0
76	1923	195276	384738	10462	11745	0
77	1922	196209	390139	11713	13435	0
78	1921	189851	398062	12732	15615	0
79	1920	171751	378308	12730	17184	0
80	1919	121719	274411	9962	13802	0
81	1918	73777	171470	6675	10030	0
82	1917	64685	154190	6475	10229	0
83	1916	64182	160269	7086	11791	0
84	1915	73654	194921	9088	16554	0
85	1914	82969	229031	11155	21681	0
86	1913	73071	208933	10830	22502	0
87	1912	62744	185745	10297	22446	0
88	1911	48990	152462	8680	21009	0
89	1910	40056	131797	7807	20521	0
90	1909	31756	108852	7142	19087	0
91	1908	24021	85684	5783	16880	0
92	1907	18425	67373	4660	14535	0
93	1906	13760	50964	3566	12122	0
94	1905	10191	36069	2828	9464	0
95	1904	25934	85205	6063	25055	0

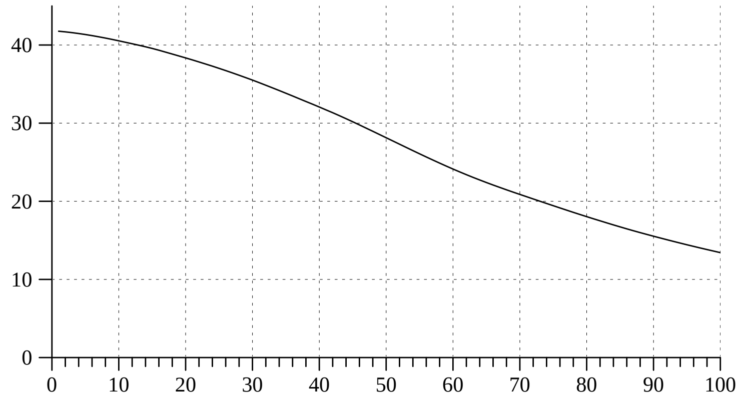


Abb. 8.1-1 Mit einem durch die Geburten- und Sterbeziffern des Jahres 1999 definierten Leslie-Modell berechnete Entwicklung der Anzahl von Frauen (in Mio.) für einen Zeitraum von 100 Jahren.

erhält man die für die Modellberechnung erforderlichen Geburtenraten für Mädchen, die mindestens das demographische Alter 1 erreichen.

Somit kann eine Leslie-Matrix $\mathbf{F}$ gebildet werden, die in diesem Fall 94 Zeilen und Spalten hat (für die Altersjahre 1 bis 94), und man kann Formel (8.2) verwenden, um sukzessive Bevölkerungsvektoren zu berechnen. Als Ausgangsvektor verwenden wir die in Tabelle 8.1-1 angegebenen Zahlen für die weibliche Bevölkerung im Jahr 1999, also

$$\mathbf{n}_0^f \equiv (\tilde{n}^f_{1999,1}, \ldots, \tilde{n}^f_{1999,94})'$$

Durch sukzessive Berechnungen erhält man $\mathbf{n}_1^f = \mathbf{F}\mathbf{n}_0^f$, $\mathbf{n}_2^f = \mathbf{F}\mathbf{n}_1^f$, $\mathbf{n}_3^f = \mathbf{F}\mathbf{n}_2^f$ usw. Die Addition der Komponenten dieser Vektoren liefert für jedes Jahr die Gesamtzahl der weiblichen Bevölkerung im Alter von 1 bis 94 Jahren. Abbildung 8.1-1 zeigt ihre Entwicklung für einen Zeitraum von 100 Jahren.

Offenbar sind die Veränderungsraten negativ. Wie Abbildung 8.1-2 zeigt, konvergieren sie gegen eine konstante intrinsische Wachstumsrate, die etwa den Wert -1.5 % hat. Wie im nächsten Abschnitt gezeigt wird, kann ihr Wert auch aus dem sogenannten dominanten Eigenwert der Leslie-Matrix berechnet werden. In unserem Beispiel findet man als dominanten Eigenwert $\lambda^* = 0.9854$ und daraus die intrinsische Wachstumsrate $\rho^* = \lambda^* - 1 \approx -1.46\,\%$.

Kennt man diese intrinsische Wachstumsrate und die altersspezifischen Mortalitätsraten, können, wie in § 5 gezeigt wurde, auch die im langfristigen Gleichgewicht erreichten stabilen Altersverteilungen berechnet werden. Abbildung 8.1-3 vergleicht sie mit den tatsächlichen Altersverteilungen des Jahres 1999, in der oberen Hälfte für die weibliche, in der unteren Hälfte

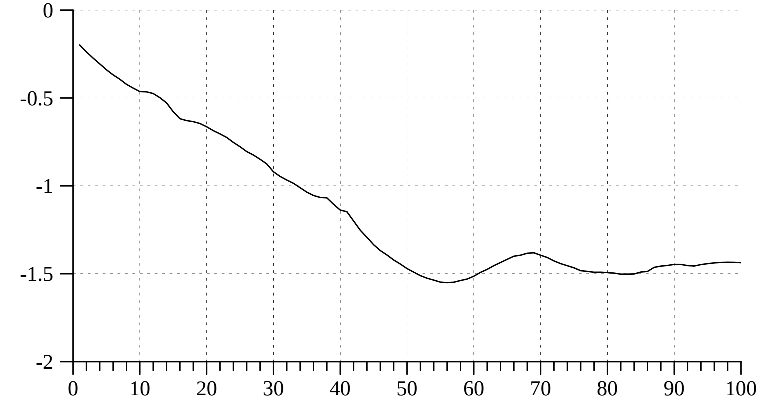


Abb. 8.1-2 Jährliche Veränderungsraten (in %) der in Abbildung 8.1-1 dargestellten weiblichen Bevölkerung.

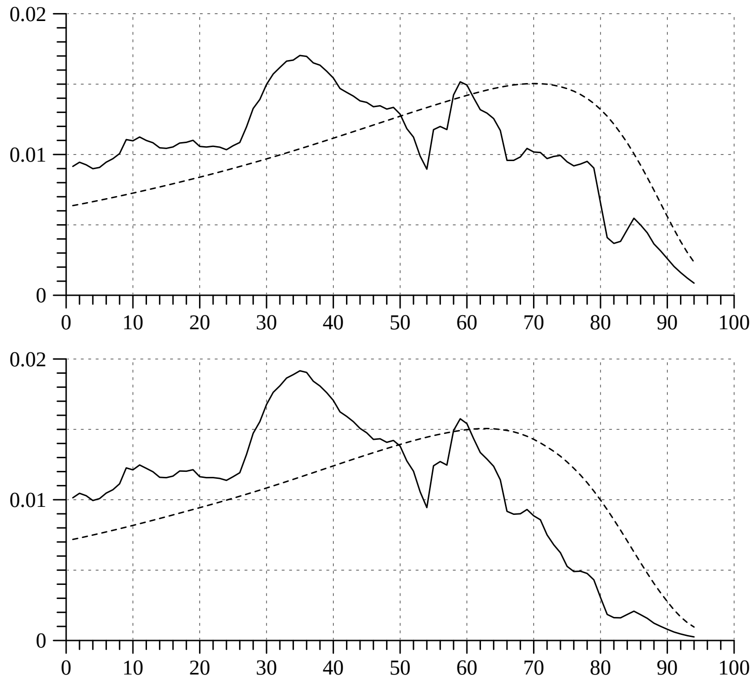


Abb. 8.1-3 Durchgezogene Linien: tatsächliche Altersverteilungen 1999 der weiblichen Bevölkerung (oberes Schaubild) und der männlichen Bevölkerung (unteres Schaubild); gestrichelte Linien: stabile Altersverteilungen bei einer intrinsischen Wachstumsrate von -1.46 %.

für die männliche Bevölkerung. Offenbar impliziert die negative intrinsische Wachstumsrate langfristig eine Umschichtung zugunsten der älteren Bevölkerung.

8.2 Mathematische Eigenschaften des Modells

In diesem Abschnitt wird besprochen, unter welchen Bedingungen bei einem einfachen Leslie-Modell eine intrinsische Wachstumsrate und eine stabile Altersverteilung existieren und ob sie auch vom anfänglichen Bevölkerungsvektor $\mathbf{n}_0^f$ oder nur von der Leslie-Matrix $\mathbf{F}$ abhängen.

1. Existenzbedingungen einer stabilen Altersverteilung. Wir beginnen mit der Frage, ob man für eine gegebene Leslie-Matrix $\mathbf{F}$ eine intrinsische Wachstumsrate und eine stabile Altersverteilung konstruieren kann. Denkt man an die im vorangegangenen Abschnitt erläuterte Struktur von $\mathbf{F}$, sind sicherlich alle Koeffizienten größer oder gleich Null, außerdem gilt $0 < \delta_\tau^f < 1$ für $\tau = 1, \ldots, \tau_m - 1$, so dass alle Koeffizienten in der Subdiagonalen größer als Null sind. Die Geburtenraten β_τ^f haben jedoch nur während der reproduktiven Phase von τ_a bis τ_b einen positiven Wert, und da $\tau_b < \tau_m$ ist, hat $\mathbf{F}$ keinen vollen Rang.

Wir gehen deshalb in zwei Schritten vor. In einem ersten Schritt betrachten wir nur die ersten τ_b Zeilen und Spalten von $\mathbf{F}$, d.h. die Teilmatrix

$$\tilde{\mathbf{F}} := \begin{bmatrix} \beta_1^f & \beta_2^f & \cdots & \beta_{\tau_b-1}^f & \sigma_f \beta_{\tau_b}^* \\ 1-\delta_1^f & 0 & \cdots & 0 & 0 \\ 0 & 1-\delta_2^f & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-\delta_{\tau_b-1}^f & 0 \end{bmatrix}$$

Dies ist jetzt eine nicht-negative Matrix mit vollem Rang.[8] Außerdem ist $\tilde{\mathbf{F}}$ eine unzerlegbare Matrix.[9] Man kann deshalb ein Theorem von G. Frobenius anwenden und schließen, dass $\tilde{\mathbf{F}}$ mindestens einen reellen positiven Eigenwert λ^* besitzt (der auch als ein *dominanter Eigenwert* von $\tilde{\mathbf{F}}$ bezeichnet wird), zu dem ein Eigenvektor $\mathbf{v}^* = (v_1^*, \ldots, v_{\tau_b}^*)'$ gehört, dessen Koeffizienten ebenfalls reell und positiv sind.[10] So gelangt man zu der

[8] Man erkennt das anhand der Determinante von $\tilde{\mathbf{F}}$:

$$\det(\tilde{\mathbf{F}}) = \pm \beta_{\tau_b}^f \prod_{\tau=1}^{\tau_b-1} (1-\delta_\tau^f) \neq 0$$

Das Vorzeichen hängt davon ab, ob τ_b gerade oder ungerade ist.

[9] Damit ist gemeint, dass man für je zwei Indizes i und j $(1 \leq i < j \leq \tau_b)$ weitere Indizes, etwa $k_1, \ldots, k_m$, finden kann, so dass $a_{ik_1} a_{k_1k_2} \cdots a_{k_mj} > 0$ ist. (Dies entspricht der Adjazenzmatrix eines unzerlegbaren Graphen.)

[10] Wir beziehen uns auf die Ausführungen bei F. R. Gantmacher (1971, Kap. xxiii).

Gleichung

$$\tilde{\mathbf{F}}\,\mathbf{v}^* = \lambda^*\,\mathbf{v}^* \tag{8.3}$$

Eine weitere Implikation des Theorems, die für die Diskussion unserer zweiten Frage verwendet wird, besteht darin, dass der Absolutbetrag (Modulus) aller übrigen Eigenwerte von $\tilde{\mathbf{F}}$ kleiner oder gleich λ^* ist.

Jetzt kann eine stabile Altersverteilung konstruiert werden. In einem ersten Schritt werden die Komponenten eines Vektors $\mathbf{n}^{f,*}$ durch

$$n_\tau^{f,*} := \begin{cases} v_\tau^* & \text{für } \tau = 1,\dots,\tau_b \\ \frac{1-\delta_{\tau-1}^f}{\lambda^*}\, n_{\tau-1}^{f,*} & \text{für } \tau = \tau_b+1,\dots,\tau_m \end{cases}$$

definiert. Aus Gleichung (8.3) und der Struktur von $\mathbf{F}$ folgt dann

$$\mathbf{F}\,\mathbf{n}^{f,*} = \lambda^*\,\mathbf{n}^{f,*} = (1+\rho_f^*)\,\mathbf{n}^{f,*} \tag{8.4}$$

wobei $\rho_f^* := \lambda^* - 1$ ist. Die durch $\mathbf{n}^{f,*}$ repräsentierte Altersverteilung verändert sich also nicht, wenn sie mit $\mathbf{F}$ multipliziert wird; alle Komponenten von $\mathbf{n}^{f,*}$ wachsen oder schrumpfen mit derselben Rate ρ_f^*. Um zur Definition einer stabilen Altersverteilung zu gelangen, ist es also nur erforderlich, $\mathbf{n}^{f,*}$ in Anteilswerte zu transformieren:

$$n_\tau^{f,p} := n_\tau^{f,*} / \textstyle\sum_{j=1}^{\tau_m} n_j^{f,*}$$

2. Ein einfaches Zahlenbeispiel. Zur Illustration verwenden wir ein einfaches Beispiel, bei dem es nur vier Altersklassen gibt ($\tau_m = 4$). Für die Geburtenziffern werden die Werte $\beta_1^f = 0$, $\beta_2^f = 1$, $\beta_3^f = 0.6$, $\beta_4^f = 0$, für die Sterbeziffern die Werte $\delta_1^f = 0.2$, $\delta_2^f = 0.3$, $\delta_3^f = 0.4$, $\delta_4^f = 1$ angenommen. Somit erhält man die Matrizen

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0.6 & 0 \\ 0.8 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \end{bmatrix} \quad \text{bzw.} \quad \tilde{\mathbf{F}} = \begin{bmatrix} 0 & 1 & 0.6 \\ 0.8 & 0 & 0 \\ 0 & 0.7 & 0 \end{bmatrix}$$

Als dominanten Eigenwert von $\tilde{\mathbf{F}}$ (und somit auch von $\mathbf{F}$) findet man $\lambda^* = 1.0573$ mit dem zugehörigen Eigenvektor $\mathbf{v}^* = (0.7405, 0.5603, 0.3710)'$. Aus diesem Eigenwert gewinnt man auch sofort die intrinsische Wachstumsrate $\rho_f^* = 0.0573$, die zur Berechnung der stabilen Altersverteilung verwendet werden kann: $\mathbf{n}^{f,p} = (0.39, 0.30, 0.20, 0.11)'$.

3. Konvergenz gegen eine stabile Altersverteilung. Jetzt wenden wir uns der zweiten Frage zu: ob die mit einem beliebigen Bevölkerungsvektor $\mathbf{n}_0^f$

beginnende Sequenz $\mathbf{n}_t^f = \mathbf{F}^t\mathbf{n}_0^f$ schließlich zu einem durch die intrinsische Wachstumsrate ρ_f^* und die Altersverteilung $\mathbf{n}^{f,p}$ definierten Gleichgewichtszustand führt. Wie sich zeigen wird, gibt es unter sehr allgemeinen Bedingungen eine positive Antwort.

Um die Argumentation zu erläutern, beginnen wir wie in § 1 mit der Teilmatrix $\tilde{\mathbf{F}}$ und definieren dazu korrespondierend den Vektor $\mathbf{n}_t^{f,a}$, der aus den ersten τ_b Elementen von $\mathbf{n}_t^f$ besteht. Somit gelangt man sofort zu der Gleichung

$$\mathbf{n}_t^{f,a} = \tilde{\mathbf{F}}^t\mathbf{n}_0^{f,a} \tag{8.5}$$

Hiervon ausgehend kann nun gezeigt werden, dass $\mathbf{n}_t^{f,a}$ gegen einen Vektor konvergiert, der proportional zum Eigenvektor $\mathbf{v}^*$ ist.

Dafür ist es erforderlich, sich auf alle Eigenwerte λ_j und Eigenvektoren $\mathbf{v}_j$ $(j = 1,\dots,\tau_b)$ der Matrix $\tilde{\mathbf{F}}$ zu beziehen. Einer dieser Eigenwerte, etwa $\lambda_{j^*} := \lambda^*$, ist der dominante Eigenwert mit dem zugehörigen Eigenvektor $\mathbf{v}_{j^*} := \mathbf{v}^*$. Somit kann man die Gleichungen $\tilde{\mathbf{F}}\mathbf{v}_j = \lambda_j\mathbf{v}_j$ (für $j = 1,\dots,\tau_b$) formulieren und sie, indem man die Definitionen $\mathbf{\Lambda} := \operatorname{diag}(\lambda_1,\dots,\lambda_{\tau_b})$ und $\mathbf{V} := (\mathbf{v}_1,\dots,\mathbf{v}_{\tau_b})$ verwendet, in Gestalt einer Matrizengleichung

$$\tilde{\mathbf{F}}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}$$

zusammenfassen. Wie bereits erwähnt wurde, hat $\tilde{\mathbf{F}}$ vollen Rang und ihre Eigenvektoren sind infolgedessen linear unabhängig. Daraus folgt, dass $\mathbf{V}$ eine invertierbare Matrix ist, so dass man auch zu den Gleichungen $\tilde{\mathbf{F}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ und $\tilde{\mathbf{F}}^t = \mathbf{V}\mathbf{\Lambda}^t\,\mathbf{V}^{-1}$ gelangt. Die zuletzt genannte Gleichung erlaubt nun die Formulierung

$$\mathbf{n}_t^{f,a} = \tilde{\mathbf{F}}^t\,\mathbf{n}_0^{f,a} = \mathbf{V}\mathbf{\Lambda}^t\mathbf{V}^{-1}\mathbf{n}_0^{f,a} = \mathbf{V}\mathbf{\Lambda}^t\mathbf{u}$$

bei der als Abkürzung $\mathbf{u} := \mathbf{V}^{-1}\mathbf{n}_0^{f,a}$ verwendet wird. Eine weitere Umformulierung führt zu

$$\mathbf{n}_t^{f,a} = (\mathbf{v}_1,\dots,\mathbf{v}_{\tau_b}) \begin{bmatrix} \lambda_1^t u_1 \\ \vdots \\ \lambda_{\tau_b}^t u_{\tau_b} \end{bmatrix} = \sum_{j=1}^{\tau_b} (\lambda_j^t u_j)\,\mathbf{v}_j$$

wodurch man erkennt, dass $\mathbf{n}_t^{f,a}$ als ein gewichteter Mittelwert der Eigenvektoren von $\tilde{\mathbf{F}}$ aufgefasst werden kann. Dividiert man schließlich durch $\lambda_{j^*}^t$, erhält man die Gleichung

$$\frac{1}{\lambda_{j^*}^t}\,\mathbf{n}_t^{f,a} = \sum_{j=1}^{\tau_b}\left(\frac{\lambda_j^t}{\lambda_{j^*}^t}\,u_j\right)\mathbf{v}_j = u_{j^*}\mathbf{v}_{j^*} + \sum_{j\neq j^*}\left(\frac{\lambda_j}{\lambda_{j^*}}\right)^t u_j\,\mathbf{v}_j \tag{8.6}$$

anhand derer man über die Konvergenzfrage nachdenken kann.

Aus dem Theorem von Frobenius weiß man bereits, dass für alle $j = 1, \ldots, \tau_b$ gilt: $\lambda_{j^*} \geq |\lambda_j|$. Jetzt nehmen wir zusätzlich an, dass $\lambda_{j^*} > |\lambda_j|$ ist, wenn $j \neq j^*$. Wenn diese Annahme (die weiter unten besprochen wird) zutrifft, folgt, dass der zweite Term auf der rechten Seite der Gleichung (8.6) gegen Null konvergiert; und dies impliziert die Konvergenz

$$\frac{1}{\lambda_{j^*}^t}\,\mathbf{n}_t^{f,a} \longrightarrow u_{j^*}\,\mathbf{v}_{j^*}$$

Wenn also t hinreichend groß ist, gilt näherungsweise $\mathbf{n}_{t+1}^{f,a} \approx \lambda_{j^*}\mathbf{n}_t^{f,a}$, so dass $\mathbf{n}_t^{f,a}$ näherungsweise proportional zum Eigenvektor $\mathbf{v}^*$ ist. Außerdem konvergieren die übrigen Komponenten von $\mathbf{n}_t^f$ ebenfalls gegen eine stabile Altersverteilung, da sie nur von der Veränderung der Anzahl von Frauen in der Altersklasse τ_b und den nach Voraussetzung konstanten Sterbeziffern abhängig sind. Wenn sich also schließlich die Anzahl der Frauen in der Altersklasse τ_b mit einer konstanten Rate verändert, verändern sich auch die Anzahlen in allen höheren Altersklassen mit derselben Rate; und somit kann auch die stabile Altersverteilung so berechnet werden, wie es in § 1 besprochen wurde.

Es bleibt also nur die Frage, ob man annehmen kann, dass der dominante Eigenwert betragsmäßig größer ist als alle anderen Eigenwerte. Das ist nicht unbedingt der Fall, wie die Matrix

$$\tilde{\mathbf{F}} := \begin{bmatrix} 0 & 1 \\ 0.8 & 0 \end{bmatrix}$$

zeigt, die zwei betragsgleiche Eigenwerte hat, nämlich 0.8944 und -0.8944. Wie Gleichung (8.6) zeigt, konvergiert in diesem Beispiel der Vektor $\mathbf{n}_t^{f,a}$ nicht gegen eine stabile Altersverteilung, sondern oszilliert zwischen zwei unterschiedlichen Verteilungen. Solche Fälle bilden jedoch Ausnahmen. Als hinreichende Bedingung für die Existenz eines dominanten Eigenwerts, der betragsmäßig größer als alle anderen Eigenwerte ist, genügt es bereits, dass es zwei aufeinanderfolgende Altersklassen mit einer positiven Geburtenrate gibt.[11] Wenn man also während der reproduktiven Phase mindestens zwei Altersgruppen unterscheidet, kann man von der Konvergenz gegen eine stabile Altersverteilung ausgehen.

[11]Dies wird bei Anton und Rorres (1991: 654) erwähnt, wo man auch eine gute Einführung in den mathematischen Hintergrund des Leslie-Modells findet. Einen expliziten Beweis der Behauptung hat Demetrius (1971) gegeben.

8.3 Berücksichtigung von Zu- und Abwanderungen

1. Erweiterung des Modellansatzes. Das bisher besprochene Modell bezieht sich auf einen demographischen Prozess ohne externe Migration. Um Zu- und Abwanderungen berücksichtigen zu können, orientieren wir uns an den in Abschnitt 5.1 (§ 4) eingeführten Notationen und an der ebenfalls dort erläuterten Buchführungsgleichung für einen demographischen Prozess mit externer Migration.

$m_{t,\tau}^{i,f}$ bzw. $m_{t,\tau}^{o,f}$ bezeichnen die Anzahlen der in der Zeitstelle t im demographischen Alter τ zu- bzw. abwandernden Frauen. Sie werden zu Vektoren

$$\mathbf{m}_t^{i,f} := \begin{bmatrix} m_{t,1}^{i,f} \\ \vdots \\ m_{t,\tau_m}^{i,f} \end{bmatrix} \quad \text{und} \quad \mathbf{m}_t^{o,f} := \begin{bmatrix} m_{t,1}^{o,f} \\ \vdots \\ m_{t,\tau_m}^{o,f} \end{bmatrix}$$

zusammengefasst, die zur Erweiterung des Modellansatzes (8.1) verwendet werden. Da aufgrund unserer Buchführungskonventionen die während einer Zeitstelle zu- bzw. abwandernden Personen als Teile der jeweiligen Bevölkerungsmengen betrachtet werden, kann der erweiterte Modellansatz folgendermaßen formuliert werden:

$$\mathbf{n}_{t+1}^f = \mathbf{F}_t\,(\mathbf{n}_t^f - \mathbf{m}_t^{o,f}) + \mathbf{m}_{t+1}^{i,f} = \mathbf{F}_t\,\mathbf{n}_t^f + (\mathbf{m}_{t+1}^{i,f} - \mathbf{F}_t\,\mathbf{m}_t^{o,f}) \qquad (8.7)$$

Vergleicht man diesen Modellansatz mit (8.1), werden jetzt die überlebenden Personen und ihre Kinder, die in der Zeitstelle t abwandern, abgezogen und die in der Zeitstelle $t+1$ zuwandernden Personen (einschließlich der in dieser Zeitstelle neugeborenen Kinder) hinzugezählt.

2. Ein Modell mit konstanter Migration. Eine besonders einfache Modellvariante entsteht, wenn man nicht nur eine konstante Leslie-Matrix $\mathbf{F}$ voraussetzt, sondern außerdem annimmt, dass sich $\mathbf{m}_{t+1}^{i,f} - \mathbf{F}\,\mathbf{m}_t^{o,f}$ im Zeitablauf nicht verändert, so dass mit einem konstanten Vektor

$$\mathbf{z}^f := \mathbf{m}_{t+1}^{i,f} - \mathbf{F}\,\mathbf{m}_t^{o,f}$$

gerechnet werden kann. Das Modell nimmt dann die Form

$$\mathbf{n}_{t+1}^f = \mathbf{F}\,\mathbf{n}_t^f + \mathbf{z}^f \qquad (8.8)$$

an. Beginnt man in einem Basisjahr $t = 0$, findet man

$$\mathbf{n}_1^f = \mathbf{F}\,\mathbf{n}_0^f + \mathbf{z}^f$$
$$\mathbf{n}_2^f = \mathbf{F}\,\mathbf{n}_1^f + \mathbf{z}^f = \mathbf{F}^2\,\mathbf{n}_0^f + \mathbf{F}\,\mathbf{z}^f + \mathbf{z}^f$$

usw., allgemein:[12]

$$\mathbf{n}_{t+1}^f = \mathbf{F}^t\mathbf{n}_0^f + \sum_{j=0}^{t-1}\mathbf{F}^j\mathbf{z}^f$$

[12]Dabei bedeutet $\mathbf{F}^0$ die Einheitsmatrix $\mathbf{I}$.

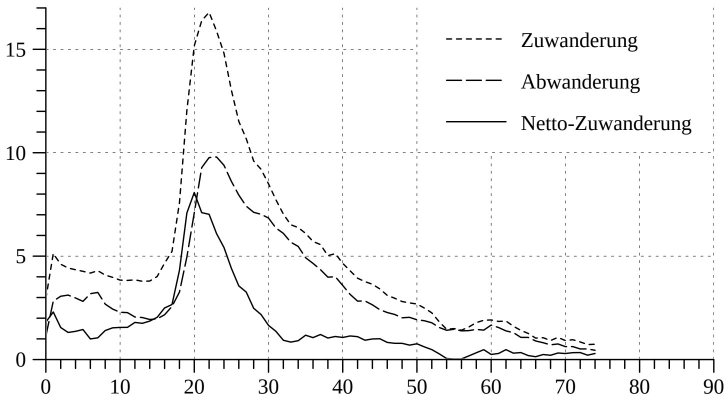


Abb. 8.3-1 Altersverteilung der weiblichen Zuwanderungen und Abwanderungen im Jahr 1999 in Deutschland. Daten aus Tabelle 8.3-1. Ordinate in 1000.

Anhand dieser Gleichung kann man sich auch überlegen, unter welchen Bedingungen ein Gleichgewichtspfad erreicht wird. Dies gilt jedenfalls dann, wenn die Leslie-Matrix $\mathbf{F}$ eine negative intrinsische Wachstumsrate aufweist und $\mathbf{z}^f \geq 0$ ist. Wird dann t immer größer, konvergiert $\mathbf{F}^t\mathbf{n}_0^f$ gegen Null, und der Bevölkerungsvektor $\mathbf{n}_t^f$ konvergiert gegen den Vektor[13]

$$\sum_{j=0}^{\infty} \mathbf{F}^j \mathbf{z}^f \;=\; (\mathbf{I} - \mathbf{F})^{-1}\, \mathbf{z}^f$$

Das heißt, langfristig wird die Bevölkerungsentwicklung nur von der Zuwanderung $\mathbf{z}^f$ bestimmt, und es entsteht allmählich ein Nullwachstum, bei dem der Bevölkerungsvektor $\mathbf{n}_t^f$ unverändert gleich $(\mathbf{I} - \mathbf{F})^{-1}\mathbf{z}^f$ ist.

3. Modellrechnungen mit konstanter Zuwanderung. Zur Illustration verwenden wir die in Tabelle 8.3-1 angegebenen Daten über die Zu- und Abwanderung von Frauen im Jahr 1999. Bei den Angaben handelt es sich um absolute Zahlen, gegliedert nach dem demographischen Alter; die letzte Altersklasse bei 75 ist nach oben offen und umfasst alle Altersgruppen ab 75 Jahren. Insgesamt sind 369049 Frauen zugewandert und 248108 Frauen abgewandert, so dass eine Nettozuwanderung von 120941 Frauen verbleibt. Abbildung 8.3-1 zeigt die Altersverteilungen bis zum Alter von 74 Jahren. Man erkennt, dass hauptsächlich jüngere Frauen zuwandern; da außerdem tendenziell eher ältere Frauen abwandern, sind bei der Nettozuwanderung die jüngeren Frauen vergleichsweise noch häufiger vertreten.

[13]Die Gleichung beruht auf folgendem mathematischen Satz: Wenn $\mathbf{A}$ eine beliebige nicht-negative quadratische Matrix mit einem dominanten Eigenwert kleiner als 1 ist, dann konvergiert die Reihe $\sum_{j=0}^{\infty} \mathbf{A}^j$ und ihr Grenzwert ist gleich $(\mathbf{I} - \mathbf{A})^{-1}$. Man vgl. J. T. Schwartz (1961: 31), wo auch auf weitere Literatur hingewiesen wird.

Tabelle 8.3-1 Nach dem demographischen Alter τ gegliederte Zuwanderung (1), Abwanderung (2) und Nettozuwanderung (3) von Frauen in Deutschland im Jahr 1999. Quelle: Statistisches Bundesamt, Fachserie 1, Reihe 1, 1999: 116f.

τ	(1)	(2)	(3)	τ	(1)	(2)	(3)
0	2942	1131	1811	38	5024	3983	1041
1	5123	2823	2300	39	5121	4006	1115
2	4609	3061	1548	40	4652	3583	1069
3	4428	3117	1311	41	4290	3150	1140
4	4342	2978	1364	42	3932	2825	1107
5	4266	2814	1452	43	3768	2832	936
6	4181	3184	997	44	3641	2646	995
7	4296	3245	1051	45	3416	2410	1006
8	4083	2680	1403	46	3103	2274	829
9	3975	2441	1534	47	2966	2180	786
10	3841	2287	1554	48	2805	2021	784
11	3826	2269	1557	49	2739	2041	698
12	3853	2058	1795	50	2686	1924	762
13	3785	2033	1752	51	2490	1877	613
14	3796	1937	1859	52	2260	1779	481
15	4015	1971	2044	53	1828	1551	277
16	4661	2166	2495	54	1466	1415	51
17	5231	2570	2661	55	1497	1466	31
18	7577	3263	4314	56	1419	1387	32
19	12043	4957	7086	57	1576	1399	177
20	15172	7095	8077	58	1787	1458	329
21	16383	9280	7103	59	1899	1422	477
22	16788	9766	7022	60	1915	1668	247
23	15885	9796	6089	61	1843	1548	295
24	14811	9393	5418	62	1863	1385	478
25	13025	8615	4410	63	1610	1301	309
26	11514	7953	3561	64	1410	1072	338
27	10679	7416	3263	65	1262	1070	192
28	9600	7119	2481	66	1036	898	138
29	9199	7023	2176	67	1057	814	243
30	8495	6842	1653	68	920	712	208
31	7724	6360	1364	69	1071	754	317
32	7035	6102	933	70	918	626	292
33	6518	5674	844	71	960	626	334
34	6381	5470	911	72	851	513	338
35	6101	4923	1178	73	726	514	212
36	5710	4650	1060	74	740	447	293
37	5559	4348	1211	75	5050	3721	1329

Um den Unterschied zwischen einer Bevölkerungsentwicklung ohne und mit Zuwanderung zu illustrieren, nehmen wir an, dass die Nettozuwanderung von Frauen des Jahres 1999, insgesamt etwa 117800 Personen (in den Altersjahren 1 bis 74), in den folgenden Jahren unverändert bestehen bleibt. Dies ist unser Vektor $\mathbf{z}^f$, der jedoch nur bis zu einem Alter von 74 Jahren berücksichtigt wird. Die Leslie-Matrix $\mathbf{F}$ und der Vektor $\mathbf{n}_0^f$ für die

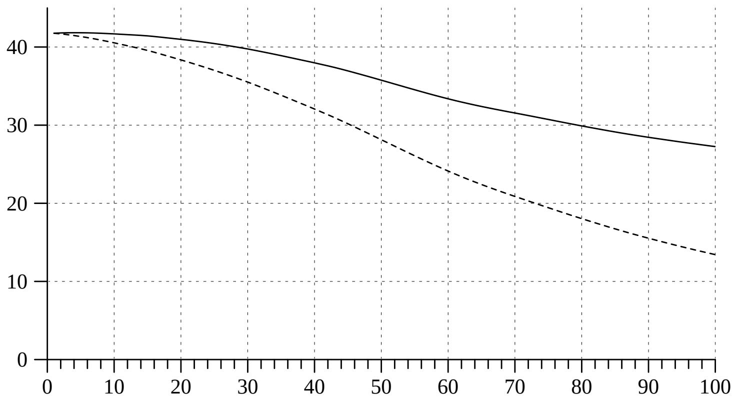


Abb. 8.3-2 Projektionen der weiblichen Bevölkerung (in Mio.) ohne Zuwanderung (gestrichelt) und mit einer jährlich konstanten Zuwanderung im Umfang von etwa 117800 Personen (durchgezogene Linie).

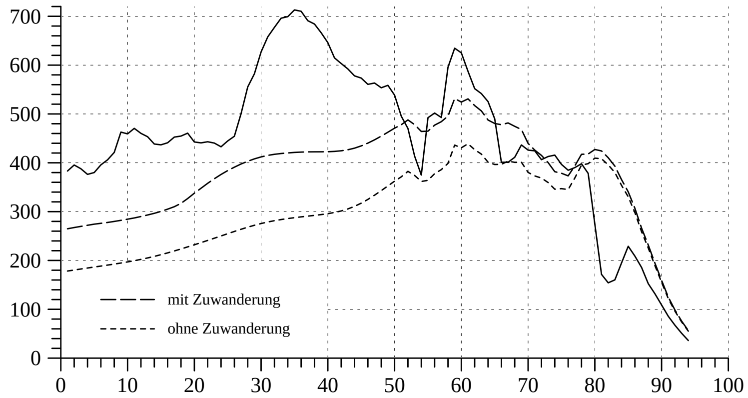


Abb. 8.3-3 Altersverteilung der weiblichen Bevölkerung 1999 (durchgezogene Linie) und im Jahr 2050 mit und ohne Zuwanderung (gestrichelte Linien). Ordinate in 1000.

weibliche Bevölkerung im Anfangsjahr 1999 werden wie in Abschnitt 8.1 gebildet. Durch iterative Anwendung der Gleichung (8.8) erhält man dann für alle folgenden Jahre einen Bevölkerungsvektor $\mathbf{n}_t^f$.

Abbildung 8.3-2 vergleicht die Bevölkerungsprojektionen mit und ohne Zuwanderung. Der gestrichelte Entwicklungspfad entspricht dem Verlauf in Abbildung 8.1-1. Offenbar würde eine konstante Zuwanderung von jährlich etwa 117800 Frauen (mit der hier vorausgesetzten Altersstruktur) den Bevölkerungsrückgang erheblich verringern.

Informativ ist auch ein Vergleich der Altersverteilungen. Abbildung 8.3-

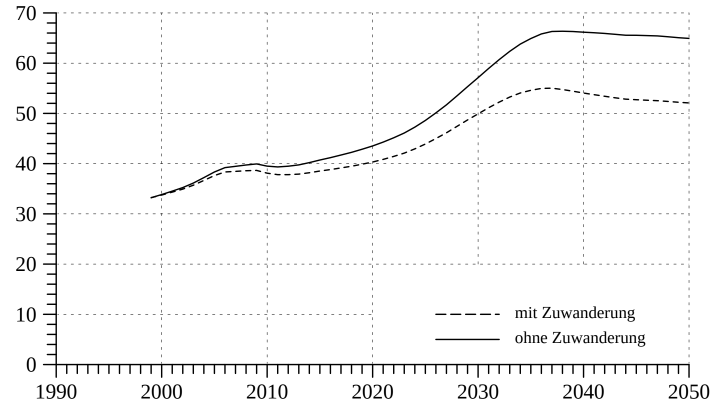


Abb. 8.3-4 Projektionen des Altenquotienten bis zum Jahr 2050 durch ein Leslie-Modell ohne Zuwanderung (durchgezogene Linie) und mit Zuwanderung (gestrichelte Linie).

3 vergleicht die jeweils mit und ohne Zuwanderung für das Jahr 2050 projektierten Altersverteilungen mit derjenigen des Jahres 1999. Da absolute Häufigkeiten verwendet werden, erkennt man zunächst das unterschiedliche Ausmaß des Bevölkerungsrückgangs. Bemerkenswert ist jedoch auch die geringere Umschichtung zugunsten älterer Personen bei einer Bevölkerungsentwicklung mit Zuwanderung. Um das sichtbar zu machen, kann auch ein sogenannter Altenquotient verwendet werden. Wir verwenden folgende Definition:

$$\text{Altenquotient} := \frac{\text{Anzahl Frauen im Alter 65 - 94}}{\text{Anzahl Frauen im Alter 20 - 64}} \quad (\text{in } \%)$$

Abbildung 8.3-4 vergleicht seine Entwicklung bei Bevölkerungsprojektionen ohne und mit Zuwanderung.

4. Das langfristige Gleichgewicht. Wie bereits gezeigt wurde, führt eine konstante positive Nettozuwanderung auch dann zu einem langfristigen Gleichgewicht mit Nullwachstum, wenn die intrinsische Wachstumsrate der Leslie-Matrix negativ ist. Der im Gleichgewicht erreichte konstante Bevölkerungsvektor ist

$$\bar{\mathbf{n}}^f = (\mathbf{I} - \mathbf{F})^{-1}\mathbf{z}^f \tag{8.9}$$

und kann somit aus der Leslie-Matrix und dem Zuwanderungsvektor $\mathbf{z}^f$ berechnet werden. Werden die Daten für das Jahr 1999 und der im vorangegangenen Paragraphen definierte Zuwanderungsvektor $\mathbf{z}^f$ verwendet, erhält man einen Vektor $\bar{\mathbf{n}}^f$, bei dem die Gesamtzahl der weiblichen Personen im Alter von 1 bis 94 Jahren etwa 19.5 Mio. beträgt.

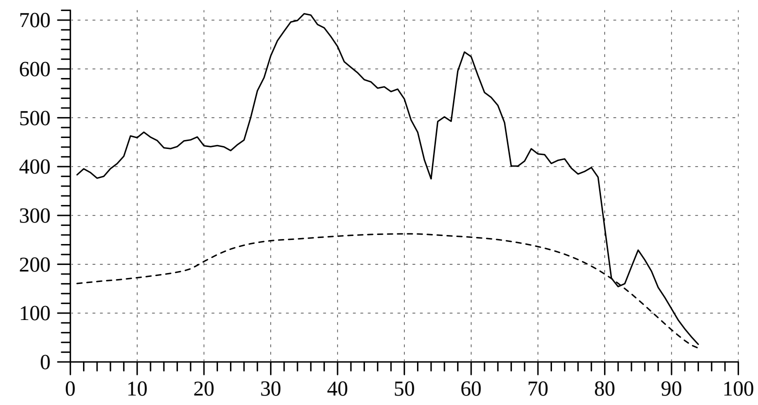


Abb. 8.3-5 Altersverteilung der weiblichen Bevölkerung 1999 (durchgezogene Linie) und stabile Altersverteilung bei einem Leslie-Modell mit konstanter Zuwanderung (gestrichelte Linie). Ordinate in 1000.

Das ist zwar deutlich weniger als der Stand im Jahr 1999; aber die Gleichung (8.9) ist linear, so dass sich eine prozentuale Erhöhung der Nettozuwanderung in einer prozentual gleichen Erhöhung der schließlich erreichten Bevölkerungszahl auswirken würde. Würde sich z.B. langfristig die Nettozuwanderung verdoppeln, würde schließlich eine Bevölkerungszahl von etwa 39 Mio. erreicht, die fast dem gegenwärtigen Stand entspricht.

Im langfristigen Gleichgewicht wird auch wieder eine stabile Altersverteilung erreicht. Abbildung 8.3-5 vergleicht sie mit der gegenwärigen Altersverteilung (in absoluten Häufigkeiten).

Kapitel 9

Haushalte und Netzwerke

9.1 Haushalte, Familien und Netzwerke
1. Definition des Haushaltsbegriffs.
2. Charakterisierung von Haushalten.
3. Haushalte als Untersuchungseinheiten.
4. Haushalte und Familien.
5. Beziehungen und Netzwerke.

9.2 Erfassung von Haushaltsstrukturen
1. Verteilung der Haushaltsgrößen.
2. Lebensalter und Haushaltsgröße.

9.3 Haushalte und persönliche Netzwerke
1. Erfassung persönlicher Netzwerke.
2. Anzahl der Bezugspersonen.
3. Bezugspersonen innerhalb und ausserhalb der Haushalte.
4. Umgangssprachliches Reden von Familien.

9.4 Varianten personeller Netzwerke
1. Personelle und personell konstituierte Netzwerke.
2. Durch Ereignisse definierte personelle Netzwerke.
3. Unterschiedliche Ansätze zur Definition von Gruppen.
4. Können Strukturen als Bedingungen interpretiert werden?
5. Knotenzentrierte Netzwerke.

Es gibt vier Abschnitte. Im ersten Abschnitt wird der Haushaltsbegriff besprochen; einige ergänzende Bemerkungen betreffen das Reden von Familien und Netzwerken. Im zweiten Abschnitt werden einige Aspekte von Haushaltsstrukturen und ihrer Veränderung anhand von Daten dargestellt; dabei beziehen wir uns zunächst nur auf die Haushaltsgröße, d.h. auf die Anzahl der Personen, die in einem Haushalt zusammen leben. Dann wird in einem dritten Abschnitt anhand von Daten aus dem Familiensurvey des Deutschen Jugendinstituts gezeigt, dass sich persönliche Netzwerke in den meisten Fällen nicht auf den jeweiligen Haushaltskontext beschränken. Schließlich werden im vierten Abschnitt in allgemeinerer Weise einige Varianten personeller bzw. personell konstituierter Netzwerke unterschieden.

9.1 Haushalte, Familien und Netzwerke

1. Definition des Haushaltsbegriffs. Das Statistische Bundesamt verwendet folgende Definition:

„Als *Haushalt* (Privathaushalt) zählt jede zusammenwohnende und eine wirtschaftliche Einheit bildende Personengemeinschaft sowie Personen, die allein wohnen und wirtschaften (z.B. Einzeluntermieter). Zum Haushalt können verwandte und familienfremde Personen gehören (z.B. Hauspersonal). Gemeinschafts- und Anstaltsunterkünfte gelten nicht als Haushalte, können aber Privathaushalte beherbergen (z.B. Haushalt des Anstaltsleiters.“ (Fachserie 1, Reihe 3, 1999: 12)

Die Definition macht deutlich, dass Haushalte als Institutionen (und nicht einfach als Mengen von Menschen) verstanden werden. Das wird besonders an den Verweisen auf Wohnungen deutlich. Folgt man der Definition, gehört zu jedem Haushalt eine Wohnung. Somit können Menschen auch mehrere Haushalte haben bzw. Mitglieder mehrerer Haushalte sein. Das zweite Kritierium, dass die Mitglieder eines Haushalts eine „wirtschaftliche Einheit“ bilden, ist vergleichsweise vage, weil es offen lässt, in welchem Ausmaß die wirtschaftlichen Ressourcen der Haushaltsmitglieder gemeinsam genutzt werden.

Die Definition bezieht sich auf Privathaushalte, aber nicht alle Menschen leben in Haushalten dieser Art. Einerseits gibt es obdachlose Personen, die keinem Haushalt angehören; andererseits gibt es Menschen, die in „Anstaltshaushalten“ leben. Im Unterschied zu privaten Haushalten gibt es für Anstaltshaushalte keine genaue Definition, man kann aber exemplarisch an folgende Beispiele denken: Hotels, Pensionen, Krankenhäuser, Sanatorien, Altersheime, Kasernen und Gefängnisse. Offenbar impliziert der Haushaltsbegriff weder bei Privat- noch bei Anstaltshaushalten eine dauerhafte Zugehörigkeit.

2. Charakterisierung von Haushalten. Man kann Haushalte unter zahlreichen Aspekten charakterisieren und daran anschließend Haushaltstypen unterscheiden. Zunächst kann man die bereits genannte Unterscheidung zwischen Privat- und Anstaltshaushalten vornehmen. Zur Charakterisierung von Privathaushalten gibt es dann folgende Möglichkeiten:

- Man kann von Merkmalen der Personen ausgehen, die in einem Haushalt zusammen leben, also Haushalte z.B. nach der Anzahl und dem Alter ihrer Mitglieder unterscheiden.
- Man kann Haushalte durch verwandtschaftliche und andere Arten von Beziehungen zwischen den Haushaltsmitgliedern charakterisieren.
- Man kann sich darauf beziehen, wie die zu den Haushalten gehörenden Wohnungen beschaffen sind. Außerdem kann ihre räumliche Lage berücksichtigt werden.

- Man kann Haushalte nach der Art und dem Umfang ihrer Ressourcen, insbesondere in Form von Einkommen und Vermögen, unterscheiden.
- Man kann Haushalte im Hinblick auf die in ihnen stattfindenden (hauswirtschaftlichen) Tätigkeiten charakterisieren, wobei sowohl die Art und der Umfang als auch die Verteilung dieser Tätigkeiten auf die Haushaltsmitglieder relevant sein können.

Im nächsten Abschnitt beziehen wir uns zur Charakterisierung von Haushalten auf die Anzahl ihrer Mitglieder; weitere Aspekte werden erst in späteren Kapiteln gelegentlich thematisiert.

3. Haushalte als Untersuchungseinheiten. Offenbar kann man anstelle von Personen auch Haushalte als Untersuchungseinheiten verwenden; beispielsweise kann man Verteilungen nach der Haushaltsgröße oder nach der Höhe des Haushaltseinkommens konstruieren. Unproblematisch ist das bei einzelnen sowie bei zeitlichen Folgen von Querschnittsbetrachtungen, die sich in jeder Zeitstelle auf eine separate Querschnittsgesamtheit beziehen. So kann man beispielsweise untersuchen, wie sich der Anteil der 1-Personenhaushalte im Zeitablauf verändert hat (s. Abschnitt 9.2).

Ob es sinnvoll ist, Haushalte auch bei Längsschnittbetrachtungen als Untersuchungseinheiten zu verwenden, muss im Einzelfall überlegt werden.[1] Insbesondere muss dann überlegt werden, wie man Veränderungen in der Anzahl und Zusammensetzung der Haushaltsmitglieder im Hinblick auf die Identität des Haushalts beurteilen möchte. Ein manchmal sinnvolles Identitätskriterium erhält man daraus, dass zu jedem Haushalt eine Wohnung gehört. Wir werden uns mit dieser Frage nicht näher beschäftigen und bei Längsschnittuntersuchungen Haushalte nur als Kontexte individueller Personen betrachten.

4. Haushalte und Familien. In der neueren Literatur werden Haushalte und Familien meistens unterschieden. Das Statistische Bundesamt verwendet folgende Definition:

„Als *Familie* im Sinne der amtlichen Statistik zählen – in Anlehnung an Empfehlungen der Vereinten Nationen – Ehepaare ohne und mit Kind(ern) sowie alleinerziehende ledige, verheiratet getrenntlebende, geschiedene und verwitwete Väter und Mütter, die mit ihren ledigen Kindern im gleichen Haushalt zusammen leben.“ (Fachserie 1, Reihe 3, 1999: 12)

Das ist offenbar eine sehr enge Definition, für die sich auch der Ausdruck *Kernfamilie* verbreitet hat. In der Literatur findet man auch weiter gefasste Definitionen; fast immer wird aber gefordert, dass es in einer Familie zumindest ein Elternteil und ein Kind geben muss. R. Nave-Herz (1994: 5f.) spricht in diesem Zusammenhang von einer „Generationsdifferenzierung“ und erläutert:

[1] Man vgl. dazu G. J. Duncan und M. S. Hill (1985).

„Es darf insofern hier [für den Familienbegriff] nur die Generationsdifferenzierung (also das Eltern- bzw. Mutter- oder Vater-Kind-Verhältnis) und nicht auch die Geschlechtsdifferenzierung, also nicht das Ehesubsystem, als essentielles Kriterium gewählt werden, weil es zu allen Zeiten und in allen Kulturen auch Familien gab (und gibt), die nie auf einem Ehesubsystem beruht haben oder deren Ehesubsystem im Laufe der Familienbiographie durch Rollenausfall, infolge von Tod, Trennung oder Scheidung, entfallen ist. Damit bilden alleinerziehende Mütter und Väter sowie nichteheliche Lebensgemeinschaften mit Kindern auch Familiensysteme.“

Offenbar wird der Familienbegriff von Nave-Herz anders gefasst als in der oben zitierten Definition des Statistischen Bundesamts. Eine weitere Frage betrifft das logische Verhältnis zwischen dem Familien- und dem Haushaltsbegriff. In der Definition des Statistischen Bundesamts wird gefordert, dass die Mitglieder einer Familie zugleich Mitglieder eines gemeinsamen Haushalts sind.[2] Orientiert man sich an den Ausführungen von Nave-Herz, aber auch am umgangssprachlichen Reden von Familien, ist das jedoch keine notwendige Bedingung, sondern die Mitglieder einer Familie können auch in unterschiedlichen Haushalten (Wohnungen) leben. Beim Reden von Familien muss also deutlich gemacht werden, in welcher Bedeutung dieser Begriff verwendet werden soll.[3]

5. Beziehungen und Netzwerke. Der Haushaltsbegriff bezieht sich auf die sachlichen und personellen Zusammenhänge, in denen Menschen wohnen und wirtschaften. Offenbar erschöpfen sich damit nicht die Beziehungen, die zwischen Menschen bestehen können. Will man also ein umfassenderes Bild von der sozialen Einbettung von Menschen gewinnen, genügt es nicht, nur ihre Haushalte zu erfassen, sondern man muss versuchen, die Beziehungen zu ermitteln, die sie zu anderen Personen sowohl innerhalb als auch außerhalb ihres Haushalts haben.

Damit beschäftigen wir uns in Abschnitt 9.3 anhand von Daten aus dem Familiensurvey des Deutschen Jugendinstituts. Anhand dieser Daten kann auch gezeigt werden, dass das übliche Reden von Familien sehr unscharf ist und sich jedenfalls nicht auf Haushalte beschränkt.

9.2 Erfassung von Haushaltsstrukturen

In diesem Abschnitt beschäftigen wir uns mit einigen Aspekten statistischer Haushaltsstrukturen. Dabei stützen wir uns zunächst auf Daten des

[2]Dies entspricht der statistischen Erfassung im Mikrozensus, bei dem zunächst Haushalte erfasst werden, bevor dann innerhalb der Haushalte Familien abgegrenzt werden.

[3]Die enge Definition der amtlichen Statistik ist bereits häufiger kritisiert worden; man vgl. etwa W. Bien und J. Marbach (1991), C. Geißler (1997). Weitere Hinweise auf neuere Diskussionen über Familienbegriffe findet man bei K. P. Strohmeier, A. Schultz und H. Strohmeier (2005: 30ff.); die Autoren sprechen von einer „Diversifikation und Dekonstruktion“ des Familienbegriffs, allerdings ohne die naheliegende Konsequenz zu ziehen, den Begriff in wissenschaftlichen Untersuchungen gar nicht zu verwenden.

Tabelle 9.2-1 Anzahl (in 1000) und Anteile (in %) der Privathaushalte mit 1, 2, 3, 4 oder 5 und mehr Mitgliedern im April 2002. Quelle: Fachserie 1, Reihe 3, Haushalte und Familien 2002, Teil 2, Tab. 1.1.

	1	2	3	4	5+	insgesamt
Alte Bundesländer	11658	10543	4295	3593	1456	31545
	37.0	33.4	13.6	11.4	4.6	100
Neue Bundesländer	2567	2517	1192	722	177	7175
	35.8	35.1	16.6	10.1	2.5	100

Mikrozensus, eine in der BRD seit 1957 (in den meisten Jahren) jährlich, seit 1991 auch in den neuen Bundesländern durchgeführte Datenerhebung der amtlichen Statistik, in der jeweils 1 % der in Privathaushalten lebenden Bevölkerung zu einer Vielzahl von Merkmalen und Lebensumständen befragt wird.[4] Da die Mikrozensusdaten nur in aggregierter Form verfügbar sind, werden zur Berechnung einiger Zusammenhänge auch Individualdaten aus dem ALLBUS verwendet.

1. Verteilung der Haushaltsgrößen. Haushalte können zunächst nach ihrer Größe, d.h. nach der Zahl ihrer Mitglieder, unterschieden werden. Tabelle 9.2-1 zeigt die durch den Mikrozensus im April 2002 ermittelten Zahlen. Man erkennt, dass die Größenverteilung in den alten und neuen Bundesländern sehr ähnlich ist. Den größten Anteil bilden 1-Personenhaushalte. Natürlich heißt dass nicht, dass die meisten Menschen in solchen Haushalten leben. Bezieht man sich auf die Bevölkerung in Privathaushalten, lebten im April 2002 in den alten Bundesländern 17.2 % (von 67.820 Mio.) und in den neuen Bundesländern 17.1 % (von 15.003 Mio.) in 1-Personenhaushalten.[5]

Ein wichtiger Prozess, der bereits in der zweiten Hälfte des 19. Jahrhunderts eingesetzt hat, besteht in der allmählichen Verringerung der Haushaltsgrößen. Abbildungs 9.2-1 illustriert diesen Prozess, wie er im Gebiet der ehemaligen BRD abgelaufen ist.[6] Man erkennt, wie die Anteile von Haushalten mit 1 und 2 Personen größer und die Anteile von Haushalten mit mehr als 2 Personen kleiner geworden sind.

2. Lebensalter und Haushaltsgröße. Ein bemerkenswerter Zusammenhang

[4]Ausführliche Informationen zum Mikrozensus findet man bei H. Rinne (1996: 69ff.) sowie – speziell zu Fragen der Stichprobenkonzeption – bei W. Krug, M. Nourney und J. Schmidt (1999, insb. S. 304ff.). Eine Zusammenstellung der Mikrozensusdaten über Haushalts- und Familienstrukturen gibt es in der Fachserie 1, Reihe 3, des Statistischen Bundesamts. Die gegenwärtig letzte Ausgabe, die sich auf den Mikrozensus im April 2002 bezieht, ist beim Statistischen Bundesamt (`www.destatis.de`) kostenlos erhältlich.

[5]Fachserie 1, Reihe 3, 2002, Teil 2, Tab. 7.12.

[6]Daten zur Entwicklung seit Mitte des 19. Jahrhunderts findet man bei M. Bretz und F. Niemeyer (1992).

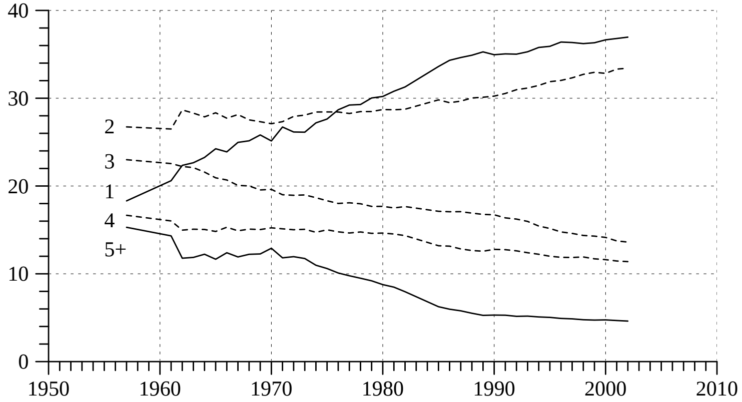


Abb. 9.2-1 Entwicklung der relativen Anteile (in %) unterschiedlicher Haushaltsgrößen an der Gesamtheit aller Haushalte im Gebiet der früheren BRD. Berechnung aus Daten des Statistischen Bundesamtes: Fachserie 1, Reihe 3, Haushalte und Familien 2002, Teil 2, Tab. 7.1.

besteht zwischen der Größe des Haushalts, in dem Menschen leben, und ihrem Alter. Um das sichtbar zu machen, verwenden wir Daten aus dem ALLBUS, eine Abkürzung für *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*. Diese Datenerhebung wird seit 1980 alle zwei Jahre durchgeführt;[7] hier verwenden wir Daten aus der Erhebung für das Jahr 2000, eine Personenstichprobe aus allen deutschsprachigen Personen, die zum Befragungszeitpunkt in Privathaushalten lebten und vor dem 1. Januar 1982 geboren sind.

Die Stichprobe umfasst insgesamt 3804 Personen, davon 2481 aus den alten Bundesländern (einschl. West-Berlin) und 1323 Personen aus den neuen Bundesländern (einschl. Ost-Berlin), im Alter von 18 bis 95 Jahren. Im Folgenden beschränken wir uns auf 2461 Personen aus den alten Bundesländern, bei denen sich die Größe ihres Haushalts feststellen lässt. Jeweils für alle Personen des gleichen Alters kann somit berechnet werden: die durchschnittliche Größe der Haushalte, in denen sie zum Befragungszeitpunkt lebten, und der Prozentanteil der Personen, die in 1-Personenhaushalten lebten. Als Untersuchungseinheiten werden hier also Personen, nicht Haushalte verwendet; die jeweilige Haushaltsgröße wird als eine Eigenschaft der jeweiligen individuellen Lebenssituation angesehen.[8]

Abbildung 9.2-2 zeigt zunächst die Abhängigkeit der durchschnittlichen Haushaltsgröße vom Alter. Als naheliegende Interpretation kann man ver-

[7] Ausführliche Informationen über diese Datenquelle findet man im Internet: `http://www.gesis.org/Datenservice/ALLBUS/index.htm`.

[8] Da es sich um eine Personenstichprobe handelt und wir uns auf die Teilstichprobe für die alten Bundesländer beschränken, ist keine Gewichtung erforderlich.

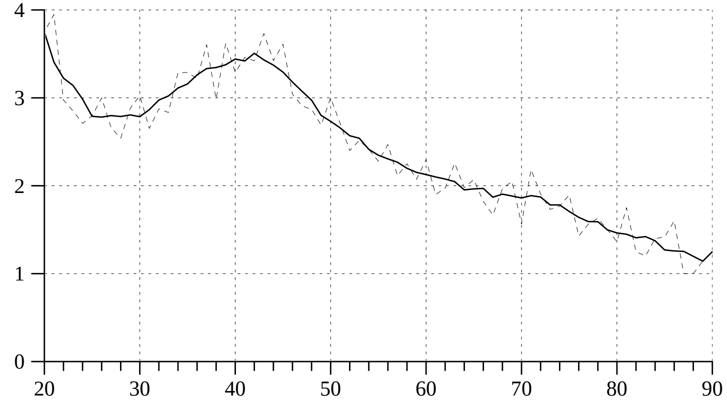


Abb. 9.2-2 Durchschnittliche Haushaltsgröße (Ordinate) in Abhängigkeit vom Alter (Abszisse), berechnet mit den Daten des ALLBUS 2000 für die alten Bundesländer. Gestrichelt: ungeglättet; durchgezogen: durch gleitende Durchschnitte geglättet.

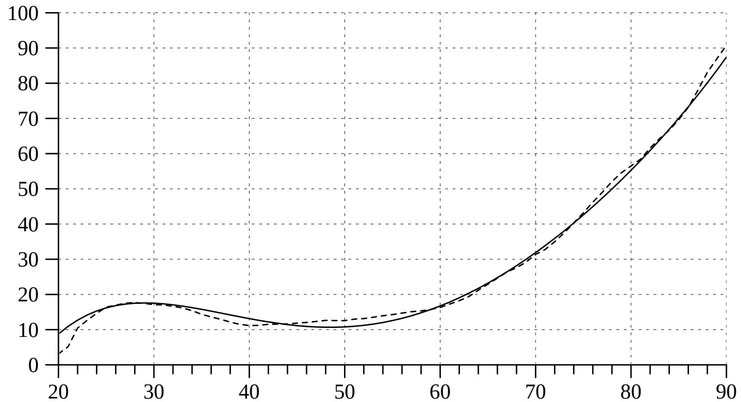


Abb. 9.2-3 Prozentanteil der in 1-Personenhaushalten lebenden Personen (Ordinate) in Abhängigkeit vom Alter (Abszisse), berechnet mit den Daten des ALLBUS 2000 für die alten Bundesländer. Geglättet mit gleitenden Durchschnitten (gestrichelt) und einer Splinefunktion (durchgezogene Linie).

muten, dass die durchschnittliche Haushaltsgröße zunächst dadurch kleiner wird, dass viele Kinder die elterlichen Haushalte verlassen und eigene Haushalte gründen; dann folgt durch Bildung von Partnerschaften und die Geburt von Kindern eine Phase größer werdender Haushalte; und schließlich werden die Haushalte wiederum durch den Auszug von Kindern, aber auch durch Trennungen und Todesfälle immer kleiner. Ganz analog kann die Altersabhängigkeit des Anteils der Personen in 1-Personenhaushalten interpretiert werden, die in Abbildung 9.2-3 gezeigt wird.[9]

9.3 Haushalte und persönliche Netzwerke

In Abschnitt 9.1 wurde bereits darauf hingewiesen, dass persönliche Beziehungen nicht nur innerhalb von Haushalten bestehen, sondern dass die Zugehörigkeit zu einem Haushalt nur einen mehr oder weniger großen Teil der sozialen Einbettung von Menschen ausmacht. Zur empirischen Illustration verwenden wir in diesem Abschnitt Daten aus dem Familiensurvey des Deutschen Jugendinstituts (DJI). Dieser Familiensurvey besteht aus einer Reihe von Datenerhebungen, die seit 1988 durchgeführt wurden.[10] In diesem Abschnitt verwenden wir nur einige Netzwerkdaten aus der 1988 in der damaligen BRD durchgeführten ersten Welle des Familiensurveys, in der 10043 Personen im Alter von 18 bis 55 Jahren befragt wurden.[11]

1. Erfassung persönlicher Netzwerke. Um die persönlichen Netzwerke zu erfassen, wurden in den Interviews der ersten Welle folgende Fragen gestellt:

– Mit wem besprechen Sie Dinge, die Ihnen persönlich wichtig sind?
– Mit wem nehmen Sie – ob an Wochenenden oder werktags – regelmäßig gemeinsame Mahlzeiten ein (ohne Kantine und Arbeitsessen)?
– Zu wem haben Sie eine sehr enge gefühlsmäßige Bindung?
– Von wem erhalten Sie ab und zu oder regelmäßig finanzielle Unterstützung?
– An wen geben Sie ab und zu oder regelmäßig finanzielle Unterstützung?
– Mit wem verbringen Sie hauptsächlich Ihre Freizeit?

Während diese Fragen gestellt wurden, wurde eine Liste aller Personen erstellt, die bei mindestens einer der Fragen erwähnt wurden; maximal konnten 20 Personen genannt werden. So entstand für jede Befragungsperson eine Liste mit den von ihr genannten Bezugspersonen, in der für jede Bezugsperson festgestellt wird, welche der in den Fragen thematisierten persönlichen Beziehungen zwischen ihr und der Befragungsperson bestanden.

[9] In dieser Abbildung wird auf eine Darstellung der ungeglätteten Anteilswerte verzichtet, da sie sehr stark schwanken. Beide Glättungsmethoden – mit gleitenden Durchschnitten und mit einer Splinefunktion (die mit der `spl`-Prozedur des Statistikprogramms TDA berechnet wurde) – führen jedoch zu ähnlichen Ergebnissen.

[10] Die Daten und ausführliche Dokumentationen sind durch das Zentralarchiv für empirische Sozialforschung in Köln erhältlich.

[11] Ausführliche Diskussionen dieser Daten findet man bei Bien und Marbach (1991) und Bien, Marbach und Templeton (1992).

Diese Liste wurde dann durch Namen weiterer Personen ergänzt (sofern vorhanden und nicht bereits in der Liste genannt):

– Eltern der Befragungsperson.
– Eltern des Ehe- oder Lebenspartners der Befragungsperson.
– Kinder der Befragungsperson.
– Weitere Personen aus dem Haushalt (ggf. auch mehreren Haushalten) der Befragungsperson.
– Weitere Personen, die von der Befragungsperson als Mitglieder ihrer Familie betrachtet werden (vgl.u. § 4).

Schließlich wurde für jede Bezugsperson, die auf der Liste genannt wurde, auch noch ermittelt:

– Ihr Geschlecht,
– ihr Alter (zum Interviewzeitpunkt) und
– welche der folgenden Beziehungen zwischen ihr und der Befragungsperson bestand: Ehepartner/Partner, geschiedener Partner/Expartner; eigenes Kind; Kind des Partners; Pflegekind; Schwiegersohn/-tochter; eigene Eltern; Eltern des Partners; eigene Geschwister; Geschwister des Partners; Großeltern/Urgroßeltern; Enkel; sonstige Verwandte; Freundeskreis; Arbeits- oder Studienkollege; Vereinsmitglied; Nachbarn; Sonstiges.

Auf diese Weise entstand für jede Befragungsperson eine Liste, die gewissermaßen ihr *persönliches Netzwerk* darstellt.[12]

2. Anzahl der Bezugspersonen. Von den befragten Personen haben 9985 mindestens eine Bezugsperson angegeben,[13] 4521 Männer mit insgesamt 29495 Bezugspersonen und 5464 Frauen mit insgesamt 37848 Bezugspersonen; Tabelle 9.3-1 zeigt die Häufigkeitsverteilungen, Abbildung 9.3-1 zeigt sie in graphischer Darstellung. Die durchschnittliche Anzahl der Bezugspersonen ist bei den Frauen mit 6.9 geringfügig größer als bei den Männern mit 6.5.

Differenziert man nach dem Alter der Befragungspersonen, ergibt sich, wie durch Abbildung 9.3-2 verdeutlicht wird, zunächst eine Zunahme, dann mit steigendem Alter eine Abnahme der Anzahl der Bezugspersonen.

3. Bezugspersonen innerhalb und ausserhalb der Haushalte. Wie erwähnt worden ist, können die Bezugspersonen danach unterschieden werden, ob sie zum Haushalt der Befragungsperson gehören oder nicht. Insgesamt sind bei den Bezugspersonen der Männer 71.3 % und bei den Bezugspersonen

[12] Man kann jedoch nur in einem eingeschränkten Sinn von einem ego-zentrierten Netzwerk sprechen, da Beziehungen zwischen den Bezugspersonen nicht erfasst wurden.

[13] Bei den übrigen 58 Befragungspersonen weiß man nicht genau, ob sie die Fragen nicht beantwortet haben oder keine Bezugspersonen hatten.

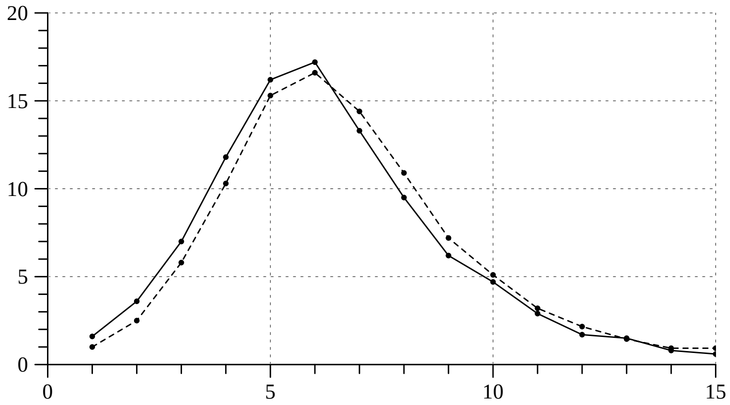


Abb. 9.3-1 Häufigkeitsverteilungen (in %) für die Anzahl der Bezugspersonen bei männlichen (durchgezogen) und weiblichen (gestrichelt) Befragungspersonen. Daten aus Tabelle 9.3-1.

Tabelle 9.3-1 Häufigkeitsverteilungen für die Anzahl der Bezugspersonen bei männlichen und weiblichen Befragungspersonen im DJI-Familiensurvey 1988.

Anzahl	Männer		Frauen	
Bezugspersonen	Anzahl	v.H.	Anzahl	v.H.
1	71	1.6	57	1.0
2	163	3.6	135	2.5
3	317	7.0	315	5.8
4	532	11.8	564	10.3
5	733	16.2	838	15.3
6	778	17.2	908	16.6
7	599	13.3	785	14.4
8	428	9.5	593	10.9
9	280	6.2	392	7.2
10	211	4.7	279	5.1
11	129	2.9	172	3.2
≥ 12	280	6.2	426	7.8
Insgesamt	4521	100	5464	100

der Frauen 71.6 % keine Mitglieder des jeweiligen Haushalts. Wie Abbildung 9.3-3 zeigt, verändern sich die Anteilswerte mit dem Alter der Befragungsperson.

Da es in dieser Hinsicht kaum Unterschiede zwischen Männern und Frauen gibt, wurde auf eine Unterscheidung verzichtet. Für jedes Alter der Befragungspersonen (18 bis 55 Jahre) wurde berechnet, wieviel Prozent der Bezugspersonen zum Haushalt der Befragungspersonen gehören. Abbildung 9.3-3 zeigt den mithilfe gleitender Durchschnitte geglätteten Verlauf. Man erkennt, dass der Anteil der Bezugspersonen innerhalb des

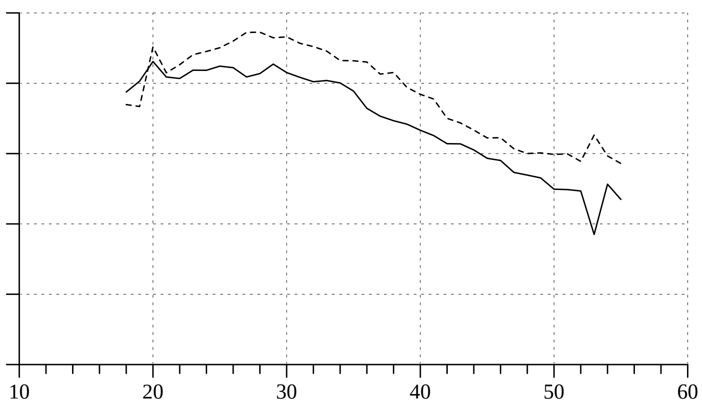


Abb. 9.3-2 Durchschnittliche Anzahl von Bezugspersonen (Ordinate) in Abhängigkeit vom Alter der Befragungspersonen (Abszisse), für Männer (durchgezogene Linie) und Frauen (gestrichelte Linie); berechnet mit den Daten des DJI-Familiensurveys 1988. Darstellung durch gleitende Durchschnitte.

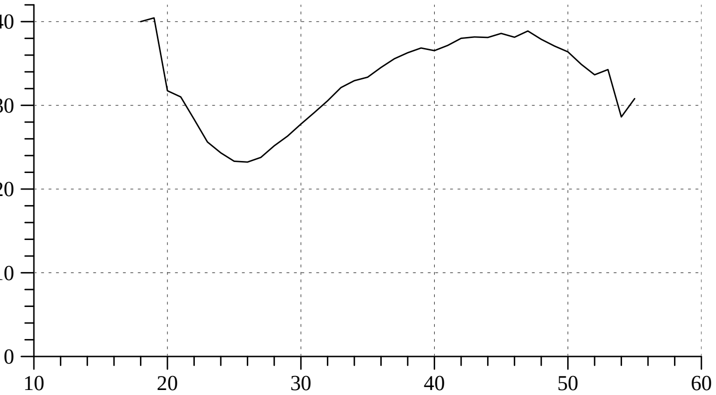


Abb. 9.3-3 Prozentanteile der Bezugspersonen, die zum Haushalt der Befragungsperson gehören (Ordinate), in Abhängigkeit vom Alter der Befragungspersonen (Abszisse); berechnet mit den Daten des DJI-Familiensurveys 1988. Darstellung durch gleitende Durchschnitte.

Haushalts zunächst kleiner (und dementsprechend der Anteil von Bezugspersonen außerhalb des Haushalts größer) wird, dann bis zu einem Alter von etwa 45 Jahren eine umgekehrte Entwicklung eintritt und schließlich der Anteil von Bezugspersonen im gleichen Haushalt erneut kleiner wird.

Bemerkenswert sind auch die Zusammenhänge zwischen der Haushaltsgröße und der Größe des persönlichen Netzwerks (Anzahl der Bezugsper-

Tabelle 9.3-2 Befragungspersonen und ihre Bezugspersonen im DJI-Familiensurvey 1988, differenziert nach der Haushaltsgröße der Befragungspersonen.

Haush.-größe	Anzahl Befragungspersonen	Anzahl Bezugspersonen pro Befragungsperson	Anteil Bezugspersonen außerhalb des Haushalts
1	1181	6.0	100.0
2	2371	6.2	79.5
3	2806	6.5	64.2
4	2483	7.3	55.5
5	813	7.8	47.6
≥ 6	331	8.8	40.3

Tabelle 9.3-3 Nach Beziehungen zu den Befragungspersonen differenzierte Bezugspersonen. Außerdem wird angegeben, wieviele der Bezugspersonen von den Befragungspersonen als Familienmitglieder betrachtet werden.

Beziehung zur Befragungsperson	Anzahl Bezugspersonen	davon Familienmitglieder Anzahl	in %
Ehepartner/Partner	8204	6935	84.5
Geschieden/Expartner	252	67	26.6
Eigenes Kind	11721	10541	89.9
Kind des Partners	432	343	79.4
Pflegekind	81	59	72.8
Schwiegersohn/-tochter	515	439	85.2
Eigene Eltern	13221	8603	65.1
Eltern des Partners	9453	3355	35.5
Eigene Geschwister	6794	5402	79.5
Geschwister des Partners	1354	854	63.1
Großeltern/Urgroßeltern	1220	940	77.1
Enkel	508	407	80.1
Sonstige Verwandte	3429	2358	68.8
Freundeskreis	8340	383	4.6
Arbeits-/Studienkollege	636	17	2.7
Vereinsmitglied	274	13	4.7
Nachbarn	206	8	3.9
Sonstiges	445	136	30.6
Keine Angabe	258	145	56.2

sonen). Wie Tabelle 9.3-2 zeigt, nimmt zwar die Netzwerkgröße mit der Haushaltsgröße zu (was bereits aufgrund der Datenerhebung zu erwarten ist), jedoch keineswegs proportional. Die persönlichen Netzwerke von Personen, die in 1-Personenhaushalten leben, sind im Durchschnitt nicht wesentlich kleiner als diejenigen von Personen, die in größeren Haushalten leben.

4. Umgangssprachliches Reden von Familien. Nachdem die persönlichen Netzwerke erfasst worden sind, wurde den Befragungspersonen auch noch

Tabelle 9.3-4 Nach Beziehungen zu den Befragungspersonen differenzierte Bezugspersonen. Jeweils gesondert für Bezugspersonen, die dem Haushalt der Befragungsperson angehören und nicht angehören, wird die Anzahl angegeben sowie der Prozentanteil derjenigen Bezugspersonen, die von den Befragungspersonen als Familienmitglieder betrachtet werden.

Beziehung zur Befragungsperson	Mitglieder des Haushalts Anzahl	% Fam.Mitglied	keine Mitglieder des Haushalts Anzahl	% Fam.Mitglied
Ehepartner/Partner	6652	93.8	1552	45.0
Geschieden/Expartner	10	80.0	242	24.4
Eigenes Kind	8173	95.7	3548	76.7
Kind des Partners	196	97.5	236	64.4
Pflegekind	33	100.0	48	54.2
Schwiegersohn/-tochter	27	92.6	488	84.8
Eigene Eltern	2438	95.8	10783	58.1
Eltern des Partners	196	78.1	9257	34.6
Eigene Geschwister	908	95.8	5886	77.0
Geschwister des Partners	25	68.0	1329	63.0
Großeltern/Urgroßeltern	136	90.4	1084	75.4
Enkel	27	92.6	481	79.4
Sonstige Verwandte	87	75.9	3342	68.6
Freundeskreis	167	24.6	8173	4.2
Arbeits-/Studienkollege	16	6.3	620	2.6
Vereinsmitglied	4	100.0	270	3.3
Nachbarn	2	50.0	204	3.4
Sonstiges	60	43.3	385	28.6
Keine Angabe	60	88.3	198	46.5

folgende Frage gestellt: „Nennen Sie mir zum Schluß bitte die Nummern der Personen, *die Sie persönlich zu ihrer Familie zählen* bzw. schreiben Sie diese Personen neu auf Ihre Liste, falls sie bisher noch nicht aufgeführt sind.“ Somit können die Netzwerkdaten des Familiensurveys auch verwendet werden, um Einsichten in das umgangssprachliche Reden von Familien zu gewinnen.

Wie Tabelle 9.3-3 zeigt, muss es sich beim umgangssprachlichen Reden von Familienmitgliedern nicht unbedingt um verwandte Personen handeln; andererseits werden verwandte Personen auch nicht unbedingt als Familienmitglieder betrachtet.

In Tabelle 9.3-4 werden die Bezugspersonen außerdem danach unterschieden, ob sie zum Haushalt der Befragungsperson gehören oder nicht. Jeweils wird der Prozentanteil derjenigen Bezugspersonen angegeben, die von den Befragungspersonen als Familienmitglieder betrachtet werden. Offenbar gibt es nicht nur zahlreiche Familienmitglieder außerhalb der jeweiligen Haushalte, sondern auch zahlreiche Personen, die zwar den Haushalten angehören, jedoch nicht als Familienmitglieder angesehen werden.

9.4 Varianten personeller Netzwerke

1. Personelle und personell konstituierte Netzwerke. In diesem Abschnitt beschäftigen wir uns in etwas allgemeinerer Weise mit einigen Varianten personeller und personell konstituierter Netzwerke. Wir beginnen mit einer Unterscheidung:

a) Einerseits gibt es Netzwerke, deren Knoten sich auf individuelle Personen beziehen; dann sprechen wir von *personellen Netzwerken*.

b) Andererseits gibt es Netzwerke, deren Knoten sich nicht auf individuelle Personen beziehen, bei denen jedoch die Beziehungen zwischen den Knoten durch Personen (nicht unbedingt, aber in vielen Fällen auch durch persönliche Beziehungen) zustande kommen. In diesen Fällen sprechen wir von *personell konstituierten Netzwerken*.[14]

Unsere Definition personeller Netzwerke setzt nur voraus, dass sich die Knoten auf Personen (individuelle Akteure) beziehen, lässt es aber offen, welcher Art die Beziehungen sind. Insbesondere impliziert die Definition nicht, dass sich die Personen, zwischen denen eine Beziehung besteht, wechselseitig kennen. Somit können zur Konstruktion personeller Netzwerke sowohl komparative als auch kontextabhängige Beziehungen verwendet werden.

Für die Interpretation personeller Netzwerke ist es offenbar wichtig, ob und ggf. wie die jeweils erfassten Beziehungen eine modale Betrachtungsweise erlauben. Wenn im Alltag von „persönlichen Beziehungen“ gesprochen wird, findet meistens eine Bezugnahme sowohl auf faktische Aspekte (man kennt sich, lebt zusammen in einem Haushalt, arbeitet in der gleichen Abteilung usw.) als auch auf modale Aspekte (man kann sich ansprechen,

[14] Man kann hauptsächlich zwei Varianten unterscheiden:

a) Einerseits können Beziehungen durch identische Personen zustande kommen, die simultan an mehreren Knoten operieren können. Als ein Beipiel kann man an personelle Unternehmensverflechtungen denken, die dadurch zustande kommen, dass Personen in zwei oder mehr Unternehmen gleichzeitig bestimmte Positionen einnehmen.

b) Andererseits können Beziehungen zwischen nicht-personellen Knoten durch personelle Beziehungen vermittelt sein. Zum Beispiel kann man an Beziehungen zwischen Unternehmen denken, die durch persönliche Beziehungen zwischen ihren Managern zustande kommen. In formaler Notation gibt es in diesem Fall einerseits ein personelles Netzwerk $(\Omega, \mathcal{K})$ und andererseits eine Knotenmenge Ω^*, deren Elemente sich auf nicht-personelle Objekte (z.B. Unternehmen oder Städte oder Regionen) beziehen. Der Zusammenhang kommt nun zunächst durch eine Abbildung $g : \Omega \longrightarrow \Omega^*$ zustande, durch die die Mitglieder des personellen Netzwerks den nicht-personellen Objekten zugeordnet werden. Dann können Beziehungen zwischen je zwei Objekten aus Ω^* durch die Gesamtheit der personellen Beziehungen zwischen den ihnen zugeordneten Personen definiert werden.

Tabelle 9.4-1 Daten über die Teilnahme von 18 Frauen an 14 sozialen Ereignissen. Quelle: Homans (1951: 83).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
ω_1	Evelyn	x	x	x	x	x	x		x	x					
ω_2	Laura	x	x	x		x	x	x	x						
ω_3	Theresa		x	x	x	x	x	x	x	x					
ω_4	Brenda	x		x	x	x	x	x	x						
ω_5	Charlotte			x	x	x		x							
ω_6	Frances			x		x	x		x						
ω_7	Eleanor					x	x	x	x						
ω_8	Pearl						x		x	x					
ω_9	Ruth					x		x	x	x					
ω_{10}	Verne							x	x	x			x		
ω_{11}	Myra								x	x	x		x		
ω_{12}	Katherine								x	x	x		x	x	x
ω_{13}	Sylvia							x	x	x	x		x	x	x
ω_{14}	Nora						x	x		x	x	x	x	x	x
ω_{15}	Helen							x	x		x	x	x		
ω_{16}	Dorothy								x	x					
ω_{17}	Olivia									x		x			
ω_{18}	Flora									x		x			

um Rat fragen, um etwas bitten usw.) statt.[15]

2. Durch Ereignisse definierte personelle Netzwerke. Zur Konstruktion personeller Netzwerke können auch ereignisförmige Beziehungen verwendet werden, deren modale Betrachtung fragwürdig sein kann. Zur Illustration betrachten wir ein in der Literatur oft diskutiertes Beispiel, in dem ein personelles Netzwerk durch eine Folge von Ereignissen definiert wird. Tabelle 9.4-1 zeigt die Daten so, wie sie zuerst von George C. Homans (1951: 83) publiziert wurden.[16] Die Daten beziehen sich auf die Teilnahme von 18 Frauen an sozialen Ereignissen (z.B. Treffen in einem Club oder bei einem kirchlichen Abendessen). Für jedes von 14 zeitlich aufeinander folgenden Ereignissen wird angegeben, welche der Frauen an ihnen teilgenommen haben (in der Tabelle durch ein Kreuz markiert).

Diese Daten können auf zwei unterschiedliche Weisen zur Definition personeller Netzwerke verwendet werden, wobei angenommen wird, dass man sich auf die Gesamtheit der Frauen $\Omega := \{\omega_1, \ldots, \omega_{18}\}$ beziehen möchte:

a) Man kann eine zeitliche Folge von Netzwerken definieren. Für jedes

[15] In dieser ambivalenten Weise wird auch in der Literatur manchmal von „sozialen Beziehungen“ gesprochen, z.B. von C. Prendergast und J. D. Knottnerus (1994: 9): „A social relationship is an opportunity for social interaction, a history of shared experience, and a means of need-satisfaction.“

[16] Die Daten selbst stammen aus einer früheren Untersuchung von A. Davis, B. Gardner und M. Gardner aus dem Jahr 1941.

Tabelle 9.4-2 Aus den Daten in Tabelle 9.4-1 berechnete Adjazenzmatrix (Anzahlen gemeinsamer Teilnahme an den sozialen Ereignissen).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	1	0	0
3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
2	1	2	1	0	1	1	2	2	2	2	2	2	1	1	2	1	1
1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

Ereignis $t = 1, \ldots, 14$ gibt es dann eine relationale Variable

$$R_t : \Omega \times \Omega \longrightarrow \{0, 1\}$$

wobei $R_t(\omega, \omega')$ den Wert 1 bekommt, wenn ω und ω' gemeinsam am t-ten Ereignis teilgenommen haben, und andernfalls den Wert 0. Somit gibt es in diesem Fall eine Folge von 14 Adjazenzmatrizen $\mathbf{A}_t$.

b) Stattdessen kann man auch die Teilnahme an allen Ereignissen betrachten und Beziehungen durch die Anzahl der Ereignisse definieren, an denen jeweils zwei Frauen gemeinsam teilgenommen haben. In diesem Fall entsteht nur ein einfaches Netzwerk, das durch eine relationale Variable

$$R : \Omega \times \Omega \longrightarrow \{0, 1, 2, \ldots\}$$

repräsentiert werden kann, wobei nun $R(\omega, \omega')$ die Anzahl der Ereignisse angibt, an denen ω und ω' gemeinsam teilgenommen haben. Als Adjazenzmatrix erhält man dann $\mathbf{A} = \Sigma_t \mathbf{A}_t$.[17] Tabelle 9.4-2 zeigt diese Adjazenzmatrix und außerdem in der Hauptdiagonalen für jede Frau die Anzahl der Ereignisse, an denen sie teilgenommen hat.

3. Unterschiedliche Ansätze zur Definition von Gruppen. Homans hat die in Tabelle 9.4-1 angegebenen Daten verwendet, um Überlegungen zur Definition sozialer Gruppen zu illustrieren (man vgl. Homans 1951: 81ff.). Zum

[17]Erfasst man die in Tabelle 9.4-1 angegebenen Daten durch eine $(18, 14)$-Inzidenzmatrix $\mathbf{B}$ (wobei $b_{ij} = 1$ ist, wenn ω_i am j-ten Ereignis teilgenommen hat, und andernfalls $b_{ij} = 0$ ist), kann man die Adjazenzmatrix auch direkt berechnen: $\mathbf{A} = \mathbf{B}\mathbf{B}'$.

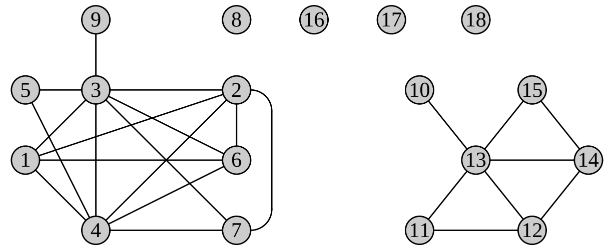


Abb. 9.4-1 Darstellung des Netzwerks mit den Daten aus Tabelle 9.4-2, wobei nur Beziehungen mit einem Wert ≥ 4 eingezeichnet sind.

Verständnis ist zunächst zu beachten, dass man dieses Definitionsproblem unterschiedlich konzipieren kann:

a) Man kann sich auf (in den meisten Fällen institutionalisierte) Kriterien für die Zugehörigkeit zu einer sozialen Gruppe beziehen. Natürlich muss man sich dafür an den Auffassungen einiger oder aller der jeweils beteiligten Personen orientieren.[18]

b) Andererseits kann man versuchen, soziale Gruppen gewissermaßen aus einer Beobachterperspektive mithilfe von äußerlich feststellbaren Daten über die Interaktion von Personen zu konstruieren.

Homans verfolgt den unter (b) genannten Ansatz; daran schließen sich seine Überlegungen zur Definition sozialer Gruppen an:

„We have been looking at the persons that participated together in social events. Our word for “participating together” is *interaction*: a group is defined by the interactions of its members. If we say that individuals A, B, C, D, E ... form a group, this will mean that at least the following circumstances hold. Within a given period of time, A interacts more often with B, C, D, E ... than he does with M, N, L, O, P ... whom we choose to consider outsiders or members of other groups. B also interacts more often with A, C, D, E ... than he does with outsiders, and so on for the other members of the group. It is possible just by counting interactions to map out a group quantitatively distinct from others.“ (Homans 1951: 84)

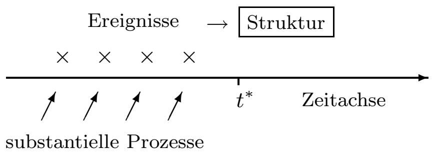
Die beiden Ansätze lassen sich mit unterschiedlichen Erkenntnisinteressen verbinden, so dass es keinen Widerspruch gibt. Im ersten Fall geht es um Gruppen, die durch ihre Wahrnehmung und normative Verankerung auch als Bedingungen für das Verhalten ihrer Mitglieder verstanden wer-

[18]Kriterien für die Mitgliedschaft in einer Gruppe können von ganz unterschiedlicher Art sein und müssen nicht auf „wechselseitige positive Gefühle“ der Gruppenmitglieder Bezug nehmen, wie dies von einigen Autoren vorgeschlagen wurde; man vgl. z.B. P. V. Marsden und E. O. Laumann (1984: 58) oder L. C. Freeman (1992: 152).

den können. Im zweiten Fall geht es um eine empirische Beobachtung von Interaktionen, deren potentielle Relevanz als Bedingungen weiterer Interaktionen zunächst gar keine Rolle spielt.

Der zweite Ansatz führt offenbar zu der Frage, wie ausgehend von einem durch Interaktionen definierten Netzwerk Gruppen konstruiert werden können. In der Literatur, die sich mit Methoden zur Darstellung und Analyse von Netzwerken beschäftigt, sind dafür zahlreiche Verfahren vorgeschlagen worden.[19] Einige dieser Verfahren wurden auch für die von Homans publizierten Daten verwendet. Zum Beispiel hat Linton C. Freeman (1992) unter Bezugnahme auf Mark Granovetter vorgeschlagen, zwischen „starken“ und „schwachen“ Beziehungen zu unterscheiden und Gruppen nur durch „starke“ Beziehungen zu bilden. Für die Daten in Tabelle 9.4-2 lautet sein Vorschlag, eine „starke“ Beziehung zwischen zwei Frauen dadurch zu definieren, dass sie sich mindestens viermal getroffen haben. Wie Abbildung 9.4-1 zeigt, gelangt man dann zu zwei Komponenten und vier isolierten Knoten. Natürlich entsprechen diesen Komponenten nicht unbedingt „soziale Gruppen“, die in der Wahrnehmung der beteiligten Personen handlungsrelevant sein könnten.

4. Können Strukturen als Bedingungen interpretiert werden? Das zuletzt angeführte Beispiel ist durchaus typisch dafür, wie man mithilfe formaler Methoden zu Charakterisierungen der Struktur eines Netzwerks gelangen kann. Bleiben wir bei diesem Beispiel, kann man auch feststellen, dass das, was als Struktur des Netzwerks beschrieben wird, das Ergebnis eines Prozesses ist und somit nicht als eine seiner Bedingungen verstanden werden kann. Das folgende Bild verdeutlicht diese Überlegung:



Die Struktur beschreibt einen Aspekt der Teilnahme der Frauen an den Ereignissen; sie kann also erst hinterher, etwa beginnend in einer Zeitstelle t^*, als ein Sachverhalt betrachtet werden. Und somit kann dieser Sachverhalt nicht als eine Bedingung der substantiellen Prozesse, die die Ereignisse hervorbringen, verstanden werden. Diese Überlegung bleibt offenbar auch bei einer dynamischen Betrachtung richtig, bei der man sich auf eine durch die Ereignisse erzeugte zeitliche Folge von Netzwerken bezieht (entsprechend der Variante (a) in § 3).

Die Frage, ob und ggf. in welcher Weise Aspekte der Struktur eines Netzwerks, wie z.B. die Existenz mehrerer Komponenten, als Bedingungen

[19]Eine umfassende Übersicht findet man bei S. Wasserman und K. Faust (1994, Kap. 7 und 8).

irgendwelcher Prozesse (insbesondere für das Verhalten beteiligter Akteure) verstanden werden können, wird tatsächlich von den formalen Methoden zur Gruppenbildung gar nicht berührt; denn dafür müsste man von einer modalen Betrachtung der durch das Netzwerk erfassten Beziehungen ausgehen oder sich mit dem Strukturbegriff auf die substantiellen Prozesse beziehen, die die Ereignisse hervorbringen, durch die das Netzwerk definiert wird.[20]

5. Knotenzentrierte Netzwerke. Zur Beantwortung der Frage, welche Bedeutung die Struktur eines personellen Netzwerks für das Verhalten der beteiligten Akteure haben kann, ist noch eine weitere Überlegung wichtig: dass es dabei für jeden Akteur nur darauf ankommt, wie er selbst in das gesamte Netzwerk eingebunden sind. Wie diese Überlegung formal präzisiert werden kann, hängt auch von der Art der durch das Netzwerk thematisierten Beziehungen ab. Wenn deren modale Interpretation voraussetzt, dass die jeweils beteiligten Personen sich kennen und direkt kommunizieren können, erscheint es plausibel, dass für jede Person nur ihre *lokale* Einbettung in das Gesamtnetzwerk relevant ist. Um dies formal zu erfassen, dienen *knotenzentrierte Netzwerke*, die bei personellen Netzwerken auch als *ego-zentrierte Netzwerke* bezeichnet werden.

Um den Begriff, der offenbar nicht nur für personelle Netzwerke verwendet werden kann, allgemein zu definieren, beziehen wir uns auf ein Netzwerk $(\Omega, \mathcal{K})$. Für jeden Knoten $\omega \in \Omega$ kann dann ein *knotenzentriertes Netzwerk* $(\Omega_\omega, \mathcal{K}_\omega)$ definiert werden, wobei gilt: Ω_ω besteht aus ω und allen denjenigen Elementen von Ω, die mit ω durch eine Kante in $\mathcal{K}$ verbunden sind; und $\mathcal{K}_\omega$ besteht aus allen Kanten aus $\mathcal{K}$, durch die Elemente von Ω_ω verbunden werden. Bezieht man sich auf das Netzwerk in Abbildung 9.4-1, ist z.B. das knotenzentrierte Netzwerk für den Knoten Nr. 13 mit der gesamten Komponente, der dieser Knoten angehört, identisch; dagegen umfasst das knotenzentrierte Netzwerk für die Nr. 15 zusätzlich nur die Knoten 13 und 14 sowie die drei Kanten, die diese Knoten verbinden.

[20]Um diese hier zunächst nur angedeutete Möglichkeit ernsthaft zu verfolgen, müssen Modelle für substantielle Prozesse konstruiert werden.

Kapitel 10

Soziale Ungleichheit

1. Eine allgemeine Definition. Der Begriff *soziale Ungleichheit*, wie er im Folgenden verwendet werden soll, bezieht sich auf Ungleichheiten, die in irgendeiner Hinsicht zwischen den Mitgliedern einer Gesellschaft festgestellt werden können. Wenn der Begriff in einer bestimmten Bedeutung verwendet werden soll, muss also angegeben werden, auf welche Merkmale der Gesellschaftsmitglieder bzw. ihrer Lebensumstände sich Feststellungen sozialer Ungleichheit beziehen sollen. Exemplarisch kann man an Ungleichheiten in der körperlichen Verfassung, in der schulischen und beruflichen Bildung, im Einkommen, in der Verfügung über Vermögen und in den Berufstätigkeiten denken. Offenbar kann man fast beliebig viele *Dimensionen sozialer Ungleichheit* unterscheiden.[1]

In der angegebenen Definition hat der Begriff keine unmittelbaren normativen Implikationen; die Definition lässt es offen, ob und ggf. wie soziale Ungleichheit auch als ein gesellschaftspolitisch zu bearbeitendes Problem verstanden werden sollte. Ich betone dies, weil man in der Literatur häufig von vornherein normative Definitionen sozialer Ungleichheit findet.[2] Natürlich soll die empirische Erforschung sozialer Ungleichheit auch dazu dienen, die jeweils thematisierten Aspekte normativen Beurteilungen zugänglich zu machen. Die dafür herangezogenen normativen Standards müssen jedoch jeweils explizit gemacht und können nicht durch einen Begriff sozialer Ungleichheit vorausgesetzt werden. Eine kritische Beurteilung sozialer Ungleichheit setzt auch nicht unbedingt eine egalitäre gesellschaftspolitische Position voraus; eine andere Möglichkeit besteht zum Beispiel darin, für Mindeststandards zu plädieren.

[1]Im Anschluss an Max Weber ist von einigen Soziologen versucht worden, drei „grundlegende Dimensionen“ sozialer Ungleichheit auszuzeichnen, nämlich ökonomische Verfügungsrechte, „Status“ und „Macht“ (man vgl. die kritische Diskussion bei Reinhard Kreckel 1992: 54ff.). Aber offenbar erschöpft diese Dreiteilung nicht alle Aspekte, unter denen relevante Ungleichheiten thematisiert werden können.

[2]Zum Beispiel hat Reinhard Kreckel (1992: 17) folgende Definition vorgeschlagen: „Soziale Ungleichheit im weiteren Sinne liegt überall dort vor, wo die Möglichkeiten des Zuganges zu allgemein verfügbaren und erstrebenswerten Gütern und/oder zu sozialen Positionen, die mit ungleichen Macht- und/oder Interaktionsmöglichkeiten ausgestattet sind, dauerhafte Einschränkungen erfahren und dadurch die Lebenschancen der betroffenen Individuen, Gruppen oder Gesellschaften beeinträchtigt bzw. begünstigt werden.“ Weitere Beispiele findet man bei S. Hradil (2000).

2. Soziale versus natürliche Ungleichheit? Von *sozialer* Ungleichheit wird gesprochen, um darauf hinzuweisen, dass die jeweils thematisierte Ungleichheit zwischen Mitgliedern einer Gesellschaft besteht, einen Aspekt ihrer gesellschaftlichen Verhältnisse bildet und durch diese in ihren Erscheinungsformen und Konsequenzen geprägt wird. Das Reden von sozialer Ungleichheit verweist also auf eine Betrachtungsweise von Ungleichheit und meint ausdrücklich nicht, dass es sich um eine besondere Art von Ungleichheit handelt.

Ich betone dies auch deshalb, weil man in der Literatur oft den Versuch findet, soziale und „natürliche“ Ungleichheit gegenüber zu stellen. Zum Beispiel beginnt Rousseaus Abhandlung über den Ursprung der Ungleichheit unter den Menschen mit folgender Überlegung:

> „Ich finde in der menschlichen Gattung zwei Arten der Ungleichheit. Die eine, die ich natürlich oder physisch nenne, weil sie von der Natur gesetzt ist und im Unterschied des Alters, der Gesundheit, der Körperkraft und der Eigenschaften des Geistes und der Seele besteht. Die andere, die man die moralische oder politische Ungleichheit nennen kann, weil sie von einer Art Übereinkunft abhängt. Sie ist durch die Zustimmung der Menschen gesetzt oder wenigstens ins Recht gesetzt worden. Diese besteht in den verschiedenen Privilegien, die einige zum Nachteil der andern genießen, wie etwa reicher, angesehener, mächtiger zu sein als andere oder Gehorsam von ihnen verlangen zu können.“ (Rousseau 1755/1978: 77)

Die Unterscheidung ist aber offenbar problematisch, da auch die „natürlichen“ Ungleichheiten ihre jeweils bestimmten Erscheinungsformen erst durch die gesellschaftlichen Verhältnisse gewinnen, in denen Menschen aufwachsen und leben; man denke etwa an die „natürliche“ Ungleichheit zwischen Männern und Frauen. Ein anderes Beispiel sind Altersunterschiede: Obwohl es vielleicht möglich ist, Altersunterschiede unabhängig von jeweils spezifischen sozialen Kontexten zu thematisieren, kann man sie sicherlich auch als einen Aspekt sozialer Ungleichheit reflektieren.

3. Statistische Erfassung sozialer Ungleichheit. Viele Formen sozialer Ungleichheit können durch statistische Variablen erfasst werden. Man erkennt dies unmittelbar anhand des Schemas

$$Y : \Omega \longrightarrow \tilde{\mathcal{Y}} \tag{10.1}$$

Repräsentiert Ω eine Gesamtheit von Menschen, wird durch die statistische Variable Y jeder zu dieser Gesamtheit gehörenden Menschen ein bestimter Wert im Merkmalsraum $\tilde{\mathcal{Y}}$, der die jeweils verwendete Ungleichheitsdimension (z.B. Bildungsabschlüsse oder Erwerbseinkommen) erfasst, zugeordnet. Die Häufigkeitsverteilung der Variablen, also $\mathrm{P}[Y]$, zeigt dann unmittelbar das Ausmaß, in dem sich die Mitglieder von Ω bzgl. des Merkmals

$\tilde{\mathcal{Y}}$ unterscheiden.

Ausgehend von dem Schema einer statistischen Variablen können Verallgemeinerungen und Varianten konzipiert werden.

– Mithilfe mehrdimensionaler statistischer Variablen können gleichzeitig mehrere Dimensionen sozialer Ungleichheit berücksichtigt werden.

– Es können unterschiedliche Referenzgesamtheiten verwendet werden, zum Beispiel Haushalte, aber auch Regionen, um regionale Ungleichheiten der Lebensbedingungen zu erfassen.

4. Soziale Ungleichheit zwischen Gruppen. Eine wichtige Verallgemeinerung betrifft den Vergleich von Merkmalsverteilungen zwischen Gruppen von Individuen (oder anderen Untersuchungseinheiten). Als Beispiel kann man an Unterschiede bei den Arbeitsverdiensten von Männern und Frauen denken. Das einfache Schema (10.1) genügt dann nicht mehr, sondern man benötigt eine zweidimensionale Variable

$$(Y, G) : \Omega \longrightarrow \tilde{\mathcal{Y}} \times \tilde{\mathcal{G}} \qquad (10.2)$$

durch die zusätzlich zur Ungleichheitsdimension Y (in diesem Beispiel der Arbeitsverdienst) eine Gruppenzugehörigkeit G (in diesem Beispiel das Geschlecht) erfasst wird. Die Feststellung sozialer Ungleichheit besteht dann in einem Vergleich der durch die Gruppenzugehörigkeiten $g \in \tilde{\mathcal{G}}$ bedingten Merkmalsverteilungen $\mathrm{P}[Y|G = g]$.

Der relevante Unterschied besteht darin, dass nicht mehr unmittelbar Ungleichheiten zwischen Individuen (oder anderen Untersuchungseinheiten) festgestellt werden, sondern Unterschiede zwischen statistischen Verteilungen. Dem entsprechen dann statistische Formulierungen wie zum Beispiel, dass die durchschnittlichen Arbeitsverdienste von Frauen niedriger sind als die der Männer.

5. Qualitative und quantitative Ungleichheitsdimensionen. Die Unterscheidung verläuft analog zu derjenigen zwischen qualitativen und quantitativen Merkmalsräumen statistischer Variablen.

– Von einer *quantitativen* Dimension sozialer Ungleichheit kann gesprochen werden, wenn sich ihre Ausprägungen auf eine sinnvoll interpretierbare Weise in eine lineare Ordnung bringen lassen; zum Beispiel: Bildung, Einkommen und Vermögen.

– Andernfalls handelt es sich um eine *qualitative* Dimension sozialer Ungleichheit, zum Beispiel: Beruf, Erwerbsstatus und Form des Zusammenlebens mit anderen Menschen.

Bei quantitativen Dimensionen kann man auch von „vertikaler“ Ungleichheit sprechen.[3]

[3] Das Reden von „vertikaler“ Ungleichheit (und Mobilität) wurde von Pitirim A. Sorokin verbreitet; die dafür relevanten Auszüge aus seinem Buch über „Social and Cultural Mobility“ findet man bei D. B. Grusky (1994: 245ff.). Offenbar verdankt sich der Ausdruck Sorokins Vorliebe für räumliche Metaphern.

6. Synthetische Ungleichheitskonstruktionen. In der empirischen Sozialforschung ist oft vorgeschlagen worden, mehrere Ungleichheitsdimensionen zu einem gemeinsamen Index sozialer Ungleichheit zusammenzufassen. Zum Beispiel ist vorgeschlagen worden, zur Konstruktion sozialer Schichten die Merkmale Bildung, Einkommen und berufliche Stellung zu kombinieren.[4] Konstruktionen dieser Art sind jedoch aus mehreren Gründen fragwürdig.

– Sie sind unvermeidlich mit einem erheblichen Informationsverlust verbunden. Fasst man zum Beispiel Bildung (X), Einkommen (Y) und berufliche Stellung (Z) zu einem additiven Index $S := X + Y + Z$ zusammen, können offenbar ganz unterschiedliche Kombinationen von Werten bei X, Y und Z zum gleichen Wert der Variablen S führen.

– Dieser Informationsverlust verstärkt sich noch, wenn die Werte des Index S anschließend gruppiert werden, um soziale Schichten zu definieren.

– Man verliert die Möglichkeit, Korrelationen zwischen den Komponenten eines synthetischen Ungleichheitsindex und anderen Aspekten sozialer Ungleichheit zu untersuchen.

– Es kann leicht der falsche Eindruck entstehen, dass mit einem synthetischen Ungleichheitsindex „vertikale“ soziale Ungleichheit erfasst werden kann. Aber selbst wenn es sich bei allen jeweils verwendeten Komponenten um quantitative Variablen handeln würde (was meistens nicht der Fall ist), lieferte eine additive Zusammenfassung nicht ohne weiteres wiederum eine quantitative Variable.

7. Quantifizierungen sozialer Ungleichheit. Eine andere Frage bezieht sich darauf, wie das Ausmaß und mithin Veränderungen sozialer Ungleichheit quantifiziert werden können. Wird soziale Ungleichheit durch statistische Variablen erfasst, zielt die Frage normalerweise darauf, Verteilungen mithilfe von Abstandsfunktionen bzw. Metriken zu vergleichen. Einige Möglichkeiten werden in Kapitel 11 besprochen.

[4] Eine Variante, auf die in der Literatur immer noch Bezug genommen wird, wurde von E. K. Scheuch und H. Daheim (1961) vorgeschlagen. Eine ausführliche Diskussion und Kritik findet man bei Rohwer und Pötter (2002a: 83ff.).

Kapitel 11

Bildungsungleichheit

11.1 Daten über Schulabschlüsse
1. Gliederungen des Bildungssystems.
2. Entwicklung der Schulanfänger.
3. Entwicklung der Schulabschlüsse.
4. Schulabschlüsse ausländischer Jugendlicher.

11.2 Kohortenanalysen zur Schulbildung
1. Daten aus dem ALLBUS.
2. Gliederung nach Geburtskohorten.
3. Veränderungen der Bildungsungleichheit.
4. Substitutionsmetriken für Verteilungen.

11.3 Schulbildung von Eltern und Kindern
1. Daten aus dem ALLBUS.
2. Veränderungen der Bildungsabstände.
3. Differenzierung nach der Schulbildung der Eltern.
4. Statistische Chancenvergleiche.

Lebensmöglichkeiten von Menschen hängen in erheblichem Maß von ihrer schulischen und beruflichen Ausbildung ab. In diesem Kapitel befassen wir uns mit der Entwicklung von Ungleichheiten bei den Schulabschlüssen. Es gibt drei Abschnitte. Im ersten Abschnitt werden einige Rahmendaten zur Schulausbildung besprochen; dann werden mithilfe von Kohortenanalysen historische Veränderungen untersucht; und schließlich wird in der Abfolge der Kohorten die Schulbildung von Eltern und Kindern verglichen.

11.1 Daten über Schulabschlüsse

1. Gliederungen des Bildungssystems. Es gibt in Deutschland viele verschiedene Schularten und Bildungswege. Die Grundstruktur kann anhand von Abbildung 11.1-1 verdeutlicht werden;[1] die Ziffern auf der linken Seite geben Jahrgangsstufen an.

Primarbereich. Die schulische Ausbildung beginnt normalerweise mit etwa sechs Jahren[2] in der Grundschule (= Primarbereich), die die ersten vier Jahrgangsstufen umfasst.

[1] Die Darstellung erfolgt in Anlehnung an ein Schema in KMK 2005: 33. In dieser Publikation findet man eine umfassende Darstellung des Bildungswesens in der BRD.

[2] Die Schulpflicht beginnt für Kinder, die bis zum 30. Juni das sechste Lebensjahr vollendet haben.

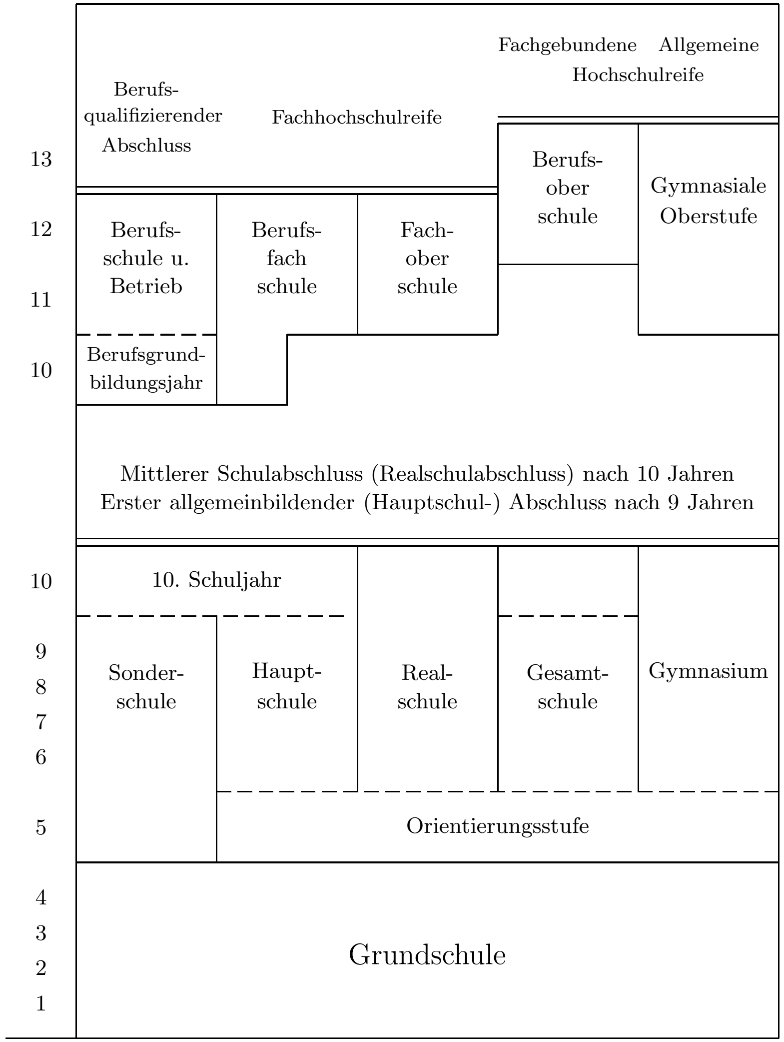


Abb. 11.1-1 Schema zur Gliederung des Bildungssystems in Deutschland (ohne tertiären Bereich).

Sekundarbereich I. Es folgt der Sekundarbereich I mit weiteren fünf oder sechs Jahrgangsstufen. Nach der 9. Jahrgangsstufe kann man die Schule mit einem Hauptschulabschluss, nach der 10. Jahrgangsstufe mit einem Realschulabschluss (mittlere Reife) verlassen.

Sekundarbereich II. Es folgt der Sekundarbereich II, in dem es folgende Möglichkeiten gibt:

- Man kann (nach einem Berufsgrundbildungsjahr) eine Berufsausbildung im Dualen System (Berufsschule und betriebliche Ausbildung) absolvieren und damit einen berufsqualifizierenden Abschluss erwerben.
- Man kann eine Berufsfachschule oder eine Fachoberschule besuchen; als Abschluss kann in beiden Fällen eine Fachhochschulreife erworben werden.
- Man kann eine Berufsoberschule besuchen und als Abschluss eine fachgebundene Hochschulreife erwerben.
- Man kann die Gymnasiale Oberstufe besuchen und als Abschluss die Hochschulreife erwerben.

Tertiärer Bereich. Schließlich gibt es noch einen sogenannten tertiären Bereich, der berufliche Ausbildungen in Hochschulen und Fachhochschulen umfasst.

Das Schema bezieht sich auf den gegenwärtigen Entwicklungsstand. Seit Beginn der Bundesrepublik hat es zahlreiche Veränderungen gegeben; die Gliederung in Schulstufen ist jedoch weitgehend erhalten geblieben, und mit der Eingliederung der ehemaligen DDR ist auch dort das westdeutsche Schulsystem eingeführt worden.[3]

2. Entwicklung der Schulanfänger. Die Anzahl der Schulanfänger hängt von der Entwicklung der Geburten und von Zu- und Abwanderungen ab; es hat infolgedessen erhebliche Veränderungen und Schwankungen gegeben. Eine grobe Orientierung liefert die Entwicklung der Anzahl der sechsjährigen Kinder, die in Abbildung 11.1-2 dargestellt wird. Durch den Geburtenrückgang hat sich offenbar insbesondere in den neuen Bundesländern die Anzahl der Schulanfänger erheblich verringert.

3. Entwicklung der Schulabschlüsse. Nach einem mehr oder weniger langen Schulbesuch verlassen die Schulanfänger das Schulsystem mit unterschiedlichen Abschlüssen. Die verfügbaren statistischen Daten unterscheiden meistens folgende Möglichkeiten:

- ohne Hauptschulabschluss
- mit Hauptschulabschluss
- mit Realschulabschluss (mittlere Reife)
- mit Fachhochschulreife
- mit Hochschulreife (Abitur)

[3]Eine zusammenfassende Darstellung geben J. Baumert, K.S. Cortina und A. Leschinsky (2005).

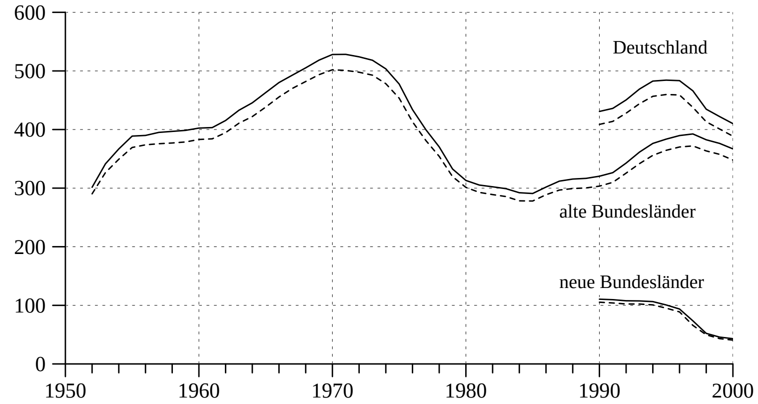


Abb. 11.1-2 Anzahl (in 1000) der 6-jährigen Jungen (durchgezogen) und Mädchen (gestrichelt) im Gebiet der früheren BRD sowie ab 1990 in Deutschland und den neuen Bundesländern. Quelle: Segmente 36 und 685 der STATIS-Datenbank.

Abbildung 11.1-3 zeigt, wie sich die Verteilung der Schulabgänger entwickelt hat. Angegeben sind Prozentanteile, die sich bis 1990 auf die frühere BRD, ab 1992 auf Deutschland beziehen;[4] Fachhochschul- und Hochschulreife wurden wegen der geringen Anzahl von Schulabgängern mit einer Fachhochschulreife zusammengefasst. Man erkennt, dass es seit etwa 1970 eine Tendenz zu höherer Schulbildung gegeben hat:[5] Der Anteil von Schulabgängern ohne oder nur mit Hauptschulabschluss ging zurück, die Anteile derjenigen mit einem Realschulabschluss oder einer Fachhochschul- oder Hochschulreife wurden größer. Diese Entwicklung hat jedoch seit Anfang der 1990er Jahre aufgehört.

Besonders bemerkenswert ist, dass im Durchschnitt Frauen die Schule mit höheren Abschlüssen verlassen als Männer. Seit etwa Anfang der 1990er Jahre gibt es insbesondere mehr weibliche als männliche Schulabgänger mit einer Hochschulreife. Wie Abbildung 11.1-4 zeigt, ist dies das Ergebnis eines langfristigen Prozesses.

4. Schulabschlüsse ausländischer Jugendlicher. Die amtliche Schulstatistik unterscheidet nicht nur nach dem Geschlecht, sondern auch zwischen Ju-

[4]Die Daten stammen aus der Fachserie 11, Reihe S.2 (Allgemeinbildende und berufliche Schulen 1950 bis 1999) und sind bis 1990 für die Jahre 1970, 1975, 1980, 1985 und 1990 verfügbar, ab 1992 jährlich. Für den Zeitraum ab 1980 gibt es vergleichbare, jedoch teilweise abweichende Daten auch im Segment 2909 der STATIS-Datenbank.

[5]Teilweise hat der Prozess schon früher begonnen, was in den Abbildungen natürlich nicht sichtbar ist.

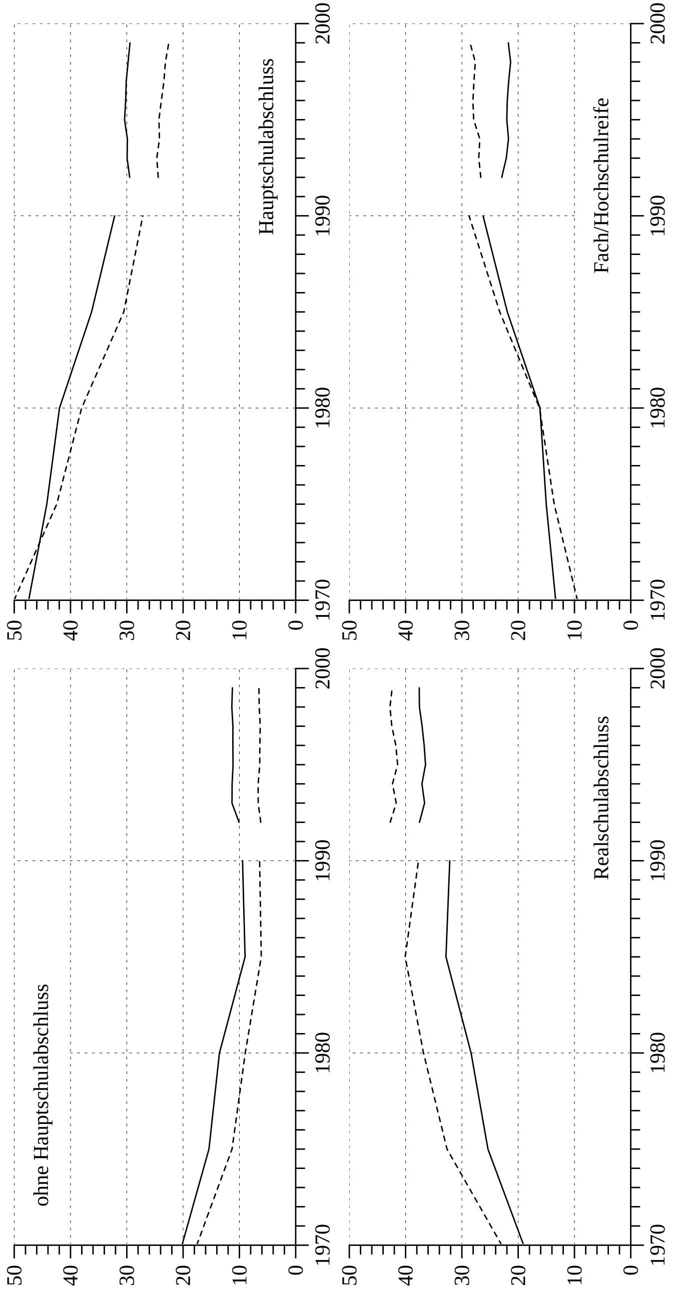


Abb. 11.1-3 Anteile (in %) der Schulabschlüsse, differenziert nach dem Geschlecht (durchgezogene Linien: männlich, gestrichelte Linien: weiblich). Bis 1990 früheres Bundesgebiet, ab 1992 Deutschland. Quelle: Fachserie 11, Reihe S.2: 34ff.

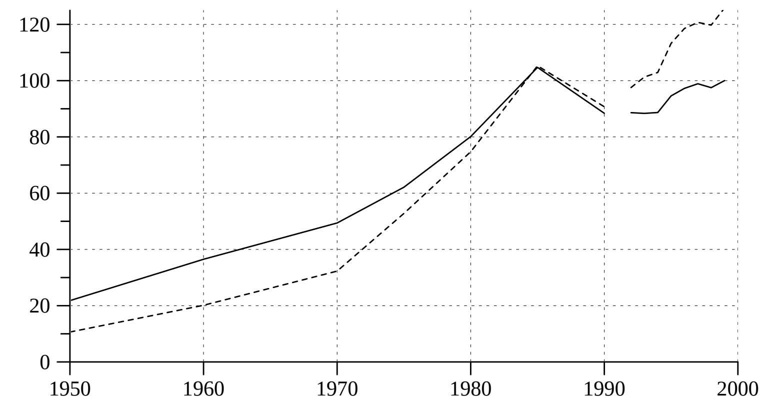


Abb. 11.1-4 Anzahl (in 1000) der männlichen (durchgezogen) und weiblichen (gestrichelt) Schulabgänger mit einer Hochschulreife. Bis 1990 früheres Bundesgebiet, ab 1992 Deutschland. Quelle: Fachserie 11, Reihe S.2: 34ff.

gendlichen mit deutscher und ausländischer Staatsbürgerschaft.[6] Wie Abbildung 11.1-5 zeigt, verlassen Jugendliche mit ausländischer Staatsbürgerschaft die Schule in erheblich größerem Umfang ohne oder nur mit einem Hauptschulabschluss, dagegen viel seltener mit einem höheren Abschluss.

11.2 Kohortenanalysen zur Schulbildung

Im vorangegangenen Abschnitt wurde anhand von Querschnittsverteilungen gezeigt, wie sich die Teilnahme an schulischer Bildung und die Verteilungen von Schulabschlüssen verändert haben. Da sich diese Querschnittsverteilungen auf einzelne Kalenderjahre beziehen, fassen sie jeweils unterschiedliche Geburtskohorten zusammen. In diesem Abschnitt wird untersucht, wie sich die Bildungsbeteiligung in der Abfolge von Geburtskohorten verändert hat. Dafür stützen wir uns zunächst auf Daten des ALLBUS.[7]

1. Daten aus dem ALLBUS. Der ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) ist ein Survey, der seit 1980 alle zwei Jahre (außerdem 1991 zwecks Ausweitung für die neuen Bundesländer) von ZUMA (Zentrum für Umfragen, Methoden und Analysen, Mannheim) und

[6] Unter analytischen Gesichtspunkten wäre es sinnvoller, Jugendliche mit einem Migrationshintergrund zu betrachten; sie können aber mit Daten der amtlichen Schulstatistik nicht ermittelt werden.

[7] Daten des ALLBUS wurden bereits von W. Müller und D. Haun (1994) zur Untersuchung von Veränderungen in der Bildungsbeteiligung verwendet. Im Vergleich zu dieser Untersuchung konzentrieren wir uns hier auf Schulabschlüsse und verwenden auch Daten aus neueren ALLBUS-Erhebungen.

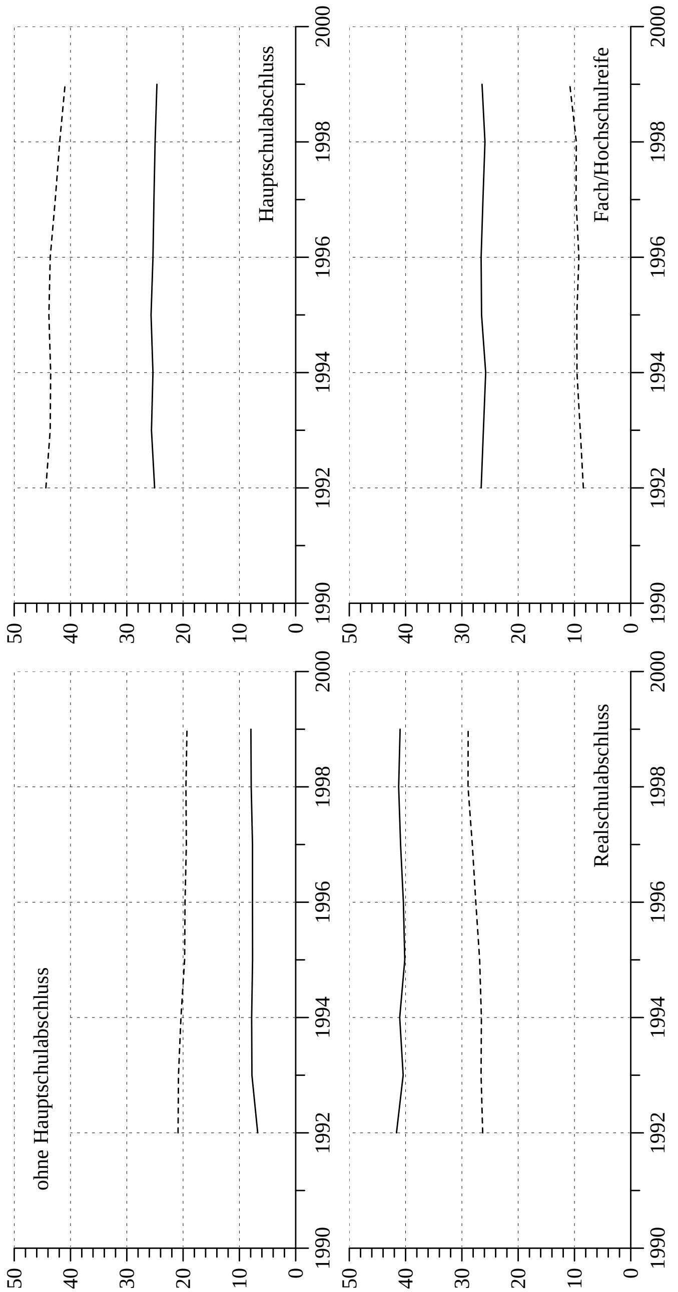


Abb. 11.1-5 Anteile (in %) der Schulabschlüsse, differenziert nach deutscher (durchgezogene Linien) und ausländischer (gestrichelte Linien) Staatsbürgerschaft. Angaben für Deutschland. Quelle: Fachserie 11, Reihe S.2: 34ff.

Tabelle 11.2-1 Aus dem kumulierten ALLBUS 1980–2002 ausgewählte Personen (vgl. § 1 zu den Auswahlkriterien) mit Angaben zu ihrem Schulabschluss.

		Männer		Frauen	
1	Kein Abschluss	190	1.4	332	2.1
2	Volks- oder Hauptschulabschluss	7348	52.1	8621	55.2
3	Mittlere Reife, Realschule	2954	20.9	4056	26.0
4	Fachhochschulreife	927	6.6	601	3.9
5	Abitur, Hochschulreife	2689	19.1	2008	12.9
		14108	100.0	15618	100.0

dem Zentralarchiv für Empirische Sozialforschung (Köln) in Zusammenarbeit mit dem ALLBUS-Ausschuss durchgeführt wird.[8] Für die folgenden Auswertungen verwenden wir den kumulierten Datensatz für die Jahre 1980–2002, der beim Zentralarchiv für Empirische Sozialforschung erhältlich ist. Es werden alle Personen mit deutscher Staatsangehörigkeit betrachtet, die in den alten Bundesländern befragt wurden, die im Zeitraum 1908 – 1977 geboren wurden und zur Zeit des Interviews mindestens 22 Jahre alt waren: insgesamt 29939 Personen.[9] Die meisten von ihnen, 14108 Männer und 15618 Frauen, haben auf die Frage nach dem allgemeinen Schulabschluss eine der in Tabelle 11.2-1 dargestellten Angaben gemacht (die Ziffern 1 – 5 dienen auch im Folgenden zum Verweis auf die unterschiedlichen Abschlüsse). Die Tabelle zeigt jeweils Anzahlen und Prozentanteile. Ersichtlich haben im Durchschnitt der hier betrachteten Geburtskohorten Frauen etwas niedrigere Schulabschlüsse als Männer.

2. Gliederung nach Geburtskohorten. Um Veränderungen in den Verteilungen der Schulabschlüsse zu untersuchen, gliedern wir die Personen nicht nur nach dem Geschlecht, sondern auch nach Geburtsjahren. Jeweils 5 Geburtsjahre werden zu einer Geburtskohorte zusammengefasst. Wie Tabelle 11.2-2 zeigt, können insgesamt 14 dieser Geburtskohorten gebildet werden. Die Tabelle zeigt außerdem die Anzahlen und Anteile der Personen mit den

[8] `http://www.gesis.org/Datenservice/ALLBUS/index.htm`

[9] Bis 1990 bezogen sich die Umfragen auf Personen mit deutscher Staatsangehörigkeit, die in Westdeutschland in Privathaushalten lebten und bis zum Befragungstag mindestens 18 Jahre alt waren; ab 1991 wurden die Befragungen auf deutschsprachige Personen in den alten und neuen Bundesländern ausgedehnt (wiederum mindestens 18 Jahre alt und in Privathaushalten lebend). Auch die für die folgenden Analysen ausgewählte Teilgesamtheit enthält vermutlich Personen, die ihre Schulbildung nicht oder nur teilweise im Schulsystem der BRD absolviert haben. Ein weiteres Problem betrifft die Tatsache, dass es sich teilweise um Personen- und teilweise um Haushaltsstichproben handelt. Es ist jedoch unklar, ob bzw. wie ein möglicherweise (via Geschwisterzahl) bestehender Zusammenhang zwischen Haushaltsgröße und Schulabschlüssen durch Gewichte berücksichtigt werden könnte. Wir verwenden die Daten deshalb in ungewichteter Form. Da wir uns auf Befragungspersonen aus den alten Bundesländern beschränken, sind keine Ost-West-Gewichte erforderlich.

Tabelle 11.2-2 Nach Geburtskohorten differenzierte Verteilungen auf Schulabschlüsse $j = 1, \ldots, 5$. n_j^m bzw. n_j^f geben die Anzahlen der Männer bzw. Frauen mit dem Abschluss j an; n^m und n^f sind die Summen, p_j^m und p_j^f die Anteilswerte in %. Quelle: ALLBUS 1980 – 2002.

Kohorte	n_1^m	p_1^m	n_2^m	p_2^m	n_3^m	p_3^m	n_4^m	p_4^m	n_5^m	p_5^m	n^m
1908 - 1912	11	2.7	284	70.5	57	14.1	10	2.5	41	10.2	403
1913 - 1917	6	1.0	402	68.6	100	17.1	14	2.4	64	10.9	586
1918 - 1922	14	1.9	497	65.6	123	16.2	30	4.0	94	12.4	758
1923 - 1927	10	1.0	655	67.0	140	14.3	37	3.8	136	13.9	978
1928 - 1932	34	3.0	757	65.8	185	16.1	50	4.3	125	10.9	1151
1933 - 1937	23	1.8	827	65.8	209	16.6	65	5.2	133	10.6	1257
1938 - 1942	9	0.6	888	57.9	334	21.8	93	6.1	210	13.7	1534
1943 - 1947	11	0.9	641	51.6	299	24.1	79	6.4	212	17.1	1242
1948 - 1952	15	1.1	685	49.1	307	22.0	103	7.4	285	20.4	1395
1953 - 1957	15	1.1	607	42.5	308	21.6	149	10.4	348	24.4	1427
1958 - 1962	17	1.2	542	36.5	369	24.9	135	9.1	421	28.4	1484
1963 - 1967	17	1.5	352	31.2	307	27.2	93	8.2	360	31.9	1129
1968 - 1972	6	1.1	166	31.1	145	27.2	48	9.0	169	31.7	534
1973 - 1977	2	0.9	45	19.6	71	30.9	21	9.1	91	39.5	230

Kohorte	n_1^f	p_1^f	n_2^f	p_2^f	n_3^f	p_3^f	n_4^f	p_4^f	n_5^f	p_5^f	n^f
1908 - 1912	24	3.3	560	77.8	102	14.2	6	0.8	28	3.9	720
1913 - 1917	28	3.6	555	71.7	138	17.8	11	1.4	42	5.4	774
1918 - 1922	43	3.9	800	72.9	172	15.7	18	1.6	64	5.8	1097
1923 - 1927	37	3.0	832	68.2	206	16.9	36	3.0	109	8.9	1220
1928 - 1932	50	4.4	778	68.9	206	18.2	26	2.3	70	6.2	1130
1933 - 1937	34	2.7	902	71.0	233	18.3	27	2.1	75	5.9	1271
1938 - 1942	26	1.7	975	62.1	399	25.4	37	2.4	133	8.5	1570
1943 - 1947	16	1.3	718	56.4	372	29.2	42	3.3	125	9.8	1273
1948 - 1952	11	0.7	841	54.5	432	28.0	69	4.5	190	12.3	1543
1953 - 1957	18	1.2	674	43.3	491	31.6	89	5.7	283	18.2	1555
1958 - 1962	20	1.3	513	32.7	584	37.2	99	6.3	354	22.6	1570
1963 - 1967	14	1.2	315	26.7	448	38.0	86	7.3	317	26.9	1180
1968 - 1972	4	0.8	126	24.3	207	39.9	42	8.1	140	27.0	519
1973 - 1977	7	3.6	32	16.3	66	33.7	13	6.6	78	39.8	196

Schulabschlüssen 1 – 5 (wie sie im vorangegangenen Paragraphen unterschieden wurden). Abbildung 11.2-1 veranschaulicht die Veränderungen. Man erkennt:

- Der Anteil der Personen, die die Schule nur mit einem Hauptschulabschluss verlassen, hat in der Abfolge der Kohorten stark abgenommen.
- Dagegen haben höhere Schulabschlüsse zugenommen. Insbesondere hat sich der Anteil der Personen, die die Schule mit einem Abitur beenden, erheblich erhöht.

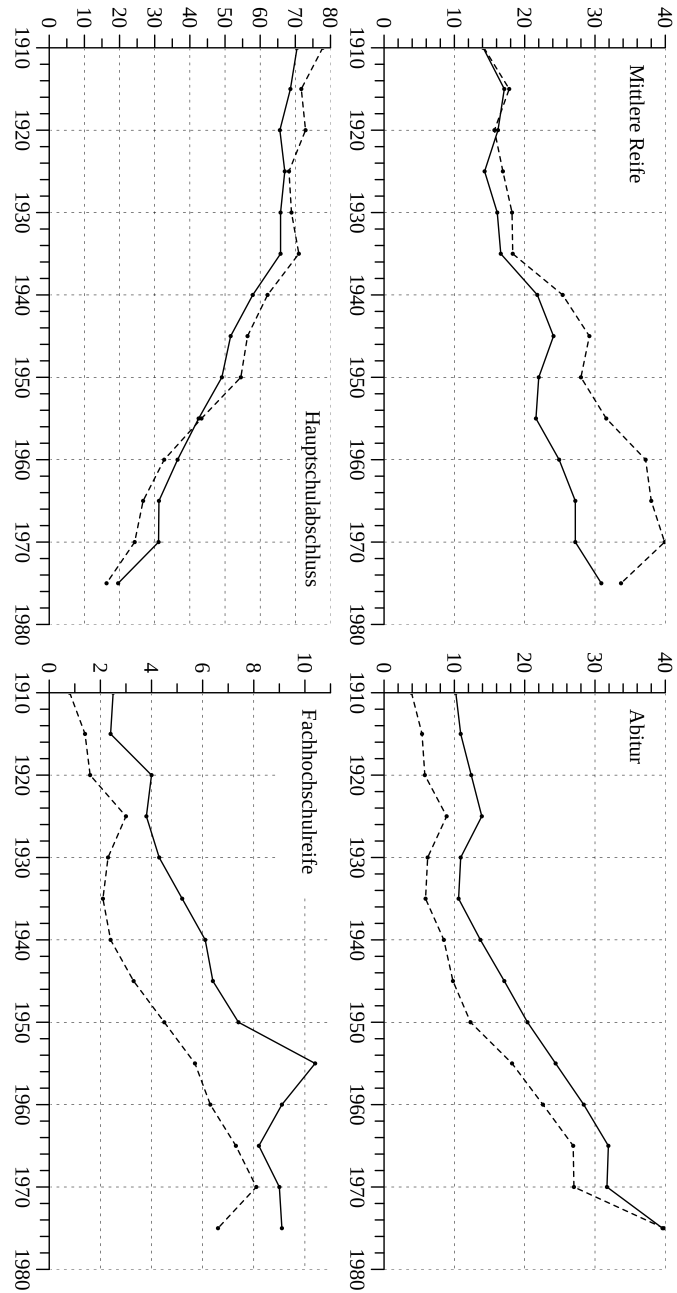


Abb. 11.2-1 Nach Geburtskohorten differenzierte Prozentanteile mit den angeführten schulischen Abschlüssen. Abszisse: Geburtsjahre der Kohorten. Daten aus Tabelle 11.2-2.

– Veränderungen in den Anteilen von Schulabgängern ohne Hauptschulabschluss können wegen der geringen Fallzahlen nicht verlässlich beurteilt werden.[10]

Die Daten zeigen außerdem, dass die „Bildungsexpansion“ zugunsten höherer Schulabschlüsse bei der Schulausbildung nicht erst in den 1960er Jahren beginnt, sondern bereits mit den um 1935 geborenen Kohorten, also mit dem Beginn der Nachkriegszeit.[11]

3. Veränderungen der Bildungsungleichheit. Dies wird auch sichtbar, wenn man Veränderungen der Bildungsungleichheit untersucht. Allerdings muss überlegt werden, wie eine solche Ungleichheit erfasst werden kann. Wir verwenden zwei Maßzahlen, die jeweils unterschiedliche Aspekte sichtbar machen.

a) Zunächst verwenden wir einen *Diversitätsindex*, der durch

$$D_t := 1 - \sum_{j=1}^{5} p_{t,j}^2$$

definiert ist, wobei $p_{t,j}$ der Anteil der Personen der Geburtskohorte t mit dem Schulabschluss j ist. Der Wert dieses Index kann näherungsweise als die Wahrscheinlichkeit interpretiert werden, dass zwei zufällig ausgewählte Personen nicht den gleichen Schulabschluss haben.[12] Wie Abbildung 11.2-2 zeigt, hat sich die durch diesen Index erfasste Ungleichheit in der Verteilung der Schulabschlüsse bis zu den um 1960 geborenen Kohorten vergrößert.

b) Die zeitweilige Vergrößerung der Bildungsungleichheit ist hauptsächlich auf eine Verringerung des Anteils der Hauptschulabschlüsse zugunsten höherer Schulabschlüsse zurückzuführen. Um das sichtbar zu machen, vergleichen wir die jeweils aktuelle Verteilung der Schulabschlüsse mit einer idealen Verteilung, bei der alle Personen die Schule mit einem Abitur verlassen. Um den Abstand zwischen der aktuellen und der idealen Verteilung zu erfassen, kann die Maßzahl

$$d_t := \sum_{j=1}^{5} (5-j)\, p_{t,j}$$

dienen. Je kleiner der Wert von d_t ist, desto kleiner ist auch der Abstand zwischen der aktuellen Verteilung der Schulabschlüsse und der idealen Verteilung, bei der alle Personen die Schule mit einem Abitur verlassen. Der maximale Wert, nämlich 4, wird erreicht, wenn alle

[10] Insgesamt ist der Anteil von Personen ohne Hauptschulabschluss wesentlich größer als in den hier verwendeten ALLBUS-Daten sichtbar ist (vgl. Abb. 11.1-3). Vermutlich liegt dies in erster Linie daran, dass wir hier nur Personen mit deutscher Staatsangehörigkeit betrachten.

[11] Dies haben bereits W. Müller und D. Haun (1994: 14f.) festgestellt. Man vgl. dazu auch die Bemerkungen von W. Müller (1998: 91).

[12] Eine Diskussion von Eigenschaften und Interpretationsmöglichkeiten des Diversitätsindex findet man bei A. und B. F Agresti (1977).

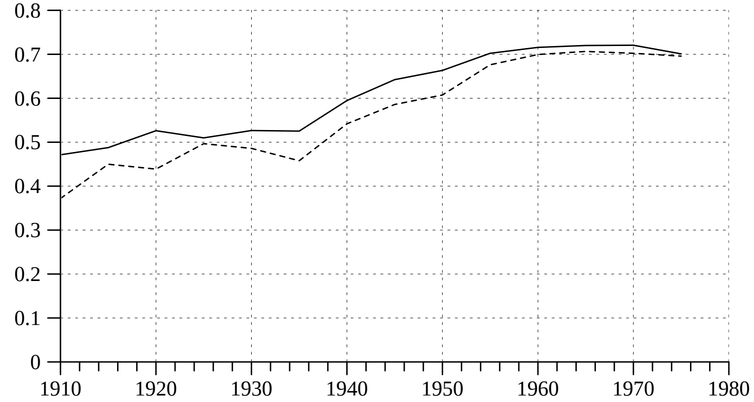


Abb. 11.2-2 Entwicklung des Diversitätsindex D_t für männliche (durchgezogene Linie) und weibliche (gestrichelte Linie) Personen, berechnet mit den Daten in Tabelle 11.2-2. Abszisse: Geburtsjahre der Kohorten.

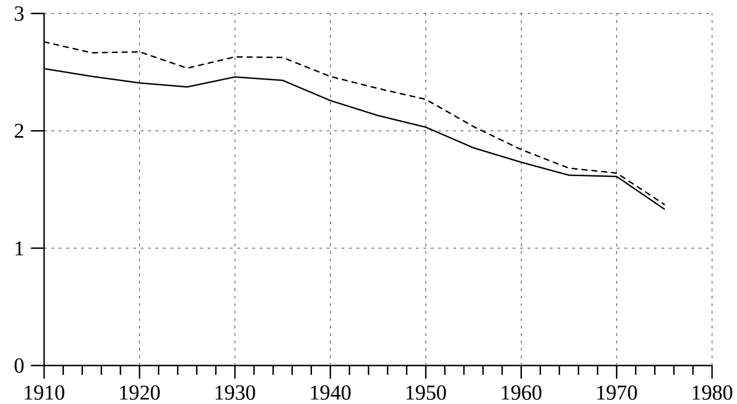


Abb. 11.2-3 Entwicklung der Maßzahl d_t für männliche (durchgezogene Linie) und weibliche (gestrichelte Linie) Personen, berechnet mit den Daten in Tabelle 11.2-2. Abszisse: Geburtsjahre der Kohorten.

Personen die Schule ohne einen Abschluss verlassen. Abbildung 11.2-3 zeigt, wie sich diese Maßzahl in der Abfolge der Geburtskohorten verändert hat. Man erkennt, dass ein relativ kontinuierlicher Übergang zu höheren Schulabschlüssen etwa mit den um 1935 geborenen Kohorten eingesetzt hat. Außerdem wird sichtbar, wie sich die geschlechtsspezifischen Unterschiede verringert haben.

4. Substitutionsmetriken für Verteilungen. Die Maßzahl d_t ist eine Variante sogenannter Substitutionsmetriken zum Vergleich von Verteilungen. Da wir später auch andere Varianten verwenden wollen, soll an dieser Stelle kurz die allgemeine Definition erklärt werden.

Die allgemeine Aufgabe besteht darin, eine Abstandsfunktion zu definieren, mit der Unterschiede zwischen Verteilungen quantifiziert werden können. Hier beziehen wir uns auf Verteilungen für Merkmalsräume mit m Kategorien: $1, \ldots, m$. Die zu vergleichenden Verteilungen sind durch

$$p'_1, \ldots, p'_m \quad \text{und} \quad p''_1, \ldots, p''_m$$

gegeben (jeweils nicht-negative Anteilswerte, deren Summe = 1 ist). Substitutionsmetriken liegt nun die Idee zugrunde, die Unterschiedlichkeit der Verteilungen durch das Ausmaß der Umschichtungen zwischen den Kategorien zu erfassen, die erforderlich sind, um die beiden Verteilungen in Übereinstimmung zu bringen.

Zur Berechnung werden Bewertungen vorausgesetzt, durch die angegeben wird, wie sich die einzelnen Kategorien unterscheiden. Metaphorisch gesprochen geben die Bewertungen die Kosten an, die bei einer Umschichtung von 1 % der Objekte aus einer Kategorie i in eine Kategorie j entstehen. Wir verwenden zur Bezeichung: c_{ij} (für $i, j = 1, \ldots, m$). Es wird vorausgesetzt, dass diese Bewertungskoeffizienten nicht-negativ und symmetrisch ($c_{ij} = c_{ji}$) sind, dass $c_{ii} = 0$ ist, und dass sie die Dreiecksungleichung erfüllen: $c_{ij} \leq c_{ik} + c_{kj}$ für beliebige i, j, k.

Es soll also eine kostenminimale Umschichtung berechnet werden, die die beiden Verteilungen in Übereinstimmung bringt. Zu diesem Zweck werden zunächst zwei Vektoren $(r_1, \ldots, r_{m_r})$ und $(s_1, \ldots, s_{m_s})$ definiert. Dabei erfasst m_r die Anzahl der Kategorien, bei denen $p'_i > p''_i$ ist, dann ist $r_i := p'_i - p''_i$; und m_s erfasst die Anzahl der Kategorien, bei denen $p''_i > p'_i$ ist, dann ist $s_i := p''_i - p'_i$. Jede Umschichtung, die die beiden Verteilungen in Übereinstimmung bringt, enspricht dann einer Matrix (u_{ij}) mit m_r Zeilen und m_s Spalten, wobei u_{ij} den Anteil der Umschichtungen von der Kategorie i in die Kategorie j angibt, die folgenden Bedingungen genügt:

$$\textstyle\sum_{j=1}^{m_s} u_{ij} \;=\; r_i \quad \text{und} \quad \sum_{i=1}^{m_r} u_{ij} \;=\; s_j$$

Da es im Allgemeinen mehrere mögliche Umschichtungen gibt, die diese Bedingungen erfüllen, wird außerdem gefordert, dass die Kosten der Umschichtung, also

$$\textstyle\sum_{i=1}^{m_r} \sum_{j=1}^{m_s} u_{ij} c_{ij}$$

minimal sein sollen.[13] Diese Minimalkosten werden schließlich zur Quantifizierung des Abstands der Verteilungen verwendet.[14]

[13] Eine Lösung kann mit der Methode der linearen Programmierung berechnet werden. Wir verwenden den `subm`-Befehl des Programms TDA, der auf dieser Methode beruht.

Bezieht man sich auf die Menge aller Verteilungen (für eine bestimmte Anzahl von Kategorien), gelangt man zu einer Abstandsfunktion, die jeweils zwei Verteilungen einen Abstand, nämlich die eben definierten Minimalkosten, zuordnet. Diese Abstandsfunktion erfüllt auch die Bedingungen einer Metrik.[15]

Die im vorangegangenen Paragraphen verwendete Maßzahl d_t ist ein einfaches Beispiel für eine Substitutionsmetrik, bei dem die Bewertungen $c_{ij} = |i-j|$ verwendet werden. Als ein weiteres Beispiel berechnen wir einen Abstand zwischen den Verteilungen der Schulabschlüsse bei männlichen und weiblichen Schulabgängern der Geburtskohorte 1973-77 (Tab. 11.2-2). Es ist also $m = 5$, und die Verteilungen sind durch die Vektoren

$$\mathbf{p}' = (0.009, 0.196, 0.309, 0.091, 0.395) \quad \text{und}$$
$$\mathbf{p}'' = (0.036, 0.16.3, 0.337, 0.066, 0.398)$$

gegeben. Somit ist $\mathbf{r} = (0.033, 0.025)$ und $\mathbf{s} = (0.027, 0.028, 0.003)$. Zur

[14] Wenn die Kosten der Umschichtung stets den Wert 1 haben ($c_{ij} = 1$ für $i \neq j$), entspricht die Substitutionsmetrik dem durch

$$\frac{1}{2} \textstyle\sum_{i=1}^{m} |p'_i - p''_i|$$

definierten *Dissimilaritätsindex*.

[15] Wir orientieren uns an den Definitionen bei Rohwer und Pötter (2002a: 135). Ausgangspunkt ist eine beliebige Menge M. Eine *Abstandsfunktion* ist eine Funktion

$$d : M \times M \longrightarrow \mathbf{R}$$

die jeweils zwei Elementen $a, b \in M$ eine Zahl $d(a, b) \in \mathbf{R}$ zuordnet und für die folgende Bedingungen gelten:

(a) für alle $a, b \in M$: $d(a, b) \geq 0$

(b) für alle $a, b \in M$: $d(a, b) = d(b, a)$

(c) für alle $a \in M$: $d(a, a) = 0$

Eine Abstandsfunktion, für die außerdem die Dreieicksungleichung

(d) für alle $a, b, c \in M$: $d(a, c) \leq d(a, b) + d(b, c)$

gilt, wird *Semi-* oder *Quasi-Metrik* genannt. Gilt schließlich auch noch

(e) für alle $a, b \in M$: $d(a, b) = 0 \Longrightarrow a = b$

spricht man von einer *Metrik*. Abstandsfunktionen ordnen also jeweils zwei Elementen einer Menge eine nicht-negative Zahl zu, die in einem allgemeinen (nicht unbedingt räumlichen) Sinn als ihr Abstand interpretiert werden kann. Je größer diese Zahl ist, desto größer ist der Abstand der beiden Elemente. Die Möglichkeiten einer inhaltlichen Interpretation der Abstände hängen natürlich von der Art der Elemente ab.

Bewertung soll jetzt folgende Matrix verwendet werden:

$$\mathbf{C} := \begin{bmatrix} 0 & 3 & 5 & 7 & 8 \\ 3 & 0 & 2 & 4 & 5 \\ 5 & 2 & 0 & 2 & 3 \\ 7 & 4 & 2 & 0 & 1 \\ 8 & 5 & 3 & 1 & 0 \end{bmatrix} \tag{11.1}$$

Die folgende Tabelle zeigt eine kostenminimale Umschichtung (von den Kategorien in den Zeilen in die Kategorien in den Spalten):

	1	3	5
2	0.027	0.006	0.000
4	0.000	0.022	0.003

woraus sich der Abstand 0.14 ergibt.

11.3 Schulbildung von Eltern und Kindern

Wovon hängt es ab, welchen Schulabschluss ein Kind erreicht? Eine dafür wichtige Bedingung ist das Bildungsniveau der Eltern. Damit beschäftigen wir uns in diesem Abschnitt. Untersucht wird, welche Zusammenhänge es zwischen der Schulbildung von Eltern und Kindern gibt und wie sie sich in der Abfolge von Geburtskohorten verändert haben.

1. Daten aus dem ALLBUS. Wir beziehen uns auf die bereits in Abschnitt 11.2 zugrunde gelegte Teilgesamtheit aus den kumulierten ALLBUS-Daten für die Jahre 1980 bis 2002. In allen Erhebungen wurde auch der Schulabschluss des Vaters, mit Ausnahme der Erhebungen 1980 und 1982 auch der Schulabschluss der Mutter erfragt; dabei wurden die gleichen Kategorien verwendet wie für die Befragungspersonen (vgl. § 1 in Abschnitt 11.2). Natürlich fehlen in einigen Fällen die Angaben. Soweit Informationen vorhanden sind, fassen wir sie auf folgende Weise zu einem *Schulabschluss der Eltern* zusammen:

– Wenn nur eine Information über den Schulabschluss des Vaters oder der Mutter vorhanden ist, wird diese verwendet.
– Wenn die Schulabschlüsse des Vaters und der Mutter angegeben sind, wird der höchste dieser Abschlüsse verwendet.

Für insgesamt 28345 der in Abschnitt 11.2 betrachteten 29726 Befragungspersonen kann auf diese Weise zusätzlich zu ihrem eigenen auch ein Schulabschluss der Eltern bestimmt werden. Tabelle 11.3-1 zeigt die gemeinsame Verteilung. Zum Beispiel hatten bei 2642 Personen der Vater und/oder die Mutter ein Abitur, und von diesen 2642 Personen haben wiederum 1600 Personen ebenfalls die Schule mit einem Abitur abgeschlossen.

Man erkennt bereits aus dieser Tabelle, dass es einen deutlichen Zusammenhang zwischen den Schulabschlüssen der Eltern und der Kinder

Tabelle 11.3-1 Anzahlen und Zeilenprozente der nach ihrer Schulbildung (in den Spalten) und der Schulbildung ihrer Eltern (in den Zeilen) gegliederten Personen. Berechnet mit den Daten des kumulierten ALLBUS 1980 –2002.

	1	2	3	4	5	Insgesamt
1 Kein Abschluss	152	359	80	17	21	629
	24.2	57.1	12.7	2.7	3.3	100.0
2 Volks- oder Hauptschulabschluss	278	13789	4463	813	1567	20910
	1.3	65.9	21.3	3.9	7.5	100.0
3 Mittlere Reife, Realschule	16	717	1346	330	1122	3531
	0.5	20.3	38.1	9.4	31.8	100.0
4 Fachhochschulreife	4	85	176	95	273	633
	0.6	13.4	27.8	15.0	43.1	100.0
5 Abitur, Hochschulreife	6	203	615	218	1600	2642
	0.2	7.7	23.3	8.3	60.6	100.0
	456	15153	6680	1473	4583	28345

gibt. Bei 60 % der Personen stimmt ihr Schulabschluss mit dem des Vaters und/oder der Mutter überein.

2. Veränderungen der Bildungsabstände. Um zu untersuchen, wie sich die Zusammenhänge zwischen der Schulbildung von Eltern und Kindern verändert haben, gliedern wir die Befragungspersonen (Kinder) wie in Abschnitt 11.2 in 14 Geburtskohorten. Für jede Geburtskohorte kann dann eine gemeinsame Verteilung der Schulabschlüsse der Befragungspersonen und ihrer Eltern gebildet werden.

Eine erste Vergleichsmöglichkeit besteht darin, für jede Kohorte den Abstand zwischen den Verteilungen der Schulabschlüsse der Befragungspersonen und ihrer Eltern (also zwischen den beiden Randverteilungen) zu berechnen. Dafür verwenden wir eine Substitutionsmetrik mit der in § 4 von Abschnitt 11.2 definierten Bewertungsmatrix **C**. Abbildung 11.3-1 zeigt, wie sich diese Abstände verändert haben. Man erkennt, dass sie etwa beginnend mit der um 1935 geborenen Kohorte größer werden. Dies entspricht der bereits festgestellten Tatsache, dass etwa ab dieser Geburtskohorte zunehmend höhere Schulabschlüsse erreicht werden. Etwa beginnend mit den in den 1960er Jahren geborenen Kohorten endet jedoch die Beschleunigungsphase, und man kann erwarten, dass mit zunehmendem Bildungsniveau der Eltern die Abstände schließlich wieder kleiner werden.

Weitere Informationen erhält man daraus, wie sich der Anteil der Kinder, die den gleichen Schulabschluss wie ihre Eltern erreicht haben, verändert hat. Wie am Ende des vorangegangenen Paragraphen festgestellt wurde, liegt dieser Anteil im Durchschnitt bei 60 %. Abbildung 11.3-2 zeigt, wie er sich in der Abfolge der Kohorten verändert hat. Offenbar ist der Zusammenhang zwischen der Schulbildung der Eltern und ihrer Kinder erheblich geringer geworden. Ähnlich wie in Abbildung 11.3-1 haben sich

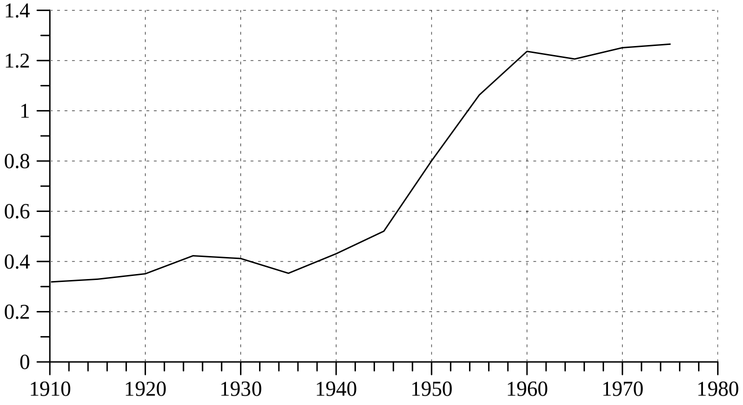


Abb. 11.3-1 Abstand zwischen den Verteilungen der Schulabschlüsse bei Befragungspersonen und ihren Eltern. Abszisse: Geburtsjahre der Kohorten. Berechnet mit Daten des kumulierten ALLBUS 1980–2002.

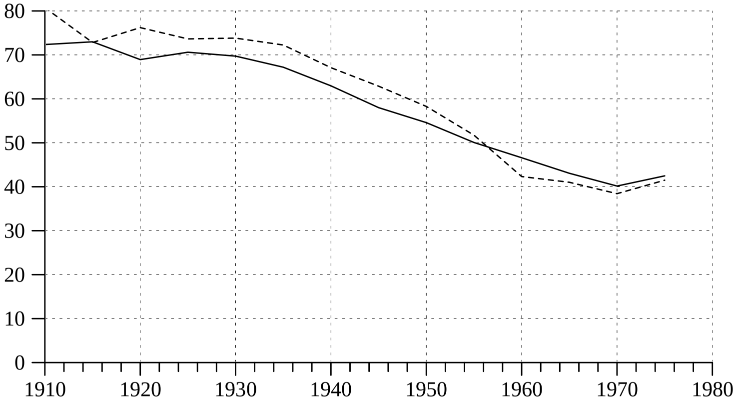


Abb. 11.3-2 Prozentanteile der männlichen (durchgezogen) und weiblichen (gestrichelt) Befragungspersonen, die den gleichen Schulabschluss wie ihre Eltern erworben haben. Abszisse: Geburtsjahre der Kohorten. Berechnet mit Daten des kumulierten ALLBUS 1980–2002.

die Veränderungen hauptsächlich bei den etwa zwischen 1935 und 1960 geborenen Kohorten vollzogen.

3. Differenzierung nach der Schulbildung der Eltern. Man kann vermuten, dass Kinder im Vergleich zu ihren Eltern nicht einfach eine andere, sondern tendenziell eine bessere Schulbildung erworben haben. Um das zu untersuchen, unterscheiden wir die Befragungspersonen nicht nur nach der Geburtskohorte, sondern auch nach der Schulbildung ihrer Eltern und

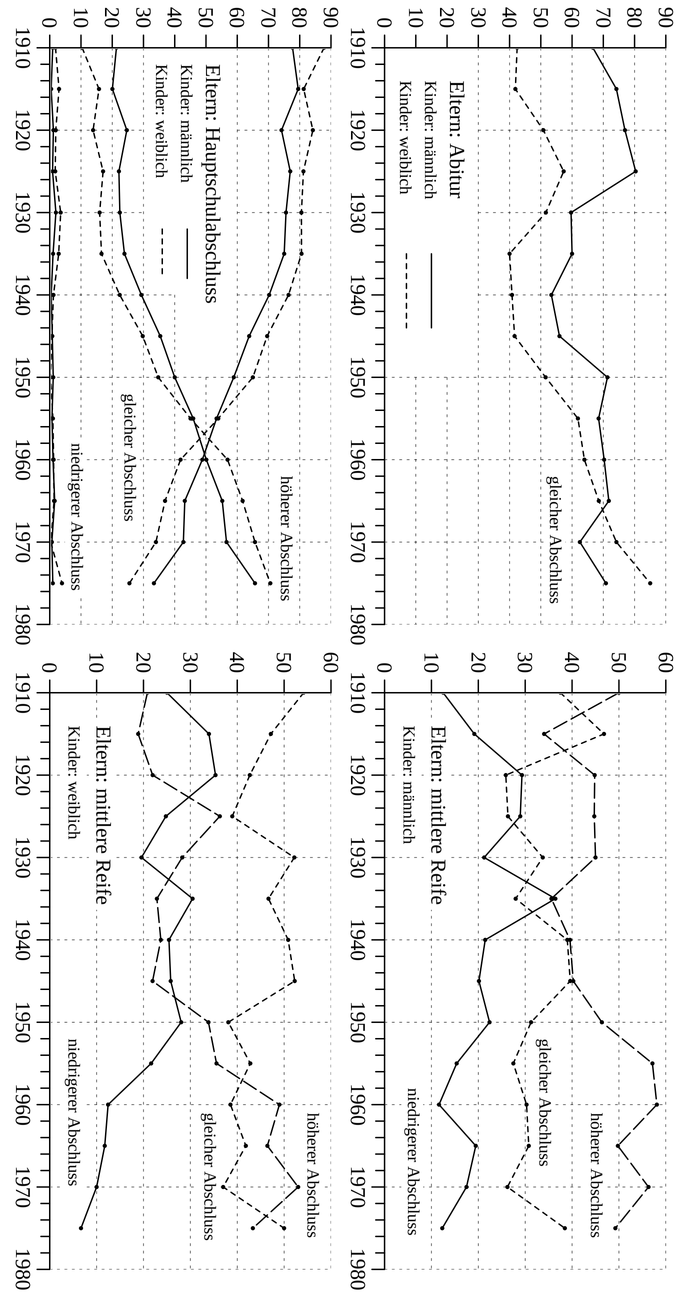


Abb. 11.3-3 Anteile (in %) der Kinder (Befragungspersonen) mit höherem, gleichem oder niedrigerem Schulabschluss als ihre Eltern, differenziert nach dem Schulabschluss der Eltern. Abszisse: Geburtsjahre der Befragungspersonen (Kohorten). Berechnet mit Daten des kumulierten ALLBUS 1980–2002.

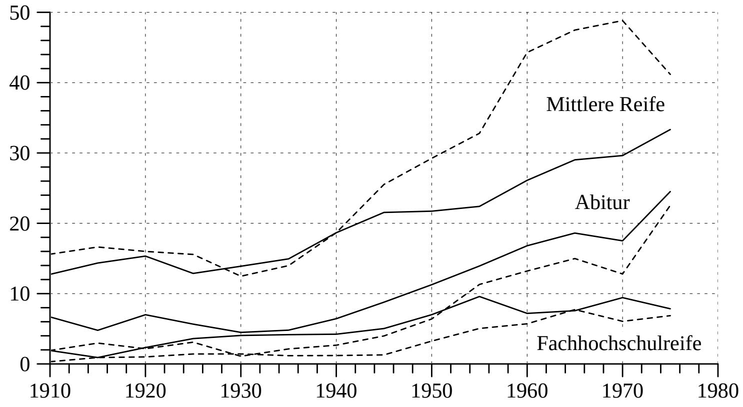


Abb. 11.3-4 Anteile (in %) der männlichen (durchgezogen) und weiblichen (gestrichelt) Befragungspersonen mit den angegebenen Schulabschlüssen an allen Befragungspersonen, bei denen der Vater und/oder die Mutter einen Hauptschulabschluss haben. Abszisse: Geburtsjahre der Kohorten. Berechnet mit Daten des kumulierten ALLBUS 1980–2002.

berechnen dann für jede Gruppe drei Anteile:

- einen Anteil der Personen, die den gleichen Schulabschluss erreichen wie ihre Eltern;
- einen Anteil der Personen, die einen höheren Schulabschluss erreichen als ihre Eltern; und
- einen Anteil der Personen, die einen niedrigeren Schulabschluss erreichen als ihre Eltern.

Abbildung 11.3-3 zeigt, wie sich einige dieser Anteile in der Abfolge der Geburtskohorten verändert haben. Offenbar gibt es bemerkenswerte Unterschiede.

a) Bei den Personen, deren Eltern nur einen Hauptschulabschluss haben, hat der Anteil mit höheren Abschlüssen erheblich zugenommen. Wie Abbildung 11.3-4 zeigt, gab es – vor allem bei Mädchen – in erster Linie eine Zunahme der Realschulabschlüsse, aber auch eine Zunahme von Abschlüssen für eine (Fach-) Hochschulausbildung.

b) Bei Personen, deren Eltern einen mittleren Schulabschluss haben, gab es – unabhängig vom Geschlecht – einen Rückgang des Anteils mit niedrigeren und eine Zunahme des Anteils mit höheren Abschlüssen.

c) Schließlich scheint es bei Personen, deren Eltern Abitur haben, deutliche Unterschiede zwischen Männern und Frauen zu geben. Während es bei den Frauen ab den um 1935 geborenen Kohorten eine deutliche Zunahme des Anteils mit einem Abitur gibt, gibt es bei den Männern

Tabelle 11.3-2 Anzahl Personen der Geburtskohorten 1950 und 1965, kreuztabelliert nach ihrem Schulabschluss ($j = 1$: kein Abitur, $j = 2$: Abitur) und dem Schulabschluss ihrer Eltern ($i = 1$: Hauptschulabschluss, $i = 2$: Abitur). Berechnet mit Daten des kumulierten ALLBUS 1980–2002.

1950	j = 1	j = 2	Insges.
i = 1	1914	174	2088
i = 2	100	159	259
Insges.	2014	333	2347

1965	j = 1	j = 2	Insges.
i = 1	1142	213	1355
i = 2	98	231	329
Insges.	1240	444	1684

keine klar erkennbare Tendenz, bestenfalls kann beginnend mit den um 1940 geborenen Kohorten ein leicht steigender Anteil von Abiturienten festgestellt werden.

4. Statistische Chancenvergleiche. Bei der Verwendung und Analyse statistischer Daten wird oft von „Chancen“ gesprochen. Verbreitet ist diese Rhetorik insbesondere in der empirischen Bildungsforschung, etwa bei der Fragestellung, wie sich die Chancen zum Erwerb höherer Bildungsabschlüsse verändert haben. Offenbar muss beachtet werden, dass sich in diesem Zusammenhang der Chancenbegriff auf statistische Häufigkeiten bezieht und nicht auf individuell interpretierbare Handlungschancen.[16] Im Folgenden sind mit „Chancen“ stets statistische Häufigkeiten gemeint.

Als Beispiel betrachten wir Veränderungen in der Verteilung der Schulabschlüsse bei den Geburtskohorten 1950 (Geburtsjahre 1948–1952) und 1965 (Geburtsjahre 1963–1968), die mit unserer Auswahl aus den ALLBUS-Daten gebildet werden können. Bei beiden Kohorten betrachten wir nur Personen, deren Eltern einen Hauptschulabschluss (Gruppe 1) oder ein Abitur (Gruppe 2) haben. Tabelle 11.3-2 zeigt die gemeinsamen Verteilungen der Schulabschlüsse in Form von Kreuztabellen.

Offenbar haben die Chancen, die Schule mit einem Abitur zu verlassen, in beiden Gruppen zugenommen. Bezeichnet $q_{t,i}$ den Abiturientenanteil in der Gruppe i für die Kohorte t, findet man

$$q_{1950,1} = 0.083 \longrightarrow q_{1965,1} = 0.157$$

für die Gruppe 1 (Eltern Hauptschulabschluss) und

$$q_{1950,2} = 0.614 \longrightarrow q_{1965,2} = 0.702$$

für die Gruppe 2 (Eltern Abitur). Zu überlegen ist, ob und ggf. wie man auch das Ausmaß der Veränderungen in den beiden Gruppen sinnvoll vergleichen kann.

In der Literatur ist von einigen Autoren vorgeschlagen worden, sich an

[16] Man vgl. dazu Rohwer und Pötter (2002b: Kap. 7) und A. Swift (2004: 4).

den relativen Chancenverhältnissen zu orientieren.[17] Dafür gibt es (mindestens) zwei Möglichkeiten. Eine Möglichkeit bezieht sich auf die Veränderung der Chancenverhältnisse $q_{t,1}/q_{t,2}$; in unserem Beispiel:

$$\frac{q_{1950,1}}{q_{1950,2}} = 0.135 \longrightarrow \frac{q_{1965,1}}{q_{1965,2}} = 0.224$$

Eine andere Möglichkeit besteht darin, von den Chancenverhältnissen (den sogenannten *Odds*) innerhalb der beiden Gruppen auszugehen und daraus ein komparatives Chancenverhältniss (eine sogenannte *Odds Ratio*) zu bilden. In unserem Beispiel sind die Odds (für ein Abitur vs. Nicht-Abitur) in der Gruppe i durch $o_{t,i} := q_{t,i}/(1-q_{t,i})$ definiert, und die komparativen Chancenverhältnisse haben sich folgendermaßen verändert:

$$\frac{o_{1950,1}}{o_{1950,2}} = \frac{174/1914}{159/100} = 0.057 \longrightarrow \frac{o_{1965,1}}{o_{1965,2}} = \frac{213/1142}{231/98} = 0.079$$

In beiden Varianten kommt man zu dem Ergebnis, dass die Chancen für ein Abitur in der ersten Gruppe (Eltern mit Hauptschulabschluss) mehr zugenommen haben als in der zweiten Gruppe (Eltern mit Abitur).

Relative Chancenverhältnisse erfassen jedoch nur einen Aspekt, der bei einem Vergleich berücksichtigt werden sollte. Ein anderer Aspekt betrifft die Größe der Chancen und ihre Differenzen. Die Bedeutung dieses Aspekts kann man sich anhand von Beispielen überlegen. Man kann sich z.B. vorstellen, dass der Anteil der Abiturienten in einer Gruppe A von 1 auf 2 % und in einer Gruppe B von 50 auf 99 % zugenommen hat; dann würde man bei einer Orientierung an den Chancenverhältnissen immer noch zu dem Ergebnis kommen, dass die Gruppe A ihre Chancen mehr steigern konnte als die Gruppe B; aber offenbar wäre es fragwürdig, daraus auf eine größere „Chancengleichheit“ zu schließen.

Eine Alternative besteht darin, sich daran zu orientieren, wie sich der Abstand der Verteilungen zwischen den Gruppen verändert hat. Wie in § 4 von Abschnitt 11.2 besprochen wurde, kann dieser Abstand mit einer Substitutionsmetrik erfasst werden. Wenn es nur zwei Kategorien gibt, sind deren Ergebnisse auch unabhängig von der gewählten Bewertung. Bezeichnet also $q_{t,i}$ den Anteil der Abiturienten in der Gruppe i (für die Geburtskohorte t), kann der Abstand der Verteilungen einfach durch $|q_{t,1} - q_{t,2}|$ berechnet werden. In unserem Beispiel:

$$|q_{1950,1} - q_{1950,2}| = |\,0.083 - 0.614\,| = 0.531 \longrightarrow$$
$$|q_{1965,1} - q_{1965,2}| = |\,0.157 - 0.702\,| = 0.545$$

Man käme also zu dem Ergebnis, dass der Abstand zwischen den Verteilungen der beiden Gruppen etwas zugenommen hat.

[17] Man vgl. beispielsweise J. Handl (1985) und W. Müller und D. Haun (1994: 16).

Kapitel 12

Rentenbeginn und Lebenserwartung

Eine für die Diskussion der Rentenversicherung wichtige Frage bezieht sich darauf, ob es einen Zusammenhang zwischen dem Renteneintrittsalter und der ferneren Lebenserwartung gibt. Gäbe es keinen Zusammenhang, würden Personen, bei denen der Rentenbezug vor dem 65.ten Lebensjahr beginnt, im Durchschnitt für längere Zeiten Renten beziehen. Zwar wäre es ein Fehler anzunehmen, dass der Rentenbezug bei Personen, die im Alter $65 - t$ Jahre verrentet werden, durchschnittlich t Jahre länger ist als bei Personen, die mit 65 verrentet werden; denn ein Teil derjenigen, bei denen ein Rentenbezug im Alter $65 - t$ beginnt, sterben bereits bis zum Alter 65. Dennoch würde die durchschnittliche Rentenbezugsdauer mit dem Wert t in nicht-linearer Weise zunehmen.

Andererseits liegt jedoch die Vermutung nahe, dass für die Frage, ob Personen bereits im Alter $65 - t$ verrentet werden, auch gesundheitliche Gründe ein Rolle spielen, was zur Folge haben könnte, dass diese Personen auch eine kürzere fernere Lebenserwartung haben. So ist es z.B. durchaus denkbar, dass diejenigen Personen, bei denen ein Rentenbezug im Alter $65-t$ beginnt, die gleiche oder sogar eine geringere Lebenserwartung haben als diejenigen, bei denen der Rentenbezug erst im Alter 65 beginnt.

In diesem Kapitel versuchen wir, anhand von zwei Datensätzen empirische Informationen über Zusammenhänge zwischen dem Alter beim Rentenbeginn und der ferneren Lebenserwartung zu gewinnen. Zuerst verwenden wir einen Datensatz aus der Gesundheitsberichterstattung der Gmünder Ersatzkasse (GEK), dann verwenden wir Daten des Sozio-ökonomischen Panels (SOEP). In beiden Fällen beziehen wir uns nur auf Männer.

12.1 Berechnungen mit GEK-Daten

1. Basisinformationen zum Datensatz. In diesem Abschnitt verwenden wir einen Datensatz der Gmünder Ersatzkasse (GEK).[1] Die uns verfügbare Version des Datensatzes stammt aus der zweiten Hälfte des Jahres 2002.[2] Der Stammdatensatz enthält Informationen über insgesamt 2669062 Versicherte, davon 1449640 männlichen und 1219422 weiblichen Geschlechts. Es handelt sich um einen historischen Datensatz, der auch Informationen über Personen enthält, die gegenwärtig nicht mehr bei der GEK versichert sind, sei es dass sie gestorben oder dass sie ausgetreten sind. Allerdings ist zu vermuten, dass die älteren Geburtsjahrgänge nicht systematisch erfasst sind; das ist im folgenden zu berücksichtigen. Box 12.1-1 zeigt, wie sich die Anzahl der männlichen Versicherten der GEK seit 1960 entwickelt hat.[3]

Bereits aus dem in Box 12.1-1 angegebenen Durchschnittsalter lässt sich ersehen, dass die GEK-Versicherten im Vergleich zum Bevölkerungsdurchschnitt etwas jünger sind.[4] Informativer ist ein Vergleich der Altersverteilungen, wie sie exemplarisch in Abbildung 12.1-1 für das Jahr 1999 dargestellt werden. Die durchgezogene Linie zeigt für jedes Alter von 0 bis 100 Jahren den Anteil (relative Häufigkeit) der männlichen GEK-Versicherten in diesem Alter. Die gestrichelte Linie bezieht sich auf die gesamte männliche Bevölkerung in Deutschland.[5] Man erkennt, dass die GEK überproportional viele männliche Versicherte im Altersbereich von Anfang 20 bis Ende 40 hat. Da uns in späteren Abschnitten die älteren, bereits Rente beziehenden Versicherten interessieren, zeigt Abbildung 12.1-2 die Altersverteilungen nur für diejenigen Personen, die mindestens 55 Jahre alt sind. Wiederum erkennt man, dass bei der GEK mehr vergleichsweise jüngere Personen – in diesem Fall bis etwa zum Alter 65 – versichert sind. Diese Feststellungen betreffen das Jahr 1999; man kann aber annehmen, dass es sich für die vorangegangenen Jahre entsprechend (und sogar noch ausgeprägter) so verhält.

Von besonderer Bedeutung für unsere Untersuchung sind die Sterbefälle. Die Angaben in Box 12.1-1 zeigen, dass man annehmen kann, dass

[1] Dieser Datenbestand liegt auch der Gesundheitsberichterstattung der GEK zugrunde; vgl. Gmünder Ersatzkasse (Hg.), GEK-Gesundheitsreport 2002, Sankt Augustin: Asgard-Verl. Hippe 2002. Wie im folgenden näher beschrieben wird, verwenden wir aus dem umfangreichen Datenbestand nur einige wenige Angaben, die sich auf das Geburtsjahr, Versicherungszeiten, Rentenbezugszeiten und Sterbefälle beziehen.

[2] Die Daten wurden im Rahmen eines am Zentrum für Sozialpolitik (Bremen) durchgeführten Projekts aufbereitet; dafür möchte ich mich bei Wolfgang Voges herzlich bedanken.

[3] Für jedes Jahr wird die Anzahl der männlichen Personen angegeben, die mindestens einen Tag des Jahres bei der GEK versichert waren.

[4] Das Alter wird in diesem Text stets als Differenz zwischen Kalenderjahr und Geburtsjahr berechnet.

[5] Statistisches Bundesamt, Fachserie 1, Reihe 1, 1999, S. 64f.

Box 12.1-1 Anzahl und Durchschnittsalter sowie Sterbefälle der männlichen Versicherten der GEK von 1960 bis 2002.

Jahr	Anzahl Versicherte	Durchschnitts-alter	Anzahl Sterbefaelle
1960	77719	33.5	1
1961	86916	33.0	0
1962	94530	32.8	1
1963	103212	32.4	0
1964	112515	32.1	1
1965	123031	31.9	0
1966	133870	31.6	2
1967	144000	31.5	3
1968	154635	31.5	12
1969	165601	31.5	12
1970	175124	31.6	16
1971	183563	31.7	26
1972	190405	31.9	29
1973	197118	32.1	37
1974	202177	32.4	799
1975	208195	32.7	1203
1976	217154	32.9	1365
1977	226670	33.0	1937
1978	233009	33.4	2074
1979	245451	33.6	2213
1980	260315	33.7	2366
1981	275554	33.9	2405
1982	290669	33.8	2522
1983	303076	33.9	2665
1984	319020	34.0	2536
1985	334149	34.0	2639
1986	353788	34.0	2672
1987	377064	33.8	2621
1988	402570	33.7	2529
1989	424042	33.7	2292
1990	459385	33.5	2474
1991	523161	33.2	2630
1992	568439	33.1	2720
1993	598987	33.3	2880
1994	619876	33.7	2900
1995	650179	33.9	3003
1996	690789	34.3	3059
1997	758743	34.3	3291
1998	810457	34.5	3443
1999	843268	34.8	3575
2000	807680	35.4	3675
2001	788750	35.9	3694
2002	783002	36.6	2735

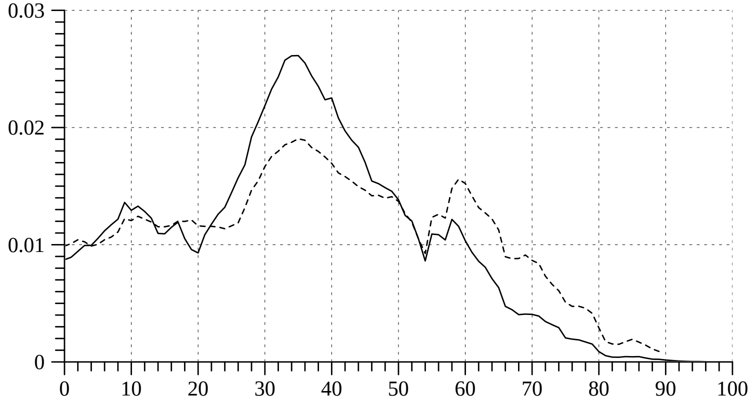


Abb. 12.1-1 Altersverteilung der männlichen GEK-Versicherten (durchgezogene Linie) und der männlichen Gesamtbevölkerung (gestrichelte Linie) im Jahr 1999.

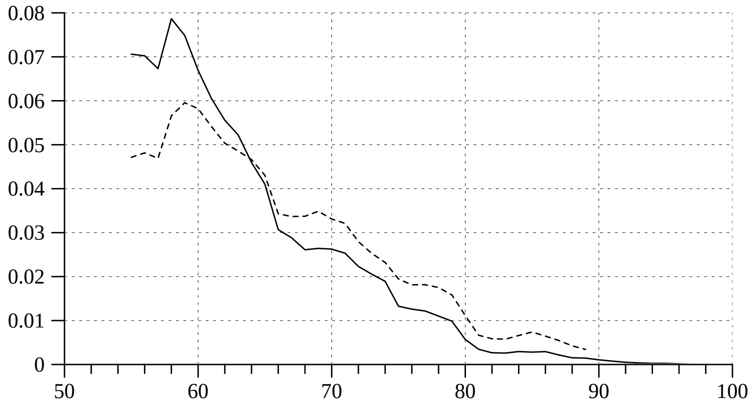


Abb. 12.1-2 Altersverteilung der männlichen GEK-Versicherten (durchgezogene Linie) und der männlichen Gesamtbevölkerung (gestrichelte Linie) im Jahr 1999, beschränkt auf Personen, die mindestens 55 Jahre alt sind.

die Sterbefälle in dem uns vorliegenden Datensatz etwa ab dem Jahr 1975 systematisch erfasst worden sind.

2. Berechnung von Lebenserwartungen. Das Mortalitätsgeschehen kann im Querschnitt oder im Längsschnitt erfasst werden. Wir folgen zunächst einer Querschnittsbetrachtung, wie sie der amtlichen Statistik, aber auch der Statistik des VDR zugrunde liegt. Man bezieht sich dann auf eine be-

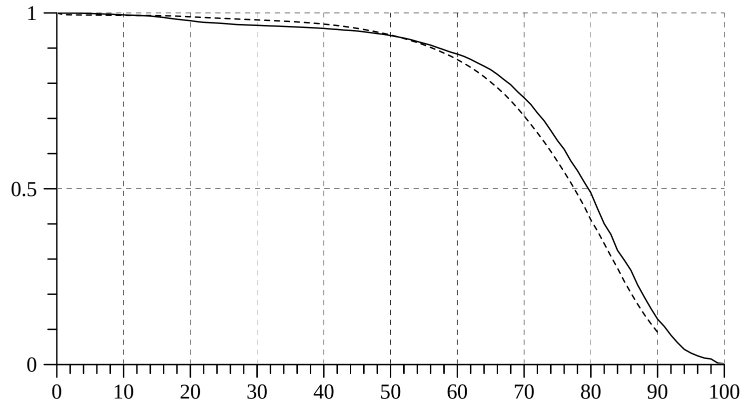


Abb. 12.1-3 Vergleich der Überlebensfunktionen der männlichen GEK-Versicherten (durchgezogene Linie) und der männlichen Gesamtbevölkerung (gestrichelte Linie) im Jahr 1999.

stimmte Periode (z.B. ein Kalenderjahr oder ein 3-Jahres-Intervall) und berechnet zunächst für diese Periode altersspezifische Sterberaten; dann wird aus diesen Sterberaten eine Sterbetafel konstruiert. In den folgenden Berechnungen gehen wir von Kalenderjahren aus. Die altersspezifische Sterberate für das Kalenderjahr t und das Alter τ wird folgendermaßen berechnet:

$$\delta_{t,\tau} \;=\; \frac{d_{t,\tau}}{n_{t,\tau}}$$

Im Zähler steht die Anzahl der Sterbefälle, die im Jahr t und im Alter τ eingetreten sind; im Nenner steht die Anzahl der männlichen Personen im Jahr t im Alter τ (bei denen also in diesem Jahr und in diesem Alter ein Sterbefall eintreten könnte). Dann erhält man die Überlebensfunktion durch

$$G_{t,\tau} \;=\; \prod_{j=0}^{\tau-1} (1-\delta_{t,j})$$

Abbildung 12.1-3 zeigt die auf diese Weise berechneten Überlebensfunktionen sowohl für die männlichen GEK-Versicherten als auch für die männliche Gesamtbevölkerung, in beiden Fällen für das Jahr 1999.[6] Man erkennt,

[6] Für die Berechnung der Überlebensfunktion für die GEK-Versicherten wurden Personen und Sterbefälle bis zum Alter 100 berücksichtigt. Die Daten zur Berechnung einer Überlebensfunktion für die männliche Gesamtbevölkerung wurden der Fachserie 1, Reihe 1 (1999) des Statistischen Bundesamtes entnommen; in diesem Fall gibt es eine nach oben offene Altersklasse, die im Alter 90 beginnt.

Box 12.1-2 Anzahl der männlichen GEK-Versicherten (N) und Anzahl der Sterbefälle (D) in den Jahren 1980 bis 2001, sowie die auf der Grundlage jährlicher Sterbetafeln berechneten ferneren Lebenserwartungen.

			fernere Lebenserwartung im Alter ...							
Jahr	N	D	50	55	60	65	70	75	80	85
1980	260315	2366	25.1	20.9	17.0	13.3	10.2	7.8	5.8	4.1
1981	275536	2404	24.9	20.7	16.7	13.2	10.1	7.5	5.4	3.7
1982	290651	2521	25.2	21.0	16.9	13.4	10.0	7.2	5.1	4.0
1983	303062	2665	25.4	21.2	17.3	13.8	10.7	8.0	6.1	4.6
1984	319005	2535	26.1	21.7	17.6	13.9	10.8	8.4	6.3	5.2
1985	334136	2639	26.1	21.7	17.7	14.2	10.9	8.2	6.1	4.5
1986	353776	2672	26.1	21.8	17.8	14.2	10.8	8.0	5.7	3.4
1987	377054	2621	26.5	22.2	18.2	14.4	11.1	8.5	6.4	5.0
1988	402561	2529	26.4	22.1	18.1	14.4	11.2	8.1	5.5	3.8
1989	424035	2292	27.1	22.7	18.6	14.9	11.4	8.6	6.5	4.5
1990	459381	2474	26.8	22.4	18.2	14.4	11.0	8.0	5.6	4.0
1991	523158	2630	27.0	22.6	18.5	14.9	11.6	8.6	6.6	5.0
1992	568435	2719	27.5	23.1	18.9	15.1	11.7	8.8	6.3	4.7
1993	598983	2880	27.2	22.8	18.6	14.9	11.5	8.6	6.3	4.3
1994	619872	2900	27.5	23.1	18.9	15.1	11.6	8.6	6.2	4.2
1995	650175	3003	27.7	23.4	19.2	15.3	11.9	9.0	6.7	4.5
1996	690784	3057	28.0	23.6	19.4	15.4	12.0	8.7	6.2	4.1
1997	758740	3291	28.1	23.6	19.5	15.6	12.0	9.0	6.7	4.5
1998	810453	3443	28.6	24.2	19.9	16.0	12.3	9.3	6.6	4.1
1999	843268	3575	28.6	24.2	20.0	16.0	12.4	9.4	6.6	4.6
2000	807676	3674	28.8	24.4	20.2	16.2	12.6	9.6	6.9	4.6
2001	788746	3693	29.1	24.7	20.6	16.6	12.9	9.7	7.0	4.7

dass die männlichen GEK-Versicherten im Vergleich zur männlichen Gesamtbevölkerung etwa ab dem Alter 60 eine etwas höhere Lebenserwartung haben.

Auf der Grundlage von Perioden-Sterbetafeln können fernere Lebenserwartungen berechnet werden. Ausgangspunkt ist die Definition

$$p_{t,\tau} = G_{t,\tau} - G_{t,\tau+1}$$

mit der der relative Anteil der im Jahr t im Alter τ gestorbenen Personen erfasst wird.[7] Dann ist

$$E_{t,\tau} = \frac{\sum_{j=\tau}^{100} j\, p_{t,j}}{\sum_{j=\tau}^{100} p_{t,j}} - \tau$$

die fernere Lebenserwartung im Jahr t für diejenigen Personen, die bereits

[7] Für das höchste Alter – in den folgenden Berechnungen 100 Jahre – wird angenommen, dass $G_{t,101} = 0$ ist.

Box 12.1-3 Verteilung des zeitlich ersten Rentenbezugs der männlichen GEK-Versicherten auf unterschiedliche Rentenarten.

Rentenart	Anzahl	Prozent
Altersrente	91980	61.1
BU-Rente	7213	4.8
EU-Rente	31888	21.2
Hinterbliebenenrente	19186	12.8
Renten nach Ost-Recht	122	0.1
	150389	100.0

ein Alter von τ Jahren erreicht haben.[8]

Box 12.1-2 zeigt, wie sich die auf diese Weise berechneten ferneren Lebenserwartungen bei den männlichen GEK-Versicherten im Zeitraum von 1980 bis 2001 entwickelt haben. Von jährlichen Schwankungen abgesehen haben diese Lebenserwartungen ersichtlich zugenommen. Zu berücksichtigen ist, dass die Berechnungen von jährlichen Perioden-Sterbetafeln ausgehen, die die tatsächliche Zunahme der Lebenserwartungen im historischen Zeitablauf unterschätzen.

3. Informationen über Rentenempfänger. Im weiteren interessieren uns nur männliche Versicherte, die eine Rente beziehen. Dafür verwenden wir einen weiteren Datensatz aus dem GEK-Datenbestand, der für jeden Versicherten einen oder mehrere Rekords mit folgenden Angaben enthält:

a) Datum des Rentenantrags

b) Datum des Rentenbeginns (wenn der Antrag bewilligt wurde)

c) Datum der Beendigung des Rentenbezugs (oder Angabe, dass die Rente noch bezogen wird[9])

d) Typ der beantragten bzw. bezogenen Rente

Bei den Rentenarten werden unterschieden: (1) Altersrenten, (2) Berufsunfähigkeitsrenten, (3) Erwerbsunfähigkeitsrenten, (4) Hinterbliebenenrenten, und (5) Renten nach bisherigem Ostrecht (Rentenbeginn vor dem 1.1.1991).

Der gegenwärtig verfügbare Datensatz enthält Einträge für insgesamt

[8] Man könnte ein halbes Jahr hinzufügen, um zu berücksichtigen, dass wir das Alter als Differenz zwischen Kalenderjahr und Geburtsjahr berechnen.

[9] In diesem Fall nehmen wir an, dass die Rente mindestens bis zum Jahr 2002 bezogen worden ist.

Box 12.1-4 Männliche GEK-Versicherte, die in dem angegebenen Jahr eine Alters- oder EU-Rente beziehen.

Jahr	Rentenbezieher insgesamt	Rentenbezieher Rentenzugangsalter 55-66
1975	13630	12329
1976	15567	13995
1977	16857	15069
1978	17559	15657
1979	18156	16111
1980	19077	16877
1981	20114	17719
1982	21404	18864
1983	23094	20336
1984	24734	21804
1985	26352	23268
1986	27850	24631
1987	29449	26069
1988	31531	27977
1989	33676	29947
1990	36010	32072
1991	38338	34180
1992	40709	36188
1993	43920	38980
1994	47610	42228
1995	51794	46072
1996	57219	50834
1997	61699	54797
1998	67877	60177
1999	73317	64902
2000	78096	69176
2001	82203	72890

164743 männliche Versicherte, davon haben 150389 Personen mindestens eine bewilligte Rente, sind also für einen gewissen Zeitraum Bezieher einer Rente. Box 12.1-3 zeigt, wie sich bei diesen Personen der zeitlich erste Rentenbezug auf die unterschiedlichen Rentenarten verteilt. Im weiteren beziehen wir uns nur auf Personen, deren erste Rente eine Alters- oder eine EU-Rente ist. Für diese Personen wird das Rentenzugangsalter als dasjenige Alter definiert, in dem eine Person zum ersten Mal eine Alters- oder eine EU-Rente bezogen hat. Box 12.1-4 zeigt für die Jahre 1975 bis 2001 die Anzahl der männlichen Personen, die in dem betreffenden Jahr bei der GEK versichert waren und bei denen der erstmalige Bezug einer Alters- oder EU-Rente vor oder in dem betreffenden Jahr begonnen hat. Wie in der dritten Spalte der Box angegeben wird, beschränken wir uns im weiteren auf Personen, bei denen der erstmalige Rentenbezug im Altersbereich

Box 12.1-5 Männliche GEK-Versicherte, die im Zeitraum 1975 bis 2001 zum ersten Mal eine Alters- oder EU-Rente bezogen haben und bei denen das Rentenzugangsalter zwischen 55 und 66 Jahren liegt.

RA	AR	EU	Insgesamt
55	16	1692	1708
56	31	1954	1985
57	30	2344	2374
58	58	2735	2793
59	83	2753	2836
60	28543	2022	30565
61	10499	1355	11854
62	5944	999	6943
63	20257	362	20619
64	4900	93	4993
65	6180	31	6211
66	603	2	605
Insg.	77144	16342	93486

von 55 bis 66 Jahren stattgefunden hat. Box 12.1-5 zeigt für den Kalenderzeitraum von 1975 bis 2001 das Rentenzugangsalter dieser Personen.[10] Das durchschnittliche Rentenzugangsalter dieser Personen wird in Box 12.1-6 angegeben. Addiert man ein halbes Jahr hinzu, ist es sehr ähnlich zu dem in der VDR-Statistik angegebenen durchschnittlichen Rentenzugangsalter.

4. Berechnung von Rentenbezugsdauern. Im weiteren interessiert uns, wie lange diese Personen maximal Rente bezogen haben könnten. Das ist die Zeitdauer von ihrem Rentenzugangsalter bis zu ihrem Tod.[11] Bei Personen, die noch nicht gestorben sind, ist diese fernere Lebensdauer, die mit dem Rentenzugangsalter beginnt, natürlich noch nicht bekannt. Wir folgen deshalb in diesem Abschnitt einer Methode, wie sie auch in der VDR-Statistik verwendet wird, um die durchschnittliche Rentenbezugsdauer zu berechnen: Für jedes Kalenderjahr werden diejenigen Personen betrachtet, die aus dem Rentenbezug ausscheiden (in den meisten Fällen durch Tod), und es wird dann für diese Personen berechnet, wie lange sie im

[10] Bei den Altersrenten, die vor dem Alter 60 begonnen haben, handelt es sich vermutlich überwiegend um Fehlbuchungen innerhalb des Datensatzes, die jedoch wegen der geringen Anzahl nicht ins Gewicht fallen. Ein Rentenzugangsalter von 59 Jahren ist im übrigen auch wegen unserer Berechnung des Alters als Differenz von Kalenderjahr und Geburtsjahr möglich.

[11] Zur Vereinfachung der Berechnungen verwenden wir also nicht die effektive Rentenbezugsdauer, die mit unseren Daten auch nur schwer zu ermitteln wäre, sondern die maximal mögliche Rentenbezugsdauer, also die fernere Lebensdauer, die mit dem Rentenzugangsalter beginnt.

Box 12.1-6 Durchschnittliches Rentenzugangsalter (DRA) bei männlichen GEK-Versicherten, die im Zeitraum 1975 bis 2001 zum ersten Mal eine Alters- oder EU-Rente bezogen haben und bei denen das Rentenzugangsalter zwischen 55 und 66 Jahren liegt. Ergänzend ist DRA1 das durchschnittliche Rentenzugangsalter bei EU-Renten ohne untere Altersbeschränkung.

	Insgesamt		AR		EU		
	Anzahl	DRA	Anzahl	DRA	Anzahl	DRA	DRA1
1975	1644	63.1	1434	63.6	210	59.8	52.6
1976	1694	62.5	1396	63.2	298	59.2	52.8
1977	1519	62.2	1213	63.1	306	58.6	52.9
1978	1299	61.9	962	63.0	337	58.6	53.7
1979	1224	61.4	874	62.6	350	58.2	53.3
1980	1541	60.9	1143	61.7	398	58.4	53.2
1981	1678	60.7	1225	61.7	453	58.3	53.4
1982	2051	61.0	1574	61.7	477	58.5	53.4
1983	2442	61.2	1922	61.9	520	58.5	53.4
1984	2460	61.1	1878	61.9	582	58.6	53.6
1985	2495	61.2	1971	61.9	524	58.5	53.5
1986	2430	61.0	1877	61.8	553	58.5	53.7
1987	2573	61.2	1998	61.9	575	58.5	53.9
1988	3101	61.3	2468	61.9	633	58.8	54.3
1989	3208	61.3	2626	61.9	582	58.7	53.5
1990	3359	61.4	2708	62.0	651	58.7	53.7
1991	3628	61.5	2849	62.3	779	58.6	53.7
1992	3489	61.4	2781	62.2	708	58.3	52.6
1993	4372	61.3	3605	61.9	767	58.3	52.4
1994	5178	61.1	4347	61.7	831	58.2	52.2
1995	6061	60.9	5174	61.4	887	58.1	52.6
1996	5746	60.6	4787	61.2	959	57.9	52.4
1997	5706	60.8	4947	61.2	759	57.8	51.4
1998	5813	60.9	4944	61.4	869	58.1	51.7
1999	6213	61.0	5287	61.5	926	58.2	51.4
2000	6507	61.1	5637	61.5	870	58.3	51.7
2001	6055	61.4	5517	61.7	538	58.2	50.6
Insg.	93486		77144		16342		

Durchschnitt eine Rente bezogen haben. Wegen der vergleichsweise kleinen Fallzahlen verwenden wir jedoch nicht einzelne Kalenderjahre, sondern 5-Jahres-Intervalle.

Box 12.1-7 zeigt die Rechenergebnisse.[12] Der obere Teil bezieht sich auf alle Personen, die mit einer Alters- oder EU-Rente begonnen haben und im angegebenen Zeitraum bei der GEK versichert waren. Von diesem Personen sind z.B. im Zeitraum 1977 bis 1981 3775 Personen gestorben;

[12] Alter und Lebensdauern werden wie stets in diesem Beitrag als Differenz zwischen Kalenderjahr (Sterbejahr) und Geburtsjahr oder dem Jahr des Rentenzugangs berechnet.

Box 12.1-7 Durchschnittliches Alter und Lebensdauer vom Rentenzugangsalter bis zum Tod bei männlichen GEK-Versicherten, die in den angegebenen Zeiträumen gestorben sind.

Zeitraum	Anzahl	durchschnittl. Alter	Lebensdauer von Verrentung bis Tod
1977 - 1981	3775	72.4	8.8
1982 - 1986	4754	73.4	10.5
1987 - 1991	5749	74.0	11.7
1992 - 1996	7226	74.2	12.4
1997 - 2001	9455	74.0	12.6

Rentenbeginn mit Altersrente

Zeitraum	Anzahl	durchschnittl. Alter	Lebensdauer von Verrentung bis Tod
1977 - 1981	3096	73.8	9.3
1982 - 1986	3820	75.5	11.6
1987 - 1991	4547	76.2	12.9
1992 - 1996	5650	76.3	13.5
1997 - 2001	7171	75.9	13.6

Rentenbeginn mit EU-Rente

Zeitraum	Anzahl	durchschnittl. Alter	Lebensdauer von Verrentung bis Tod
1977 - 1981	679	65.7	6.1
1982 - 1986	934	65.0	6.1
1987 - 1991	1202	65.6	7.1
1992 - 1996	1576	66.6	8.2
1997 - 2001	2284	67.9	9.5

diese Personen sind im Durchschnitt 72.4 Jahre alt geworden und haben noch durchschnittliche 8.8 Jahre nach ihrem Rentenzugangsalter gelebt. In Bezug auf das durchschnittliche Lebensalter sind diese Zahlen ähnlich zu denjenigen, die in der VDR-Statistik veröffentlicht worden sind.[13] Die in der VDR-Statistik angegebenen durchschnittlichen Rentenbezugsdauern sind etwas länger als die Werte in Box 12.1-7, was daran liegen mag, dass wir uns auf Alters- und EU-Renten sowie auf Rentenzugänge im Altersbereich 55 bis 66 Jahre beschränkt haben.

Bemerkenswert sind die deutlichen Unterschiede zwischen Personen, die mit einer Alters- oder einer EU-Rente begonnen haben (untere Teile von Box 12.1-7). Personen, die mit einer EU-Rente begonnen haben,

[13] Vgl. VDR: Rentenversicherung in Zeitreihen, Ausgabe 2002, S. 128.

Box 12.1-8 Durchschnittliches Alter und Lebensdauer vom Rentenzugangsalter bis zum Tod bei männlichen GEK-Versicherten, die in den angegebenen Zeiträumen gestorben sind, differenziert nach dem Rentenzugangsalter (RA).

Zeitraum	RA	Anzahl	durchschnittl. Alter	Lebensdauer von Verrentung bis Tod
1977 - 1981	55 - 56	119	59.8	4.3
	57 - 58	131	62.8	5.3
	59 - 60	192	64.9	5.3
	61 - 62	274	67.6	6.1
	63 - 64	744	68.8	5.4
	65 - 66	2315	75.9	10.8
1982 - 1986	55 - 56	206	61.6	6.0
	57 - 58	261	63.1	5.6
	59 - 60	452	65.0	5.4
	61 - 62	523	68.6	7.1
	63 - 64	1148	71.8	8.5
	65 - 66	2164	79.6	14.5
1987 - 1991	55 - 56	292	62.3	6.8
	57 - 58	372	64.9	7.4
	59 - 60	864	66.4	6.6
	61 - 62	789	70.2	8.8
	63 - 64	1661	74.5	11.2
	65 - 66	1771	82.6	17.6
1992 - 1996	55 - 56	361	63.0	7.5
	57 - 58	518	65.5	8.0
	59 - 60	1428	68.4	8.6
	61 - 62	1139	72.1	10.7
	63 - 64	2382	76.3	13.1
	65 - 66	1398	84.2	19.2
1997 - 2001	55 - 56	525	63.9	8.3
	57 - 58	729	66.9	9.4
	59 - 60	2477	69.5	9.7
	61 - 62	1708	73.5	12.1
	63 - 64	2781	78.1	14.9
	65 - 66	1235	82.6	17.5

haben nicht nur eine deutlich kürzere Lebensdauer, sondern auch ihre fernere Lebensdauer nach dem Rentenzugangsalter (und somit die maximale Rentenbezugsdauer) ist deutlich kürzer.

Weitere Aufschlüsse lassen sich gewinnen, wenn man nach dem Rentenzugangsalter (RA) differenziert. Das geschieht in Box 12.1-8. Man erkennt, dass Personen, die vor dem Alter 65 verrentet werden, keineswegs eine längere fernere Lebensdauer aufweisen.

Box 12.1-9 Durchschnittliches Alter und Lebensdauer vom Rentenzugangsalter bis zum Tod bei männlichen GEK-Versicherten, die in den angegebenen Zeiträumen gestorben sind und deren Rentenbezug mit einer Altersrente begonnen hat, differenziert nach dem Rentenzugangsalter (RA).

Zeitraum	RA	Anzahl	durchschnittl. Alter	Lebensdauer von Verrentung bis Tod
1977 - 1981	59 - 60	24	67.0	7.1
	61 - 62	123	67.2	5.5
	63 - 64	644	68.1	4.8
	65 - 66	2298	75.9	10.8
1982 - 1986	59 - 60	205	65.4	5.5
	61 - 62	360	68.6	7.0
	63 - 64	1075	71.6	8.3
	65 - 66	2148	79.6	14.5
1987 - 1991	59 - 60	544	66.4	6.4
	61 - 62	584	70.3	8.8
	63 - 64	1608	74.5	11.3
	65 - 66	1764	82.7	17.6
1992 - 1996	59 - 60	1021	68.4	8.7
	61 - 62	894	72.3	10.9
	63 - 64	2316	76.4	13.1
	65 - 66	1394	84.2	19.2
1997 - 2001	59 - 60	1832	69.6	9.6
	61 - 62	1375	73.5	12.1
	63 - 64	2703	78.3	15.0
	65 - 66	1229	82.6	17.5

Wegen der deutlichen Unterschiede zwischen Personen, die mit einer Alters- oder eine EU-Rente begonnen haben, ist es sinnvoll, auch noch ausschließlich diejenigen zu betrachten, die mit einer Altersrente begonnen haben. Analog zu Box 12.1-8 bezieht sich Box 12.1-9 nur auf diese Personen. Wiederum zeigt sich auch bei diesen Personen eine deutliche Abhängigkeit der ferneren Lebensdauer vom Rentenzugangsalter.

5. Berechnungen mit zensierten Daten. Anstatt von den Sterbefällen auszugehen, kann man auch eine prospektive Berechnung durchführen. Wir betrachten hierfür alle männlichen Personen, die im Zeitraum 1975 – 2001 und im Alter von 55 bis 66 Jahren mit einem Alters- oder EU-Rentenbezug begonnen haben (vgl. Box 12.1-5). Dann werden zwei Variablen D und T definiert. Für jede Person bekommt die Variable D den Wert 1, wenn die Person bis zum Ende des Beobachtungszeitraums im Jahr 2002 gestor-

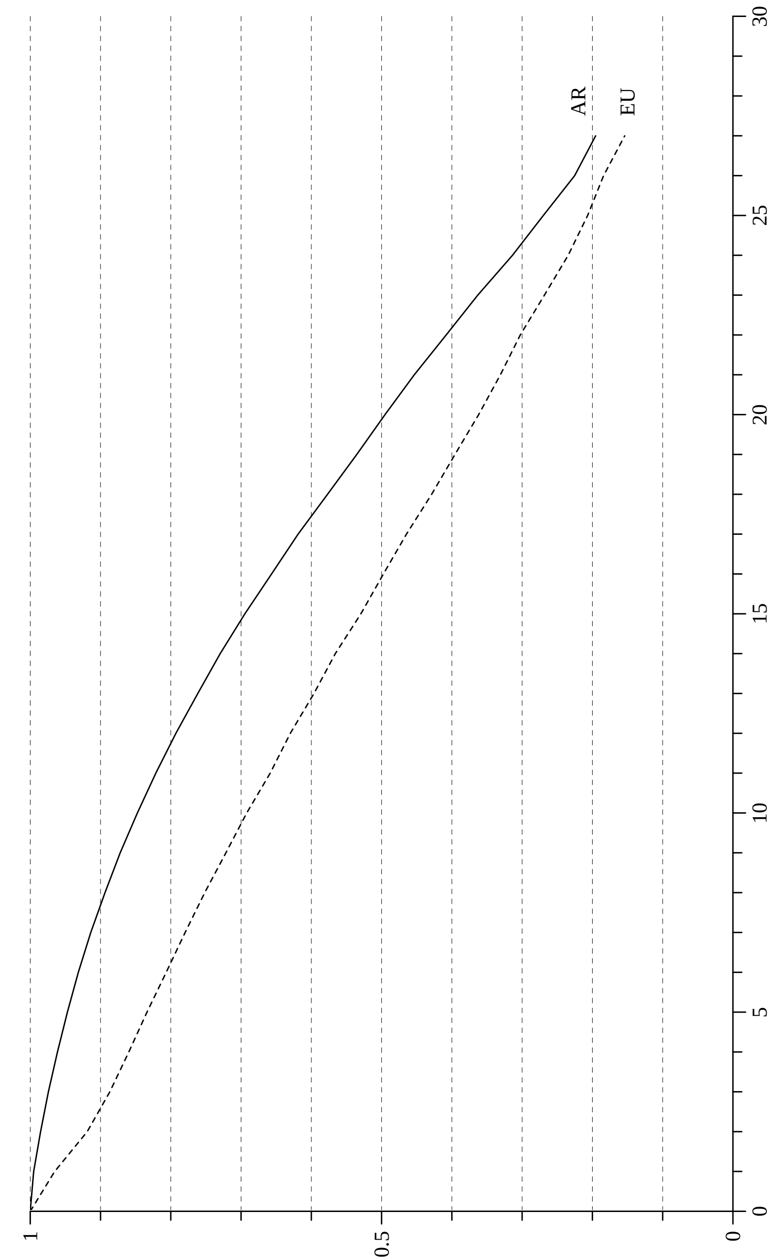


Abb. 12.1-4 Überlebensfunktionen der männlichen GEK-Versicherten, die im Zeitraum von 1975 bis 2001 mit einer Alters- oder EU-Rente begonnen haben.

ben ist, andernfalls den Wert 0; und je nachdem werden die Werte von T bestimmt: Wenn die Person gestorben ist, wird für T die Lebensdauer vom Rentenzugangsalter bis zum Tod verwendet, andernfalls die Lebensdauer vom Rentenzugang bis zum Ende der Beobachtungsperiode. Somit hat man es mit teilweise rechts zensierten Beobachtungen zu tun. Deshalb berechnen wir in diesem Fall Sterberaten durch

$$r(\tau) \;=\; \frac{\text{Anzahl Personen mit } T = \tau}{\text{Anzahl Personen mit } T \geq \tau}$$

und mit Hilfe dieser Raten Überlebensfunktionen

$$G^*(\tau) \;=\; \prod_{j=0}^{\tau-1} (1 - r(j))$$

Da unsere Beobachtungsperiode im Jahr 1975 beginnt und im Jahr 2002 endet, können auf diese Weise Überlebensfunktionen für maximal 27 Jahre nach dem Beginn der Verrentung geschätzt werden.

Abbildung 12.1-4 zeigt auf diese Weise berechnete Überlebensfunktionen für Personen, die mit einer Alters- oder einer EU-Rente begonnen haben. Es bestätigt sich das bereits im vorangegangenen Abschnitt erzielte Ergebnis, dass Personen, die mit einer EU-Rente beginnen, eine deutlich kürzere fernere Lebenserwartung haben als Personen, die mit einer Altersrente beginnen. Beschränkt man sich auf eine maximale fernere Lebensdauer von 27 Jahren, kann man auch Durchschnittswerte berechnen (im Prinzip auf die gleiche Weise wie in § 2 dargestellt worden ist). Für Personen, die mit einer Altersrente beginnen, beträgt dann die durchschnittliche fernere Lebenserwartung 18.2 Jahre, für Personen, die mit einer EU-Rente beginnen, 15.0 Jahre.[14] Dabei ist natürlich zu berücksichtigen, dass wir uns auf EU-Renten beschränkt haben, die frühestens im Alter 55 beginnen.

Wiederum kann man auch nach dem Rentenzugangsalter differenzieren. Abbildung 12.1-5 zeigt die Überlebensfunktionen, diesmal auf einer Lebenszeitachse: die Überlebensfunktionen beginnen mit dem Rentenzugangsalter, so dass ein direkter Vergleich möglich ist. Auch in diesem Fall bestätigt sich im wesentlichen das Ergebnis des vorangegangenen Abschnitts. Einen bemerkenswerten Unterschied gibt es jedoch beim Vergleich der Rentenzugangsalter 63-64 und 65-66, die in diesem Fall fast identische Überlebenskurven aufweisen. Das könnte ein Hinweis darauf sein, dass sich Verrentungen im Alter 63 – im Unterschied zu solchen, die in einem früheren Alter erfolgen – nicht in erster Linie gesundheitlichen Beeinträchtigungen verdanken. Ergänzend zeigt Abbildung 12.1-6 nach dem Rentenzugangsalter differenzierte Überlebensfunktionen nur für diejenigen Personen, die mit einer Altersrente begonnen haben.

[14] Ergänzend könnte man auch die aus Abbildung 12.1-4 direkt ablesbaren Median-Werte heranziehen.

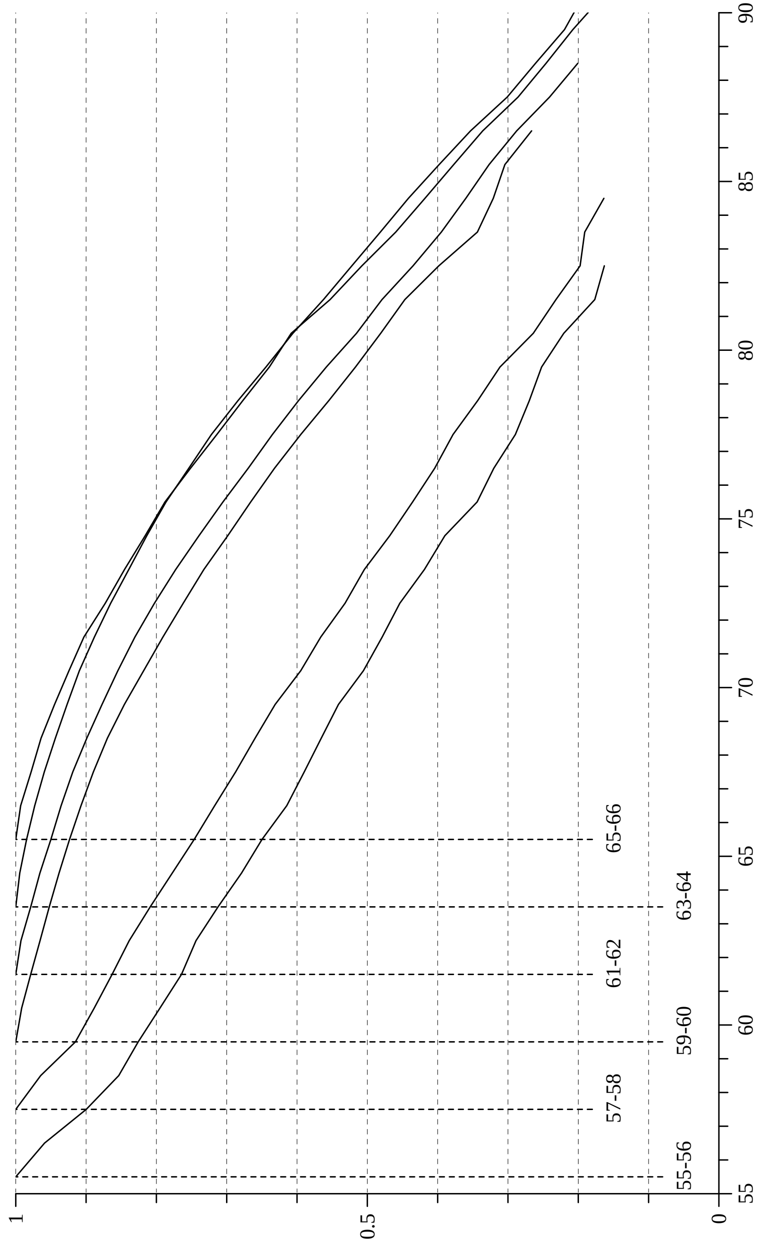


Abb. 12.1-5 Überlebensfunktionen der männlichen GEK-Versicherten, die im Zeitraum von 1975 bis 2001 mit einer Alters- oder EU-Rente begonnen haben, differenziert nach dem Rentenzugangsalter.

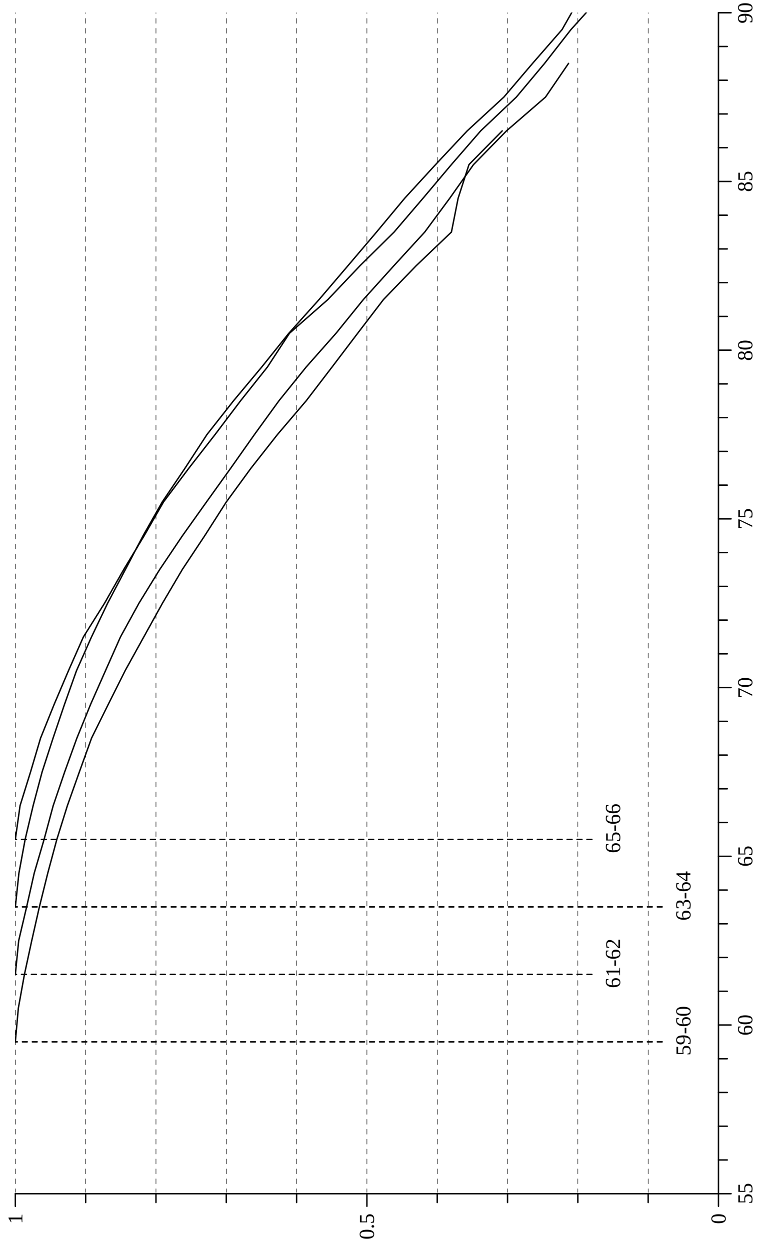


Abb. 12.1-6 Überlebensfunktionen der männlichen GEK-Versicherten, die im Zeitraum von 1975 bis 2001 mit einer Altersrente begonnen haben, differenziert nach dem Rentenzugangsalter.

12.2 Berechnungen mit SOEP-Daten

In diesem Abschnitt soll mit Hilfe von Daten des Sozio-ökonomischen Panels (SOEP) versucht werden, Hinweise auf mögliche Zusammenhänge zwischen dem Verrentungsalter und der Mortalität zu gewinnen. Dabei gibt es zwei mögliche Zugänge. Einerseits können mit dem sog. Kalendarium Personen ermittelt werden, bei denen während der Laufzeit des Panels ein Rentenbezug begonnen hat; diesen Ansatz verfolgen wir im ersten Abschnitt. Andererseits kann man sich auf Personen beziehen, die in der zweiten Welle des SOEP (1985) danach gefragt worden sind, seit wann sie eine Rente beziehen; damit beschäftigt sich der zweite Abschnitt. Bei beiden Ansätzen beschränken wir uns auf männliche Rentenbezieher.

1. Daten des Kalendariums. In diesem Abschnitt verwenden wir das Kalendarium, in dem die Teilnehmer des SOEP für jeden Monat des jeweils vergangenen Jahres nach einer Reihe von Sachverhalten gefragt werden; und zwar verwenden wir das Kalendarium für Einkommensarten, in dem nach unterschiedlichen Arten von Einkommen gefragt wird. Zwei Fragen beziehen sich auf Renten: einerseits die Rubrik „Altersrente–EU–BU“ und andererseits die Rubrik „Witwen/Witwer-Rente“. Für unsere Untersuchung ist nur die erste Rubrik relevant. Tatsächlich wäre es auch wünschenswert, wenn man sich auf Alters- und EU-Renten beschränken könnte; aber diese Spezifizierung ist mit dem Kalendarium nicht möglich.

Das Kalendarium beginnt im Januar 1983. Allerdings hängt der tatsächliche Beginn bei jeder Person davon ab, wann sie zum ersten Mal an einer Personenbefragung teilgenommen hat. Wenn eine Person zum ersten Mal im Jahr t einen Personenfragebogen ausgefüllt hat, beginnen ihre Einträge im Kalendarium im Januar des Jahres $t-1$. Also kann man für jede Person feststellen, ob sie mindestens einmal und, wenn ja, in welchem Jahr sie zum ersten Mal den Bezug einer Alters-, EU- oder BU-Rente angegeben hat. Allerdings ist dies nicht unbedingt das Jahr, in dem der Rentenbezug tatsächlich begonnen hat; denn bei Personen, die zum ersten Mal im Jahr t befragt worden sind und die bereits für den Januar des Jahres $t-1$ einen Rentenbezug angegeben haben, weiß man nicht, wann ihr Rentenbezug begonnen hat. Diese Fälle sind links zensiert, und wir werden sie im folgenden ausschließen. Anders formuliert: Wir beschränken uns auf diejenigen männlichen Personen, bei denen ein Rentenbezug während ihrer Teilnahme am SOEP begonnen hat. Da das Einkommenskalendarium in der Welle 12 des SOEP (1995) umgestellt worden ist, berücksichtigen wir das Kalendarium nur bis zum Ende des Jahres 1993. Diese Einschränkung ist auch deshalb sinnvoll, damit es eine hinreichende Zeitdauer nach der Verrentung gibt, in der Sterbefälle ermittelt werden können.

Insgesamt haben 3796 Personen (Männer und Frauen) bis zur Welle 11, d.h. bis zum Befragungsjahr 1994, in mindestens einem Monat einen

Box 12.2-1 Jahre des Rentenbeginns bei 702 männlichen Personen.

Jahr	Anzahl
1984	65
1985	76
1986	53
1987	60
1988	46
1989	52
1990	75
1991	104
1992	84
1993	87
Insgesamt	702

Rentenbezug angegeben.[15] In 2268 Fällen sind die Angaben links zensiert, so dass 1528 Personen übrig bleiben, bei denen der Rentenbezug seit 1983 begonnen hat; davon 738 Männer und 790 Frauen. Da uns die Mortalität interessiert, die erst ab der zweiten Welle des SOEP (1985) erfasst werden kann, schließen wir auch diejenigen Personen aus, deren Rentenbezug im Jahr 1983 begonnen hat. Dann verbleiben 702 männliche Personen, bei denen der Rentenbezug im Zeitraum 1984 bis 1993 begonnen hat. Box 12.2-1 zeigt, in welchen Jahren bei diesen Personen der Rentenbezug begonnen hat. Box 12.2-2 zeigt dazu korrespondierend das Alter, in dem der Rentenbezug begonnen hat. Wir beschränken uns im weiteren auf diejenigen 569 Personen, bei denen der Rentenbezug im Altersbereich 55 bis 66 Jahre begonnen hat.

Was kann nun über die Lebensdauer dieser 569 Personen ausgesagt werden? Als Datengrundlage dienen einerseits die im SOEP bereits erfassten Sterbefälle, außerdem eine von Infratest im Jahr 2001 durchgeführte Verbleibstudie.[16] Die Zusatzinformationen aus dieser Verbleibstudie sind insofern sehr nützlich, weil sie nicht nur die im SOEP bereits vorhandenen Angaben über Sterbefälle ergänzen, sondern weil sie es auch erlauben, über die bisher nicht gestorbenen Personen zu sagen, dass sie mindestens bis zum Jahr 2001 gelebt haben. Infolgedessen vermindert sich das Problem der rechts zensierten Beobachtungen. — Insgesamt sind 127 der 569 Personen bis zum Jahr 2001 gestorben. Box 12.2-3 zeigt für diese Personen, in welchem Alter sie gestorben sind. Dabei entspricht jede Zeile einem bestimmten Verrentungsalter. Zum Beispiel kann man der Box 12.2-2 entnehmen, dass 17 Personen im Alter 55 verrentet wurden, und dann ersieht

[15] Der Ausdruck „Rentenbezug“ bezieht sich hier und im folgenden stets auf Alters-, EU- und BU-Renten.

[16] Infratest Sozialforschung, Verbesserung der Datengrundlagen für Mortalitäts- und Mobilitätsanalysen: Verbleibstudie bei Panelausfällen im SOEP, München 2002.

Box 12.2-2 Verteilung der 702 Personen auf Verrentungsalter.

Alter	Anzahl	Alter	Anzahl	Alter	Anzahl	Alter	Anzahl
20	3	38	1	51	8	64	35
22	2	39	3	52	14	65	58
23	1	40	1	53	17	66	11
25	2	41	3	54	8	67	2
26	1	42	1	55	17	69	2
27	1	43	4	56	20	70	3
28	5	44	1	57	22	71	2
29	2	45	3	58	27	75	1
30	2	46	6	59	57	83	1
33	1	47	5	60	113	85	1
34	2	48	7	61	69	86	1
35	1	49	4	62	55		
36	3	50	8	63	85		

Box 12.2-3 Verteilung der 127 Sterbefälle auf das Rentenzugangsalter (Zeilen) und das Lebensalter (Spalten).

	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	78	79	Insg.
55	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4
56	0	0	0	0	0	2	2	0	0	1	0	0	0	1	0	0	0	0	0	0	0	6
57	0	1	0	0	1	0	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	6
58	0	0	1	0	3	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	8
59	0	0	0	1	1	1	1	2	2	2	2	2	1	2	0	0	0	0	0	0	0	17
60	0	0	0	0	1	0	1	2	3	3	4	1	0	0	4	1	0	0	0	0	0	20
61	0	0	0	0	0	1	1	2	2	2	1	0	1	3	1	0	0	1	0	0	0	15
62	0	0	0	0	0	0	0	1	2	0	0	4	0	0	1	2	1	0	0	0	0	11
63	0	0	0	0	0	0	0	2	1	0	2	3	3	3	1	1	2	0	0	0	0	18
64	0	0	0	0	0	0	0	0	0	3	0	1	1	1	1	0	0	0	0	0	1	8
65	0	0	0	0	0	0	0	0	0	2	2	1	0	0	0	0	2	0	2	1	1	11
66	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	3
Insg.	1	2	2	1	6	4	9	11	12	13	12	12	6	10	8	4	6	2	2	2	2	127

man aus Box 12.2-3, dass vier dieser Personen bis zum Jahr 2001 gestorben sind; ausserdem sieht man, in welchem Alter diese Personen gestorben sind.

Es ist klar, dass diese Informationen nicht genügen, um – differenziert nach dem Rentenzugangsalter – die fernere Lebenserwartung zu berechnen; denn für die meisten Personen ist ja nicht bekannt, wie lange sie noch leben werden. Man kann jedoch für einen gewissen Zeitraum, der mit der Verrentung beginnt, Survivorfunktionen schätzen, die einen gewissen Vergleich des Mortalitätsgeschehens im Anschluss an die Verrentung erlauben. Dabei können auch diejenigen Personen berücksichtigt werden, die bis zum Jahr 2001 noch nicht gestorben sind. Wegen der kleinen Fallzahl von Sterbefällen betrachten wir vier Gruppen:

a) Verrentungsalter 55 – 57,

b) Verrentungsalter 58 – 60,

c) Verrentungsalter 61 – 63,

d) Verrentungsalter 64 – 66.

Für jede dieser vier Gruppen wird nun gesondert eine Survivorfunktion berechnet, wobei sich die Zeitdauervariable auf die fernere Lebensdauer nach dem Eintritt der Verrentung bezieht. Ist z.B. eine Person im Jahr 1985 im Alter 60 verrentet worden und im Jahr 1999 gestorben, beträgt ihre fernere Lebensdauer 14 Jahre; ist sie dagegen bis zum Jahr 2001 nicht gestorben, liegt eine rechts zensierte Beobachtung vor, so dass man nur weiß, dass die fernere Lebensdauer dieser Person mindestens 16 Jahre beträgt.

Abbildung 12.2-1 zeigt den Verlauf der Survivorfunktionen für die vier Gruppen. Man erkennt, wie sich die Mortalität in den ersten 15 Jahren nach dem Beginn der Verrentung entwickelt. Zumindest für diesen Zeitraum kann man also sagen, dass Personen, die bereits in einem jüngeren Alter verrentet werden, keine niedrigere Mortalität aufweisen.

2. Daten aus der zweiten Welle. In diesem Abschnitt gehen wir von Daten aus, die in der zweiten Welle des SOEP (1985) erhoben worden sind. In dieser Welle wurden alle Personen gefragt, ob sie gegenwärtig eine Rente oder Pension beziehen und, wenn ja, welcher Art die Rente ist und wann der Rentenbezug begonnen hat. Dabei wurde zwischen (a) Altersrenten und Pensionen, (b) Erwerbs-, Berufsunfähigkeits- und Verletztenrenten und (c) Renten und Pensionen für Witwen und Waisen unterschieden. Für die weiteren Untersuchungen beziehen wir uns auf männliche Personen, die einen Rentenbezug in den Kategorien (a) und/oder (b) angegeben haben. Insgesamt gibt es 735 Personen, die eine Angabe über den Beginn ihres Rentenbezugs gemacht haben.[17] Dabei erstreckt sich dieser Beginn über die Lebensalter von 21 bis 75 Jahren. Wie in § 1 betrachten wir im weiteren nur Personen, bei denen der Rentenbezug im Alter von 55 bis 66 Jahren begonnen hat. Dann verbleiben noch 553 Personen. Box 12.2-4 zeigt die Verteilung auf die Rentenzugangsalter.

Wiederum wollen wir versuchen, Survivorfunktionen für die fernere Lebensdauer nach dem Beginn der Verrentung zu schätzen. Dabei muss jedoch berücksichtigt werden, dass bei allen Personen schon eine mehr oder weniger lange Zeit seit ihrem Rentenbeginn verstrichen ist, bevor sie in der zweiten Welle des SOEP befragt werden konnten. Box 12.2-5 zeigt, wie sich die Rentenzugänge auf Kalenderjahre verteilen. Zum Beispiel gibt es eine Person, die im Jahr 1954 verrentet wurde, also mindestens 31 Jahre überlebt hat, bevor sie im Jahr 1985 im SOEP befragt worden ist und bevor – in den nachfolgenden Jahren – ein Sterbefall beobachtet werden könnte. Umgekehrt: Über Personen, die nicht bis zum Jahr 1985 überlebt haben,

[17] Tatsächlich haben wesentlich mehr Personen einen Rentenbezug angegeben; aber in vielen Fällen gibt es keine gültige Angabe über den Beginn des Rentenbezugs.

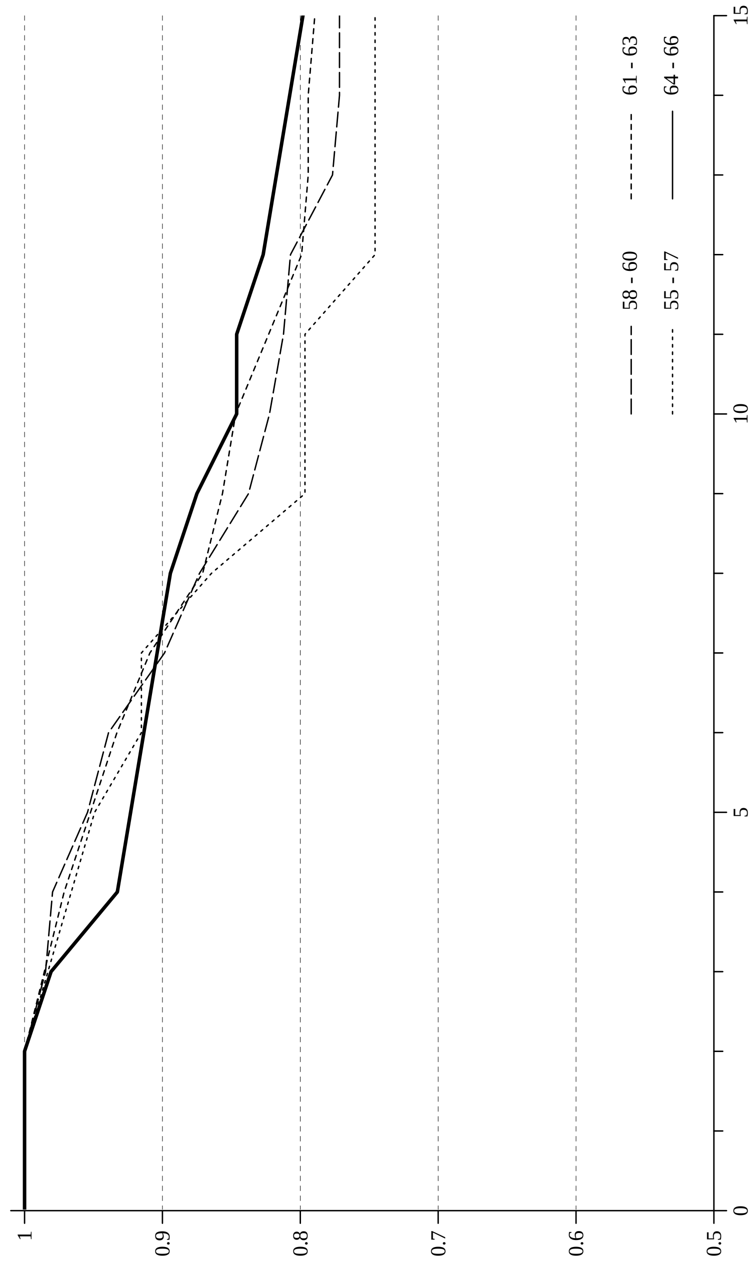


Abb. 12.2-1 Survivorfunktionen für einen Zeitraum von 15 Jahren nach der Verrentung.

Box 12.2-4 Verteilung der 553 männlichen Personen, die in der zweiten Welle des SOEP einen Rentenbeginn angegeben haben, auf die Rentenzugangsalter.

Rentenzugangsalter	Anzahl
55	19
56	16
57	27
58	18
59	30
60	71
61	54
62	39
63	92
64	44
65	117
66	26
Insgesamt	553

gibt es keine Informationen. Man hat es also, technisch gesprochen, nicht nur mit rechts zensierten, sondern auch mit „links abgeschnittenen" Daten zu tun. Um dies bei der Schätzung von Survivorfunktionen zu berücksichtigen, definieren wir drei Variablen:

a) S_i erfasst die Zeitdauer (in Jahren) zwischen dem Rentenbeginn der Person i und dem Jahr 1985.

b) D_i ist der Zensierungsindikator. Er bekommt den Wert 1, wenn die Person i bis zum Jahr 2001 gestorben ist, andernfalls den Wert 0.

c) T_i erfasst die beobachtete Lebensdauer der Person i seit ihrem Rentenbeginn. Wenn $D_i = 1$ ist, ist T_i die tatsächliche Lebensdauer seit dem Rentenbeginn bis zu dem Jahr, in dem die Person i gestorben ist. Wenn $D_i = 0$ ist, ist T_i die Zeitdauer vom Rentenbeginn bis zum Jahr 2001; man weiß dann nur, dass die Person i nach ihrem Rentenbeginn noch mindestens T_i Jahre gelebt hat.

Mit Hilfe dieser Variablen können nun wiederum Survivorfunktionen für die fernere Lebensdauer geschätzt werden.[18] Da es mit den verfügbaren Daten kaum möglich ist, zwischen nahe beieinander liegenden Verrentungsaltern – etwa den Verrentungsaltern 63 und 65 – auf statistisch aussagekräftige Weise zu unterscheiden, bilden wir in diesem Fall zwei Gruppen:

[18] Dabei wird ein modifiziertes Kaplan-Meier-Verfahren verwendet; man vgl. Rohwer, Pötter, Grundzüge der sozialwissenschaftlichen Statistik, Weinheim: Juventa 2001, S. 207ff.

Box 12.2-5 Verteilung der Rentenzugänge auf Kalenderjahre.

Jahr	Anzahl	Jahr	Anzahl	Jahr	Anzahl
1954	1	1966	2	1976	21
1956	1	1967	11	1977	28
1958	2	1968	14	1978	38
1959	1	1969	15	1979	30
1960	1	1970	19	1980	44
1961	2	1971	14	1981	37
1962	2	1972	23	1982	34
1963	4	1973	31	1983	48
1964	4	1974	16	1984	46
1965	10	1975	32	1985	22

a) Verrentungsalter 55 bis 61 Jahre, dies sind insgesamt 235 Personen, von denen im Beobachtungszeitraum 133 gestorben sind; und

b) Verrentungsalter 63 bis 66 Jahre, dies sind insgesamt 279 Personen, von denen im Beobachtungszeitraum 186 gestorben sind.

Abbildung 12.2-2 zeigt die Survivorfunktionen für diese beiden Gruppen, wobei es sich um konditionale Survivorfunktionen handelt, bei denen die fernere Lebensdauer ab dem Verrentungsalter betrachtet wird. Zur Darstellung wird ein Koordinatensystem verwendet, dessen Abszisse eine Lebenszeitachse darstellt. Die Survivorfunktionen in dieser Abbildung beginnen beim durchschnittlichen Verrentungsalter in den beiden Gruppen: 58.9 Jahre in der Gruppe (a) und 64.3 Jahre in der Gruppe (b). Man erkennt, dass die Personen in der Gruppe (a) eine deutlich kürzere Lebensdauer haben als die Personen in der Gruppe (b).

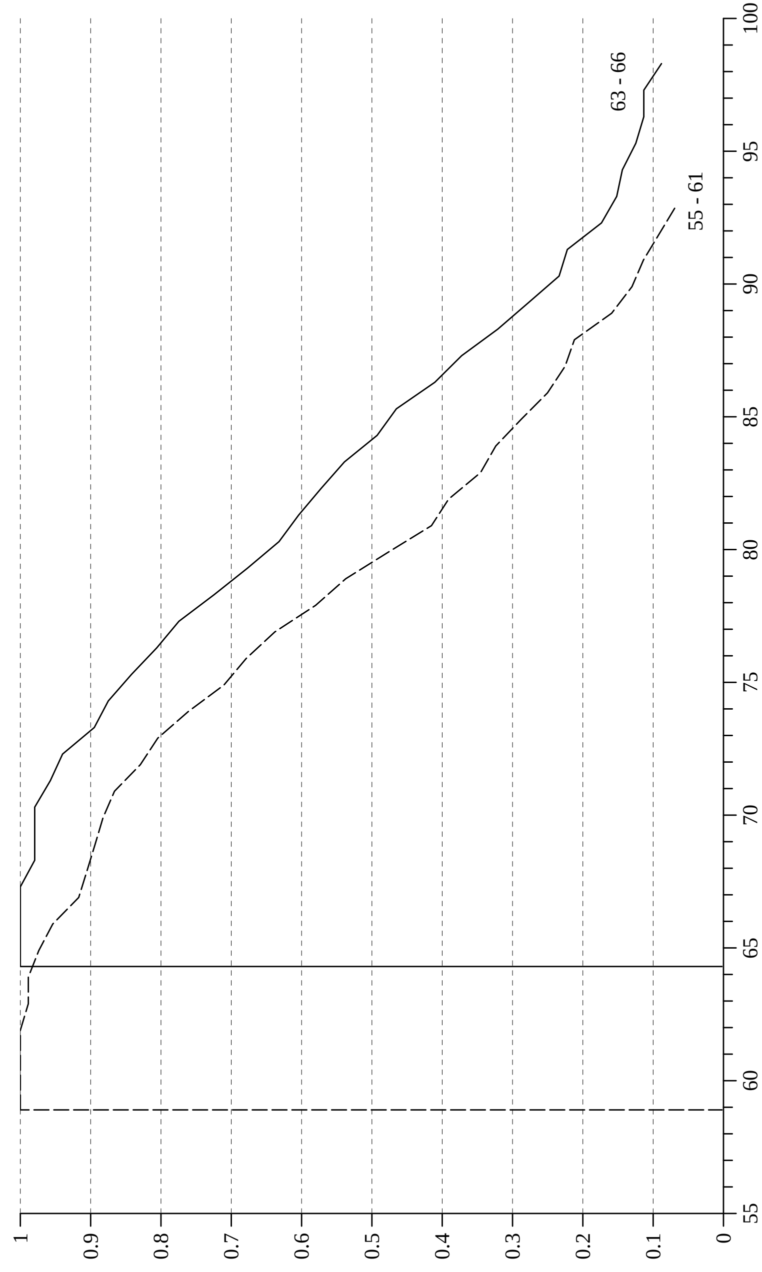


Abb. 12.2-2 Survivorfunktionen für die fernere Lebensdauer nach dem Beginn der Verrentung.

Kapitel 13

Daten der Sozialhilfestatistik

In diesem Kapitel beschäftigen wir uns mit Daten der amtlichen Sozialhilfestatistik. Bisher gibt es nur einen Abschnitt, in dem besprochen wird, wie mithilfe von Mikrodaten aus der Sozialhilfestatistik Bezugsdauern berechnet werden können.

13.1 Berechnung von Bezugsdauern

1. Mikrodaten der Sozialhilfestatistik. Seit einigen Jahren publiziert das Statistische Bundesamt Mikrodaten aus der amtlichen Sozialhilfestatistik. Jeweils zum Jahresende wird eine 25 %-Stichprobe aus der Gesamtheit der Haushalte bzw. Personen gebildet, die zu diesem Zeitpunkt Leistungen aus der Sozialhilfe beziehen. Der erste dieser Datensätze bezieht sich auf den 31. Dezember 1997. In diesem Abschnitt verwenden wir drei aufeinander folgende Datensätze, die sich auf das Ende der Jahre 1997, 1998 und 1999 beziehen. Die Datensätze haben den gleichen Aufbau: Einzelne Zeilen beziehen sich auf Personen, die am jeweiligen Stichtag Leistungen der Sozialhilfe beziehen. Entsprechend den Regeln der Sozialhilfegewährung sind diese Personen innerhalb der Datensätze als Mitglieder von Haushalten (Bedarfsgemeinschaften) angeordnet. Somit können sowohl Aussagen über einzelne Personen als auch über Haushalte gemacht werden. Die folgende Tabelle zeigt die Anzahl der Personen und Haushalte in den drei Datensätzen.

	1997	1998	1999
Personen	723177	719642	698037
Haushalte	372121	371910	363158

Für diese Personen bzw. Haushalte enthalten die Datensätze eine Vielzahl von Informationen. Insbesondere gibt es zwei Angaben über bisherige Bezugsdauern:

Box 13.1-1 Anzahl der Haushalte mit einer bisherigen Bezugsdauer von 0 bis 120 Monaten.

S97	N97	S98	N98	S99	N99	S97	N97	S98	N98	S99	N99
				0	14509	26	4480	38	3259	50	2397
				1	15515	27	4161	39	3067	51	2158
				2	18492	28	4075	40	2996	52	2163
				3	11412	29	3833	41	2853	53	1994
				4	9759	30	3345	42	2337	54	1665
				5	9187	31	3189	43	2314	55	1735
				6	7723	32	3296	44	2374	56	1771
				7	6872	33	3503	45	2523	57	1822
				8	7283	34	3536	46	2544	58	2027
				9	6805	35	4877	47	3815	59	2813
				10	6904	36	3550	48	2642	60	2066
				11	7691	37	2822	49	2092	61	1676
		0	15397	12	6724	38	2792	50	2176	62	1658
		1	15499	13	6184	39	2287	51	1677	63	1280
		2	14776	14	6448	40	2163	52	1566	64	1167
		3	11402	15	5712	41	2130	53	1585	65	1161
		4	10889	16	5279	42	2043	54	1563	66	1214
		5	9463	17	5037	43	2031	55	1457	67	1205
		6	8652	18	4783	44	2546	56	1927	68	1642
		7	7949	19	4474	45	2024	57	1580	69	1238
		8	8184	20	4749	46	1912	58	1455	70	1131
		9	7795	21	4353	47	3674	59	2587	71	2027
		10	7101	22	4048	48	2500	60	2826	72	1426
		11	8538	23	4941	49	1952	61	1531	73	1219
0	16654	12	7730	24	4924	50	1941	62	1570	74	1266
1	18199	13	7616	25	4688	51	1615	63	1358	75	1072
2	17621	14	7791	26	4889	52	1662	64	1376	76	1200
3	12901	15	6242	27	3984	53	1632	65	1296	77	1039
4	11062	16	5620	28	3513	54	1459	66	1247	78	1003
5	11853	17	6336	29	4163	55	1257	67	1061	79	806
6	9640	18	5440	30	3460	56	1764	68	1434	80	1234
7	8927	19	5038	31	3217	57	1405	69	1224	81	991
8	9103	20	5276	32	3486	58	1087	70	901	82	748
9	8120	21	4941	33	3215	59	1387	71	1166	83	976
10	7983	22	4965	34	3175	60	1177	72	939	84	849
11	10356	23	6711	35	4450	61	1287	73	1060	85	859
12	8199	24	5268	36	3335	62	1193	74	1034	86	787
13	7952	25	5184	37	3458	63	1037	75	875	87	720
14	7835	26	4932	38	3339	64	925	76	835	88	673
15	7119	27	4627	39	3308	65	967	77	843	89	721
16	6483	28	4230	40	2932	66	905	78	737	90	591
17	6115	29	4163	41	2817	67	890	79	723	91	661
18	5621	30	3757	42	2593	68	823	80	714	92	566
19	5297	31	3596	43	2449	69	902	81	637	93	546
20	5469	32	3888	44	2569	70	794	82	630	94	498
21	5445	33	3970	45	2610	71	927	83	758	95	626
22	4775	34	3261	46	2184	72	760	84	609	96	491
23	5874	35	4104	47	2928	73	748	85	593	97	525
24	5107	36	3596	48	2581	74	854	86	750	98	628
25	4789	37	3437	49	2553	75	600	87	522	99	421

Box 13.1-1 (Fortsetzung) Anzahl der Haushalte mit einer bisherigen Bezugsdauer von 0 bis 120 Monaten.

S97	N97	S98	N98	S99	N99	S97	N97	S98	N98	S99	N99
76	666	88	517	100	450	99	415	111	280		
77	770	89	638	101	542	100	305	112	279		
78	514	90	449	102	340	101	432	113	421		
79	621	91	515	103	403	102	304	114	236		
80	645	92	564	104	448	103	390	115	327		
81	635	93	493	105	404	104	373	116	329		
82	495	94	414	106	317	105	316	117	252		
83	763	95	628	107	512	106	293	118	254		
84	523	96	446	108	371	107	483	119	376		
85	596	97	470	109	410	108	300	120	19693		
86	637	98	522	110	412	109	272				
87	488	99	429	111	344	110	305				
88	481	100	391	112	324	111	287				
89	617	101	484	113	376	112	321				
90	817	102	674	114	507	113	336				
91	493	103	438	115	336	114	281				
92	502	104	400	116	296	115	306				
93	520	105	478	117	400	116	559				
94	2278	106	1914	118	1696	117	418				
95	587	107	528	119	378	118	258				
96	393	108	316	120	21036	119	338				
97	558	109	465			120	18663				
98	349	110	288								

a) Eine Information über die „längste bisherige Dauer der Hilfegewährung an die Bedarfsgemeinschaft in der aktuellen Zusammensetzung", und

b) eine Information über die „bisherige Dauer der ununterbrochenen Hilfegewährung für mindestens ein Mitglied der Bedarfsgemeinschaft".

In beiden Fällen beziehen sich die Angaben auf Bedarfsgemeinschaften, die entsprechenden Variablen haben also für alle Personen der jeweiligen Bedarfsgemeinschaft den gleichen Wert. Dementsprechend beziehen sich auch alle folgenden Untersuchungen auf Bedarfsgemeinschaften (wir sprechen in diesem Abschnitt gleichbedeutend von Haushalten); und zwar verwenden wir die Bezugsdauerinformation in der Variante (b), da sich Veränderungen in der Zusammensetzung der Bedarfsgemeinschaften mit den verfügbaren Daten nicht erfassen lassen.

Die Bezugsdauer ist in Monaten ausgewiesen, und zwar von 0 bis 120 Monaten. Dabei ist jedoch zu berücksichtigen, dass die Kategorie 120 alle Haushalte mit einer Bezugsdauer von 120 oder mehr Monaten umfasst. Box 13.1-1 zeigt die Daten für die drei Stichtage. Dabei geben S_{97}, S_{98} und S_{99} die Bezugsdauern an und N_{97}, N_{98} und N_{99} die jeweiligen Fallzahlen. Zum Beispiel gab es am 31.12.1997 5469 Haushalte mit einer bisherigen Bezugsdauer von 20 Monaten, und 16654 Haushalte hatten eine bisherige

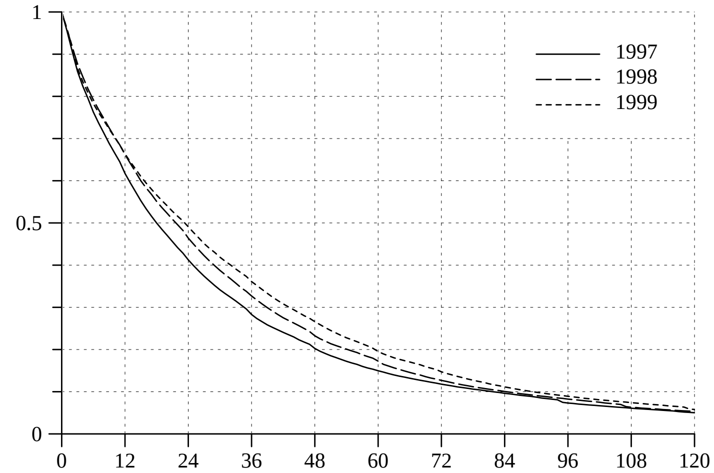


Abb. 13.1-1 Darstellung der Verteilungen der bisherigen Bezugsdauern durch Survivorfunktionen.

Bezugsdauer von 0 Monaten, d.h. bei ihnen begann der Sozialhilfebezug im Dezember 1997.

2. Überlegungen zum methodischen Vorgehen. Die Daten liefern unmittelbar nur Informationen über *bisherige* Bezugsdauern. Um deren Verteilung darzustellen, kann man z.B. Survivorfunktionen verwenden:

$$H_j(t) := \text{Anteil der Haushalte mit einer bisherigen Bezugsdauer von mindestens } t \text{ Monaten am Ende des Jahres } j.$$

Abb. 13.1-1 zeigt diese Survivorfunktionen am Ende der Jahre 1997, 1998 und 1999. Man erkennt, wie sich der Anteil der Haushalte mit längeren Bezugsdauern in diesem Zeitraum etwas vergrößert hat. Dabei kann man sich z.B. an den Medianwerten orientieren: 18 Monate Ende 1997, 21 – 22 Monate Ende 1998 und 23 – 24 Monate Ende 1999. Daraus folgt aber nicht ohne weiteres, dass sich die durchschnittlichen Bezugsdauern verlängert haben. Der Anteil von Haushalten mit längeren Bezugsdauern könnte auch deshalb zugenommen haben, weil weniger neue Haushalte mit einem Sozialhilfebezug begonnen haben.

Die Orientierung an bisherigen Bezugsdauern, die zu einem Stichtag retrospektiv ermittelt werden, hat einen weiteren Mangel: sie überschätzt die *tatsächlichen* Bezugsdauern der Sozialhilfe. Das wird deutlich, wenn man von Haushalten ausgeht, bei denen zu irgendeinem Zeitpunkt der Sozialhilfebezug beginnt, und dann fragt, wie lange sie im Sozialhilfebezug verbleiben. In diesem Abschnitt verfolgen wir eine solche prospektive Betrachtungsweise und kontrastieren sie mit den Angaben über bisherige

Bezugsdauern. Insbesondere wird besprochen, wie sich durch eine Verknüpfung mehrerer Datensätze aus Angaben über bisherige Bezugsdauern Informationen für eine prospektive Betrachtungsweise gewinnen lassen.

Zunächst präzisieren wir den gedanklichen Ansatz. Die prospektive Betrachtungsweise bezieht sich auf Haushalte, bei denen während eines gewissen Zeitraums, z.B. während eines Jahres, der Sozialhilfebezug begonnen hat. Zur expliziten Formulierung ist es sinnvoll, Bezugsdauervariablen

$$T_j : \Omega_j \longrightarrow \{0, 1, 2, 3, \ldots\}$$

zu definieren. Hierbei ist Ω_j die Gesamtheit derjenigen Haushalte, bei denen während des Jahres j der Sozialhilfebezug begonnen hat, und T_j ist die Bezugsdauervariable, d.h. für jeden Haushalt ω aus der Gesamtheit Ω_j gibt $T_j(\omega)$ die Zeitdauer bis zum Ausscheiden aus dem Sozialhilfebezug an. Dann wäre der Mittelwert von T_j die durchschnittliche Bezugsdauer derjenigen Haushalte, bei denen der Sozialhilfebezug im Jahr j begonnen hat. Zwar erhält man aus unseren Daten weder Informationen über die Gesamtheiten Ω_j noch über die Variablen T_j, so dass eine direkte Berechnung nicht möglich ist; man kann aber versuchen, auf einem Umweg dennoch zu einer Schätzung zu gelangen. Der Umweg besteht darin, zunächst Raten zu berechnen und daraus die Verteilung der Bezugsdauervariablen zu rekonstruieren. Raten können folgendermaßen definiert werden:

$$r_j(t) := \frac{\text{Anzahl Haushalte in } \Omega_j \text{ mit Bezugsdauer} = t}{\text{Anzahl Haushalte in } \Omega_j \text{ mit Bezugsdauer} \geq t}$$

Aus diesen Raten lässt sich die Survivorfunktion der Bezugsdauervariablen T_j, also

$$G_j(t) := \text{Anteil der Haushalte in } \Omega_j \text{ mit Bezugsdauer} \geq t$$

auf folgende Weise berechnen:[1]

$$G_j(t) = \prod_{k=0}^{t-1} (1 - r_j(k))$$

Allerdings treten bei den hier verfügbaren Daten noch zwei weitere Komplikationen auf. Die erste Komplikation besteht darin, dass sich Informationen über Raten für die Beendigung des Sozialhilfebezugs bestenfalls während des Zeitraums von Ende 1997 bis Ende 1999 ermitteln lassen. Es ist also nicht möglich, nach dem Jahr j des Sozialhilfebeginns differenzierte Bezugsdauerverteilungen zu schätzen. Stattdessen muss man sich darauf beschränken, aus den während eines beschränkten Zeitraums ermittelbaren Raten Verteilungen für fiktive Bezugsdauervariablen zu schätzen. Im weiteren beziehen wir uns auf zwei solche Variablen:

[1] Nähere Ausführungen zu den formalen Zusammenhängen findet man bei Rohwer und Pötter (2001, Kap. 12).

a) Eine Bezugsdauervariable T^*_{97}, deren Verteilung aus den Veränderungen zwischen Ende 1997 und Ende 1998 geschätzt wird, und

b) eine Bezugsdauervariable T^*_{98}, deren Verteilung aus den Veränderungen zwischen Ende 1998 und Ende 1999 geschätzt wird.

Die Kennzeichnung durch Sternchen soll darauf hinweisen, dass es sich um konstruierte Bezugsdauervariablen handelt. Ihre Survivorfunktionen werden im weiteren durch G^*_j, die korrespondierenden Raten durch r^*_j bezeichnet.

Die zweite Komplikation besteht darin, dass sich aus den gegebenen Daten auch die monatlichen Raten r^*_j nicht ohne weiteres berechnen lassen, sondern dass man stattdessen nur Jahresraten ermitteln kann. Um unsere Vorgehensweise zu erläutern, nehmen wir zunächst an, dass es sich bei den Daten nicht um Stichproben, sondern um Vollerhebungen handelt. Dann kann man sich auf folgende Größen beziehen:

$$n_j(t) := \begin{array}{l}\text{Anzahl der Haushalte, die Ende des Jahres } j \text{ eine} \\ \text{Bezugsdauer von } t \text{ Monaten hatten}\end{array}$$

und aus diesen Größen gewinnt man dann die Jahresraten

$$u_j(t) := \frac{n_j(t) - n_{j+1}(t+12)}{n_j(t)}$$

Im Nenner steht die Anzahl der Haushalte, die Ende des Jahres j eine Bezugsdauer von t Monaten hatten, und im Zähler steht die Anzahl derjenigen dieser Haushalte, die während des Jahres $j+1$ aus dem Sozialhilfebezug ausgeschieden sind.

Diese Überlegung gilt strenggenommen natürlich nur für die Grundgesamtheit aller Sozialhilfebezieher. Unsere Daten stammen dagegen aus Stichproben, die außerdem unabhängig voneinander in den einzelnen Jahren gezogen worden sind. Dennoch erscheint der Auswahlsatz hinreichend groß, um die Annahme zu rechtfertigen, dass man aus diesen Daten nicht nur Schätzungen der Größen $n_j(t)$ sondern auch der Jahresraten $u_j(t)$ gewinnen kann. Diese Annahme wird im weiteren vorausgesetzt.

Es ist auch klar, dass zur Schätzung der Jahresraten gar keine explizite „Hochrechnung“ erforderlich ist, da der Auswahlsatz in allen drei Jahren gleichermaßen 25 % beträgt. Man kann also unmittelbar mit den aus den Stichprobendaten ermittelbaren Werten für $n_j(t)$ rechnen. Zum Beispiel ist

$$u_{97}(0) \approx \frac{16654 - 7730}{16654} = 0.536$$

so dass man sagen kann, dass etwa 53.6 % derjenigen Haushalte, die im Dezember 1997 mit dem Sozialhilfebezug begonnen haben, bis Ende 1998 bereits wieder ausgeschieden waren.

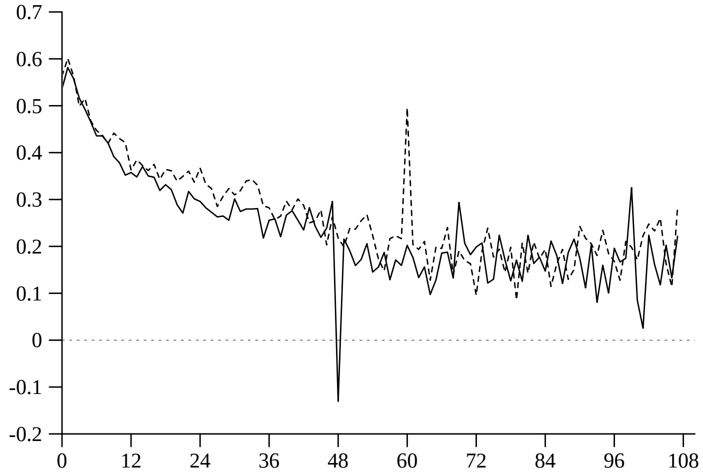


Abb. 13.1-2 Darstellung der Jahresraten $u_{97}(t)$ (durchgezoge Linie) und $u_{98}(t)$ (gestrichelte Linie).

3. Betrachtung der Jahresraten. Bevor überlegt wird, wie man mit den Jahresraten Bezugsdauerverteilungen schätzen kann, sollten sie selbst etwas genauer betrachtet werden. Abb. 13.1-2 zeigt $u_{97}(t)$ und $u_{98}(t)$. Man erkennt, dass es bei $t = 48$ eine negative Jahresrate gibt, was aufgrund der Begriffsbildungen eigentlich nicht vorkommen sollte. Anhand der Daten in Box 13.1-1 erkennt man den Grund. Ende 1997 gab es 2500 Haushalte mit einer Bezugsdauer von 48 Monaten, und Ende 1998 gab es *mehr* Haushalte, nämlich 2826, mit einer Bezugsdauer von 60 Monaten. Dies impliziert eine negative Jahresrate. Eine weitere unmittelbare Folge ist, dass die Jahresrate für $t = 60$ deutlich über den Werten in ihrer näheren Umgebung liegt.

Es ist zwar nicht ausgeschlossen, aber eher unwahrscheinlich, dass die negative Jahresrate bei $t = 48$ eine Folge der Stichprobenziehungen ist. In jedem Fall ist zu vermuten, dass es sich um einen Artefakt handelt, der aus der Datenaufbereitung und/oder Datenkonstruktion resultiert. Wir haben uns deshalb entschlossen, den Wert $n_{98}(60)$ so zu verändern, dass $u_{97}(48)$ in etwa dem Durchschnitt von $u_{97}(47) = 0.296$ und $u_{97}(49) = 0.216$, also dem Wert 0.256 entspricht. In allen weiteren Verwendungen der Daten aus Box 13.1-1 ändern wir also den Wert für $n_{98}(48)$ von bisher 2826 in den neuen Wert 1860. Dies führt auch zu einer plausiblen Jahresrate $u_{98}(60)$.

Eine weitere bemerkenswerte Tatsache betrifft das Niveau der Jahresraten: Im Bereich von etwa 1 bis 5 Jahren liegen die Jahresraten für das Jahr 1998 tendenziell über denen des Jahres 1997, d.h. dass in diesem Bereich von Bezugsdauern während des Jahres 1999 vergleichsweise mehr Haushalte aus dem Sozialhilfebezug ausgeschieden sind als während des

Tabelle 13.1-1 Berechnung von Stützstellen für die Survivorfunktionen G^*_{97} und G^*_{98} aus den Jahresraten u_{97} bzw. u_{98}.

t	$u_{97}(t)$	$G^*_{97}(t)$	$u_{98}(t)$	$G^*_{98}(t)$
0	0.5358	1.0000	0.5633	1.0000
12	0.3575	0.4642	0.3630	0.4367
24	0.2959	0.2982	0.3669	0.2782
36	0.2558	0.2100	0.2823	0.1761
48	0.2560	0.1563	0.2180	0.1264
60	0.2022	0.1163	0.2333	0.0988
72	0.1987	0.0928	0.0958	0.0758
84	0.1472	0.0743	0.1938	0.0685
96	0.1959	0.0634	0.1682	0.0552
108		0.0510		0.0460

Jahres 1998. Somit zeigt sich bereits an dieser Stelle, dass eine Längsschnittbetrachtung mit Hilfe von Raten zu anderen Ergebnissen führt als eine Betrachtung von bisherigen Bezugsdauern (man vgl. die Medianberechnungen in § 2).

4. Verkettung der Jahresraten. Es bleibt zu überlegen, wie man aus den Jahresraten die Verteilungen der Bezugsdauervariablen rekonstruieren kann. Dafür gehen wir zunächst separat von den Jahresraten $u_{97}(t)$ und $u_{98}(t)$ aus. Die aus diesen Raten berechenbaren Survivorfunktionen werden mit G^*_{97} bzw. G^*_{98} bezeichnet. Dann gilt folgender Zusammenhang:

$$G^*_j(t+12) \;=\; G^*_j(t)(1-u_j(t))$$

wobei $j = 97$ oder $j = 98$ ist. Diese Beziehung kann man verwenden, um zunächst auf einfache Weise Stützpunkte für die Survivorfunktionen zu berechnen. Ausgehend von $G^*_j(0) = 1$ findet man $G^*_j(12)$, ausgehend von $G^*_j(12)$ findet man $G^*_j(24)$ usw. Tabelle 13.1-1 zeigt die Rechenergebnisse.

Abb. 13.1-3 zeigt eine graphische Darstellung der Survivorfunktionen. Man erkennt, dass die tendenziell höheren Jahresraten für 1998 zu einer vergleichsweise kürzeren Bezugsdauer führen. Bereits ein oberflächlicher Vergleich der Abbildungen 13.1-1 und 13.1-3 zeigt auch, dass die aus Jahresraten berechneten Bezugsdauern deutlich kürzer sind als die zu einem Stichtag ermittelten bisherigen Bezugsdauern.

Bei dieser Vorgehensweise werden nur Jahresraten $u_j(t)$ verwendet, bei denen t ein Vielfaches von 12 ist. Aus zwei Gründen sollte versucht werden, auch die Information aus den anderen Jahresraten zu nutzen. Erstens kann man dadurch versuchen, auch Werte der Survivorfunktionen zwischen den Stützstellen zu ermitteln. Zweitens kann man vielleicht die bei der bisherigen Methode unvermeidliche Akkumulation von Stichprobenfehlern, die aus der Verkettung ungenauer Jahresraten entsteht, verringern.

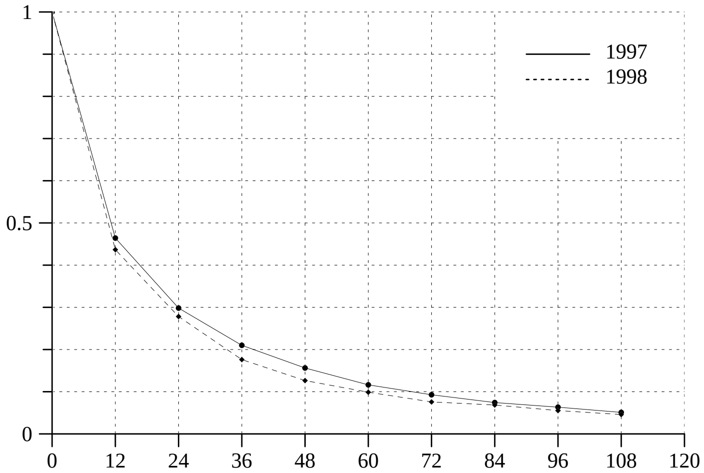


Abb. 13.1-3 Aus Jahresraten berechnete Stützpunkte der Survivorfunktionen G_{97} und G_{98}.

Gesucht ist also eine Survivorfunktion G_j^*, die möglichst gut zu den gegebenen Jahresraten $u_j(t)$ passt. Angenommen, man hat eine solche Funktion gefunden, dann können monatliche Raten

$$r_j^*(t) \;=\; \frac{G_j^*(t) - G_j^*(t+1)}{G_j^*(t)} \;=\; 1 - \frac{G_j^*(t+1)}{G_j^*(t)}$$

berechnet und zu Jahresraten

$$u_j^*(t) \;:=\; \prod_{k=t}^{t+11} (1 - r_j^*(k))$$

akkumuliert werden. Somit kann präzisiert werden: Es sollte eine Survivorfunktion G_j^* gefunden werden, so dass sich die durch sie implizierten Jahresraten $u_j^*(t)$ und die durch die Daten gegebenen Jahresraten $u_j(t)$ möglichst wenig unterscheiden. Eine mögliche Zielfunktion, aus deren Minimierung die gesuchten Survivorfunktionen berechenbar sind, sieht also folgendermaßen aus:

$$f_j(\alpha_{j,1}, \ldots, \alpha_{j,119}) \;:=\; \textstyle\sum_{t=0}^{107} (u_j^*(t) - u_j(t))^2$$

wobei die Parameter $\alpha_{j,t}$ den Werten der gesuchten Survivorfunktion entsprechen,[2] so dass die Größen $u_j^*(t)$ auf folgende Weise von den Parameter abhängen:

[2] Also $\alpha_{j,t} \equiv G_j(t)$. Zusätzlich wird $\alpha_{j,0} = 1$ fest vorgegeben.

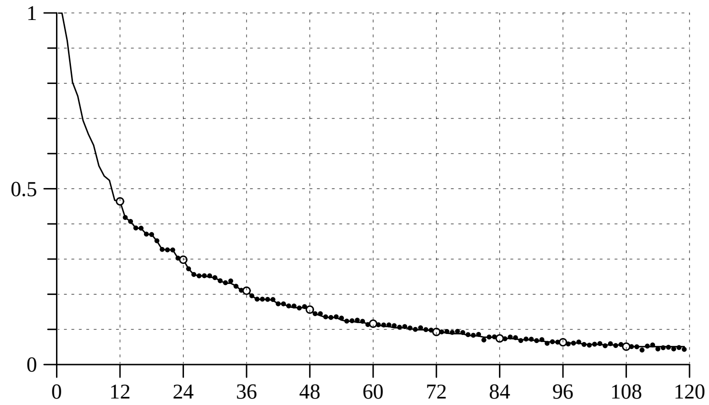


Abb. 13.1-4 Aus der Minimierung der im Text angegebenen Zielfunktion berechnete Survivorfunktion für die Jahresraten $u_{97}(t)$.

Abb. 13.1-5 Aus der Minimierung der im Text angegebenen Zielfunktion berechnete Survivorfunktion für die Jahresraten $u_{98}(t)$.

$$u_j^*(t) \;=\; \prod_{k=t}^{t+11} \big(1 - \frac{\alpha_{j,k+1}}{\alpha_{j,k}}\big)$$

Die Zielfunktion ist natürlich unter der Nebenbedingung zu minimieren, dass die Survivorfunktion monoton fällt: $1 \geq \alpha_{j,1} \geq \alpha_{j,2} \geq \cdots \geq \alpha_{j,119}$. Infolgedessen ist die Zielfunktion zwar nicht stetig differenzierbar (nicht einmal in der Nähe ihre Minimums, denn es gibt keine Survivorfunktion, die mit den gegebenen Jahresraten perfekt vereinbar ist); man kann aber mit einem einfachen Verfahren der direkten Suche zumindest ein lokales Minimum der Zielfunktion finden.

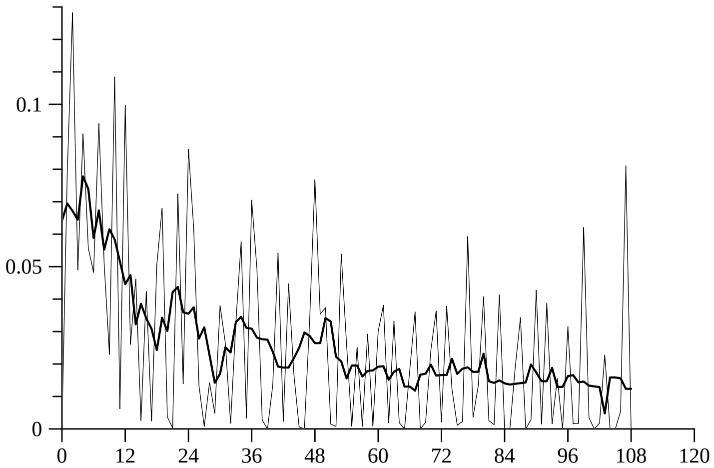


Abb. 13.1-6 Ungeglättete und durch gleitende Durchschnitte geglättete Raten r^*_{97}, korrespondierend zur Survivorfunktion G^*_{97} in Abbildung 13.1-4.

Die Abbildungen 13.1-4 und 13.1-5 zeigen die durch die Minimierung der Zielfunktion für die Jahre $j = 1997$ bzw. $j = 1998$ gewonnenen Survivorfunktionen. Eingezeichnet sind außerdem die Stützpunkte aus der Abbildung 13.1-3 in Form von Kreisen und die den Jahresraten korrespondierenden Positionen. Für jede Jahresrate $u_j(t)$ wurde ein Punkt an der Stelle $t+12$ (X-Achse) und $\hat{\alpha}_{j,t} u_j(t)$ (Y-Achse) eingetragen, wobei $\hat{\alpha}_{j,t}$ der Wert der geschätzten Survivorfunktion im Monat t ist. Man erkennt, dass das Verfahren eine Schätzung der Survivorfunktionen liefert, die mit den vorgegebenen Jahresraten weitgehend konsistent ist.

5. Die Gestalt der Ratenfunktion. Aus den Survivorfunktionen G^*_j lassen sich die zugehörigen monatlichen Raten r^*_j berechnen. Abb. 13.1-6 zeigt als Beispiel die Raten $r^*_{97}(t)$. Es gibt ersichtlich große Schwankungen, die – zwar nicht nur, aber auch – sowohl aus dem Konstruktionsprozess der Survivorfunktion als auch aus der Verwendung von Stichprobendaten resultieren. Um die zeitliche Entwicklung der Raten sichtbar zu machen, ist deshalb eine Glättung sinnvoll. Das kann zum Beispiel durch gleitende Durchschnitte erreicht werden. Verwendet man zur Durchschnittsbildung an jeder Stelle jeweils drei benachbarte linke und rechte Werte, erhält man den in Abb. 13.1-6 eingezeichneten geglätteten Ratenverlauf. Man erkennt, dass die Raten für die Beendigung des Sozialhilfebezugs anfangs vergleichsweise groß sind (etwa 6 % pro Monat), dann jedoch immer kleiner werden.

Es ist evident, dass man aus der geglätteten Ratenfunktion wiederum eine Survivorfunktion berechnen könnte. Sie wäre dann ebenfalls glatter als die ursprüngliche Survivorfunktion. Anders als bei Ratenfunktionen würde

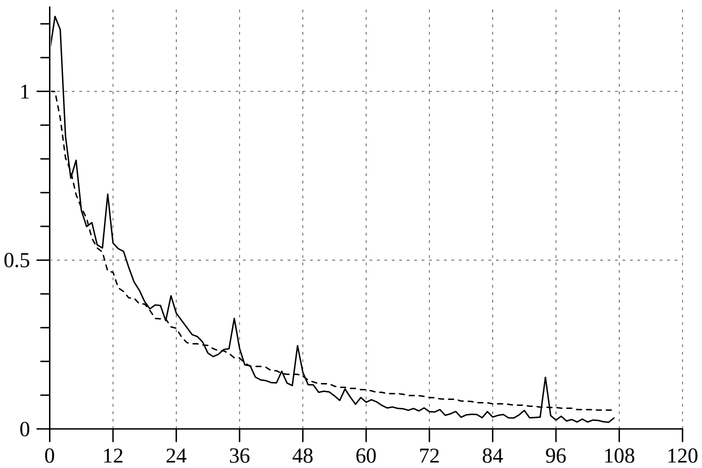


Abb. 13.1-7 Verteilung der bisherigen Bezugsdauern $\alpha\, n_{97}(t)$ (durchgezogene Linie) und der Survivorfunktion G^*_{97} (gestrichelte Linie).

man jedoch bei Survivorfunktionen das Resultat der Glättung visuell kaum erkennen, so dass hier auf eine Illustration verzichtet werden soll.

6. Vergleich mit bisherigen Bezugsdauern. Ausgehend von einer Survivorfunktion G^*_j kann man eine hypothetische Überlegung anstellen. Man kann sich einen stationären Prozess vorstellen, bei dem in jedem Monat eine gleichbleibende Anzahl von Haushalten mit dem Sozialhilfebezug beginnt und entsprechend der Survivorfunktion G^*_j im Bezug bleibt bzw. ausscheidet. Dann kann man für einen solchen Prozess zu irgendeinem Stichtag die Verteilung der bisherigen Bezugsdauern ermitteln. Sie ist jedoch zur Survivorfunktion G^*_j proportional. Denn angenommen, in jedem Monat beginnen n Haushalte mit dem Sozialhilfebezug. Dann ist an jeder beliebigen Stelle des Prozesses die Anzahl der Haushalte mit einer bisherigen Bezugsdauer von t Monaten gleich $n\, G^*_j(t)$.

Diese Überlegung kann man verwenden, um die durch G^*_{97} implizierte stationäre Verteilung bisheriger Bezugsdauern mit der tatsächlich am Ende des Jahres 1997 durch die Daten gegebenen Verteilung bisheriger Bezugsdauern zu vergleichen. Es sei also $n_{97}(t)$ die Anzahl der Haushalte, die Ende 1997 eine bisherige Bezugsdauer von t Monaten hatten (vgl. Box 13.1-1). Für den Vergleich beschränken wir uns auf die Bezugsdauern $t = 0, \ldots, 107$ und berechnen Anteilswerte $\alpha\, n_{97}(t)$ durch die Normierung

$$\textstyle\sum_{t=0}^{107} \alpha\, n_{97}(t) \;=\; \sum_{t=0}^{107} G^*_{97}(t)$$

Abb. 13.1-7 vergleicht die beiden Verteilungen bisheriger Bezugsdauern. Es ist bemerkenswert, dass die Verteilungen sehr ähnlich sind. Natürlich

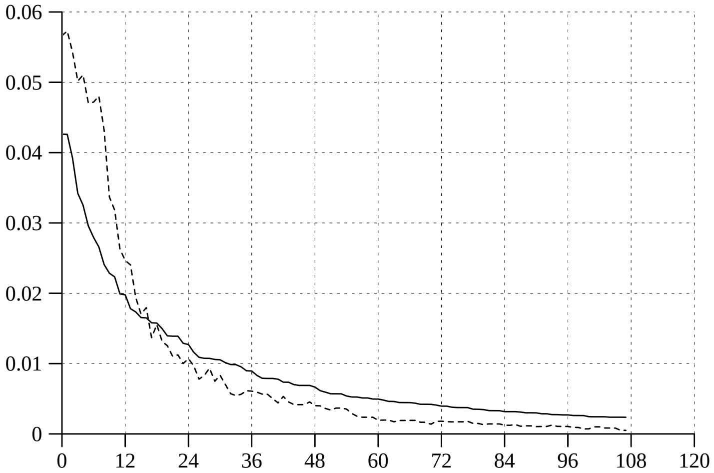


Abb. 13.1-8 Vergleich von G^*_{97} (durchgezogene Linie) mit der geglätteten Verteilung der Häufigkeiten $p^*_{97}(t)$ (gestrichelte Linie).

kann man nicht erwarten, dass sie sich vollständig entsprechen, da die Daten nicht aus einem stationären Prozess resultieren, sondern aus einem historischen Prozess, in dem sich sowohl die Anzahlen der neu hinzukommenden Haushalte als auch die Bezugsdauerverteilungen fortwährend verändern. Mit den uns verfügbaren Daten ist jedoch eine Interpretation der Unterschiede in den Verteilungen nicht möglich.

Die gedankliche Bezugnahme auf einen stationären Prozess erlaubt es auch, den Unterschied zwischen den Verteilungen für bisherige und tatsächliche Bezugsdauern genauer kenntlich zu machen. Als Beispiel gehen wir wieder von der Survivorfunktion G^*_{97} aus. Bei einem stationären Prozess ist dann $G^*_{97}(t)$ der Anteil der Haushalte mit einer bisherigen Bezugsdauer von t Monaten. Andererseits ist

$$p^*_{97}(t) := G^*_{97}(t) - G^*_{97}(t+1)$$

der Anteil der Haushalte mit einer tatsächlichen Bezugsdauer von t Monaten. Zum Vergleich beschränken wir uns wieder auf die Bezugsdauern $t = 0, \ldots, 107$, so dass $p^*_{97}(0) + \cdots + p^*_{97}(107) = 1$ ist. Dementsprechend werden auch die Anteile der bisherigen Bezugsdauern so normiert, dass ihre Summe für die Monate 0 bis 107 gleich 1 ist. Abb. 13.1-8 vergleicht die beiden Verteilungen, wobei die Verteilung der tatsächlichen Bezugsdauern geglättet worden ist.[3] Man erkennt deutlich, dass anfangs die tatsächlichen Bezugsdauern viel häufiger auftreten als die bisherigen Bezugsdauern,

[3]Wieder mit gleitenden Durchschnitten, wobei diesmal zu beiden Seiten jeweils 6 Nachbarwerte verwendet worden sind.

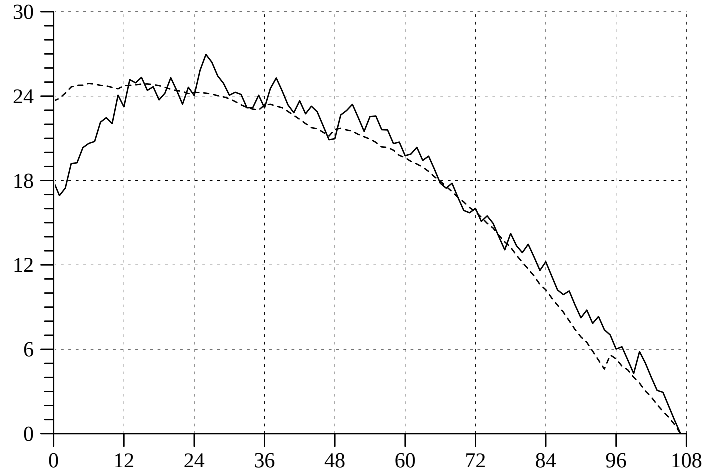


Abb. 13.1-9 Fernere durchschnittliche Bezugsdauern bei tatsächlichen (durchgezogene Linie) und bisherigen Bezugsdauern (gestrichelte Linie).

während es sich bei den längeren Bezugsdauern umgekehrt verhält.

Im nächsten Abschnitt wird besprochen, welche Folgen dies für die Berechnung durchschnittlicher Bezugsdauern hat. An dieser Stelle soll nur noch erwähnt werden, dass der beschriebene Sachverhalt aus der in § 5 dargestellten Form der Ratenfunktion resultiert. Nur bei einer unabhängig von der Bezugsdauer konstanten Rate wären die Verteilungen der tatsächlichen und der bisherigen Bezugsdauern identisch, wie man unmittelbar aus der Beziehung $p^*_j(t) = G^*_j(t) - G^*_j(t+1) = G^*_j(t)\,(1 - r^*_j(t))$ erkennt.

7. Durchschnittliche Bezugsdauern. Je nachdem ob Häufigkeiten bisheriger oder tatsächlicher Bezugsdauern verwendet werden, erhält man auch sehr unterschiedliche Werte für durchschnittliche Bezugsdauern. Einen ersten Hinweis liefern bereits die Medianwerte. Orientiert man sich an der Verteilung bisheriger Bezugsdauern in Abb. 13.1-1, erhält man Medianwerte, die von etwa 17 – 18 Monaten Ende 1997 auf 23 – 24 Monate Ende 1999 ansteigen. Dagegen liegen die Medianwerte der Verteilungen G^*_{97} und G^*_{98} bei etwa 10 – 11 Monaten und weisen keine Zunahme auf.

Stattdessen kann man auch Mittelwerte berechnen. Dabei beschränken wir uns auf Bezugsdauern von maximal 107 Monaten (9 Jahre), verwenden also die Definition

$$M^*_j := \frac{\sum_{t=0}^{107} t\, p^*_j(t)}{\sum_{t=0}^{107} p^*_j(t)}$$

für tatsächliche Bezugsdauern und eine analoge Definition, bei der anstelle der Häufigkeiten $p^*_j(t)$ die Häufigkeiten $n_j(t)$ aus Box 13.1-1 verwendet werden. Dann findet man folgende Mittelwerte:

a) Die Mittelwerte der tatsächlichen Bezugsdauern sind von 17.9 Monaten für 1997 auf 17.0 Monaten für 1998 etwas gesunken; dagegen sind

b) die Mittelwerte der bisherigen Bezugsdauern von 23.7 Monaten für 1997 auf 26.2 Monate für 1998 angestiegen und wiederum deutlich größer.

Der Grund für die Unterschiede liegt in den unterschiedlichen Verteilungen der tatsächlichen und bisherigen Bezugsdauern, die im vorangegangenen Abschnitt besprochen wurden. Das wird auch deutlich, wenn man durchschnittliche fernere Bezugsdauern berechnet:

$$M_j^*(t) := \frac{\sum_{k=t}^{107} k\, p_j^*(k)}{\sum_{k=t}^{107} p_j^*(k)} - t$$

Analog zum Begriff einer ferneren Lebenserwartung geben diese Größen an, wie lange Haushalte noch im Durchschnitt im Sozialhilfebezug bleiben, wenn sie bereits t Monate Sozialhilfe bezogen haben. Die durchgezogene Linie in Abb. 13.1-9 zeigt die Werte von $M_{97}^*(t)$ in Abhängigkeit von t. Man erkennt, dass diese Kurve – im Unterschied zu der gestrichelten Kurve, die sich auf bisherige Bezugsdauern bezieht – zunächst ansteigt. D.h., während die durchschnittliche fernere Bezugsdauer anfangs, bei $t = 0$, der durchschnittlichen Bezugsdauer von 17.9 Monaten entspricht, haben Haushalte, die bereits 12 Monate Sozialhilfe beziehen, eine durchschnittliche fernere Bezugsdauer von etwa 24 Monaten, bleiben also im Durchschnitt drei Jahre im Sozialhilfebezug.

Grundsätzlich sind natürlich Berechnungen von Verteilungen und Mittelwerten für tatsächliche und bisherige Bezugsdauern gleichermaßen berechtigt. Eine prospektive Betrachtungsweise verlangt jedoch eine Orientierung an tatsächlichen Bezugsdauern. Dies ist auch dann erforderlich, wenn man Aussagen über die Verteilung der Kosten der Sozialhilfe auf Haushalte mit unterschiedlichen Bezugsdauern machen möchte. Dann wird eine Orientierung an bisherigen Bezugsdauern irreführend, weil sie suggeriert, dass der größte Teil der Kosten für Haushalte mit besonders langen Bezugsdauern aufgewendet werden muss.

Box 13.1-2 Anzahl von Haushalten in den neuen Bundesländern, gegliedert nach der bisherigen Bezugsdauer von Sozialhilfe.

S97	N97	S98	N98	S99	N99	S97	N97	S98	N98	S99	N99
0	0	0	0	0	2684	31	245	43	134	55	109
0	0	0	0	1	2854	32	209	44	113	56	66
0	0	0	0	2	2735	33	239	45	136	57	82
0	0	0	0	3	2124	34	207	46	108	58	79
0	0	0	0	4	1892	35	249	47	155	59	103
0	0	0	0	5	1809	36	186	48	110	60	74
0	0	0	0	6	1398	37	160	49	109	61	81
0	0	0	0	7	1227	38	176	50	131	62	76
0	0	0	0	8	1356	39	134	51	66	63	54
0	0	0	0	9	1132	40	126	52	81	64	61
0	0	0	0	10	1025	41	121	53	101	65	41
0	0	0	0	11	1281	42	93	54	74	66	48
0	0	0	3000	12	992	43	111	55	65	67	56
0	0	1	2738	13	937	44	114	56	63	68	57
0	0	2	2810	14	1126	45	110	57	82	69	69
0	0	3	2035	15	902	46	86	58	60	70	52
0	0	4	1800	16	761	47	135	59	113	71	94
0	0	5	1698	17	831	48	82	60	44	72	45
0	0	6	1322	18	669	49	91	61	57	73	44
0	0	7	1241	19	637	50	79	62	52	74	42
0	0	8	1270	20	637	51	62	63	54	75	32
0	0	9	1301	21	652	52	62	64	40	76	27
0	0	10	1151	22	534	53	68	65	58	77	31
0	0	11	1437	23	692	54	58	66	45	78	35
0	2899	12	1072	24	573	55	46	67	38	79	32
1	2622	13	995	25	531	56	60	68	42	80	27
2	2712	14	1099	26	546	57	81	69	65	81	47
3	2205	15	1010	27	502	58	41	70	30	82	18
4	1876	16	823	28	437	59	69	71	56	83	37
5	2043	17	1012	29	533	60	41	72	25	84	25
6	1549	18	752	30	407	61	36	73	33	85	21
7	1315	19	652	31	346	62	41	74	35	86	20
8	1556	20	777	32	450	63	35	75	16	87	26
9	1260	21	641	33	361	64	22	76	20	88	17
10	1209	22	642	34	339	65	45	77	48	89	32
11	2295	23	1463	35	931	66	32	78	28	90	15
12	969	24	483	36	279	67	17	79	17	91	17
13	979	25	505	37	287	68	26	80	18	92	15
14	987	26	551	38	312	69	43	81	34	93	29
15	774	27	411	39	241	70	32	82	33	94	19
16	690	28	380	40	218	71	62	83	51	95	41
17	801	29	445	41	262	72	16	84	12	96	9
18	584	30	340	42	202	73	26	85	22	97	18
19	510	31	300	43	173	74	37	86	30	98	25
20	611	32	315	44	210	75	27	87	22	99	21
21	419	33	270	45	170	76	31	88	19	100	17
22	423	34	214	46	134	77	42	89	29	101	25
23	556	35	296	47	242	78	20	90	24	102	16
24	336	36	207	48	114	79	28	91	11	103	12
25	411	37	221	49	144	80	20	92	23	104	11
26	376	38	209	50	135	81	24	93	20	105	14
27	330	39	192	51	134	82	27	94	27	106	25
28	315	40	154	52	118	83	97	95	68	107	60
29	280	41	175	53	111	84	20	96	19	108	10
30	235	42	131	54	67						

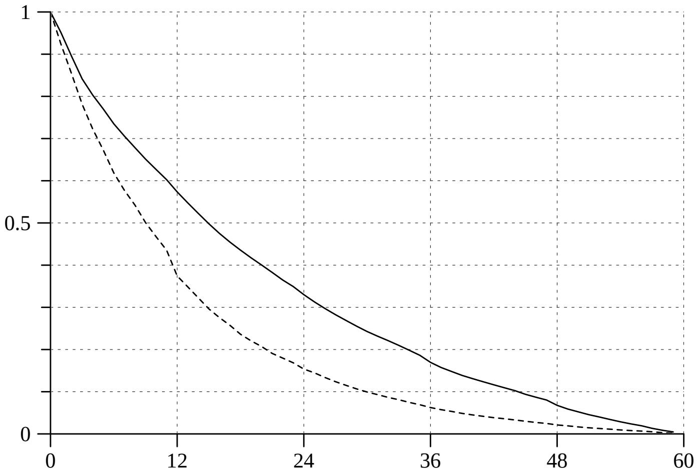


Abb. 13.1-10 Verteilungen der bisherigen Bezugsdauern (maximal 59 Monate) in den alten Bundesländern (durchgezogene Linie) und in den neuen Bundesländern (gestrichelte Linie).

8. Alte und neue Bundesländer. Die hier verwendeten Daten der Sozialhilfestatistik erlauben eine Unterscheidung von Haushalten in den alten Bundesländern (einschl. Berlin-Ost) und in den neuen Bundesländern (ohne Berlin-Ost). Wir können also versuchen, dementsprechend auch die Bezugsdauerverteilungen zu unterscheiden. Folgende Tabelle zeigt die Anzahl der Haushalte Ende 1997, 1998 und 1999.

	1997	1998	1999
Alte Bundesländer	333439	330252	320485
Neue Bundesländer	38682	41658	42673

Zu berücksichtigen ist natürlich, dass Haushalte in den neuen Bundesländern erst ab 1990 mit einem regulären Sozialhilfebezug beginnen konnten. Wir beschränken uns deshalb auf bisherige Bezugsdauern, die im Dezember 1997 maximal 84 Monate betragen. Box 13.1-2 zeigt analog zu Box 13.1-1 die Anzahl der Haushalte in den neuen Bundesländern, die Ende 1997, 1998 und 1999 eine bestimmte bisherige Dauer des Sozialhilfebezugs hatten.

Wiederum kann man Jahresraten für die Beendigung des Sozialhilfebezugs berechnen. Für die alten Bundesländer tritt erneut das in § 3 besprochene Datenproblem auf, das wir auf analoge Weise korrigieren. Bei den Daten für die neuen Bundesländer macht sich aufgrund der insgesamt geringen Fallzahlen noch deutlicher die Tatsache bemerkbar, dass es sich um Stichprobendaten handelt. Bei Bezugsdauern ab etwa 60 Monaten treten zahlreiche Inkonsistenzen in Gestalt negativer Jahresraten auf. Wir beschränken uns deshalb auf Bezugsdauern von maximal 59 Monaten.

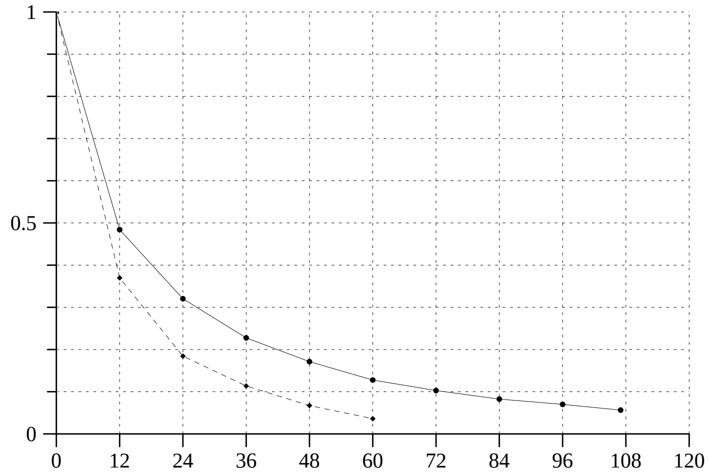


Abb. 13.1-11 Stützpunkte der Survivorfunktionen für tatsächliche Bezugsdauern in den alten Bundesländern (durchgezogene Linie) und in den neuen Bundesländern (gestrichelte Linie).

Zunächst vergleichen wir die Verteilungen der bisherigen Bezugsdauern. Um einen sinnvollen Vergleich zu ermöglichen, beschränken wir uns auch in den alten Bundesländern auf Bezugsdauern von maximal 59 Monaten. Abb. 13.1-10 zeigt die Verteilungen in Gestalt von Survivorfunktionen (analog zu Abb. 13.1-1). Ersichtlich sind die bisherigen Bezugsdauern in den neuen Bundesländern erheblich kürzer als in den alten Bundesländern.

Daraus dass die bisherigen Bezugsdauern kürzer sind, folgt jedoch nicht unbedingt, dass auch die tatsächlichen Bezugsdauern kürzer sind. Um auch die tatsächlichen Bezugsdauern zu vergleichen, konstruieren wir mit Hilfe der Jahresraten Stützpunkte für die Survivorfunktionen (vgl. § 4). Abb. 13.1-11 zeigt die so konstruierten Survivorfunktionen, wobei die Jahresraten $u_{97}(t)$ verwendet werden, die sich auf die Abgänge zwischen Ende 1997 und Ende 1998 beziehen.

Die Abbidlung macht deutlich, dass auch die tatsächlichen Bezugsdauern in den neuen Bundesländern deutlich kürzer sind als in den alten Bundesländern.[4] Es ist auch bemerkenswert, dass man bei Verwendung der Jahresraten $u_{98}(t)$ zu im wesentlichen identischen Ergebnissen gelangt. Tatsächlich lassen sich die Survivorfunktionen G^*_{97} und G^*_{98} für die neuen Bundesländer im Rahmen der durch die Stichprobendaten bedingten Fehler nicht unterscheiden.

[4] Wiederum könnte man, wie in § 4 besprochen wurde, die Gesamtheit der Jahresraten in die Berechnung einbeziehen. Für einen Vergleich ist das jedoch nicht unbedingt erforderlich.

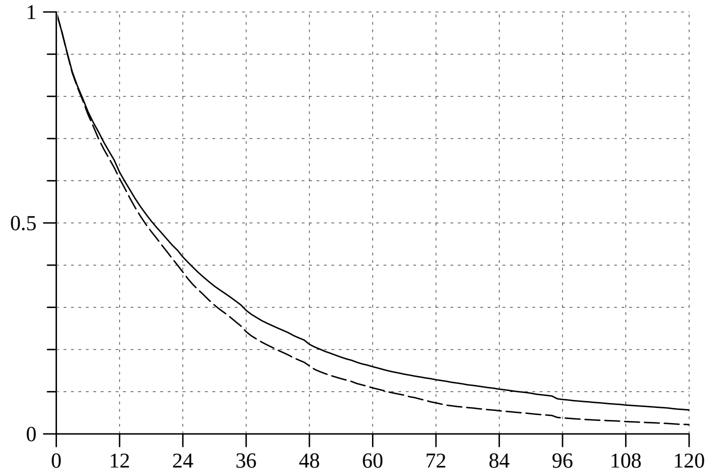


Abb. 13.1-12 Verteilungen der bisherigen Bezugsdauern bei Haushalten mit deutscher (durchgezogene Linie) und ausländischer (gestrichelte Linie) Bezugsperson.

Die Sozialhilfestatistik enthält zahlreiche weitere Haushaltsmerkmale, die man verwenden könnte, um Bezugsdauerverteilungen zu unterscheiden und Unterschiede zwischen den alten und neuen Bundesländern zu erklären. An dieser Stelle beschränken wir uns auf die Staatsangehörigkeit bzw. den Migrationsstatus, da sich Flüchtlinge, Zuwanderer und Personen mit ausländischer Staatsangehörigkeit überwiegend in den alten Bundesländern aufhalten. Die Sozialhilfestatistik unterscheidet fünf Personengruppen.[5] Die folgende Tabelle zeigt die in der Statistik verwendeten Bezeichnungen und die Fallzahlen:

	1997	1998	1999
Deutscher	300680	300020	295030
EG-Ausländer	8149	8305	8228
Asylberechtigter	8594	8208	7748
Bürgerkriegsflüchtling	1258	1105	1102
Sonstiger Ausländer	53440	54272	51050

Wegen der teilweise geringen Fallzahlen fassen wir die Haushalte, bei denen die Bezugsperson keine deutsche Staatsangehörigkeit hat, zu einer Gruppe zusammen.

[5]Die Zuordnung erfolgt auf der Ebene von Personen. Wir gruppieren die Haushalte entsprechend der jeweils ersten Person in jedem Haushaltsblock, d.i. in der Regel der Haushaltsvorstand.

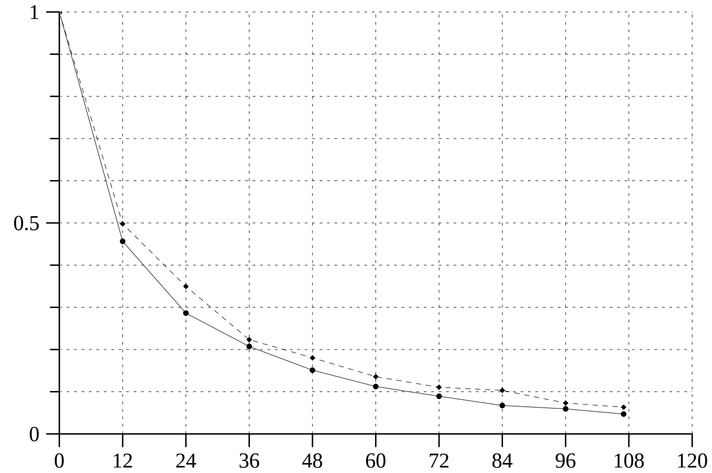


Abb. 13.1-13 Stützpunkte der Survivorfunktionen für tatsächliche Bezugsdauern bei Haushalten mit deutscher (durchgezogene Linie) und ausländischer Bezugsperson (gestrichelte Linie).

Zunächst vergleichen wir die Verteilungen der bisherigen Bezugsdauern Ende des Jahres 1997. Wie Abb. 13.1-12 zeigt, haben Haushalte mit einer ausländischen Bezugsperson tendenziell kürzere bisherige Bezugsdauern. Wie durch die bisherigen Ausführungen deutlich geworden sein sollte, erlaubt dies jedoch keinen unmittelbaren Schluß auf die Verteilung der tatsächlichen Bezugsdauern. Deshalb berechnen wir wiederum Stützwerte für die Survivorfunktionen der tatsächlichen Bezugsdauern, wobei wir die Jahresraten aus dem Übergang von Ende 1997 zum Ende 1998 verwenden. Abb. 13.1-13 zeigt das Ergebnis. Man erkennt, dass die tatsächlichen Bezugsdauern bei Haushalten mit einer ausländischen Bezugsperson etwas länger sind als bei Haushalten mit einer deutschen Bezugsperson. Diese Differenzen können aber sicherlich die vergleichsweise großen Unterschiede zwischen den neuen und den alten Bundesländern nicht erklären.

Kapitel 14

Episoden mit mehreren Folgezuständen

1. Einleitung.
2. Der begriffliche Rahmen.
3. Mehrere Folgezustände.
4. Eine numerische Illustration.
5. Verwendung zensierter Daten.
6. Durchschnittliche Verweildauern.
7. Retrospektiv- und Periodendaten.
8. Scheidungen und Todesfaelle.

In diesem Kapitel werden einige methodische Probleme bei der Verwendung von Episodendaten mit mehreren möglichen Folgezuständen besprochen. Der Text entspricht dem Beitrag von Rohwer (2006).

1. Einleitung. In vielen Bereichen der Sozialforschung interessiert man sich für Zeitdauern, während der sich Menschen oder andere Untersuchungseinheiten in bestimmten Zuständen aufhalten, z.B. für Ausbildungsdauern oder Dauern der Arbeitslosigkeit oder Ehedauern. Oft interessiert auch nicht nur die Dauer in einem Ausgangszustand, sondern die Zeitdauer bis zum Übergang in einen bestimmten Folgezustand, z.B. Arbeitslosigkeitsdauern bis zum Erreichen einer neuen Beschäftigung. Dann muss offenbar berücksichtigt werden, dass es mehrere mögliche Folgezustände gibt, in die gewechselt werden kann.

In der Literatur wird oft von „konkurrierenden Risiken" gesprochen. Zur Darstellung geeignete statistische Methoden wurden zuerst in der Demographie entwickelt: Sterbetafeln, bei denen zwei oder mehr unterschiedliche Arten von Todesfällen unterschieden werden. In der englischsprachigen Literatur wird von „multiple-decrement (life) tables" gesprochen;[1] als kurze deutsche Bezeichung hat U. Mueller (1993: 205) vorgeschlagen, von „Multi-Exit-Tafeln" zu sprechen. Solche Tafeln können natürlich nicht nur zur Untersuchung von Lebensdauern verwendet werden, sondern lassen sich immer anwenden, wenn Episoden durch unterschiedliche Arten von Ereignissen beendet werden können.

Der Grundgedanke der meisten statistischen Methoden für konkurrierende Risiken besteht darin, von zustandsspezifischen Übergangsraten auszugehen, die sich jeweils gesondert auf den Übergang in einen bestimmten Folgezustand beziehen. Hier setzen auch Regressionsmodelle an, die diese

[1] Man vgl. etwa Namboodiri und Suchindran (1987: 93ff.) oder Smith (1992: 139ff.).

zustandsspezifischen Raten als Funktionen von Kovariablen konzeptualisieren. Darauf wird im Folgenden nicht näher eingegangen. Hier konzentrieren wir uns auf die Frage, wie bei Episoden mit mehreren möglichen Folgezuständen von Verweildauern gesprochen werden kann. In vielen praktischen Anwendungen werden hierfür Pseudo-Survivorfunktionen verwendet, bei deren Berechnung davon ausgegangen wird, dass alle Episoden, die nicht in dem jeweils interessierenden Folgezustand enden, „zensiert" sind. Es wird also angenommen, dass diese Episoden gleichwohl noch in dem interessierenden Folgezustand enden könnten. Diese Annahme ist aber offenbar problematisch, insbesondere in denjenigen Fällen, in denen man weiß oder wissen könnte, dass eine Episode bereits durch einen anderen Folgezustand tatsächlich beendet worden ist. Wenn man z.B. schon weiß, dass eine Ehe durch den Tod eines Ehepartners beendet wurde, wäre es wenig plausibel anzunehmen, dass sie später noch durch eine Scheidung enden könnte.

In den folgenden Ausführungen diskutieren wir dieses Problem. Dabei geht es nicht nur um Begriffsbildungen, sondern wir versuchen auch, eine Vorstellung der Fehler und ihrer Größenordnungen zu gewinnen, die durch das übliche Verfahren bei der Berechnung von Pseudo-Survivorfunktionen entstehen können. Um zu zeigen, dass beliebig große Fehler entstehen können, verwenden wir zunächst ein fiktives Beispiel. Dann untersuchen wir das Problem anhand von Ehen, die durch Scheidungen oder durch den Tod eines Ehepartners enden können.

Als Alternative zur Berechnung von Pseudo-Survivorfunktionen schlagen wir Anteilsfunktionen vor, die sich auf die Anteile an einer Ausgangsgesamtheit beziehen, die in die unterschiedlichen Folgezustände wechseln. Es wird gezeigt, dass auch solche Anteilsfunktionen mit teilweise zensierten Daten geschätzt werden können.

2. Der begriffliche Rahmen. Wir gehen von einer Prozesszeitachse aus, die wir uns als eine Folge von Zeitstellen $t = 0, 1, 2, 3, \ldots$ vorstellen. Zeitstellen können z.B. Tage, Wochen, Monate oder Jahre sein. Die Zeitachse soll dazu dienen, Verweildauern zu erfassen, während der sich Menschen oder andere Untersuchungseinheiten in einem bestimmten Zustand befinden. Im folgenden sprechen wir von Personen, nehmen also an, dass wir uns auf eine Gesamtheit von Personen beziehen können, die zum Beginn der Prozesszeit in einen gleichen Ausgangszustand geraten sind, z.B. durch ihre Geburt oder ihre Heirat oder dadurch, dass sie arbeitslos geworden sind. Diese Gesamtheit repräsentieren wir durch eine Menge $\Omega := \{\omega_1, \ldots, \omega_n\}$. Jedes Element ω_i repräsentiert eine Person aus der Gesamtheit Ω. Insgesamt gibt es n Personen.

Um die Verweildauern im Ausgangszustand zu erfassen, wird eine statistische Variable $T : \Omega \longrightarrow \{0, 1, 2, 3, \ldots\}$ definiert, deren Wertebereich durch die Prozesszeitachse gegeben ist. Wenn ω eine Person aus Ω ist, soll also $T(\omega)$ die Zeitstelle angeben, in der diese Person den Ausgangszustand verlässt. So kann man auch sagen, dass $T(\omega)$ die Anzahl der vollendeten Zeitstellen angibt, während der sich ω im Ausgangszustand aufgehalten hat.

Weiter sei n_t die Anzahl der Personen, die sich zum Beginn der Zeitstelle t noch im Ausgangszustand befinden, so dass also $n_0 = n$ ist. Somit erhält man die Survivorfunktion, die die Verteilung der Verweildauervariablen T darstellt, durch $G(t) := n_t/n$.[2]

Definiert man außerdem durch $w_t := n_t - n_{t+1}$ die Anzahl der Personen, die während der Zeitstelle t den Ausgangszustand verlassen, kann man die Rate für das Verlassen des Ausgangszustands in der Zeitstelle t durch $r(t) := w_t/n_t$ definieren. Offenbar kann man Raten aus der Survivorfunktion berechnen:

$$r(t) = \frac{w_t}{n_t} = \frac{p(t)}{G(t)} = 1 - \frac{G(t+1)}{G(t)}$$

wobei $p(t) := w_t/n$ die relative Häufigkeit für einen Wechsel in der Zeitstelle t ist. Umgekehrt kann man aber auch die Survivorfunktion aus den Raten berechnen, denn es gilt der Zusammenhang

$$G(t) = \prod_{k=0}^{t-1} (1 - r(k))$$

Allerdings benötigt man hierfür die gesamte Ratenfunktion, die jeder Zeitstelle t die entsprechende Rate $r(t)$ zuordnet.[3]

3. Mehrere Folgezustände. Solange man sich nur für die Aufenthaltsdauer im Ausgangszustand interessiert, sind die Begriffsbildungen unproblematisch. In den meisten Anwendungen richtet sich die Aufmerksamkeit jedoch auf bestimmte Folgezustände, z.B. interessiert man sich für Ehedauern bis zu einer Scheidung oder für Arbeitslosigkeitsdauern bis zur Aufnahme einer neuen Beschäftigung. Dann muss berücksichtigt werden, dass es stets mehrere mögliche Folgezustände gibt. Insbesondere muss auch berücksichtigt werden, dass die Verweildauer im Ausgangszustand durch den Tod einer Person beendet werden kann.

Wir betrachten im weiteren zwei Folgezustände, die durch a und b indiziert werden. Es sei also $w_t = w_t^a + w_t^b$, wobei w_t^a und w_t^b die Anzahlen der Personen angeben, die in der Zeitstelle t in den Folgezustand a bzw. b wechseln. Somit kann man die Häufigkeiten $p^a(t) := w_t^a/n$ und $p^b(t) := w_t^b/n$ definieren, außerdem die *zustandsspezifischen Raten*

$$r^a(t) := \frac{w_t^a}{n_t} = \frac{p^a(t)}{G(t)} \quad \text{und} \quad r^b(t) := \frac{w_t^b}{n_t} = \frac{p^b(t)}{G(t)}$$

[2] Eine alternative Möglichkeit besteht darin, die Survivorfunktion durch $1 - F$ zu definieren, wobei F die Verteilungsfunktion ist. Die oben angegebene Definition vereinfacht jedoch bei einer diskreten Zeitachse die Notationen.

[3] Eine ausführliche Darlegung dieser formalen Zusammenhänge findet man bei Rohwer und Pötter 2001, Kap. 12.

Offenbar gilt $r(t) = r^a(t) + r^b(t)$, wobei $r(t)$ wie in § 2 die Rate für das Verlassen des Ausgangszustands ist.[4]

Wie kann man aber jetzt von Verweildauern sprechen? Man könnte versuchen, Verweildauerverteilungen durch Survivorfunktionen

$$\tilde{G}^a(t) := \prod_{k=0}^{t-1} (1 - r^a(k))$$

zu definieren. Diese Funktion liefert aber nicht die Verteilung der Verweildauern bis zum Erreichen des Folgezustands a. Denn ist man an diesen Verweildauern interessiert, kann man sich offenbar nur auf Personen beziehen, die diesen Folgezustand auch tatsächlich erreichen. Man muss also zunächst die Gesamtheit Ω in zwei Teilgesamtheiten zerlegen, wobei sich die Unterscheidung an dem Folgezustand orientiert, der schließlich erreicht wird. Beziehen wir uns auf die Teilgesamtheit Ω^a, die aus denjenigen Personen besteht, die schließlich den Folgezustand a erreichen, ist die Anzahl ihrer Mitglieder

$$n^a := \sum_{t=0}^{\infty} w_t^a$$

Bezugnehmend auf diese Gesamtheit kann man nun erneut eine Verweildauervariable $T^a : \Omega^a \longrightarrow \{0, 1, 2, 3, \ldots\}$ definieren, die die Verweildauer bis zum Erreichen des Folgezustands a erfasst. Ihr entsprechen die relativen Häufigkeiten w_t^a/n^a und die *zustandsspezifische Survivorfunktion*

$$G^a(t) := \sum_{k=t}^{\infty} \frac{w_k^a}{n^a} = 1 - \sum_{k=0}^{t-1} \frac{w_k^a}{n^a} = \prod_{k=0}^{t-1} \Big(1 - \frac{w_k^a/n^a}{G^a(k)}\Big)$$

die sich ersichtlich von der Survivorfunktion $\tilde{G}^a$ unterscheidet. Wenn man G^a berechnen möchte, genügt es auch nicht, nur die Ratenfunktion r^a zu kennen; man benötigt außerdem die Ratenfunktion r^b oder die allgemeine Ratenfunktion $r = r^a + r^b$ oder, damit äquivalent, die Survivorfunktion G.

Bei dieser Formulierung ist wichtig, dass man jeweils die gesamten Funktionen bis zum Ende der Prozesszeit kennen muss, oder anders formuliert: Zur Berechnung der Survivorfunktion G^a muss man den Anteil der Personen kennen, die schließlich in den Folgezustand a wechseln. Da dieser Anteil in den meisten praktischen Anwendungen – insbesondere dann, wenn Retrospektivdaten verwendet werden – nicht bekannt ist, wird stattdessen fast immer die Survivorfunktion $\tilde{G}^a$ berechnet. Wie im nächsten

[4] Es ist bemerkenswert, dass diese Additivität ohne weitere Annahmen unmittelbar aus den Definitionen und den Rechenregeln für relative Häufigkeiten folgt. Insbesondere braucht man nicht anzunehmen, dass die konkurrierenden Risiken „stochastisch unabhängig“ sind. Annahmen über Zusammenhänge zwischen konkurrierenden Risiken sind nur erforderlich, wenn man hypothetische Überlegungen über Situationen anstellen möchte, in denen einige der konkurrierenden Risiken wegfallen. Der in der Literatur dominierende Diskussionskontext betrifft Veränderungen in Lebensdauern, wenn bestimmte Mortalitätsrisiken hypothetisch ausgeschlossen werden (man vgl. etwa Keyfitz 1977: 48ff.).

Paragraph anhand eines Beispiels gezeigt wird, bezieht sich diese Funktion jedoch nicht auf sinnvoll interpretierbare Verweildauern, $\tilde{G}^a$ wird deshalb oft eine *Pseudo-Survivorfunktion* genannt.

Wenn man sich für Übergänge in einen Folgezustand a interessiert und nicht bekannt ist, wieviele Personen schließlich in diesen Folgezustand wechseln werden, kann man die Survivorfunktion G^a nicht berechnen. Man kann aber sinnvoll interpretierbare *Anteilsfunktionen*

$$H^a(t) := \frac{1}{n} \sum_{k=0}^{t-1} w_k^a = \sum_{k=0}^{t-1} p^a(k) = \sum_{k=0}^{t-1} r^a(k)\, G(k)$$

berechnen. $H^a(t)$ gibt den Anteil der Personen an der Ausgangsgesamtheit an, die bis zur Zeitstelle t in den Folgezustand a gewechselt haben. Bezieht man sich z.B. auf Ehen, die durch eine Scheidung (a) oder durch den Tod eines Ehepartners (b) enden können, würde $H^a(t)$ den Anteil der bis zur Ehedauer t geschiedenen Ehen angeben.

In der Praxis wird stattdessen oft die Pseudo-Survivorfunktion $\tilde{G}^a$ berechnet und so interpretiert, *als ob* es sich um eine Anteilsfunktion handeln würde. Diese Interpretation ist aber bei mehreren Folgezuständen nicht korrekt, da $1 - \tilde{G}^a$ keineswegs mit $H^a(t)$ identisch ist.[5] Man sieht das, wenn man $H^a(t)$ analog zur Definition von $\tilde{G}^a$ in der Form

$$H^a(t) = 1 - \prod_{k=0}^{t-1} \left(1 - \frac{w_k^a}{\bar{n}_k^a}\right)$$

schreibt, wobei $\bar{n}_k^a$ die Anzahl der Personen ist, die bis zur Zeitstelle k *nicht* in den Folgezustand a gewechselt haben. Bei der Berechnung der Pseudo-Survivorfunktion werden anstelle der Quotienten $w_k^a/\bar{n}_k^a$ die Raten w_k^a/n_k verwendet. Im Nenner steht dann nicht die Anzahl der Personen, die bisher nicht in den Folgezustand a gewechselt haben, sondern die Anzahl der Personen, die sich noch im Ausgangszustand befinden. Da $n_k \leq \bar{n}_k^a$ ist, findet man

$$H^a(t) \leq 1 - \tilde{G}^a(t)$$

Abhängig von den Anteilen der Personen, die in den Folgezustand b wechseln, kann es also durch die Verwendung der Pseudo-Survivorfunktion $\tilde{G}^a$ zu beliebig großen Überschätzungen der Wechsel in den Folgezustand a kommen.

Es ist auch bemerkenswert, dass Anteilsfunktionen im Unterschied zu Pseudo-Survivorfunktionen additiv sind: $H(t) = H^a(t) + H^b(t)$ liefert den Anteil der Personen, die bis zur Zeitstelle t in irgendeinen Folgezustand

[5] Es sei angemerkt, dass eine solche Interpretation auch in Teilen der Lehrbuchliteratur nahegelegt wird, man vgl. z.B. Blossfeld, Hamerle und Mayer (1986: 135f.) und Blossfeld und Rohwer (2002: 84).

Tabelle 14-1 Fiktive Daten für ein Beispiel mit zwei Folgezuständen a und b.

t	w_t^a	w_t^b	t	w_t^a	w_t^b	t	w_t^a	w_t^b	t	w_t^a	w_t^b
0	100	50	5	80	60	10	60	70	15	40	80
1	100	50	6	80	60	11	60	70	16	20	80
2	100	50	7	80	60	12	40	80	17	20	90
3	100	50	8	60	70	13	40	80	18	20	90
4	80	60	9	60	70	14	40	80	19	20	

gewechselt haben. Schließlich sei auch noch angemerkt, dass es einen einfachen Zusammenhang zwischen der Anteilsfunktion H^a und der zustandsspezifischen Survivorfunktion G^a gibt:

$$G^a(t) = 1 - \sum_{k=0}^{t-1} \frac{w_k^a}{n^a} = 1 - \sum_{k=0}^{t-1} \frac{w_k^a}{n} \frac{n}{n^a} = 1 - \frac{n}{n^a} H^a(t)$$

Würde man den Anteil der Personen, die schließlich in den Folgezustand a wechseln, kennen, könnte man G^a aus H^a berechnen.

4. Eine numerische Illustration. Zur Illustration der Begriffsbildungen und Verdeutlichung des Problems betrachten wir ein Zahlenbeispiel mit fiktiven Daten. Wir gehen von einer Gesamtheit von 2500 arbeitslos gewordenen Personen aus, und es gibt zwei mögliche Folgezustände: (a) Übergang in eine neue Beschäftigung und (b) Verlassen des Arbeitsmarktes (einschließlich Sterbefälle). Tabelle 14-1 zeigt die fiktiven Daten. Die Zeitstellen der Zeitachse $t = 0, \ldots, 19$ kann man sich als Quartale vorstellen. w_t^a und w_t^b geben die Anzahl der Personen an, die in der Zeitstelle t in den Zustand a bzw. b wechseln.

Da in diesem Beispiel vollständige Daten gegeben sind, kann man alle Begriffsbildungen nachvollziehen:

a) Man findet, dass von den 2500 Personen 1200 (48 %) eine neue Beschäftigung annehmen und 1300 (52 %) aus dem Arbeitsmarkt ausscheiden.
b) Man kann die Survivorfunktionen G^a und G^b für die beiden Folgezustände berechnen (Abb. 14-1).
c) Man kann die durchschnittlichen Verweildauern berechnen. Die durchschnittliche Verweildauer im Ausgangszustand beträgt 8.5, bis zu einer neuen Beschäftigung 6.8 und bis zum Ausscheiden aus dem Arbeitsmarkt 10.0 Quartale.[6]

Jetzt berechnen wir eine Survivorfunktion für den Übergang in eine Beschäftigung, wobei wir diejenigen Episoden, die nicht mit einer neuen Beschäftigung enden, als rechts zensiert betrachten. Dafür kann das Kaplan-Meier-Verfahren verwendet werden. Tabelle 14-2 zeigt die Rechenschritte.

[6] Natürlich könnte man berücksichtigen, dass die Prozesszeitachse bei 0 beginnt und jeweils ein halbes Quartal hinzurechnen.

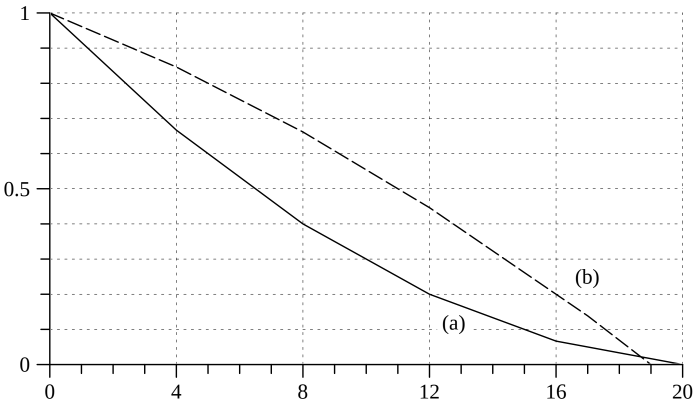


Abb. 14-1 Survivorfunktionen G^a und G^b für die durch Tabelle 14-1 gegebenen Daten.

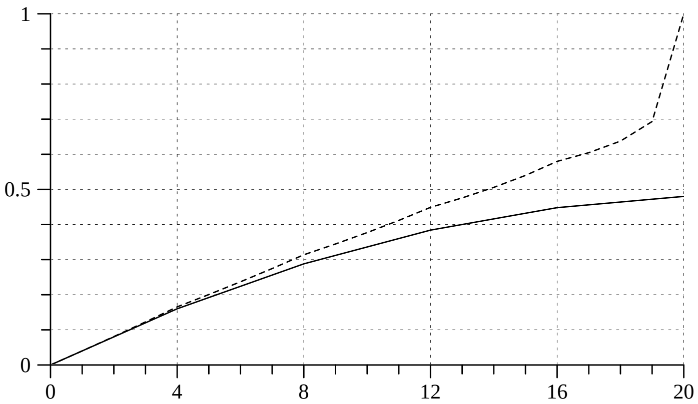


Abb. 14-2 Darstellung der Funktionen H^a (durchgezogene Linie) und $1 - \tilde{G}^a$ (gestrichelte Linie) für die durch Tabelle 14-1 gegebenen Daten.

Als Ergebnis erhält man die Pseudo-Survivorfunktion $\tilde{G}^a$.

Aber welche Bedeutung kann dieser Survivorfunktion gegeben werden? Sie liefert jedenfalls keine sinnvollen Aussagen über die Dauer der Arbeitslosigkeit, und zwar weder über die Dauer im Ausgangszustand noch über die Dauer bis zum Erreichen einer Beschäftigung. So liegt zum Beispiel der Median von $\tilde{G}^a$ bei 13 bis 14 Quartalen und der Mittelwert beträgt 11.9 Quartale. Beide Werte sind deutlich größer als die durchschnittliche Dauer bis zum Erreichen einer neuen Beschäftigung (6.8 Quartale) und sogar größer als die durchschnittliche Verweildauer im Ausgangszustand (8.5 Quartale).

Tabelle 14-2 Berechnung der Survivorfunktion $\tilde{G}^a$ mit dem Kaplan-Meier-Verfahren aus den Daten in Tabelle 14-1.

t	n_t	w_t^a	w_t^b	$r^a(t)$	$\tilde{G}^a(t)$
0	2500	100	50	0.0400	1.0000
1	2350	100	50	0.0426	0.9600
2	2200	100	50	0.0455	0.9191
3	2050	100	50	0.0488	0.8774
4	1900	80	60	0.0421	0.8346
5	1760	80	60	0.0455	0.7994
6	1620	80	60	0.0494	0.7631
7	1480	80	60	0.0541	0.7254
8	1340	60	70	0.0448	0.6862
9	1210	60	70	0.0496	0.6555
10	1080	60	70	0.0556	0.6230
11	950	60	70	0.0632	0.5884
12	820	40	80	0.0488	0.5512
13	700	40	80	0.0571	0.5243
14	580	40	80	0.0690	0.4944
15	460	40	80	0.0870	0.4603
16	340	20	80	0.0588	0.4202
17	240	20	90	0.0833	0.3955
18	130	20	90	0.1538	0.3626
19	20	20	0	1.0000	0.3068
20	0	0	0		0.0000

Die Funktion $\tilde{G}^a$ liefert aber auch keine Information über die Anteile der Personen, die noch nicht in eine Beschäftigung gewechselt haben. Wie im vorangegangenen Paragraph ausgeführt wurde, erhält man diese Anteile durch die Funktion $H^a \neq 1 - \tilde{G}^a$. Abb. 14-2 illustriert die Differenz für unser Beispiel. Es ist deutlich sichtbar, dass die Pseudo-Survivorfunktion keine adäquaten Aussagen über die Anteile derjenigen Personen macht, die in eine Beschäftigung wechseln.

5. Verwendung zensierter Daten. Manchmal wird die Verwendung einer Pseudo-Survivorfunktion mit dem Argument begründet, dass bei ihrer Berechnung zensierte Beobachtungen berücksichtigt werden können (z.B. Diekmann und Klein 1991). Dies ist zwar richtig, das eigentliche Problem besteht jedoch darin, dass zugleich unterstellt wird, dass alle zensierten Episoden schließlich in dem speziellen Folgezustand enden werden, auf den sich die Berechnung der Pseudo-Survivorfunktion bezieht. Wenn z.B. eine Pseudo-Survivorfunktion für Ehedauern bis zu einer Scheidung berechnet wird, wird zugleich angenommen, dass alle zensierten Ehe-Episoden schließlich in einer Scheidung enden werden. Eine solche Annahme ist aber sicherlich fragwürdig.

Andererseits ist es auch nicht richtig, dass bei der Berechnung von Anteilsfunktionen zensierte Beobachtungen grundsätzlich nicht verwendet werden können. Es gilt nämlich folgender Zusammenhang, der bereits in § 3 erwähnt wurde:

$$H^a(t) \;=\; \sum_{k=0}^{t-1} p^a(k) \;=\; \sum_{k=0}^{t-1} r^a(k)\, G(k) \tag{14.1}$$

Wenn man also glaubt, dass man mit partiell zensierten Daten die Survivorfunktion G und die Rate r^a (oder, damit äquivalent, die Pseudo-Survivorfunktion $\tilde{G}^a$) einigermaßen korrekt schätzen kann, dann kann man sie sogleich auch verwenden, um die Anteilsfunktion H^a zu schätzen.

6. Durchschnittliche Verweildauern. An mehreren Stellen wurde bereits über durchschnittliche Verweildauern gesprochen. Es ist nützlich, sich noch einmal die unterschiedlichen Begriffsbildungen zu verdeutlichen. In jedem Fall muss betont werden, dass man nur aus einer retrospektiven Perspektive Feststellungen über Verweildauern machen kann. Natürlich kann man Vermutungen darüber anstellen, wie lange eine noch nicht abgeschlossene Episode dauern wird; aber erst nach ihrer Beendigung kann man wissen, wie lange sie tatsächlich gedauert hat.

Dies gilt auch dann, wenn man sich zunächst nur für die durchschnittliche Verweildauer im Ausgangszustand interessiert. Zu berechnen ist in diesem Fall der Mittelwert der Verweildauervariablen T, und das setzt eine Kenntnis der gesamten Verteilung voraus. Verwendet man z.B. die Survivorfunktion, muss

$$\mathrm{M}(T) \;=\; \sum_{t=0}^{\infty} t\,(G(t) - G(t+1))$$

berechnet werden. Da man bei vielen praktischen Anwendungen nicht die gesamte Verteilung kennt (oder schätzen kann), werden anstelle des Mittelwerts oft Mediane oder andere Quantile angegeben. Dabei sollte jedoch bedacht werden, dass sich der Median einer Verteilung von ihrem Mittelwert erheblich unterscheiden kann. Zum Beispiel liegt bei Bezugsdauern der Sozialhilfe der Median unterhalb eines Jahres, die durchschnittliche Bezugsdauer ist jedoch wesentlich länger als ein Jahr.

Der begriffliche Unterschied wird deutlich, wenn man bedenkt, dass sich Quantile einer Verweildauerverteilung gar nicht auf die Gesamtheit aller Episoden beziehen, sondern nur auf Teilgesamtheiten mit den jeweils kürzesten Episodendauern. Zum Beispiel gibt der Median an, in welcher Zeit die Hälfte der Personen mit den kürzesten Verweildauern den Ausgangszustand verlassen hat; dabei bleiben die Verweildauern der übrigen Personen vollständig unberücksichtigt.

Es erscheint uns deshalb sinnvoll, nicht nur Quantile, sondern auch Mittelwerte zu berechnen. Wenn man nicht die gesamte Verteilung kennt, ist es dennoch oft möglich, bedingte Mittelwerte zu berechnen, die sich auf die Gesamtheit der Episoden bis zu einer maximalen Verweildauer t_0 beziehen, also

$$\mathrm{M}(T \,|\, T < t_0) \;=\; \sum_{t=0}^{t_0-1} t\, \frac{G(t) - G(t+1)}{1 - G(t_0)}$$

Zum Beispiel könnte man auf diese Weise die durchschnittliche Lebensdauer von Menschen berechnen, die maximal 90 oder 100 Jahre alt geworden sind.

Eine scheinbare begriffliche Unklarheit tritt auf, wenn man sich nicht auf Verweildauern im Ausgangszustand, sondern auf Verweildauern bis zum Erreichen eines bestimmten Folgezustands a interessiert, also z.B. bei Arbeitslosigkeitsepisoden für die durchschnittliche Dauer bis zum Wechsel in eine neue Beschäftigung. Es wurde bereits darauf hingewiesen, dass es irreführend wäre, hierfür die Pseudo-Survivorfunktion $\tilde{G}^a$ zu verwenden, denn bereits die Begriffsbildung impliziert, dass man sich nur auf Episoden beziehen kann, die tatsächlich im Zustand a enden. Zu berechnen ist deshalb der Mittelwert der Verweildauervariablen T^a, die für die Teilgesamtheit derjenigen Episoden definiert ist, die im Zustand a enden.[7]

Auch wenn man nicht weiß, welcher Anteil der Episoden schließlich im Zustand a enden wird, ist es oft möglich und sinnvoll, wenigstens bedingte Mittelwerte zu berechnen, die sich auf eine Teilgesamtheit von Episoden beziehen, die bis zu einer bestimmten Dauer t_0 den Zustand a, z.B. eine neue Beschäftigung, erreicht haben. Solche bedingten Mittelwerte können nämlich bereits aus der Anteilsfunktion H^a berechnet werden:

$$\begin{aligned} \mathrm{M}(T^a \,|\, T^a < t_0) \;&=\; \sum_{t=0}^{t_0-1} t\, \frac{G^a(t) - G^a(t+1)}{1 - G^a(t_0)} \\ &=\; \sum_{t=0}^{t_0-1} t\, \frac{H^a(t+1) - H^a(t)}{H^a(t_0)} \;=\; \frac{\sum_{t=0}^{t_0-1} t\, w_t^a}{\sum_{t=0}^{t_0-1} w_t^a} \end{aligned}$$

Auf diese Weise wurden z.B. in einer Publikation des Statistischen Bundesamts (1990: 129f.) „durchschnittliche Ehedauern bis zu einer Scheidung“ berechnet. Unter Verwendung von Periodendaten, die sich auf das Jahr 1988 beziehen, und mit $t_0 = 26$ wurde so ein Wert von 12.1 errechnet, d.h. die durchschnittliche Ehedauer derjenigen Ehen, die 1988 geschieden wurden und die maximal 25 Jahre dauerten, lag bei 12.1 Jahren. Das ist zugleich eine untere Grenze für die durchschnittliche Dauer aller Ehen bis zu einer Scheidung.

7. Retrospektiv- und Periodendaten. In § 4 wurde gezeigt, dass der Pseudo-Survivorfunktion $\tilde{G}^a$, die sich auf den Übergang in einen bestimmten Folgezustand a bezieht und bei deren Berechnung alle nicht in diesem Zustand endenden Episoden als rechts zensiert angenommen werden, nicht ohne weiteres eine sinnvolle Bedeutung gegeben werden kann. Das Problem tritt natürlich nur auf, wenn man sich für einen bestimmten Folgezustand

[7] Alternativ könnten bedingte Mittelwerte der Form $\mathrm{M}(T^a \,|\, T^a \geq x)$ betrachtet werden; ihnen entspricht im Kontext unterschiedlicher Mortalitätsrisiken eine „durchschnittliche todesursachenspezifische Lebenserwartung“ bzw. „die restliche Lebenserwartung einer Person im Alter x, falls diese Person an der spezifischen Ursache j [in unserer Notation: a] sterben wird.“ (Mueller 1993: 207)

interessiert und nicht nur für die undifferenzierte Verweildauer im Ausgangszustand. Aber das ist in den meisten praktischen Anwendungen der Fall. Eine Ausnahme bilden Untersuchungen der Lebensdauer, bei denen der Tod der einzige mögliche Endzustand ist. Aber in allen anderen Fällen ist mindestens der Tod einer Person eine reale Alternative, in der eine Episode enden kann.

Ob bzw. wie alternative Folgezustände zu berücksichtigen sind, hängt auch von der Art der Datengewinnung ab. In zahlreichen Studien werden retrospektiv erhobene Daten verwendet, bei denen von vornherein nur Personen betrachtet werden können, die mindestens bis zum Interviewzeitpunkt überlebt haben. Infolgedessen liegt die Vermutung nahe, dass man die Tatsache, dass Episoden auch durch Tod enden können – meistens (abgesehen von Datenerhebungen über bereits sehr alte Menschen) – ignorieren kann. Das ist jedoch in vielen Fällen zumindest fragwürdig.

Als Beispiel beziehen wir uns auf die Untersuchung von Ehen, die in einer Scheidung enden. Dabei ist offenbar zu berücksichtigen, dass Ehen entweder durch eine Scheidung (Folgezustand a) oder durch den Tod eines Ehepartners (Folgezustand b) enden können. Eine Retrospektivbefragung von Personen, die in der Vergangenheit geheiratet haben, kann drei Arten von Beobachtungen liefern:

a) Ehen, bei denen bis zum Interviewzeitpunkt eine Scheidung stattgefunden hat;

b) Ehen, bei denen bis zum Interviewzeitpunkt einer der Ehepartner gestorben ist, und

c) Ehen, die zum Interviewzeitpunkt noch bestehen (rechts zensierte Beobachtungen).

In allen Fällen muss auch mindestens einer der (ehemaligen) Ehepartner mindestens bis zum Interviewzeitpunkt leben.

Mit Daten dieser Art kann offenbar die Survivorfunktion G^a für die Ehedauern bis zu einer Scheidung nicht berechnet werden, da man nicht weiß, wieviele Ehen schließlich in einer Scheidung enden. Deshalb wird in den meisten empirischen Untersuchungen (man vgl. z.B. einige der Beiträge in Klein und Kopp 1999) eine Pseudo-Survivorfunktion $\tilde{G}^a$ berechnet. Das kann man natürlich tun; sie liefert jedoch nicht die Anteile der Ehen, die in einer Scheidung enden, wie bei Interpretationen oft unterstellt wird. Denn in ihrer Berechnung wird angenommen, dass auch diejenigen Ehen, von denen man weiß oder wissen könnte, dass sie durch den Tod eines Ehepartners enden, schließlich durch eine Scheidung enden könnten.

Wenn man sich für die Anteile der Ehen interessiert, die in einer Scheidung enden, müsste man stattdessen die in § 3 definierte Anteilsfunktion H^a berechnen. Wie in § 5 gezeigt wurde, ist dies grundsätzlich auch mit Retrospektivdaten möglich, bei denen die ermittelbaren Episoden teilweise zensiert sind. Allerdings muss bei der Interpretation beachtet werden,

dass man bei Retrospektiverhebungen nur Informationen von Personen gewinnen kann, die mindestens bis zum Interviewzeitpunkt überleben.

Im Unterschied zu retrospektiv erhobenen Daten liefert die amtliche Statistik in erster Linie Periodendaten, die sich auf einzelne Kalenderjahre beziehen. Um bei unserem Beispiel zu bleiben, betrachten wir ehedauerspezifische Scheidungsraten und -ziffern. Zur Definition werden folgende Größen verwendet:

$n_{j,t}$:= Anzahl der Ehen, die im Kalenderjahr j eine Dauer von t Jahren hatten.

$w^a_{j,t}$:= Anzahl der Ehen, die im Kalenderjahr j eine Dauer von t Jahren hatten und in diesem Kalenderjahr geschieden wurden.

Ehedauerspezifische Scheidungs*raten* sind dann durch $q^a_j(t) := w^a_{j,t}/n_{j,t}$ definiert.[8] Sie beziehen sich auf den Anteil der Scheidungen im Kalenderjahr j bei denjenigen Ehen, die in diesem Jahr eine Ehedauer von t Jahren hatten. Man kann sie auf zwei unterschiedliche Weisen zur Berechnung von Survivorfunktionen verwenden. In einer Periodenbetrachtung bezieht man sich durchgehend auf die Scheidungsraten eines Kalenderjahrs j und erhält

$$\breve{G}^a_j(t) := \prod_{k=0}^{t-1} (1 - q^a_j(k))$$

Bei einer Quasi-Kohortenbetrachtung geht man von den Ehen aus, die in einem Kalenderjahr j geschlossen worden sind, und verwendet Scheidungsraten für die jeweilige Ehedauer.[9] So gelangt man zur Survivorfunktion

$$\tilde{G}^a_j(t) := \prod_{k=0}^{t-1} (1 - q^a_{j+k}(k))$$

In beiden Fällen handelt es sich jedoch um Pseudo-Survivorfunktionen, bei deren Konstruktion unberücksichtigt bleibt, dass Ehen auch durch den Tod eines Ehepartners enden können. Sie weisen also die gleichen Mängel auf wie die Pseudo-Survivorfunktionen, die üblicherweise aus Retrospektivdaten berechnet werden.

Eine Alternative liegt in der Verwendung von ehedauerspezifischen Scheidungs*ziffern*, die folgendermaßen definiert sind: $p^a_j(t) := w^a_{j,t}/n_{j-t,0}$. In diesem Fall wird die Anzahl der Scheidungen, die im Jahr j bei Ehen mit einer bisherigen Dauer von t Jahren aufgetreten sind, auf die Anzahl der Eheschließungen im Jahr $j-t$ bezogen. Tabelle 14-3 zeigt diese Scheidungsziffern für das Kalenderjahr $j = 1999$ und Ehedauern $t = 0, \ldots, 25$. Zum Beispiel wurden in diesem Jahr 11139 Ehen geschieden, die im Jahr

[8] Diese Definition entspricht derjenigen von I. Esenwein-Rothe (1982: 293).

[9] Da die Kohortenkonstruktion auf einer zeitlichen Verkettung ehedauerspezifischer Periodendaten beruht, sprechen wir von Quasi-Kohorten.

Tabelle 14-3 Ehedauerspezifische Scheidungsziffern für das Jahr 1999, multipliziert mit 1000 (Statistisches Bundesamt 1999: 268).

t	$p^a_{1999}(t)$	t	$p^a_{1999}(t)$	t	$p^a_{1999}(t)$	t	$p^a_{1999}(t)$
0	0.3	7	25.7	14	13.8	21	7.7
1	3.5	8	22.9	15	12.6	22	7.0
2	13.4	9	21.6	16	11.3	23	6.2
3	19.2	10	20.6	17	10.5	24	5.6
4	24.2	11	17.9	18	9.8	25	5.1
5	27.2	12	16.1	19	9.3		
6	27.1	13	15.0	20	8.5		

1990 geschlossen wurden. Andererseits kennt man aus früheren Erhebungen die Anzahl der Ehen, die 1990 geschlossen wurden, nämlich 516388. So erhält man $p^a_{1999}(9) = 11139/516388 = 0.0216$. Ein Problem besteht offenbar darin, dass in einer Gesellschaft mit Zu- und Abwanderung die Ehen, die im Jahr j bei einer Dauer von t geschieden werden, keine Teilmenge derjenigen Ehen sind, die im Jahr $j - t$ geschlossen wurden.[10] Ist man jedoch bereit, von diesem Problem in erster Näherung zu abstrahieren, liefern ehedauerspezifische Scheidungsziffern eine durchaus sinnvolle Möglichkeit, um zu schätzen, wieviele Ehen eines Eheschließungsjahrgangs in einer Scheidung enden. Ausgehend von den Ehen, die im Jahr j geschlossen worden sind, ist der Anteil der bis zu einer Ehedauer von t geschiedenen Ehen

$$H^a_j(t) := \frac{1}{n_{j,0}} \sum_{k=0}^{t-1} w^a_{j+k,k} = \sum_{k=0}^{t-1} p^a_{j+k}(k)$$

Zur Berechnung benötigt man natürlich die ehedauerspezifischen Scheidungsziffern für eine Folge von Jahren. Kennt man sie bis zum Jahr 1999, könnte man z.B. $H^a_{1974}(26)$ berechnen, also – unter der genannten Einschränkung – den Anteil der 1974 geschlossenen Ehen, die bis zum Ende des Jahres 1999 geschieden worden sind. Abb. 14-3 zeigt diese Funktion H^a_{1974} als durchgezogene Linie.[11] Andererseits könnte man auch eine hypothetische Betrachtung anstellen, bei der man von den ehedauerspezifischen Scheidungsziffern eines bestimmten Kalenderjahrs ausgeht, also

$$\breve{H}^a_j(t) := \sum_{k=0}^{t-1} p^a_j(k)$$

berechnet. Mit den Angaben aus Tabelle 14-3 findet man z.B. $\breve{H}^a_{1999}(26) = 0.362$. Unter Zugrundelegung der Scheidungsziffern des Jahres 1999 würden also bei einem hypothetischen Eheschließungsjahrgang bis zu einer

[10] Eine weitere Fehlerquelle für Deutschland ist die Gebietsveränderung seit 1990.

[11] Die Daten entstammen der Tabelle VIII B/181 des Statistischen Bundesamts, die uns freundlicherweise von Herrn H.-P. Bosse zur Verfügung gestellt wurde.

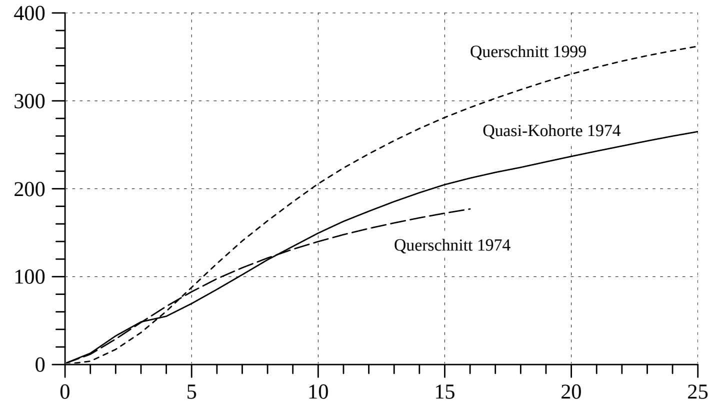


Abb. 14-3 Kumulierte ehedauerspezifische Scheidungsziffern (multipliert mit 1000).

Ehedauer von einschließlich 25 Jahren 36.2 % der Ehen geschieden werden. Abb. 14-3 zeigt die entsprechende Funktion als gestrichelte Linie.

8. Scheidungen und Todesfälle. In § 4 wurde gezeigt, dass es sehr große Differenzen zwischen der Pseudo-Survivorfunktion $\breve{G}^a$ und der Anteilsfunktion H^a geben kann. Wie groß die Unterschiede sind, hängt insbesondere von den Häufigkeiten ab, mit denen die unterschiedlichen Folgezustände auftreten, und von ihrer zeitlichen Lagerung in der Prozesszeitachse. In diesem Paragraph untersuchen wir das Problem bei Ehedauern, wobei die Ehen entweder durch eine Scheidung oder durch den Tod eines Ehepartners enden können. Zuerst konstruieren wir mit Hilfe von Periodendaten der amtlichen Statistik einen hypothetischen Entwicklungsprozess für Ehen, dann verwenden wir diese Daten, um die unterschiedlichen Funktionen zu vergleichen.

Wir beziehen uns auf Ehen, die 1999 in Deutschland geschlossen wurden, insgesamt 430674 (Statistisches Bundesamt 1999: 258). In der Quelle findet man auch eine Aufgliederung dieser Ehen nach dem Alter des Mannes und der Frau bei der Heirat. Es sei $n_{\tau,\tau'}$ die Anzahl der Ehen, die 1999 geschlossen wurden, bei denen der Mann bei der Heirat τ Jahre und die Frau τ' Jahre alt war. Im weiteren beschränken wir uns auf diejenigen Ehen, bei denen die Altersangaben nicht gruppiert, sondern in Jahren ausgewiesen sind:

$$n := \sum_{\tau=18}^{64} \sum_{\tau'=18}^{54} n_{\tau,\tau'} = 414030$$

Es sei n_t die Anzahl der Ehen, die bei einer Ehedauer von t Jahren noch bestehen. Anfangs ist $n_0 = n$. Ehen können durch Scheidungen oder durch den Tod eines Ehepartners enden. Die Scheidungen simulieren wir mit

Tabelle 14-4 Hypothetischer Entwicklungsprozess von 414030 Ehen, die 1999 geschlossen wurden. Zu den Annahmen vgl. man die Erläuterungen im Text.

t	n_t	w_t^a	w_t^b	t	n_t	w_t^a	w_t^b
0	414030	124	947	26	210791	2056	3767
1	412959	1449	1010	27	204968	1999	3952
2	410500	5548	1080	28	199017	1941	4137
3	403872	7949	1147	29	192940	1882	4323
4	394775	10020	1215	30	186735	1821	4510
5	383540	11262	1283	31	180403	1759	4698
6	370995	11220	1350	32	173946	1696	4885
7	358425	10641	1421	33	167365	1632	5069
8	346364	9481	1496	34	160664	1567	5248
9	335387	8943	1578	35	153849	1500	5421
10	324865	8529	1667	36	146928	1433	5583
11	314670	7411	1758	37	139913	1365	5733
12	305501	6666	1857	38	132816	1295	5868
13	296978	6210	1963	39	125652	1225	5989
14	288805	5714	2074	40	118438	1155	6086
15	281017	5217	2191	41	111196	1084	6162
16	273610	4679	2309	42	103950	1014	6209
17	266622	4347	2436	43	96728	943	6231
18	259839	4057	2567	44	89554	873	6224
19	253214	3850	2703	45	82456	804	6183
20	246661	3519	2840	46	75469	736	6108
21	240302	3188	2981	47	68625	669	5993
22	234133	2898	3128	48	61962	604	5840
23	228107	2567	3280	49	55519	541	5646
24	222260	2319	3438	50	49331	481	5411
25	216504	2112	3601	51	43439		

Hilfe der in Tabelle 14-1 angegebenen Scheidungsziffern für 1999. Um den Simulationszeitraum über 25 Jahre hinaus ausdehnen zu können, wird für $t > 25$ eine konstante Ehescheidungs*rate* angenommen, die dem Wert für $t = 25$ entspricht.[12]

Um zu simulieren, dass Ehen auch durch den Tod eines Ehepartners enden können, verwenden wir die alters- und geschlechtsspezifischen Ster-

[12] Für unsere Zwecke wäre es zwar besser, von vornherein ehedauerspezifische Scheidungsraten zu verwenden; solche Scheidungsraten werden jedoch vom Statistischen Bundesamt nicht publiziert, da es für die Gliederung der Ehen nach ihrer Dauer keine Periodendaten gibt.

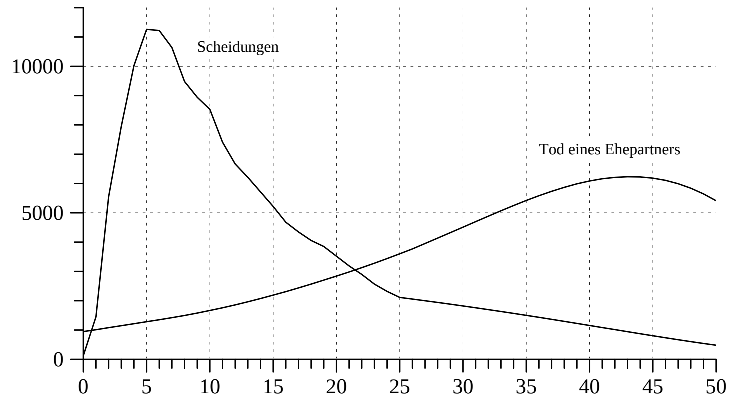


Abb. 14-4 Absolute Häufigkeiten von Ehen, die durch eine Scheidung oder durch den Tod eines Ehepartners beendet wurden, entsprechend den Daten in Tabelle 14-4.

beraten des Jahres 1999:

$$\delta_\tau^m := \frac{\text{Anzahl der 1999 gestorbenen Männer im Alter } \tau}{\text{jahresdurchschnittliche Anzahl von Männern im Alter } \tau}$$

$$\delta_\tau^f := \frac{\text{Anzahl der 1999 gestorbenen Frauen im Alter } \tau}{\text{jahresdurchschnittliche Anzahl von Frauen im Alter } \tau}$$

Da uns geeignete Daten nur bis zu einem Alter von 94 Jahren zur Verfügung stehen,[13] nehmen wir an, dass die Sterberaten im Alter 95 den Wert 1 haben. In unserem simulierten Datensatz sterben also alle Personen spätestens im Alter von 95 Jahren.

Es sei nun $n_{\tau,\tau',t}$ die Anzahl der Ehen bei einer Ehedauer von t Jahren, bei denen bei der Heirat der Mann τ und die Frau τ' Jahre alt gewesen sind. Von diesen Ehen enden

$$w_{\tau,\tau',t}^b := n_{\tau,\tau',t}\,(\delta_{\tau+t}^m + \delta_{\tau'+t}^f - \delta_{\tau+t}^m \delta_{\tau'+t}^f)$$

durch den Tod mindestens eines Ehepartners. Insgesamt enden also bei einer Ehedauer von t Jahren

$$w_t^b := \sum_{\tau=18}^{64} \sum_{\tau'=18}^{54} w_{\tau,\tau',t}^b$$

Ehen durch den Tod eines Ehepartners. Außerdem treten Scheidungen auf. Wenn $t \leq 25$ ist, gibt es

$$w_{\tau,\tau',t}^a := n_{\tau,\tau',0}\, p_{1999}^a(t)$$

[13] Die erforderlichen Daten haben wir den Segmenten 685 und 1124-26 der STATIS Datenbank des Statistischen Bundesamts entnommen.

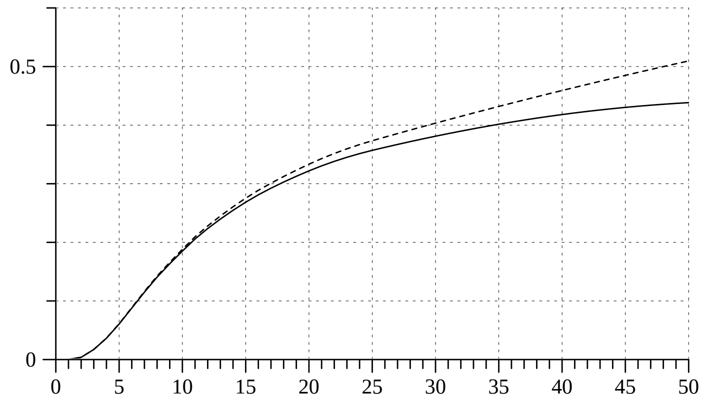


Abb. 14-5 Anteilsfunktion H^a (durchgezogene Linie) und Pseudo-Survivorfunktion $1 - \tilde{G}^a$ (gestrichelte Linie) für die Daten in Tabelle 14-4

Scheidungen, wobei $p^a_{1999}(t)$ die Scheidungsziffer für 1999 bei einer Ehedauer von t Jahren ist; und wenn $t > 25$ ist, gibt es

$$w^a_{\tau,\tau',t} := n_{\tau,\tau',t} \frac{w^a_{\tau,\tau',25}}{n_{\tau,\tau',25}}$$

Scheidungen. Somit kann man den Bestand der Ehen auf folgende Weise fortschreiben:

$$n_{\tau,\tau',t+1} = n_{\tau,\tau',t} - w^a_{\tau,\tau',t} - w^b_{\tau,\tau',t}$$

Durch Aggregation gewinnt man die in Tabelle 14-4 angegebenen Zahlen. Abb. 14-4 zeigt, wie sich die Häufigkeiten von Ehelösungen durch Scheidung bzw. durch den Tod eines Ehepartners bis zu einer Ehedauer von 50 Jahren verteilen.

Die Daten in Tabelle 14-4 können verwendet werden, um Anteils- und Pseudo-Survivorfunktionen zu berechnen. Abb. 14-5 zeigt als durchgezogene Linie die Funktion H^a und als gestrichelte Linie die Funktion $1 - \tilde{G}^a$. Man erkennt, dass die Unterschiede bei Ehedauern bis zu 20 Jahren nur klein sind, dann jedoch mit zunehmender Ehedauer immer größer werden. Die Pseudo-Survivorfunktion suggeriert sogar die sicherlich falsche Vorstellung, dass bis zu einer Ehedauer von etwa 47 Jahren die Hälfte aller Ehen durch eine Scheidung beendet wurde.

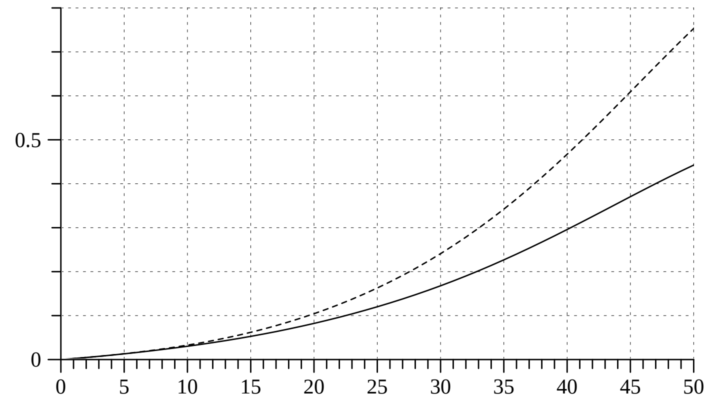


Abb. 14-6 Anteilsfunktion H^b (durchgezogene Linie) und Pseudo-Survivorfunktion $1 - \tilde{G}^b$ (gestrichelte Linie) für die Daten in Tabelle 14-4.

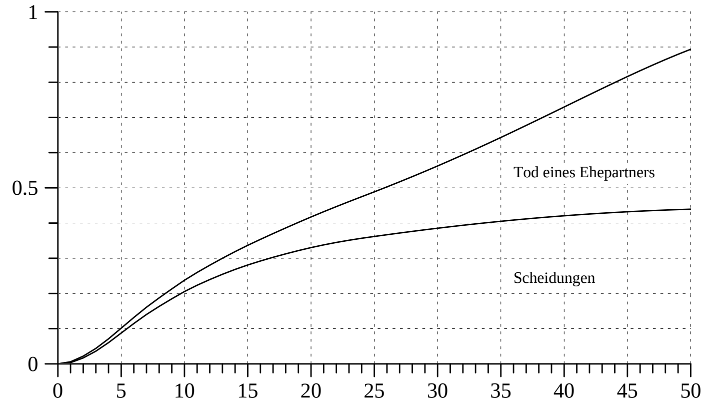


Abb. 14-7 Anteilsfunktionen H^a (unterer Linie) und $H^a + H^b$ (obere Linie) für die Daten in Tabelle 14-4.

Noch deutlicher werden die Diskrepanzen, wenn man sich für diejenigen Ehen interessiert, die nicht durch eine Scheidung sondern durch den Tod eines Ehepartners enden. Abb. 14-6 zeigt wiederum die Anteilsfunktion H^b und die Pseudo-Survivorfunktion $1 - \tilde{G}^b$.

Da Anteilsfunktionen im Unterschied zu Pseudo-Survivorfunktionen additiv sind, können sie sinnvoll zusammengefasst werden. Das wird in Abb. 14-7 illustriert. Die Abbildung zeigt, wie sich in Abhängigkeit von der Ehedauer die Gesamtheit der jeweils bisherigen Ehelösungen auf Scheidungen und den Tod eines Ehepartners verteilt.

Kapitel 15

Statistische Bedingungsanalysen

In diesem Kapitel beschäftigen wir uns mit der Frage, wie statistische Begriffsbildungen und Methoden zur Ermittlung von Bedingungsverhältnissen verwendet werden können. Der erste Abschnitt erklärt zunächst den grundlegenden Begriff einer bedingten Verteilung und daran anschließend Regressionsfunktionen. Im zweiten Abschnitt werden Regressionsmodelle zur Darstellung von Regressionsfunktionen besprochen und durch einfache Beispiele illustriert. (Komplexere Varianten und Erweiterungen dieser einfachen Beispiele werden in späteren Kapiteln zur Datenanalyse verwendet.) Schließlich wird im dritten Abschnitt die Frage aufgegriffen, ob bzw. in welcher Bedeutung auch statistische Strukturen als Bedingungen (sozialer Prozesse) betrachtet werden können.

15.1 Bedingte Verteilungen

1. Definition bedingter Verteilungen. Grundlegend für alle analytischen Verwendungen statistischer Methoden ist der Begriff einer *bedingten Verteilung*. Wir verwenden allgemein die Notation

$$\mathrm{P}[\text{Variablen} \,|\, \text{Bedingungen}]$$

womit die Verteilung derjenigen Variablen gemeint ist, die vor dem Bedingungsstrich | stehen, und zwar eingeschränkt auf diejenige Teilgesamtheit der Referenzmenge Ω, bei der die hinter dem Bedingungsstrich genannten Bedingungen erfüllt sind. Bedingte Verteilungen sind also, wie alle Häufigkeitsverteilungen, Funktionen; bestimmte numerische Werte (Häufigkeiten) erhält man erst, wenn man sich auf den Wert der Funktion bei einem bestimmten Argument bezieht (das in runden Klammern an den Funktionsnamen angehängt wird).

Zur Illustration beziehen wir uns auf die im Abschnitt 2.1 verwendete zweidimensionale Variable

$$(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$$

bei der X das Geschlecht und Y das Alter von 10 Personen erfasst (die Daten findet man in Tabelle 2.1-1). Einige Beispiele für bedingte Verteilungen, die in diesem Fall gebildet werden können, sind die folgenden:

– $\mathrm{P}[Y|X=0]$. Dies ist die Verteilung der Variablen Y bei derjenigen Teilgesamtheit, bei der die Variable X den Wert 0 hat; es handelt sich also um die Altersverteilung der Männer. Wie gesagt, ist dies eine Funktion. Ein bestimmter Wert wäre etwa $\mathrm{P}[Y|X=0](Y \geq 25) = 0.5$, d.h. 50% der Männer sind mindestens 25 Jahre alt. In analoger Bedeutung ist $\mathrm{P}[Y|X=1]$ die Altersverteilung der Frauen.

– $\mathrm{P}[X|Y \geq 25]$ ist die Verteilung von X bei den Personen, die mindestens 25 Jahre alt sind. Also ist z.B. $\mathrm{P}[X|Y \geq 25](1) = 0.5$, d.h. 50 % der mindestens 25 Jahre alten Personen sind weiblich.

– $\mathrm{P}[Y|Y \geq 25, X=1]$ ist die Altersverteilung der Frauen, die mindestens 25 Jahre alt sind. Dieses Beispiel zeigt auch, dass Variablen gleichzeitig vor und hinter dem Bedingungsstrich verwendet werden können.

2. Statistische Regressionsrechnung. Wenn von statistischer Regressionsrechnung gesprochen wird, sind allgemein Methoden zur Darstellung bedingter Verteilungen gemeint. Ein Verständnis gewinnt man durch eine Unterscheidung allgemeiner und spezieller Regressionsfunktionen.[1] Ausgangspunkt ist eine zweidimensionale Variable $(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ (wobei sowohl X als auch Y mehrdimensional sein können). Dann wird eine

[1] Eine ausführliche Diskussion findet man bei Rohwer und Pötter (2001, Teil II).

der beiden Komponenten als *unabhängige Variable* (auch *Regressorvariable* genannt), die andere als *abhängige Variable* bestimmt; wir wählen für die folgende Darstellung X als unabhängige und Y als abhängige Variable. Schließlich kann eine *allgemeine Regressionsfunktion* definiert werden, die jedem möglichen Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ die bedingte Verteilung $\mathrm{P}[Y|X = \tilde{x}]$ zuordnet.

Der Zweck dieser Konstruktion besteht darin, dass man ermitteln möchte, wie die Verteilung der abhängigen Variablen durch jeweils spezielle Werte der unabhängigen Variablen bedingt wird. Allerdings sind allgemeine Regressionsfunktionen gewissermaßen Funktionen zweiter Ordnung, da es sich bei ihren Werten um bedingte Verteilungen, also wiederum um Funktionen handelt. Folgendes Bild veranschaulicht dies:

$$\underbrace{\tilde{x} \longrightarrow \overbrace{(\tilde{y} \longrightarrow \mathrm{P}[Y|X = \tilde{x}](\tilde{y}))}^{\text{bedingte Verteilung}}}_{\text{allgemeine Regressionsfunktion}}$$

Eine allgemeine Regressionsfunktion ordnet jedem möglichen Wert $\tilde{x}$ der unabhängigen Variablen X eine bedingte Verteilung zu, die jedoch selbst eine Funktion ist, die jedem möglichen Wert $\tilde{y}$ der abhängigen Variablen Y eine bedingte Häufigkeit $\mathrm{P}[Y|X = \tilde{x}](\tilde{y})$ zuordnet.

Somit entsteht die Frage, wie man Regressionsfunktionen darstellen kann. Die meisten Vorschläge folgen einem einfachen Grundgedanken, der darin besteht, als Werte einer Regressionsfunktion nicht bedingte Verteilungen im Sinne von Funktionen zu verwenden, sondern Zahlen, durch die die bedingten Verteilungen charakterisiert werden können. Regressionsfunktionen nehmen dann folgende Form an:

$$\tilde{x} \longrightarrow \text{Charakterisierung von } \mathrm{P}[Y|X = \tilde{x}]$$

Dadurch werden Regressionsfunktionen zu gewöhnlichen Funktionen; wir nennen sie *spezielle Regressionsfunktionen*. Durch eine spezielle Regressionsfunktion wird jedem Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ eine bestimmte Zahl zugeordnet, die die bedingte Verteilung $\mathrm{P}[Y|X = \tilde{x}]$ charakterisiert.

Offenbar können spezielle Regressionsfunktionen auf fast beliebig viele Weisen definiert werden. Hauptsächlich werden jedoch die folgenden drei Charakterisierungen verwendet:

– Bedingte Mittelwerte; spezielle Regressionsfunktionen sehen dann folgendermaßen aus: $\tilde{x} \longrightarrow \mathrm{M}(Y|X = \tilde{x})$. Jedem Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ wird der durch ihn bedingte Mittelwert der abhängigen Variablen zugeordnet.

– Bedingte Quantile; spezielle Regressionsfunktionen haben dann die Gestalt: $\tilde{x} \longrightarrow \mathrm{Q}_p(Y|X = \tilde{x})$. Jedem Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ werden ein oder mehrere durch ihn bedingte Quantile der Verteilung der abhängigen Variablen zugeordnet.[2]

– Bedingte Häufigkeiten; spezielle Regressionsfunktionen haben in diesem Fall die Form: $\tilde{x} \longrightarrow \mathrm{P}[Y|X = \tilde{x}](\tilde{y})$. Hier gibt es für jeden Merkmalswert $\tilde{y} \in \tilde{\mathcal{Y}}$ eine spezielle Regressionsfunktion, die jedem Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ die durch ihn bedingte Häufigkeit für den Merkmalswert $\tilde{y} \in \tilde{\mathcal{Y}}$ zuordnet.

3. Beispiel: Autofahrer an einer Ampel. Um die Verwendung von Regressionsfunktionen zu verdeutlichen, beginnen wir mit einem einfachen Beispiel: Autofahrer, die sich einer Straßenkreuzung nähern, an der es eine Ampel gibt. Angenommen, wir beobachten für einen gewissen Zeitraum die Straßenkreuzung. Die Beobachtungen können als Werte einer zweidimensionalen Variablen $(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ repräsentiert werden. Die Elemente von Ω sind die Situationen, in denen sich ein Autofahrer der Ampel nähert. X erfaßt, ob die Ampel rot (1) oder nicht rot (0) ist; Y erfaßt, ob das Auto anhält (1) oder nicht anhält (0). Insgesamt sind 100 Situationen beobachtet und folgende Werte festgestellt worden:

$\tilde{x}$	$\tilde{y}$	Anzahl
0	0	47
0	1	0
1	0	3
1	1	50

D.h. bei ‘rot’ haben 50 Autos angehalten, 3 sind jedoch weitergefahren; und bei ‘nicht rot’ sind die Autos ausnahmslos weitergefahren.

In diesem Beispiel sind die Daten so überschaubar, dass es nicht erforderlich ist, Regressionsfunktionen zu berechnen; trotzdem kann man es tun. Man muss sich dann zunächst entscheiden, welche der beiden Variablen als unabhängig und welche als abhängig betrachtet werden soll. Trotz der formalen Symmetrie liegt es in diesem Beispiel natürlich nahe, X als unabhängige und Y als abhängige Variable zu betrachten. Eine allgemeine Regressionsfunktion ordnet dann jedem möglichen Wert von X (dem Ampelsignal) eine durch ihn bedingte Verteilung von Y (des Verhaltens der Autofahrer) zu. Stattdessen kann man auch spezielle Regressionsfunktionen bilden, wobei es sich in diesem Beispiel anbietet, Anteilswerte zu verwenden, z.B. den Anteil der Autofahrer, die ihr Auto anhalten.

Bereits an diesem einfachen Beispiel kann man sich auch verdeutlichen, dass Regressionsfunktionen keine Abbildungen in den Merkmalsraum der abhängigen Variablen sind; sie unterscheiden sich also von Funktionen der Form $g : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$. Bei diesem Schema liefert die Funktion g für jeden

[2] Sei Y irgendeine statistische Variable mit der Verteilungsfunktion $F_{[Y]}$. Dann ist das durch $\mathrm{Q}_p(Y)$ bezeichnete p-Quantil der Verteilung von Y eine Zahl, für die näherungsweise gilt: $F_{[Y]}(Q_p(Y)) \approx p$. Dementsprechend ist $\mathrm{Q}_p(Y|X = \tilde{x})$ das p-Quantil der durch $X = \tilde{x}$ bedingten Verteilung von Y.

Tabelle 15.1-1 Ausgaben privater Haushalte für Nahrungsmittel (einschl. Getränke und Tabakwaren), differenziert nach dem klassifizierten monatlichen Haushaltsnettoeinkommen in DM. Angaben nach der Einkommens- und Verbrauchsstichprobe 1998. Quelle: Statistisches Jahrbuch 2001: 573.

Einkommensklasse von ... bis unter	Ausgabefähiges Einkommen	Ausgaben für Nahrungsmittel in DM	in %
unter 1800	1383	269	19.45
1800 – 2500	2196	341	15.53
2500 – 3000	2788	391	14.02
3000 – 4000	3543	473	13.35
4000 – 5000	4566	584	12.79
5000 – 7000	6057	677	11.18
7000 – 10000	8422	775	9.20
10000 – 35000	13843	894	6.46

Wert $\tilde{x} \in \tilde{\mathcal{X}}$ einen Wert $g(\tilde{x})$, der ein Element des Merkmalsraums $\tilde{\mathcal{Y}}$ ist. Offenbar gilt dies nicht für allgemeine Regressionsfunktionen, deren Werte bedingte Verteilungen sind. Aber auch die Werte spezieller Regressionsfunktionen können im Allgemeinen nicht sinnvoll als Werte des Merkmalsraums der abhängigen Variablen interpretiert werden. Das ist wiederum unmittelbar deutlich, wenn (wie in unserem Beispiel) die bedingten Verteilungen durch Häufigkeiten (Anteilswerte) charakterisiert werden. Ein irreführender Eindruck kann höchstens bei der Mittelwertregression entstehen, bei der jedem Wert der unabhängigen Variablen der Mittelwert der durch ihn bedingten Verteilung der abhängigen Variablen zugeordnet wird. Aber auch wenn es sich dabei um einen bestimmten Wert im Merkmalsraum der abhängigen Variablen handelt (was nicht unbedingt der Fall sein muss), muss ein Mittelwert in seiner Bedeutung (nämlich als Charakterisierung einer statistischen Verteilung) von einem individuell zurechenbaren Merkmalswert unterschieden werden.

4. Beispiel: Ausgaben privater Haushalte. Jetzt betrachten wir ein anderes Beispiel, in dem es sowohl für die unabhängige als auch für die abhängige Variable einen quantitativen Merkmalsraum gibt. Die Variable (X, Y) bezieht sich in diesem Beispiel auf eine Gesamtheit privater Haushalte; die unabhängige Variable X erfasst das in einem bestimmten Monat für Ausgaben verfügbare Einkommen, und die abhängige Variable Y erfasst die in diesem Monat getätigten Ausgaben für Nahrungsmittel (einschließlich Getränke und Tabakwaren). Da es sich um eine quantitative abhängige Variable handelt, bietet sich eine Mittelwertregression $\tilde{x} \longrightarrow \mathrm{M}(Y \mid X = \tilde{x})$ an, durch die jedem möglichen Wert $\tilde{x}$ der unabhängigen Variablen der Mittelwert der Nahrungsausgaben der Haushalte mit dem verfügbaren Einkommen $\tilde{x}$ zugeordnet wird.

Für die Berechnung dieser Regressionsfunktion wären allerdings Individualdaten erforderlich, die uns nicht zur Verfügung stehen. Stattdes-

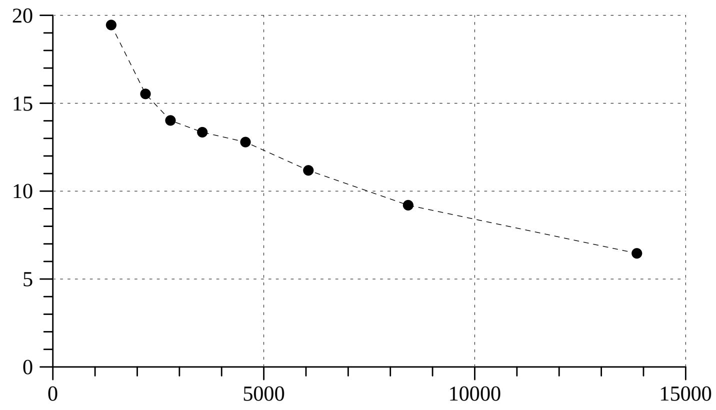


Abb. 15.1-1 Anteile der Ausgaben für Nahrungsmittel (y-Achse: in %) als Funktion des für Ausgaben verfügbaren Haushaltseinkommens (x-Achse: in DM). Daten aus der EVS 1998 (Tabelle 15.1-1).

sen verwenden wir die in Tabelle 15.1-1 angegebenen Daten, die aus der Einkommens- und Verbrauchsstichprobe (EVS) des Jahres 1998 stammen. Die Haushalte wurden vom Statistischen Bundesamt nach ihrem monatlichen Haushaltsnettoeinkommen den in der Tabelle angegebenen Einkommensklassen zugeordnet, dann wurden für jede Einkommensklasse das durchschnittliche für Ausgaben verfügbare Einkommen und der Durchschnittswert der Ausgaben für Nahrungsmittel berechnet. Somit wird durch diese Rechnungen bereits eine Mittelwertregression durchgeführt: den durchschnittlichen verfügbaren Einkommen werden durchschnittliche Ausgaben für Nahrungsmittel zugeordnet. Natürlich kann man stattdessen auch die durchschnittlichen Anteile der Ausgaben für Nahrungsmittel als eine Funktion des verfügbaren Einkommens betrachten; Abb. 15.1-1 zeigt dies in Form einer graphischen Darstellung. Durch die Daten der Tabelle 15.1-1 sind zwar nur die als Punkte eingezeichneten Funktionswerte begründbar. (Man kann vermuten, dass eine Regressionsrechnung mit den zugrundeliegenden Individualdaten einen zur gestrichelt eingezeichneten Linie sehr ähnlichen Funktionsverlauf liefern würde.)

5. Statistische und substantielle Bedingungen. Schließlich stellt sich die Frage, ob und ggf. wie die bei der Bildung bedingter Verteilungen formal zur Konditionierung verwendeten Variablen auch in irgendeinem substantiellen Sinn als Bedindungen interpretiert werden können. Man sollte sich klarmachen, dass der begriffliche Rahmen der Regressionsrechnung dies weder voraussetzt noch impliziert. Denn bei der Bildung einer bedingten Verteilung $\mathrm{P}[Y \mid \text{Bedingung}]$ bezieht sich die hinter dem Bedingungsstrich angeführte Bedingung nicht auf einen Prozess, sondern spezifiziert nur eine Referenzmenge, nämlich diejenige Teilmenge der Objektmenge Ω, für die die Verteilung von Y ausgewiesen werden soll. Um dagegen bei einer Regressionsrechnung Werte der unabhängigen Variablen als effektive Bedingungen interpretieren zu können, wäre es erforderlich, zunächst einen substantiellen Prozess zu konzipieren, durch den Werte der abhängigen Variablen entstehen – um dann darlegen zu können, wie dieser substantielle Prozess von Werten der unabhängigen Variablen abhängt.

Allerdings setzt die Regressionsrechnung nicht einmal die Existenz eines solchen substantiellen Prozesses voraus, was man schon daraus erkennt, dass unabhängige und abhängige Variablen beliebig vertauscht werden können. Insbesondere liefert die Regressionsrechnung keine begrifflichen Hilfsmittel zur Repräsentation substantieller Prozesse. Sofern man sich gedanklich auf solche Prozesse beziehen möchte, müssen sie vielmehr jenseits der statistischen Begriffsbildungen und Rechnungen vorstellbar gemacht werden; oder anders formuliert: die Ergebnisse einer Regressionsrechnung müssen im Hinblick auf einen irgendwie vorstellbar gemachten substantiellen Prozess „interpretiert“ werden.

6. Bezugnahmen auf substantielle Prozesse. Betrachten wir das oben angeführte Beispiel, in dem sich Autofahrer einer Ampel nähern. Ob die Autos anhalten oder nicht, wird offenbar nicht durch die Ampelsignale bewirkt, sondern hängt vom Verhalten der Autofahrer ab. Ein Ampelsignal bildet nur einen Aspekt einer Situation, in der sich ein Autofahrer so oder so verhält, und tatsächlich schränkt es nicht einmal seine effektiven Handlungsmöglichkeiten ein. Um zu erklären, wie Ampelsignale gleichwohl als Bedingungen verstanden werden können, muss man sich also in irgendeiner Weise darauf beziehen, wie sie von Autofahrern wahrgenommen und verarbeitet werden.

Ebenso muss in dem zweiten Beispiel auf menschliche Tätigkeiten Bezug genommen werden, um zu verstehen, wie Ausgaben für Nahrungsmittel zustande kommen und u.a. vom jeweils verfügbaren Haushaltseinkommen abhängen. Entsprechend gilt bei vielen Anwendungen der Regressionsrechnung in der Sozialstrukturforschung, dass bei der Konzeption substantieller Prozesse auf Akteure Bezug genommen werden muss. Nicht immer, aber in diesen Fällen kann man sich an einem Schema der folgenden Art orientieren:

$$\begin{array}{ccc} \boxed{\text{Substantielle Prozesse (durch Akteure)}} & \rightarrow & \tilde{y} \\ \uparrow & & \\ \tilde{x} & & \end{array} \tag{15.1}$$

An den Prozessen, durch die bestimmte Werte ($\tilde{y}$) der abhängigen Variablen Y entstehen, sind Akteure beteiligt; und es stellt sich dann die Frage, ob und ggf. wie bestimmte Werte ($\tilde{x}$) der unabhängigen Variablen X als Bedingungen für Verhaltensweisen und Tätigkeiten der Akteure, soweit diese am Zustandekommen von Werten der abhängigen Variablen beteiligt sind, interpretiert werden können.

Somit stellt sich auch die Frage, ob es möglich ist, an dieser Stelle mehr zu erreichen als eine weitgehend unverbindliche nachträgliche Interpretation der durch eine Regressionsfunktion ermittelten Zusammenhänge. Zu überlegen wäre, ob und ggf. wie für die substantiellen Prozesse, auf die sich eine Regressionsrechnung immer nur *implizit* beziehen kann, durch substantielle (nicht statistische) Modelle eine *explizite* Repräsentation gefunden werden kann.[3] Wir werden uns mit dieser Idee später ausführlicher beschäftigen.[4]

15.2 Statistische Regressionsmodelle

1. Regressionsmodelle. Regressionsmodelle dienen zur Berechnung und Darstellung spezieller Regressionsfunktionen. Wir unterscheiden zwei Konzeptionen:

– *Statistische Regressionsmodelle* beziehen sich auf Regressionsfunktionen, die (wie im vorangegangenen Abschnitt) für statistische Variablen definiert werden können.

– *Stochastische Regressionsmodelle* beziehen sich auf Regressionsfunktionen, die gedanklich auf Zufallsvariablen (reale oder fiktive Zufallsgeneratoren) bezogen werden.

Da Zufallsvariablen bzw. die ihnen korrespondierenden Wahrscheinlichkeitsverteilungen unterschiedlich verstanden und in unterschiedlichen Kontexten verwendet werden können, gibt es auch unterschiedliche Varianten und Deutungen stochastischer Regressionsmodelle. Darauf wird später genauer eingegangen. Zunächst beschäftigen wir uns nur mit statistischen Regressionsmodellen.[5] Für ihre Verwendung als Hilfsmittel zur Darstellung spezieller Regressionsfunktionen gibt es hauptsächlich folgende Gründe. Sie werden verwendet,

– um zu einfacheren Darstellungen von Regressionsfunktionen zu gelangen;

– um Regressionsfunktionen auch mit unvollständigen Daten berechnen zu können;

[3]Von einem *substantiellen Modell* wird in diesem Text gesprochen, wenn das Modell eine explizite Repräsentation eines substantiellen Prozesses in der in Abschnitt 2.2 (§ 7) erläuterten Bedeutung intendiert. Im Vergleich zu einem allgemeinen Reden von „theoretischen Modellen“ ist also eine spezifischere Bedeutung gemeint.

[4]Wichtige Hinweise findet man auch in der Literatur, u.a. in den Bemerkungen von Raymond Boudon (1979) über „generating models“ und in Überlegungen von Peter Hedström und Richard Swedberg über „social mechanisms“ (1996).

[5]Sie sind auch stets gemeint, wenn in diesem Text ohne ohne nähere Bestimmung von Regressionsmodellen gesprochen wird; andernfalls wird ausdrücklich von *stochastischen* Regressionsmodellen gesprochen.

– um mithilfe von (ggf. unvollständigen) Daten theoretisch vermutete Regressionsfunktionen berechenbar und beurteilbar zu machen; und
– um Schätzwerte für eine abhängige Variable zu gewinnen, die bei bestimmten (beobachteten oder hypothetisch angenommenen) Werten einer unabhängigen Variablen vermutet werden können.

2. Parametrische Modellansätze. Es gibt unterschiedliche Ansätze zur Konstruktion von Regressionsmodellen. Oft werden *parametrische Regressionsmodelle* verwendet. Zur Erläuterung beziehen wir uns auf eine zweidimensionale statistische Variable $(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ und betrachten Y als abhängige, X als unabhängige Komponente. Eine spezielle Regressionsfunktion ist dann eine Funktion $g : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$, so dass $g(x)$ eine spezifische Charakterisierung der bedingten Verteilung $\mathrm{P}[Y|X = x]$ liefert (vgl. § 2 im vorangegangenen Abschnitt). Die Konstruktion eines parametrischen Regressionsmodells besteht nun darin, die Funktion g näherungsweise durch eine einfachere Modellfunktion $\tilde{g} : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$ darzustellen, die ihrerseits von gewissen Parametern abhängt. Wir schreiben dies in der Form $\tilde{g}(x; \theta)$, wobei vor dem Semikolon das Argument der Funktion steht und dahinter ein Verweis auf die Parameter der Funktion folgt.[6] Das Ziel besteht darin, einen bestimmten Parameterwert $\hat{\theta}$ zu finden, so dass gilt:

$$\text{für alle } x \in \tilde{\mathcal{X}}: \quad \tilde{g}(x; \hat{\theta}) \approx g(x)$$

Dann kann $\tilde{g}$ als eine Modellfunktion aufgefasst werden, die die ursprüngliche Regressionfunktion g mehr oder weniger gut repräsentiert.

Ein besonders einfaches Beispiel ist die lineare Mittelwertregression. In diesem Fall hat die Regressionsfunktion die Gestalt $g(x) := \mathrm{M}(Y|X = x)$, und man verwendet zur Darstellung die Modellfunktion

$$\tilde{g}(x; \alpha, \beta) \; := \; \alpha + x\beta$$

so dass der durch die Modellfunktion geschätzte durch x bedingte Mittelwert von Y zu einer linearen Funktion von x wird.

3. Berechnungsverfahren. Die Konstruktion eines parametrischen Regressionsmodells besteht somit aus folgenden Schritten:

a) Zuerst wird eine parametrische Modellfunktion $\tilde{g}(x; \theta)$ spezifiziert.[7]
b) Dann wird ein Verfahren festgelegt, mit dem aufgrund von Daten ein bestimmter Parameterwert $\hat{\theta}$ berechnet werden kann.

[6] Für die Notation der Parameter werden meistens griechische Buchstaben verwendet; beispielsweise θ, aber auch α, β und andere.

[7] Je nachdem handelt es sich um eine Funktionsform, wenn θ als eine logische Variable aufgefasst wird, oder um eine Funktion, wenn für θ irgendein bestimmer Wert angenommen wird.

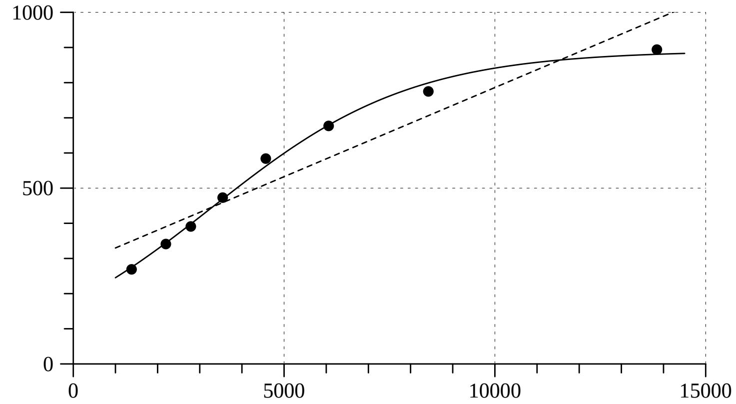


Abb. 15.1-2 Ausgaben für Nahrungsmittel (y-Achse: in DM) als Funktion des für Ausgaben verfügbaren Haushaltseinkommens (x-Achse: in DM). Daten aus der EVS 1998 (Tabelle 15.1-1). Ausderdem eingezeichnet sind die Modellfunktionen $\tilde{g}_1$ (gestrichelt) und $\tilde{g}_2$ (durchgezogene Linie).

c) Schließlich wird die Berechnung mit den vorhandenen Daten praktisch durchgeführt, so dass eine bestimmte Modellfunktion $\tilde{g}(x; \hat{\theta})$ entsteht.

Zur Berechnung von Parameterwerten werden hauptsächlich zwei Verfahren verwendet: die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode.[8] Im nächsten Paragraphen wird die erste dieser Methoden anhand eines Beispiels erläutert.

4. Modellfunktionen für Mittelwertregressionen. Die Methode der kleinsten Quadrate, auch LS-Methode genannt (wobei LS als Abkürzung für Least Squares dient), kann am besten anhand von Modellen für Mittelwertregressionen erläutert werden. Als Beispiel verwenden wir noch einmal die Ausgaben privater Haushalte für Nahrungsmittel (Daten in Tab. 15.1-1). Als unabhängige Variable (X) wird das ausgabefähige Einkommen, als abhängige Variable werden die Ausgaben für Nahrungsmittel betrachtet.

Wir beginnen mit einer linearen Modellfunktion $\tilde{g}(x; \alpha, \beta) := \alpha + x\beta$. Um bestimmte Parameterwerte mit der LS-Methode zu berechnen, wird die Zielfunktion

$$f_{LS}(\alpha, \beta) \; := \; \sum_i \left(y_i - \tilde{g}_1(x_i; \alpha, \beta)\right)^2$$

minimiert, wobei sich der Index i auf die zu verwendenden Daten bezieht. Die Parameterwerte, die diese Zielfunktion, das sog. LS-Kriterium, minimieren, werden durch $\hat{\alpha}$ und $\hat{\beta}$ bezeichnet. Verwendet man die Daten aus Tabelle 15.1-1, erhält man die Parameterwerte $\hat{\alpha} = 279.10$ und $\hat{\beta} = 0.0507$,

[8] Eine ausführliche Diskussion findet man bei Rohwer und Pötter (2001: Kap. 9).

also die Modellfunktion

$$\tilde{g}_1(x) \;=\; 279.10 + 0.0507\,x$$

Für jedes mögliche Haushaltseinkommen x liefert $\tilde{g}_1(x)$ einen Schätzwert für den entsprechenden bedingten Mittelwert $\mathrm{M}(Y|X=x)$ der abhängigen Variablen, in diesem Fall der Ausgaben für Nahrungsmittel. Abb. 15.1-2 zeigt den Verlauf der Modellfunktion als gestrichelte Linie.

Diese lineare Funktion eignet sich offensichtlich nicht besonders gut, um die Daten zu repräsentieren. Eine bessere Modellfunktion sollte ausdrücken können, dass – wie durch die Daten nahegelegt wird – der Anteil der Ausgaben mit steigendem Einkommen kleiner wird. Als ein Beispiel betrachten wir die Modellfunktion

$$\tilde{g}_2(x;\alpha,\beta,\gamma) \;:=\; \gamma\,\frac{\exp(\alpha + x\beta)}{1+\exp(\alpha + x\beta)}$$

in der die Logit-Funktion verwendet wird.[9] Bestimmte Werte für die Paramter α, β und γ können wieder mit der LS-Methode berechnet werden. Mit den Daten aus Tab. 15.1-1 erhält man die Werte $\hat{\alpha} = -1.3900$, $\hat{\beta} = 0.000422$ und $\hat{\gamma} = 890.926$. berechnet werden.[10] Abb. 15.1-2 zeigt die diesen Parameterwerten entsprechende Modellfunktion als durchgezogene Linie. Offenbar erlaubt sie eine bessere Darstellung der Daten.

15.3 Statistische Strukturen als Bedingungen?

In diesem Abschnitt gehen wir etwas genauer auf die bereits in Abschnitt 2.2 erwähnte Frage ein: ob bzw. in welcher Weise auch statistische Strukturen als Bedindungen für soziale Prozesse verstanden werden können.

1. Erläuterungen zur Fragestellung. Die Regressionsrechnung kann zeigen, wie statistische Verteilungen (Strukturen) von Werten statistischer Variablen abhängig sind. Eine andere Frage ist, ob und ggf. in welcher Weise statistische Strukturen selbst als Bedingungen für soziale Prozesse aufgefasst werden können. Dass mit einem allgemeinen Reden von Strukturen oft Vorstellungen dieser Art verbunden werden, braucht wohl kaum ausführliche Belege. Allerdings sollte man darauf achten, wie in diesem Zusammenhang sinnvoll von Bedingungen gesprochen werden kann. Zur Verdeutlichung des Problems können folgende Ausführungen von William H. Sewell (1992: 1f.) dienen:

[9]Die Definition der *Logit-Funktion* ist

$$\mathrm{logit}(u) := \frac{\exp(u)}{1+\exp(u)}$$

wobei u ein beliebiges Argument ist.

[10]Die Berechnungen wurden mit der TDA-Prozedur `freg` durchgeführt.

„“Structure” is one of the most important and most elusive terms in the vocabulary of current social science. [...] The term structure empowers what it designates. Structure, in its nominative sense, always implies structure in its transitive verbal sense. Whatever aspect of social life we designate as structure is posited as “structuring” some other aspect of social existence – whether it is class that structures politics, gender that structures employment opportunities, rhetorical conventions that structure texts or utterances, or modes of production that structure social formations.“

In gewisser Weise scheint Sewell in diesen Ausführungen betonen zu wollen, dass er Strukturen für wichtige Bedingungen des gesellschaftlichen Lebens hält. Seine Formulierungen legen allerdings die Vorstellung nahe, dass Strukturen nicht Bedingungen, sondern Kräfte sind, die bestimmte Wirkungen ausüben können. Dieses mögliche Missverständnis sollte von vornherein vermieden werden. Sicherlich sind statistische Strukturen (Verteilungen) keine Kräfte, die irgendetwas bewirken könnten (und wir werden später sehen, dass sich auch relational definierte Strukturen nicht als Kräfte interpretieren lassen).[11]

Es ist nicht nur wichtig, Bedingungen und Ursachen zu unterscheiden, sondern auch darauf zu achten, *wofür* irgendwelche Sachverhalte als Bedingungen verstanden werden sollen. Im gegenwärtigen Kontext geht es zunächst darum, ob statistische Strukturen zumindest in einigen Fällen als Darstellungen von *Handlungsbedingungen* von Akteuren verstanden werden können. Anhand von Beispielen kann man sich dann auch den Unterschied zu Determinanten menschlichen Verhaltens verdeutlichen.

2. Reflexion einer möglichen Verwechslung. Unsere Fragestellung bezieht sich darauf, ob und ggf. wie statistische Strukturen, also *Verteilungen*, als Handlungsbedingungen interpretiert werden können. Davon ist zu unterscheiden, dass es sicherlich möglich ist, mithilfe statistischer Variablen Handlungsbedingungen zu erfassen. Ein einfaches Beispiel kann das verdeutlichen. Als statistische Gesamtheit Ω werden Haushalte betrachtet, so dass mit einer statistischen Variablen X erfasst werden kann, ob es in den Haushalten eine Waschmaschine gibt oder nicht:

$$X(\omega) \;:=\; \begin{cases} 1 & \text{wenn es im Haushalt } \omega \text{ eine Waschmaschine gibt} \\ 0 & \text{andernfalls} \end{cases}$$

Offenbar bezieht sich diese Variable auf Handlungsbedingungen der Menschen, die in den Haushalten leben. Andererseits ist jedoch klar, dass *die*

[11]Dass es bei Sewell dieses Missverständnis gibt, zeigt sich nicht nur in der oben zitierten Passage, sondern auch in einer seiner Fragestellungen: „The most fundamental problem is that structural or structuralist arguments tend to assume a far too rigid causal determinism in social life.“ Offenbar unterstellt Sewell, dass es zumindest in einem gewissen Ausmaß einen solchen „kausalen Determinismus“ geben könnte. (Man erinnere sich auch an die in Abschnitt 2.2 (§ 4) zitierte Bemerkung von Blau, in der er davon spricht, dass Makrostrukturen einen „impact on social life“ haben können.)

Verteilung der Variablen X *keine* Handlungsbedingung ist. Ob in einem Haushalt eine Waschmaschine genutzt werden kann, hängt (in diesem statistisch konstruierten Beispiel) nur davon ab, ob es *in diesem* Haushalt eine Waschmaschine gibt, nicht jedoch von dem Anteil der Haushalte mit einer Waschmaschine in der Gesamtheit Ω.

Ähnlich verhält es sich in vielen anderen Fällen, in denen sich eine statistische Variable auf Handlungsbedingungen von Menschen bezieht. Die Darstellung einer statistischen Struktur liefert dann Informationen über die Verteilung von Handlungsbedingungen in einer Gesellschaft. Zugleich ist jedoch klar, dass in diesen Fällen die statistische Struktur selbst keine Handlungsbedingung ist. Zu überlegen bleibt nur, ob es auch Fälle gibt, in denen statistische Verteilungen selbst als „strukturelle Bedingungen" für Handlungsmöglichkeiten interpretiert werden können.[12]

3. Wie man von strukturellen Bedingungen sprechen kann. Dies hängt auch von dem jeweils verwendeten Strukturbegriff ab. Hier beziehen wir uns auf einen statistischen Strukturbegriff, also auf die Verteilung einer statistischen Variablen $X : \Omega \longrightarrow \tilde{\mathcal{X}}$. Wie kann davon gesprochen werden, dass *die Verteilung* $\mathrm{P}[X]$ eine Bedingung für ... ist? Man kann drei Fälle unterscheiden:

– Zunächst kann man an irgendeinen bestimmten Akteur denken, nennen wir ihn A, und sich vorstellen, dass durch $\mathrm{P}[X]$ eine Handlungsbedingung dieses Akteurs charakterisiert wird. Dies setzt zwar nicht unbedingt voraus, dass der Akteur zur Referenzmenge der Variablen X gehört, wohl aber muss angenommen werden, dass sich $\mathrm{P}[X]$ auf eine Situation bezieht, in der sich A befindet oder befinden kann. Als Beispiel kann man sich vorstellen, dass $\mathrm{P}[X]$ die Verteilung von Mädchen und Jungen in einer Schulklasse angibt, die von A unterrichtet wird.

– Dies Beispiel führt auch zu einer zweiten Möglichkeit, von strukturellen Bedingungen zu sprechen. Denn man kann offenbar sagen, dass $\mathrm{P}[X]$ für jedes Mitglied der Schulklasse Ω eine Charakterisierung seiner Situation liefert. Zwar mag die Art und Weise, in der $\mathrm{P}[X]$ eine Handlungsbedingung darstellt, für jedes $\omega \in \Omega$ unterschiedlich sein, abhängig von ω's Eigenschaften (in diesem Beispiel u.a. von ω's Geschlecht); wichtig ist jedoch, dass man in diesem Fall von einer Handlungsbedingung für die *individuellen* Mitglieder von Ω sprechen kann.

– Eine dritte Möglichkeit entsteht, wenn man $\mathrm{P}[X]$ nicht mehr auf Situationen beziehen kann, in denen sich die Mitglieder von Ω befinden. Zur Illustration bleiben wir bei unserem Beispiel, nehmen jetzt aber an, dass Ω aus zwei Schulklassen besteht, aus Ω_1, in der es nur Mädchen gibt, und aus Ω_2, in der es nur Jungen gibt. Offenbar liefert in diesem Beispiel $\mathrm{P}[X]$, also das Verhältnis von Mädchen und Jungen in der Gesamtheit $\Omega = \Omega_1 \cup \Omega_2$, keine informative Charakterisierung der jeweiligen Handlungsbedingungen in den beiden Schulklassen.

[12] Dabei soll sich das Reden von Bedingungen für Handlungsmöglichkeiten normalerweise auf die Mitglieder der jeweils betrachteten statistischen Gesamtheit beziehen; im Unterschied zu externen Akteuren, etwa Unternehmen oder staatlichen Einrichtungen, die sich in ihren Aktivitäten auf statistische Gesamtheiten beziehen können und insofern an Strukturaussagen über diese Gesamtheiten interessiert sein können (z.B. an Informationen über die Ausstattung von Haushalten mit Waschmaschinen, um entsprechende Werbekampagnen zu planen).

4. Das Problem der Mikro-Relevanz statistischer Strukturen. Die eben angeführten Möglichkeiten orientierten sich an der Frage, ob und ggf. wie statistische Strukturen als Handlungsbedingungen von Akteuren verstanden werden können. Allgemeiner kann man sich darauf beziehen, ob und ggf. wie statistische Strukturen als Bedingungen dafür verstanden werden können, wie sich irgendwelche Objekte (Elemente einer Referenzmenge Ω) verhalten oder was mit ihnen geschieht (wenn man z.B. an strukturelle Bedingungen für eine Verbreitung von Krankheiten durch Infektionen denkt). Um eine solche Annahme zu begründen, muss man zeigen können, dass und wie die statistische Struktur für jedes individuelle Objekt (aus Ω) eine Bedingung *für sein* Verhalten ist. – Ich nenne dies das *Problem der Mikro-Relevanz statistischer Strukturen*; die folgende schematische Darstellung kann zur Verdeutlichung dienen:

$$\mathrm{P}[X] \begin{array}{l} \nearrow \\ \\ \searrow \end{array} \begin{array}{l} \omega_1 \dashrightarrow Y(\omega_1) \\ \vdots \\ \omega_n \dashrightarrow Y(\omega_n) \end{array} \tag{15.2}$$

Das Problem betrifft die Pfeile, die von der statistischen Struktur $\mathrm{P}[X]$ zu den Objekten $\omega_1, \ldots, \omega_n$ (bzw. zu den Situationen, in denen sie sich befinden) führen. Um zu zeigen, dass $\mathrm{P}[X]$ eine Bedingung für die individuellen Objekte ist, muss diesen Pfeilen eine bestimmte Bedeutung gegeben werden. Dabei ist natürlich die in § 2 besprochene Unterscheidung wichtig, auf die hier deshalb noch einmal verwiesen werden soll:

a) Der Wert, den eine abhängige Variable Y bei einem Objekt ω annimmt, hängt auch *von einem Wert* $X(\omega)$ einer unabhängigen Variable X ab. Diese Vermutung entspricht dem üblichen Ansatz einer Regressionsfunktion durch die Formulierung $x \longrightarrow \mathrm{P}[Y \mid X = x]$.

b) Der Wert, den eine abhängige Variable Y bei einem Objekt ω annimmt, hängt auch *von der Verteilung* einer unabhängigen Variable X ab. Damit eine solche Vermutung in einer Regressionsfunktion erfasst werden kann, muss allerdings in irgendeiner Weise eine neue statistische Variable, etwa X^* konstruiert werden, so dass $X^*(\omega)$ angibt, wie die Verteilung von X *in einer Umgebung von* ω beschaffen ist.

Man kann es auch so sagen: Im ersten Fall (a) hängt der bei einem Objekt ω realisierte Wert von Y davon ab, welche Position ω innerhalb der statistischen Struktur $\mathrm{P}[X]$ einnimmt; im zweiten Fall (b) hängt jedoch $Y(\omega)$

von der Beschaffenheit dieser statistischen Struktur selbst ab (soweit durch sie eine reale Umgebung von ω charakterisiert wird). Offenbar schließen sich beide Möglichkeiten nicht aus.

Literatur

Agresti, A., Agresti, B. F. 1977. Statistical Analysis of Qualitative Variation. In: K. F. Schuessler (ed.), Sociological Methodology 1978, 204–237. San Francisco: Jossey-Bass.

Anton, H., Rorres, C. 1991. Elementary Linear Algebra. Applications Version. New York: Wiley.

Bahrdt, H. P. 1994. Schlüsselbegriffe der Soziologie. 6. Aufl. München: Beck.

Balzer, W. 1997. Die Wissenschaft und ihre Methoden. Grundsätze der Wissenschaftstheorie. München: Alber.

Bates, F. L., Peacock, W, G. 1989. Conceptualizing Social Structure: The Misuse of Classification in Structural Modeling. American Sociological Review 54, 565–577.

Baumert, J., Cortina, K. S., Leschinsky, A. 2005. Grundlegende Entwicklungen und Strukturprobleme im allgemein bildenden Schulwesen. In: K. Cortina u.a. (Hg.), Das Bildungswesen in der Bundesrepublik Deutschland. 2. Aufl., 52–147. Reinbek: Rowohlt Taschenbuch Verlag.

Bien, W., Marbach, J. 1991. Haushalt – Verwandtschaft – Beziehungen: Familienleben als Netzwerk, In: H. Bertram (Hg.), Die Familie in Westdeutschland, 3–44. Opladen: Leske+Budrich.

Bien, W., Marbach, J., Templeton, R. 1992. Social Networks of Single-person Households. In: C. Marsh, S. Arber (eds.), Families and Households, 157–173. London: Macmillan.

Blau, P. M. 1974. Parameters of Social Structure. American Sociological Review 39, 615–635.

Blau, P. M. 1977. A Macrosociological Theory of Social Structure. American Journal of Sociology 83, 26–54.

Blau, P. M. 1994a. Structural Contexts of Opportunities. Chicago: University of Chicago Press.

Blaut, J. M. 1971. Space, Structure and Maps. Tijdschrift voor Economische en Sociale Geografie 62, 18–21.

Blossfeld, H.-P., Hamerle, A., Mayer, K. U. 1986. Ereignisanalyse. Statistische Theorie und Anwendungen in den Sozialwissenschaften. Frankfurt: Campus.

Blossfeld, H.-P., Huinink, J. 1989. Die Verbesserung der Bildungs- und Berufschancen von Frauen und ihr Einfluß auf den Prozeß der Familienbildung. Zeitschrift für Bevölkerungswissenschaft 15, 383 – 404.

Blossfeld, H.-P., Rohwer, G. 2002. Techniques of Event History Modeling (2nd ed.). Mahwah: Lawrence Erlbaum.

Böckenförde, E.-W. 1972. Die Bedeutung der Unterscheidung von Staat und Gesellschaft im demokratischen Sozialstaat der Gegenwart. In: Ders. (Hg.), Staat und Gesellschaft, 395–431. Darmstadt: Wissenschaftliche Buchgesellschaft 1976.

Bodzenta, E. 1979. Sozialstruktur im Wandel – am Beispiel Österreichs. In: K. Salamun (Hg.), Sozialphilosophie als Aufklärung, 379–398. Tübingen: Mohr.

Bossel, H. 1992. Modellbildung und Simulation. Brauschweig: Vieweg.

Boudon, R. 1979. Generating Models as a Research Strategy. In: R. K. Merton, J. S. Coleman, P. H. Rossi (eds.), Qualitative and Quantitative Social Research, 51–64. New York: Free Press.

Boudon, R., Bourricaud, F. 1992. Soziologische Stichworte. Opladen: Westdeutscher Verlag.

Bretz, M., Niemeyer, F. 1992. Private Haushalte gestern und heute. Ein Rückblick auf die vergangenen 150 Jahre. Wirtschaft und Statistik 40, 73–81.

Busche, H. 1998. Teleologie; teleologisch. Historisches Wörterbuch der Philosophie, Bd. 10, 970–977. Darmstadt: Wissenschaftliche Buchgesellschaft.

Büschges, G., Abraham, M., Funk, W. 1998. Grundzüge der Soziologie. 3. Aufl. München: Oldenbourg.

Cantor, G. 1962. Gesammelte Abhandlungen mathematischen und philosophischen Inhalts. Hrsg. von E. Zermelo. Hildesheim: Georg Olms.

Danto, A. C. 1965. Analytical Philosophy of History. Dt.: Analytische Philosophie der Geschichte. Frankfurt: Suhrkamp 1980

Daston, L. 1998. Die Kultur der wissenschaftlichen Objektivität. In: O. G. Oexle (Hg.), Naturwissenschaft, Geisteswissenschaft, Kulturwissenschaft: Einheit – Gegensatz – Komplementarität? 9–39. Göttingen: Wallstein.

Diekmann, A., Klein, T. 1991. Bestimmungsgründe des Ehescheidungsrisikos. Kölner Zeitschrift für Soziologie und Sozialpsychologie 43, 271–290.

Dinkel, R. H. 1984. Sterblichkeit in Perioden- und Kohortenbetrachtung. Zeitschrift für Bevölkerungswissenschaft 10, 477–500.

Dinkel, R. H. 1992. Kohortensterbetafeln für die Geburtsjahrgänge ab 1900 bis 1962 in den beiden Teilen Deutschlands. Zeitschrift für Bevölkerungswissenschaft 18, 96–116.

Duncan, G. J., Hill, M. S. 1985. Conceptions of Longitudinal Households. Fertile or Futile? Journal of Economic and Social Measurement 13, 361–375.

Durkheim, E. 1888. Einführung in die Sozialwissenschaft an der Universität Bordeaux 1887-1888. In: Frühe Schriften zur Begründung der Sozialwissenschaft (hrsg. u. übers. v. L. Heisterberg), 25–52. Darmstadt: Luchterhand 1981.

Edinger, K.-E. 1998. Sozialstruktur, soziale Ungleichheit, soziale Schichtung. Paderborn: Schöningh.

Elias, N. 1996. Was ist Soziologie? 9. Aufl. Weinheim: Juventa.

Esenwein-Rothe, I. 1982. Einführung in die Demographie. Wiesbaden: Franz Steiner.

Esser, H. 1993. Soziologie. Allgemeine Grundlagen. Frankfurt: Campus.

Faber, K.-G. 1971. Theorie der Geschichtswissenschaft. München: Beck.

Freeman, L. C. 1992. The Sociological Concept of „Group": An Empirical Test of Two Models. American Journal of Sociology 98, 152–166.

Frey, G. 1961. Symbolische und Ikonische Modelle. In: H. Freudenthal (ed.), The Concept and the Role of the Model in Mathematics and Natural and Social Sciences, 89 – 97. Dordrecht: Reidel.

Friedrichs, J. 1977. Stadtanalyse. Soziale und räumliche Organisation der Gesellschaft. Reinbek: Rowohlt.

Fürstenberg, F. 1966. „Sozialstruktur" als Schlüsselbegriff der Gesellschaftsanalyse. Kölner Zeitschrift für Soziologie und Sozialpsychologie 18, 439–453.

Geiger, T. 1931. Gesellschaft. In: A. Vierkandt (Hg.), Handwörterbuch der Soziologie, 201–211. Stuttgart: Enke.

Geißler, C. 1997. Netzwerke als soziale Infrastruktur – Generationen und Haushalte im Leistungsverbund. In: U. Meier (Hg.), Vom Oikos zum modernen Dienstleistungshaushalt, 163–183. Frankfurt: Campus.

Glenn, N. D. 1977. Cohort Analysis. Beverly Hills: Sage.

Grusky, D. B. (ed.) 1994. Social Stratification. Class, Race, and Gender in Sociological Perspective. Boulder: Westview Press.

Hall, A. D., Fagen, R. E. 1956. Definition of System. General Systems 1, 18–28.

Handl, J. 1985. Mehr Chancengleichheit im Bildungssystem. Erfolg der Bildungsreform oder statistischer Artefakt? Kölner Zeitschrift für Soziologie und Sozialpsychologie 37, 698–722.

Hedström, P., Swedberg, R. 1996. Social Mechanisms. Acta Sociologica 39, 281–308.

Hendry, D. F., Richard, J.-F. 1982. On the Formulation of Empirical Models in Dynamic Econometrics. Journal of Econometrics 20, 3 – 33.

Hernes, G. 1976. Structural Change in Social Processes. American Journal of Sociology 82, 513–547.

Homans, G. C. 1951. The Human Group. London: Routledge & Kegan Paul.

Homans, G. C. 1961. Social Behavior. Its Elementary Forms. New York: Harcourt, Brace & World.

Homans, G. C. 1976. What Do We Mean by Social „Structure"? In: P. M. Blau (ed.), Approaches to the Study of Social Structure, 53–65. London: Open Books.

Hradil, S. 2000. Soziale Ungleichheit. In: G. Reinhold, S. Lamnek, H. Recker (Hg.), Soziologie-Lexikon, 589–593. München: Oldenbourg.

Huinink, J. 1987. Soziale Herkunft, Bildung und das Alter bei der Geburt des ersten Kindes. Zeitschrift für Soziologie 16, 367 – 384.

Huinink, J. 1988. Die demographische Analyse der Geburtenentwicklung mit Lebensverlaufsdaten. Allgemeines Statistisches Archiv 72, 359 – 377.

Huinink, J. 1989. Das zweite Kind. Sind wir auf dem Weg zur Ein-Kind-Familie? Zeitschrift für Soziologie 18, 192 – 207.

Huinink, J. 1998. Ledige Elternschaft junger Frauen und Männer in Ost und West. In: R. Metze, K. Mühler, K.-D. Opp (eds.), Der Transformationsprozess. Analysen und Befunde aus dem Leipziger Institut für Soziologie, 301 – 320. Leipzig: Leipziger Universitätsverlag.

International Statistical Institute 1986. Declaration on Professional Ethics. International Statistical Review 54, 227–242.

Jauss, H. R. 1973. Versuch einer Ehrenrettung des Ereignisbegriffs. In: R. Koselleck, W.-D. Stempel (Hg.), Geschichte – Ereignis und Erzählung, 554–560. München: Fink.

Joas, H. 2001. Die soziologische Perspektive. In: Ders. (Hg.), Lehrbuch der Soziologie, 11–38 Frankfurt: Campus.

Keyfitz, N. 1977. Applied Mathematical Demography, 2nd ed. Berlin: Springer-Verlag.

Kendall, M., Stuart, A. 1977. The Advanced Theory of Statistics, Vol. 1 (4th ed.). London: Charles Griffin & Comp.

Klein, T., Kopp, J. 1999. Scheidungsursachen aus soziologischer Sicht. Würzburg: Ergon.

KMK 2005. Das Bildungswesen in der Bundesrepublik Deutschland 2003. Hrsg. vom Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der BRD, bearbeitet von G. Jonen und T. Eckhardt.

König, R. 1958a. Grundformen der Gesellschaft: Die Gemeinde. Reinbek: Rowohlt Taschenbuch Verlag.

König, R. (Hg.) 1958b. Fischer-Lexikon Soziologie. Frankfurt: Fischer Bücherei.

Kottmann, P. 1987. Verrechtlichung und Bevölkerungsweisen im industriellen Deutschland. Historical Social Research, No. 41, 28–39.

Krackhardt, D. 1987. Cognitive Social Structures. Social Networks 9, 109–134.

Kreckel, R. 1992. Politische Soziologie der sozialen Ungleichheit. Frankfurt: Campus.

Krug, W., Nourney, M., Schmidt, J. 1999. Wirtschafts- und Sozialstatistik. Gewinnung von Daten. München: Oldenbourg.

Lambert, J. H. 1988. Texte zur Systematologie und zur Theorie der wissenschaftlichen Erkenntnis. Hrsg. von G. Siegwart. Hamburg: Meiner.

Laumann, E. O. 1966. Prestige and Association in an Urban Community. New York: Bobbs-Merrill.

Leff, G. 1969. History and Social Theory. London: Merlin Press.

Lepsius, M. R. 1976. Zum Verhältnis von Geschichtswissenschaft und Soziologie. In: H. M. Baumgartner, J. Rüsen (Hg.), Seminar: Geschichte und Theorie, 118–138. Frankfurt: Suhrkamp.

Leslie, P. H. 1945. On the Use of Matrices in Certain Population Mathematics. Biometrika 33, 183 – 212.

Lévi-Strauss, C. 1953. Social Structure. In: A. L. Kroeber (ed.), Anthropology Today, 524–553. Chicago: University of Chicago Press.

Linde, H. 1972. Sachdominanz in Sozialstrukturen. Tübingen: Mohr-Siebeck.

Lindenberg, S. 1977. Individuelle Effekte, kollektive Phänomene und das Problem der Transformation. In: K. Eichner, W. Habermehl (Hg.), Probleme der Erklärung sozialen Verhaltens, 46–84. Meisenheim: Hain.

Lindner, F. 1900. Die unehelichen Geburten als Sozialphänomen. Ein Beitrag zur Statistik der Bevölkerungsbewegung im Königreich Bayern. Leipzig: Deichert'sche Verlagsbuchhandlung.

Marsden, P. V., Laumann, E. O. 1984. Mathematical Ideas in Social Structural Analysis. Journal of Mathematical Sociology 10, 271–294.

Marsden, P. V., Lin, N. 1982. Introduction. In: Dies. (Hg.), Social Structure and Network Analysis, 9–11. Beverly Hills: Sage.

Meier, C. 1978. Fragen und Thesen zu einer Theorie historischer Prozesse. In: K.-G. Faber, C. Meier (Hg.), Historische Prozesse, 11–66. München: Deutscher Taschenbuch-Verlag.

Merton, R. K. 1957. Social Theory and Social Structure. Glencoe: Free Press.

Mueller, U. 1993. Bevölkerungsstatistik und Bevölkerungsdynamik. Berlin: de Gruyter.

Müller, W., Haun, D. 1994. Bildungsungleichheit im sozialen Wandel. Kölner Zeitschrift für Soziologie und Sozialpsychologie 46, 1–42.

Namboodiri, K., Suchindran, C. M. 1987. Life Table Techniques and their Applications. New York: Academic Press.

Nikles, B., Weiß, J. (Hg.) 1975. Gesellschaft. Organismus – Totaltlität – System. Hamburg: Hoffmann und Campe.

Oakeshott, M. 1983. On History and Other Essays. Oxford: Basil Blackwell.

Oaklander, L. N., Smith, Q. (eds.) 1994. The New Theory of Time. New Haven: Yale University Press.

Olkin, I., Gleser, L. J., Derman, C. 1980. Probability Models and Applications. New York: Macmillan Publ.

Opp, K.-D., Hummell, H. J. 1973. Soziales Verhalten und soziale Systeme. Frankfurt: Athenäum.

Oppitz, M. 1975. Notwendige Beziehungen. Abriß der strukturalen Anthropologie. Frankfurt: Suhrkamp.

Pappi, F. U. 1987. Die Netwerkanalyse aus soziologischer Perspektive. In: Ders. (Hg.), Methoden der Netzwerkanalyse, 11–37. München: Oldenbourg.

Popitz, H. 1980. Die normative Konstruktion von Gesellschaft. Tübingen: Mohr-Siebeck.

Popitz, H. 1995. Der Aufbruch zur Artifiziellen Gesellschaft. Zur Anthropologie der Technik. Tübingen: Mohr-Siebeck.

Prendergast, C., Knottnerus, J. D. 1994. Recent Developments in the Theory of Social Structure: Introduction and Overview. In:

J. D. Knottnerus, C. Prendergast (eds.), Current Perspectives in Social Theory, Suppl. 1, 1–26. Greenwich: JAI Press.

Pressat, R. 1972. Demographic Analysis. Methods, Results, Applications. Transl. from French by J. Matras. Foreword by N. Keyfitz. Chicago: Aldine & Atherton.

Proebsting, H. 1984. Entwicklung der Sterblichkeit. Wirtschaft und Statistik, Heft 1, 13 – 24, 438* – 440*.

Radcliffe-Brown, A. R. 1940. On Social Structure. In: Ders., Structure and Function in Primitive Society, 188–211. London: Cohen & West 1952.

Richards, E. G. 1998. Mapping Time. The Calendar and its History. Oxford: Oxford University Press.

Riedel, M. 1990. System, Struktur. In: O. Brunner, W. Conze, R. Koselleck (Hg.), Geschichtliche Grundbegriffe, Bd. 6, 285–322. Stuttgart: Klett-Cotta.

Rinne, H. 1996. Wirtschafts- und Bevölkerungsstatistik. 2. Aufl. München: Oldenbourg.

Ritsert, J. 2000. Gesellschaft. Ein unergründlicher Grundbegriff der Soziologie. Frankfurt: Campus.

Rohwer, G. 2006. Verweildauern und Übergangsraten bei mehreren Folgezuständen. In: A. Diekmann (Hg.), Methoden der Sozialforschung, 348–367. Wiesbaden: VS Verlag für Sozialwissenschaften.

Rohwer, G., Pötter, U. 2001. Grundzüge der sozialwissenschaftlichen Statistik. Weinheim: Juventa.

Rohwer, G., Pötter, U. 2002a. Methoden sozialwissenschaftlicher Datenkonstruktion. Weinheim: Juventa.

Rohwer, G., Pötter, U. 2002b. Wahrscheinlichkeit. Begriff und Rhetorik in der Sozialforschung. Weinheim: Juventa.

Rohwer, G., Pötter, U. 2003. Demography of Germany.

Röttgers, K. 1983. Der Ursprung der Prozessidee aus dem Geiste der Chemie. Archiv für Begriffsgeschichte 27, 93–157.

Rousseau, J.-J. 1755. Abhandlung über den Ursprung und die Grundlagen der Ungleichheit unter den Menschen. In: Schriften zur Kulturkritik (ed. K. Weigand). Hamburg: Meiner 1978.

Scheuch, E. K., Daheim, H. 1961. Sozialprestige und soziale Schichtung. In: D. V. Glass, R. König (Hg.), Soziale Schichtung und soziale Mobilität, 65–103. Opladen: Westdeutscher Verlag.

Scheuch, E. K., Kutsch, T. 1975. Grundbegriffe der Soziologie, Band 1: Grundlegung und elementare Phänomene. 2. Aufl. Stuttgart: Teubner.

Schmid, C. 2000. Zugang zu den Daten der Demographie. In: U. Mueller, B. Nauck, A. Diekmann (Hg.), Handbuch der Demographie, Band 1, 476 – 523. Berlin: Springer-Verlag.

Schütz, W. 1977. 100 Jahre Standesämter in Deutschland. Kleine Geschichte der bürgerlichen Eheschließung und der Buchführung des Personenstandes. Frankfurt: Verlag für Standesamtswesen.

Schwartz, J. T. 1961. Lectures on the Mathematical Method in Analytical Economics. New York: Gordon and Breach.

Sewell, W. H. 1992. A Theory of Structure: Duality, Agency, and Transformation. American Journal of Sociology 98, 1–29.

Siegwart, G. 1988. Einleitung. In: J. H. Lambert, Texte zur Systematologie und zur Theorie der wissenschaftlichen Erkenntnis, vii–xcvii. Hamburg: Meiner.

Simmel, G. 1903. Soziologie des Raumes. In: Schriften zur Soziologie, 221–242. Frankfurt: Suhrkamp 1983.

Simmel, G. 1908. Soziologie. Untersuchungen über die Formen der Vergesellschaftung. Berlin: Duncker & Humblot.

Smith, B. C. 1996. On the Origin of Objects. Cambridge: MIT Press.

Smith, D. P. (1992). Formal Demography. New York: Plenum Press.

Statistisches Bundesamt 1990. Familien heute. Strukturen, Verläufe und Einstellungen. Ausgabe 1990. Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt 1999. Fachserie 1, Reihe 1. Gebiet und Bevölkerung 1999. Stuttgart: Metzler-Poeschel.

Stein, A. von der 1968. Der Systembegriff in seiner geschichtlichen Entwicklung. In: A. Diemer (Hg.), System und Klassifikation in Wissenschaft und Dokumentation, 1–14, Meisenheim: Hain.

Steiner, H.-G. 1969. Aus der Geschichte des Funktionsbegriffs. Der Mathematikunterricht 15, 13–39.

Strohmeier, K., Schultz, A., Strohmeier, H. 2005. Familienforschung für die Familienpolitik. Wandel der Familie und sozialer Wandel als politische Herausforderungen. Bochum.

Swift, A. 2004. Would Perfect Mobility be Perfect? European Sociological Review 20, 1–11.

Tuma, N. B., Huinink, J. 1990. Postwar Fertility Patterns in the Federal Republic of Germany. In: K. U. Mayer, N. B. Tuma (eds.), Event History Analysis in Life Course Research, 146 – 169. Madison: University of Wisconsin Press.

Wagner, M. 1996. Lebensverläufe und gesellschaftlicher Wandel: Die westdeutschen Teilstudien. ZA-Informationen 38, 20 – 27.

Wasserman, S., Faust, K. 1994. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press.

Weber, M. 1921. Wirtschaft und Gesellschaft. 1. Halbband, 5. Aufl., hrsg. von J. Winckelmann. Tübingen: Mohr-Siebeck 1976.

Weymann, A. 2001. Interaktion, Sozialstruktur und Gesellschaft. In: H. Joas (Hg.), Lehrbuch der Soziologie, 93–121. Frankfurt: Campus.

Wirth, E. 1979. Theoretische Geographie. Grundzüge einer Theoretischen Kulturgeographie. Stuttgart: Teubner.

Wittgenstein, L. 1921. Tractatus logico-philosophicus. Frankfurt: Edition Suhrkamp 1963.

Würzberger, P., Störtzbach, B., Stürmer, B. 1986. Volkszählung 1987. Rechtliche Grundlagen nach dem Urteil des Bundesverfassungsgerichts vom 15. Dezember 1983. Wirtschaft und Statistik, Heft 12, 927 – 957.

Zapf, W. 1995. Entwicklung und Sozialstruktur moderner Gesellschaften. In: H. Korte, B. Schäfers (Hg.), Einführung in Hauptbegriffe der Soziologie, 181–193. Opladen: Leske + Budrich.

Namenverzeichnis

Stichwortverzeichnis