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

G. Rohwer

Version 1

März 2006

Vorbemerkung

Dieser Text enthält Materialien zum zweiten Teil des Moduls „Methoden der Demographie, Wirtschafts- und Sozialstatistik“. Eine Kenntnis des Textes zum ersten Teil wird vorausgesetzt.

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 von Gleichheitsbehauptungen 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

1	Soziale Ungleichheit	5
2	Bildungsungleichheit	9
2.1	Daten über Schulabschlüsse	9
2.2	Kohortenanalysen zur Schulbildung	14
2.3	Schulbildung von Eltern und Kindern	23
3	Rentenbeginn und Lebenserwartung	30
3.1	Berechnungen mit GEK-Daten	31
3.2	Berechnungen mit SOEP-Daten	47
4	Daten der Sozialhilfestatistik	55
4.1	Berechnung von Bezugsdauern	55
5	Episoden mit mehreren Folgezuständen	75
	Literatur	93

Kapitel 1

Soziale Ungleichheit

1. Eine allgemeine Definition.
2. Soziale versus natürliche Ungleichheit?
3. Statistische Erfassung sozialer Ungleichheit.
4. Soziale Ungleichheit zwischen Gruppen.
5. Qualitative und quantitative Ungleichheitsdimensionen.
6. Synthetische Ungleichheitskonstruktionen.
7. Quantifizierungen sozialer 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 Beurtei-

¹Im 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.

²Zum 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).

lung sozialer Ungleichheit setzt auch nicht unbedingt eine egalitäre gesellschaftspolitische Position voraus; eine andere Möglichkeit besteht zum Beispiel darin, für Mindeststandards zu plädieren.

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

$$Y : \Omega \longrightarrow \tilde{\mathcal{Y}} \quad (1.1)$$

Repräsentiert Ω eine Gesamtheit von Menschen, wird durch die statistische Variable Y jeder zu dieser Gesamtheit gehörenden Menschen ein bestimmter Wert im Merkmalsraum $\tilde{\mathcal{Y}}$, der die jeweils verwendete Ungleichheitsdimension (z.B. Bildungsabschlüsse oder Erwerbseinkommen) erfasst, zugeordnet. Die Häufigkeitsverteilung der Variablen, also $P[Y]$, 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 (1.1) genügt dann nicht mehr, sondern man benötigt eine zweidimensionale Variable

$$(Y, G) : \Omega \longrightarrow \tilde{\mathcal{Y}} \times \tilde{\mathcal{G}} \quad (1.2)$$

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 $P[Y|G = g]$.

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

³Das Reden von „vertikaler“ Ungleichheit (und Mobilität) wurde von Pitirim A. Sorokin verbreitet; die dafür relevanten Auszüge aus seinem Buch über „Social and Cultural

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 können. 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 2 besprochen.

Mobility“ findet man bei D. B. Grusky (1994: 245ff.). Offenbar verdankt sich der Ausdruck Sorokins Vorliebe für räumliche Metaphern.

⁴Eine 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 2

Bildungsungleichheit

2.1 Daten über Schulabschlüsse

1. Gliederungen des Bildungssystems.
2. Entwicklung der Schulanfänger.
3. Entwicklung der Schulabschlüsse.
4. Schulabschlüsse ausländischer Jugendlicher.

2.2 Kohortenanalysen zur Schulbildung

1. Daten aus dem ALLBUS.
2. Gliederung nach Geburtskohorten.
3. Veränderungen der Bildungsungleichheit.
4. Substitutionsmetriken für Verteilungen.

2.3 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.

2.1 Daten über Schulabschlüsse

1. *Gliederungen des Bildungssystems.* Es gibt in Deutschland viele verschiedene Schularten und Bildungswege. Die Grundstruktur kann anhand von Abbildung 2.1-1 verdeutlicht werden;¹ die Ziffern auf der linken Seite geben Jahrgangsstufen an.

Primarbereich. Die schulische Ausbildung beginnt normalerweise mit etwa sechs Jahren² in der Grundschule (= Primarbereich), die die ersten vier Jahrgangsstufen umfasst.

¹Die Darstellung erfolgt in Anlehnung an ein Schema in KMK 2005: 33. In dieser Publikation findet man eine umfassende Darstellung des Bildungswesens in der BRD.

²Die Schulpflicht beginnt für Kinder, die bis zum 30. Juni das sechste Lebensjahr vollendet haben.

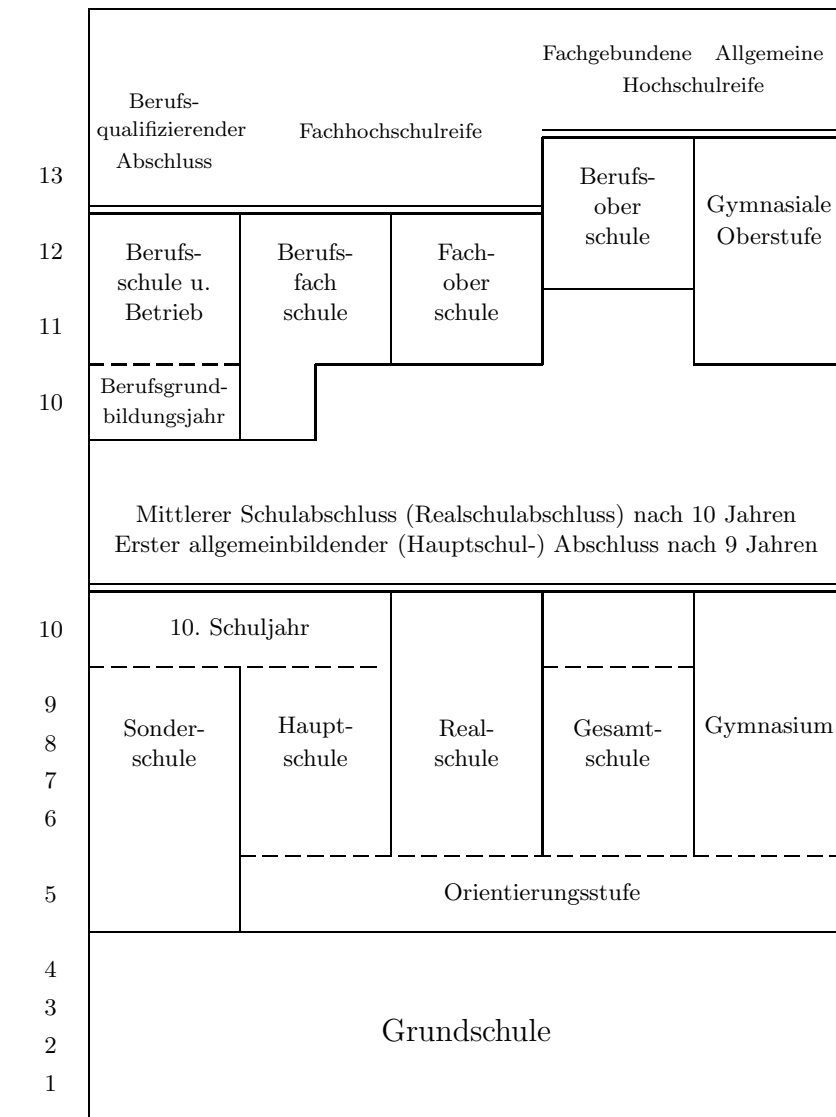


Abb. 2.1-1 Schema zur Gliederung des Bildungssystems in Deutschland (ohne tertiären Bereich).

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 Berufsober-
schule 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 2.1-2 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 langem 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)

³Eine zusammenfassende Darstellung geben J. Baumert, K.S. Cortina und A. Lechinsky (2005).

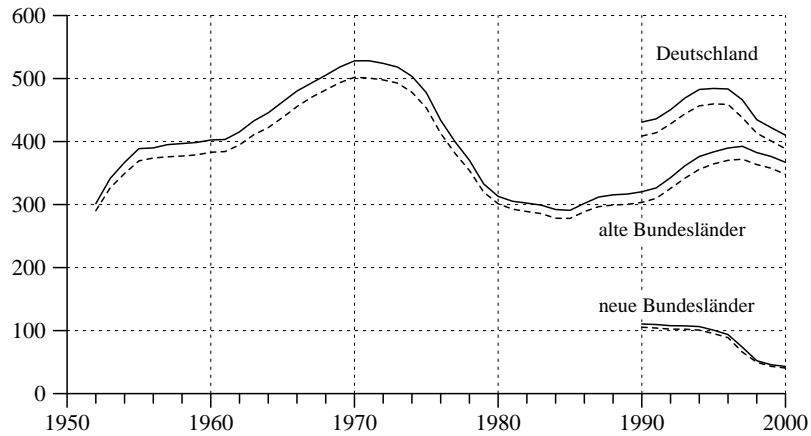


Abb. 2.1-2 Anzahl (in 1000) der 6-jährigen Jungen (durchgezogen) und Mädchen (gestrichelt) im Gebiet der früheren BRD sowie ab 1990 in Deutschland und den neuen Bundesländern. Quelle: Segmente 36 und 685 der STATIS-Datenbank.

Abbildung 2.1-3 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 2.1-4 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-

⁴Die 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.

⁵Teilweise hat der Prozess schon früher begonnen, was in den Abbildungen natürlich nicht sichtbar ist.

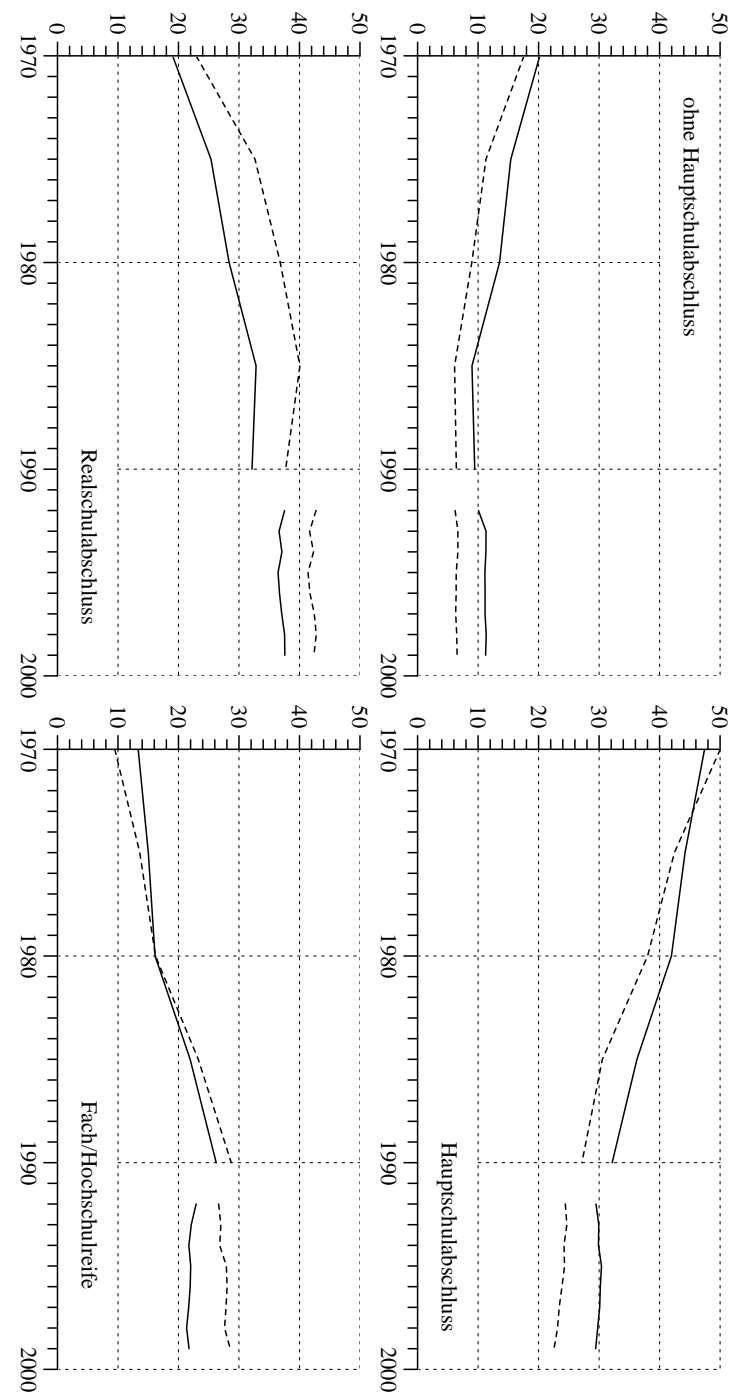


Abb. 2.1-3 Anteile (in %) der Schulabschlüsse, differenziert nach dem Geschlecht (durchgezogene Linien: männlich, gestrichelte Linien: weiblich). Bis 1990 früheres Bundesgebiet, ab 1992 Deutschland. Quelle: Fachserie 11, Reihe S.2: 34ff.

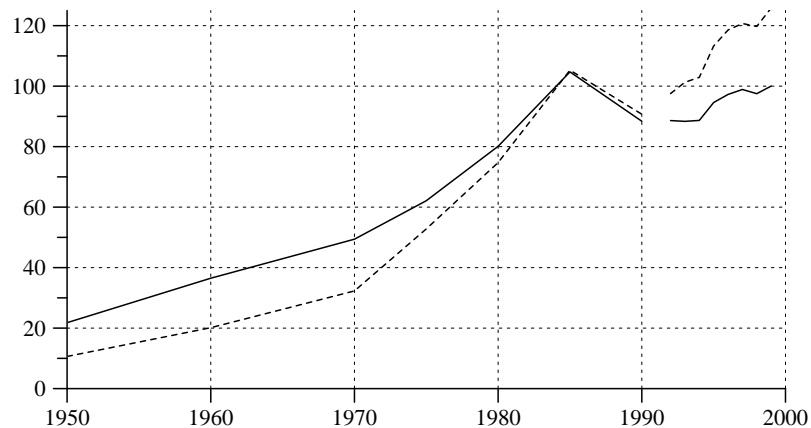


Abb. 2.1-4 Anzahl (in 1000) der männlichen (durchgezogen) und weiblichen (gestrichelt) Schulabgänger mit einer Hochschulreife. Bis 1990 früheres Bundesgebiet, ab 1992 Deutschland. Quelle: Fachserie 11, Reihe S.2: 34ff.

gendlichen mit deutscher und ausländischer Staatsbürgerschaft.⁶ Wie Abbildung 2.1-5 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.

2.2 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

⁶Unter analytischen Gesichtspunkten wäre es sinnvoller, Jugendliche mit einem Migrationshintergrund zu betrachten; sie können aber mit Daten der amtlichen Schulstatistik nicht ermittelt werden.

⁷Daten 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.

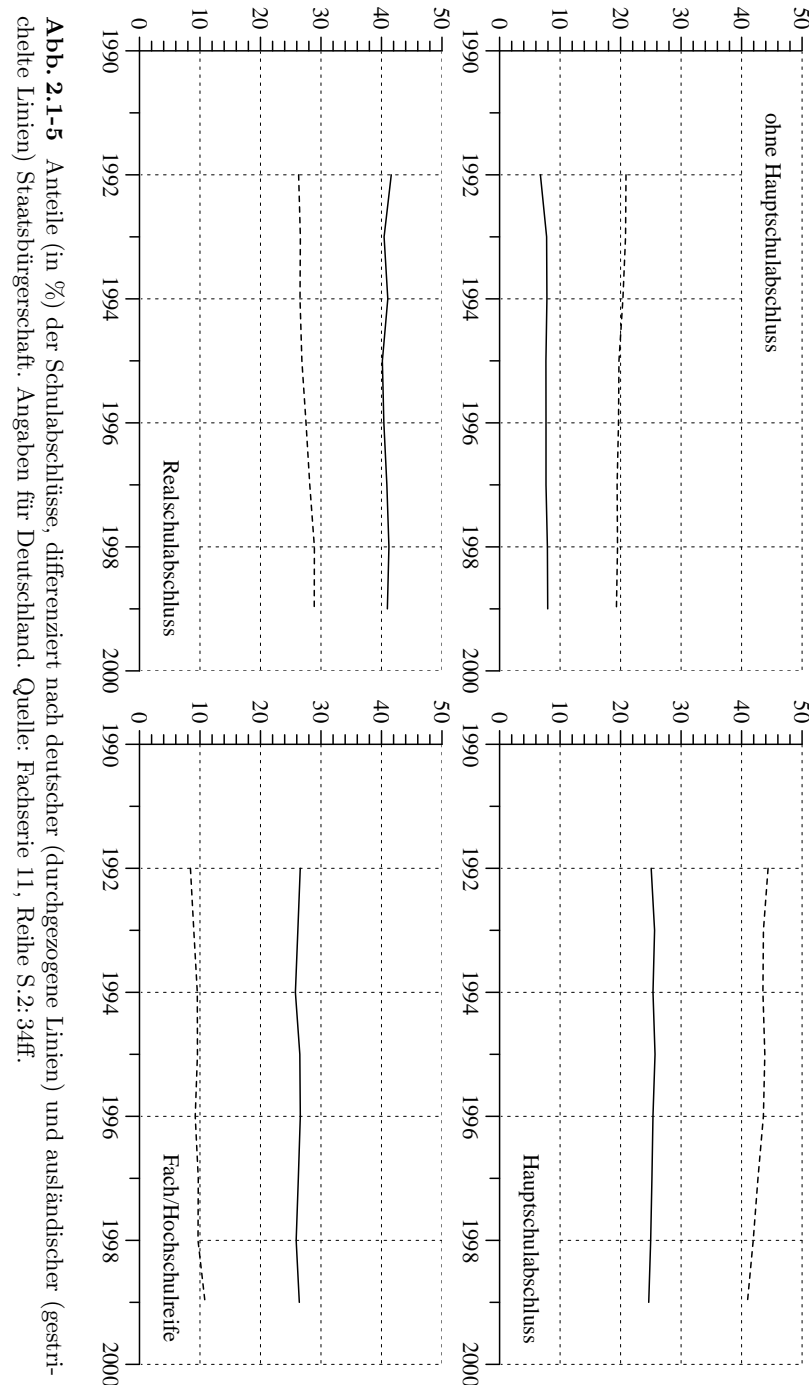


Abb. 2.1-5 Anteile (in %) der Schulabschlüsse, differenziert nach deutscher (durchgezogene Linien) und ausländischer (gestrichelte Linien) Staatsbürgerschaft. Angaben für Deutschland. Quelle: Fachserie 11, Reihe S.2: 34ff.

Tabelle 2.2-1 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 2.2-1 dargestellten Angaben gemacht (die Ziffern 1 – 5 dienen auch im Folgenden zum Verweis auf die unterschiedlichen Abschlüsse). Die Tabelle zeigt jeweils Anzahlen und Prozentanteile. Evident ist 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 2.2-2 zeigt, können insgesamt 14 dieser Geburtskohorten gebildet werden. Die Tabelle zeigt außerdem die Anzahlen und Anteile der Personen mit

⁸<http://www.gesis.org/Datenservice/ALLBUS/index.htm>

⁹Bis 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 2.2-2 Nach Geburtskohorten differenzierte Verteilungen auf Schulabschlüsse $j = 1, \dots, 5$. 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

den Schulabschlüssen 1 – 5 (wie sie im vorangegangenen Paragraphen unterschieden wurden). Abbildung 2.2-1 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.

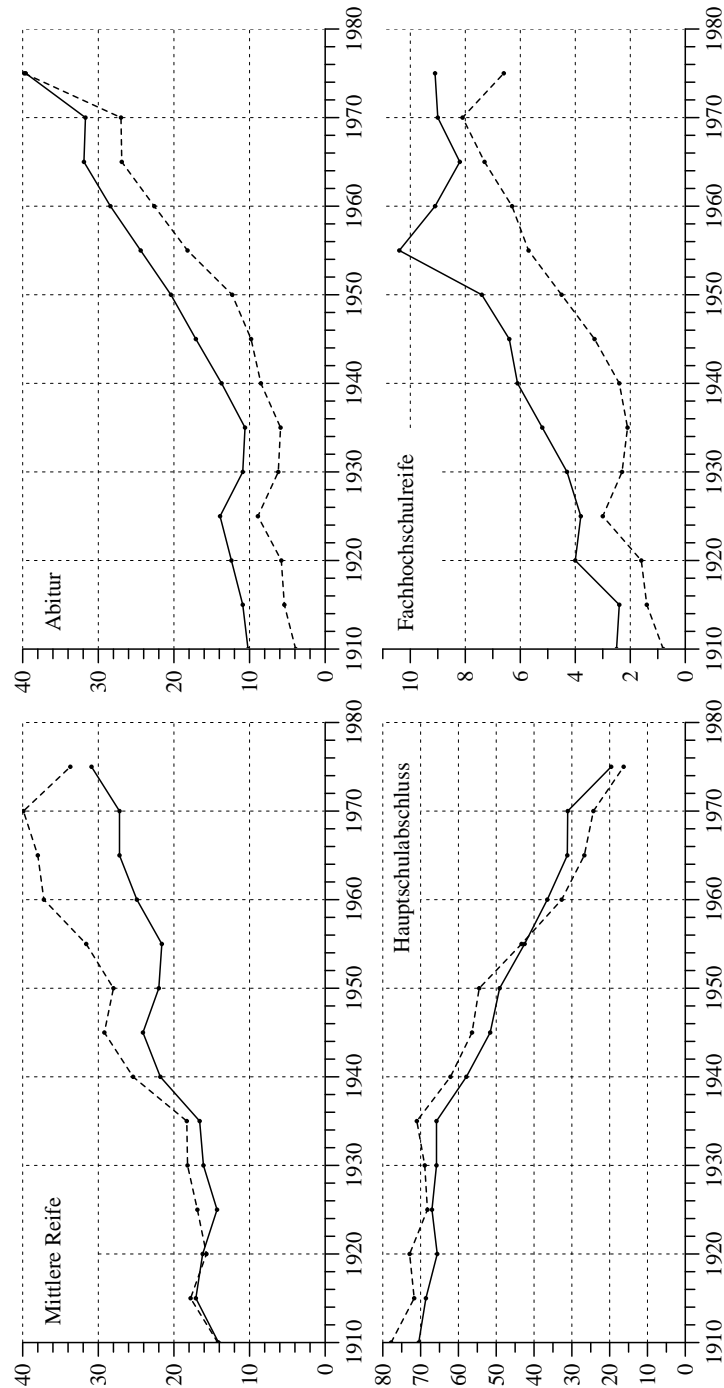


Abb. 2.2-1 Nach Geburtskohorten differenzierte Prozentanteile mit den angeführten schulischen Abschlüssen. Abszisse: Geburtsjahre der Kohorten. Daten aus Tabelle 2.2-2.

- 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

$$D_t := \sum_{j=1}^5 p_{t,j}^2$$

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 den gleichen Schulabschluss haben.¹² Wie Abbildung 2.2-2 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 jedoch 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

¹⁰Insgesamt ist der Anteil von Personen ohne Hauptschulabschluss wesentlich größer als in den hier verwendeten ALLBUS-Daten sichtbar ist (vgl. Abb. 2.1-3). Vermutlich liegt dies in erster Linie daran, dass wir hier nur Personen mit deutscher Staatsangehörigkeit betrachten.

¹¹Dies haben bereits W. Müller und D. Haun (1994: 14f.) festgestellt. Man vgl. dazu auch die Bemerkungen von W. Müller (1998: 91).

¹²Eine Diskussion von Eigenschaften und Interpretationsmöglichkeiten des Diversitätsindex findet man bei A. und B. F. Agresti (1977).

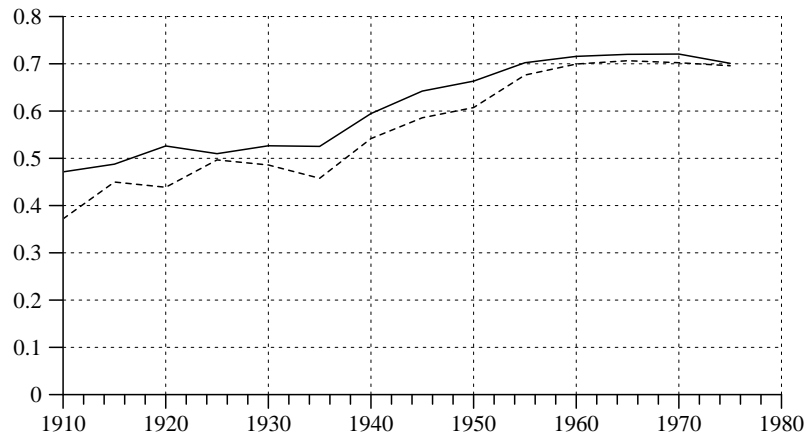


Abb. 2.2-2 Entwicklung des Disparitätsindex D_t für männliche (durchgezogene Linie) und weibliche (gestrichelte Linie) Personen, berechnet mit den Daten in Tabelle 2.2-2. Abszisse: Geburtsjahre der Kohorten.

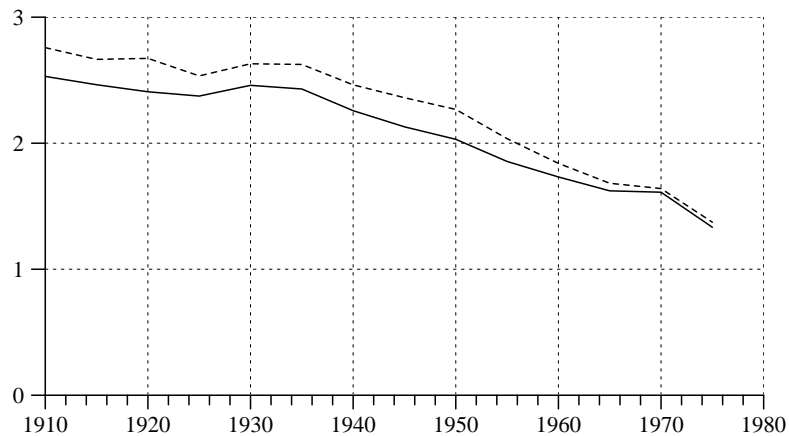


Abb. 2.2-3 Entwicklung der Maßzahl d_t für männliche (durchgezogene Linie) und weibliche (gestrichelte Linie) Personen, berechnet mit den Daten in Tabelle 2.2-2. Abszisse: Geburtsjahre der Kohorten.

verlassen. Der maximale Wert, nämlich 4, wird erreicht, wenn alle Personen die Schule ohne einen Abschluss verlassen. Abbildung 2.2-3 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: $1, \dots, m$. Die zu vergleichenden Verteilungen sind durch

$$p'_1, \dots, p'_m \quad \text{und} \quad p''_1, \dots, p''_m$$

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 Bezeichnung: c_{ij} (für $i, j = 1, \dots, m$). Es wird vorausgesetzt, dass diese Bewertungskoeffizienten nicht-negativ und symmetrisch ($c_{ij} = c_{ji}$) sind und 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, \dots, r_{m_r}) und (s_1, \dots, 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, entspricht 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:

$$\sum_{j=1}^{m_s} u_{ij} = r_i \quad \text{und} \quad \sum_{i=1}^{m_r} u_{ij} = s_j$$

Da es im Allgemeinen mehrere mögliche Umschichtungen gibt, die diese Bedingungen erfüllen, wird außerdem gefordert, dass die Kosten der Umschichtung, also

$$\sum_{i=1}^{m_r} \sum_{j=1}^{m_s} u_{ij} c_{ij}$$

minimal sein sollen.¹³ Diese Minimalkosten werden schließlich zur Quan-

¹³Eine Lösung kann mit der Methode der linearen Programmierung berechnet werden. Wir verwenden den `subm`-Befehl des Programms TDA, der auf dieser Methode beruht.

tifizierung des Abstands der Verteilungen verwendet.¹⁴

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. 2.2-2). Es ist also $m = 5$, und die Verteilungen sind durch die Vektoren

$$\mathbf{p}' = (0.009, 0.196, 0.309, 0.091, 0.395) \quad \text{und}$$

$$\mathbf{p}'' = (0.036, 0.163, 0.337, 0.066, 0.398)$$

gegeben. Somit ist $\mathbf{r} = (0.033, 0.025)$ und $\mathbf{s} = (0.027, 0.028, 0.003)$. Zur

¹⁴Wenn die Kosten der Umschichtung stets den Wert 1 haben ($c_{ij} = 1$ für $i \neq j$), entspricht die Substitutionsmetrik dem durch

$$\frac{1}{2} \sum_{i=1}^m |p'_i - p''_i|$$

definierten *Dissimilaritätsindex*.

¹⁵Wir 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 Dreiecksungleichung

$$(d) \quad \text{für alle } a, b, c \in M: d(a, c) \leq d(a, b) + d(b, c)$$

gilt, wird *Semi-* oder *Quasi-Metrik* genannt. Gilt schließlich auch noch

$$(e) \quad \text{für alle } a, b \in M: d(a, b) = 0 \implies a = b$$

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:

$$\mathbf{C} := \begin{bmatrix} 0 & 3 & 5 & 7 & 8 \\ 3 & 0 & 2 & 4 & 5 \\ 5 & 2 & 0 & 2 & 3 \\ 7 & 4 & 2 & 0 & 1 \\ 8 & 5 & 3 & 1 & 0 \end{bmatrix} \quad (2.1)$$

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.

2.3 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 2.2 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 2.2). 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 2.2 betrachteten 29726 Befragungspersonen kann auf diese Weise zusätzlich zu ihrem eigenen auch ein Schulabschluss der Eltern bestimmt werden. Tabelle 2.3-1 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 2.3-1 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 2.2 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 2.2 definierten Bewertungsmatrix **C**. Abbildung 2.3-1 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 2.3-2 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 2.3-1 haben sich

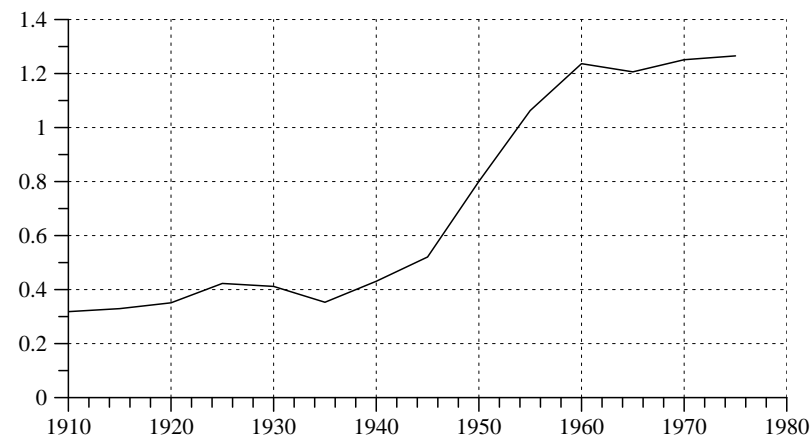


Abb. 2.3-1 Abstand zwischen den Verteilungen der Schulabschlüsse bei Befragungspersonen und ihren Eltern. Abszisse: Geburtsjahre der Kohorten. Berechnet mit Daten des kumulierten ALLBUS 1980–2002.

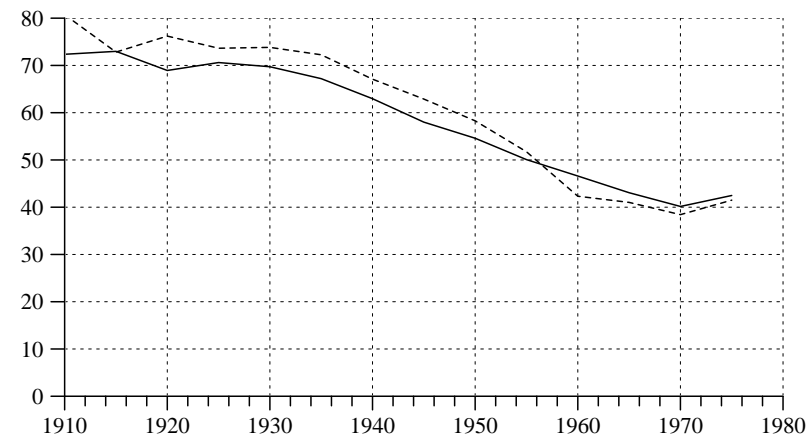


Abb. 2.3-2 Prozentanteile der männlichen (durchgezogen) und weiblichen (gestrichelt) Befragungspersonen, die den gleichen Schulabschluss wie ihre Eltern erworben haben. Abszisse: Geburtsjahre der Kohorten. Berechnet mit Daten des kumulierten ALLBUS 1980–2002.

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

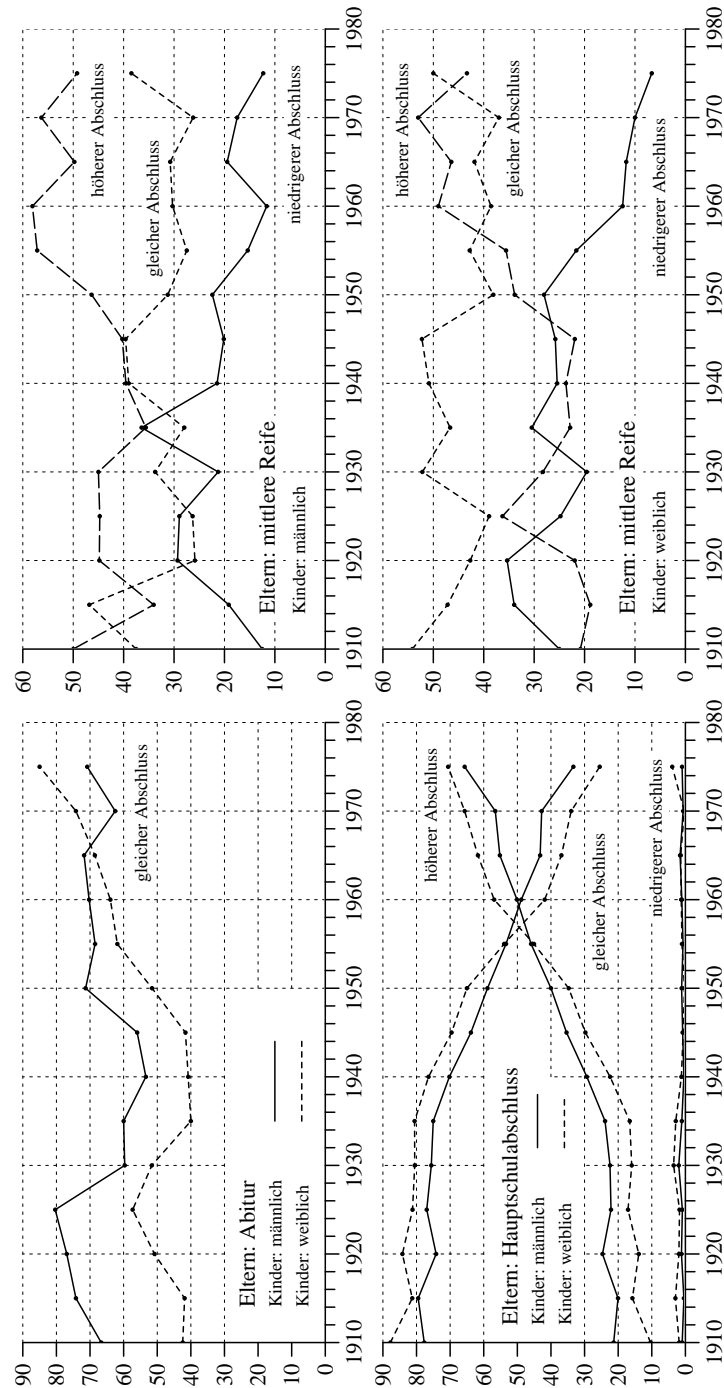


Abb. 2.3-3 Anteile (in %) der Kinder (Befragungspersonen) mit höherem, gleichem oder niedrigerem Schulabschluss als ihre Eltern, differenziert nach dem Schulabschluss der Eltern. Abszisse: Geburtsjahre der Befragungspersonen (Kohorten). Berechnet mit Daten des kumulierten ALLBUS 1980–2002.

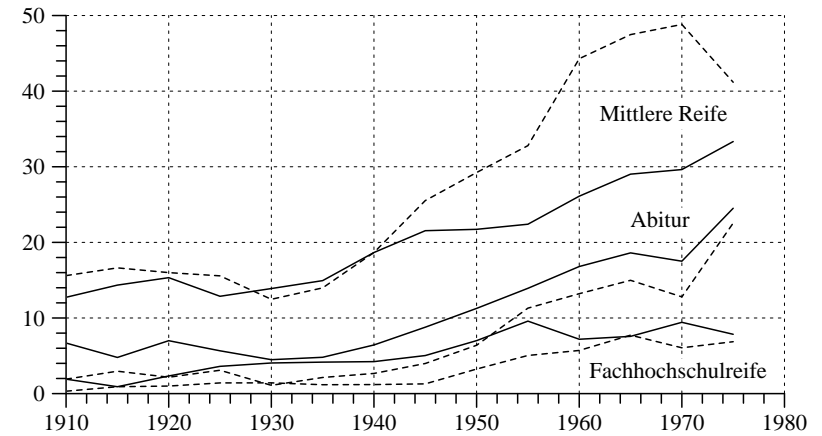


Abb. 2.3-4 Anteile (in %) der männlichen (durchgezogen) und weiblichen (gestrichelt) Befragungspersonen mit den angegebenen Schulabschlüssen an allen Befragungspersonen, bei denen der Vater und/oder die Mutter einen Hauptschulabschluss haben. Abszisse: Geburtsjahre der Kohorten. Berechnet mit Daten des 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 2.3-3 zeigt, wie sich einige dieser Anteile in der Abfolge der Geburtskohorten verändert haben. Offenbar gibt es bemerkenswerte Unterschiede.

- Bei den Personen, deren Eltern nur einen Hauptschulabschluss haben, hat der Anteil mit höheren Abschlüssen erheblich zugenommen. Wie Abbildung 2.3-4 zeigt, gab es – vor allem bei Mädchen – in erster Linie eine Zunahme der Realschulabschlüsse, aber auch eine Zunahme von Abschlüssen für eine (Fach-)Hochschulausbildung.
- 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.
- 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 2.3-2 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.	1965	j = 1	j = 2	Insges.
i = 1	1914	174	2088	i = 1	1142	213	1355
i = 2	100	159	259	i = 2	98	231	329
Insges.	2014	333	2347	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 2.3-2 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

¹⁶Man 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 Chancenverhältnis (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 2.2 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:

$$\begin{aligned} |q_{1950,1} - q_{1950,2}| &= |0.083 - 0.614| = 0.531 \longrightarrow \\ |q_{1965,1} - q_{1965,2}| &= |0.157 - 0.702| = 0.545 \end{aligned}$$

Man käme also zu dem Ergebnis, dass der Abstand zwischen den Verteilungen der beiden Gruppen etwas zugenommen hat.

¹⁷Man vgl. beispielsweise J. Handl (1985) und W. Müller und D. Haun (1994: 16).

Kapitel 3

Rentenbeginn und Lebenserwartung

3.1 Berechnungen mit GEK-Daten

1. Basisinformationen zum Datensatz.
2. Berechnung von Lebenserwartungen.
3. Informationen über Rentenempfänger.
4. Berechnung von Rentenbezugsdauern.
5. Berechnungen mit zensierten Daten.

3.2 Berechnungen mit SOEP-Daten

1. Daten des Kalendariums.
2. Daten aus der zweiten Welle.

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 eine 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.

3.1 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 3.1-1 zeigt, wie sich die Anzahl der männlichen Versicherten der GEK seit 1960 entwickelt hat.³

Bereits aus dem in Box 3.1-1 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 3.1-1 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 3.1-2 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 3.1-1 zeigen, dass man annehmen kann, dass

¹Dieser 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.

²Die Daten wurden im Rahmen eines am Zentrum für Sozialpolitik (Bremen) durchgeführten Projekts aufbereitet; dafür möchte ich mich bei Wolfgang Voges herzlich bedanken.

³Für jedes Jahr wird die Anzahl der männlichen Personen angegeben, die mindestens einen Tag des Jahres bei der GEK versichert waren.

⁴Das Alter wird in diesem Text stets als Differenz zwischen Kalenderjahr und Geburtsjahr berechnet.

⁵Statistisches Bundesamt, Fachserie 1, Reihe 1, 1999, S. 64f.

Box 3.1-1 Anzahl und Durchschnittsalter sowie Sterbefälle der männlichen Versicherten der GEK von 1960 bis 2002.

Jahr	Anzahl Versicherte	Durchschnittsalter	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

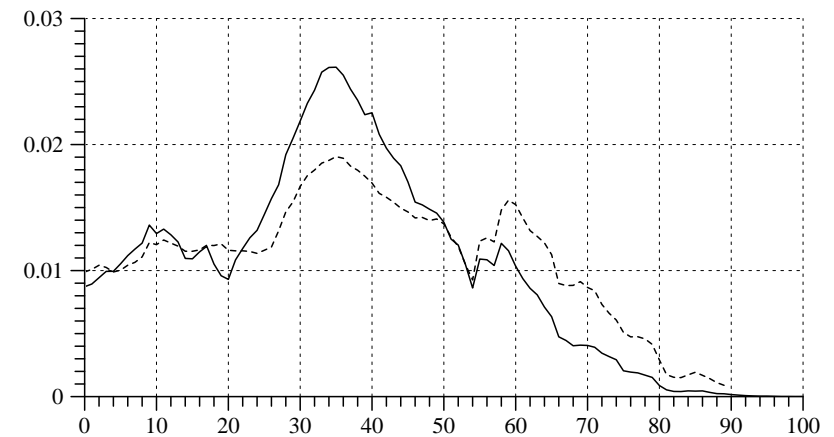


Abb. 3.1-1 Altersverteilung der männlichen GEK-Versicherten (durchgezogene Linie) und der männlichen Gesamtbevölkerung (gestrichelte Linie) im Jahr 1999.

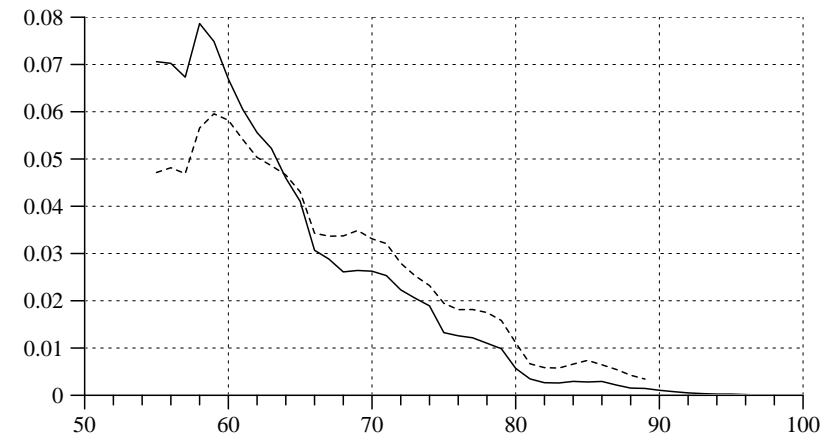


Abb. 3.1-2 Altersverteilung der männlichen GEK-Versicherten (durchgezogene Linie) und der männlichen Gesamtbevölkerung (gestrichelte Linie) im Jahr 1999, beschränkt auf Personen, die mindestens 55 Jahre alt sind.

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 bestimmte Periode (z.B. ein Kalenderjahr oder ein 3-Jahres-Intervall) und

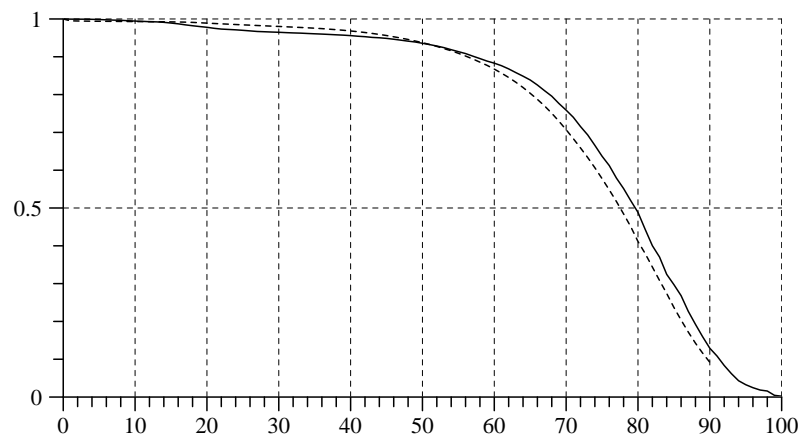


Abb. 3.1-3 Vergleich der Überlebensfunktionen der männlichen GEK-Versicherten (durchgezogene Linie) und der männlichen Gesamtbevölkerung (gestrichelte Linie) im Jahr 1999.

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 3.1-3 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,

⁶Für 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 3.1-2 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.

Jahr	N	D	fernere Lebenserwartung im Alter ...							
			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

⁷Für das höchste Alter – in den folgenden Berechnungen 100 Jahre – wird angenommen, dass $G_{t,101} = 0$ ist.

Box 3.1-3 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 3.1-2 zeigt, wie sich die auf diese Weise berechneten fernerer 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:

- Datum des Rentenanspruchs
- Datum des Rentenbeginns (wenn der Antrag bewilligt wurde)
- Datum der Beendigung des Rentenbezugs (oder Angabe, dass die Rente noch bezogen wird⁹)
- 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

⁸Man könnte ein halbes Jahr hinzufügen, um zu berücksichtigen, dass wir das Alter als Differenz zwischen Kalenderjahr und Geburtsjahr berechnen.

⁹In diesem Fall nehmen wir an, dass die Rente mindestens bis zum Jahr 2002 bezogen worden ist.

Box 3.1-4 Männliche GEK-Versicherte, die in dem angegebenen Jahr eine Alters- oder EU-Rente beziehen.

Jahr	Rentenbezieher	
	insgesamt	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 3.1-3 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 3.1-4 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 von

Box 3.1-5 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

55 bis 66 Jahren stattgefunden hat. Box 3.1-5 zeigt für den Kalenderzeitraum von 1975 bis 2001 das Rentenzugangsalter dieser Personen.¹⁰ Das durchschnittliche Rentenzugangsalter dieser Personen wird in Box 3.1-6 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

¹⁰Bei 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.

¹¹Zur 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 3.1-6 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 3.1-7 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;

¹²Alter und Lebensdauern werden wie stets in diesem Beitrag als Differenz zwischen Kalenderjahr (Sterbejahr) und Geburtsjahr oder dem Jahr des Rentenzugangs berechnet.

Box 3.1-7 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 3.1-7, 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 3.1-7). Personen, die mit einer EU-Rente begonnen haben, haben

¹³Vgl. VDR: Rentenversicherung in Zeitreihen, Ausgabe 2002, S. 128.

Box 3.1-8 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

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 3.1-8. Man erkennt, dass Personen, die vor dem Alter 65 verrentet werden, keineswegs eine längere fernere Lebensdauer aufweisen.

Box 3.1-9 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 3.1-8 bezieht sich Box 3.1-9 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 3.1-5). 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-

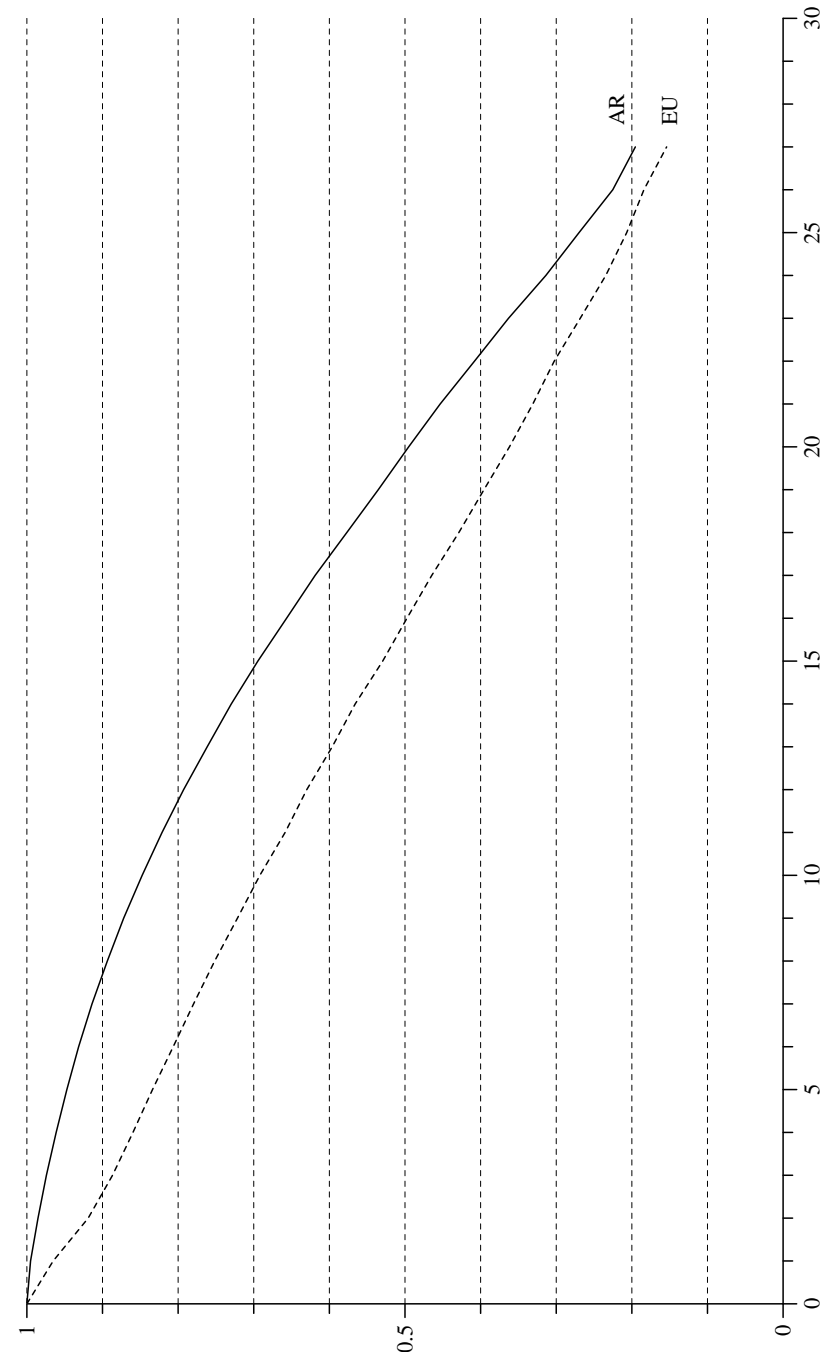


Abb. 3.1-4 Überlebensfunktionen der männlichen GEK-Versicherten, die im Zeitraum von 1975 bis 2001 mit einer Alters- oder EU-Rente begonnen haben.

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 3.1-4 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 3.1-5 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 3.1-6 nach dem Rentenzugangsalter differenzierte Überlebensfunktionen nur für diejenigen Personen, die mit einer Altersrente begonnen haben.

¹⁴Ergänzend könnte man auch die aus Abbildung 3.1-4 direkt ablesbaren Median-Werte heranziehen.

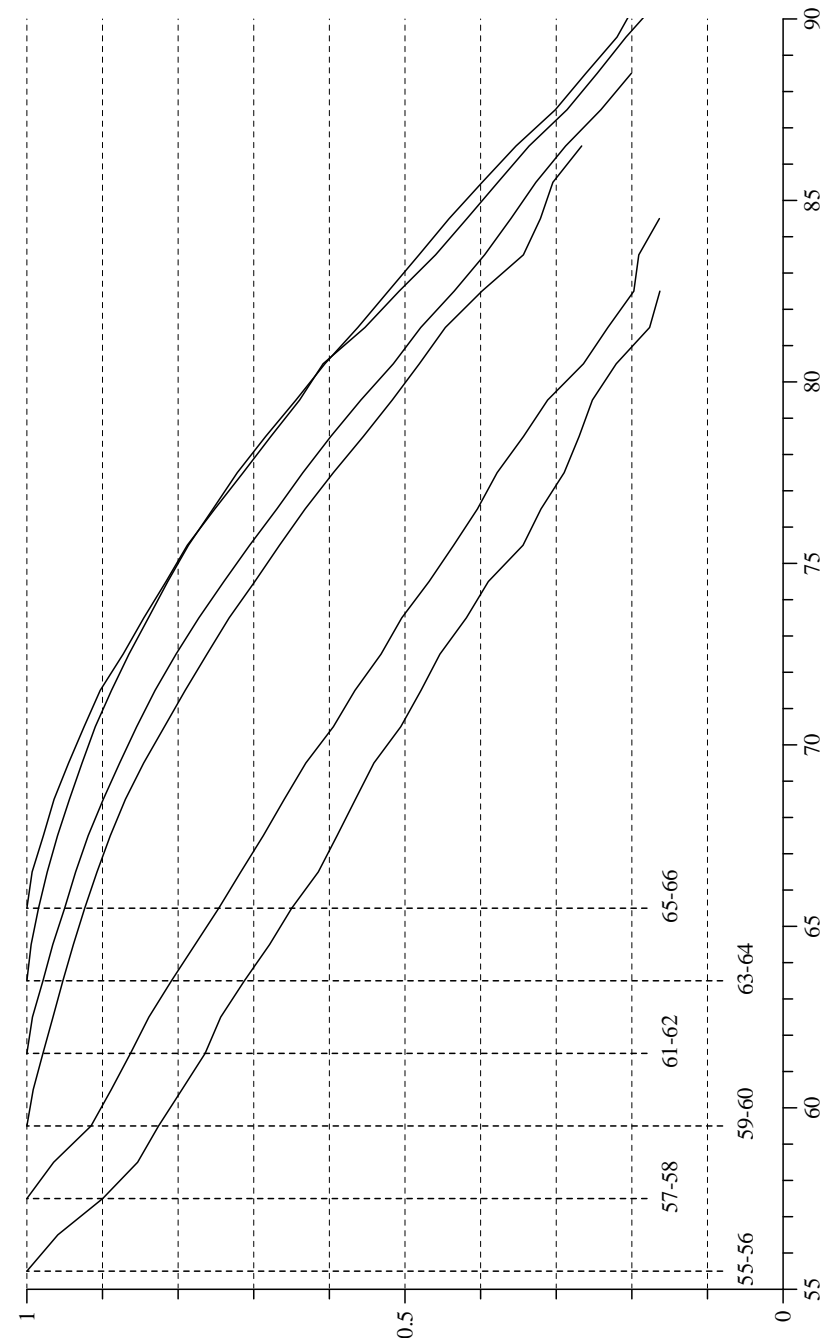


Abb. 3.1-5 Überlebensfunktionen der männlichen GEK-Versicherten, die im Zeitraum von 1975 bis 2001 mit einer Alters- oder EU-Rente begonnen haben, differenziert nach dem Rentenzugangsalter.

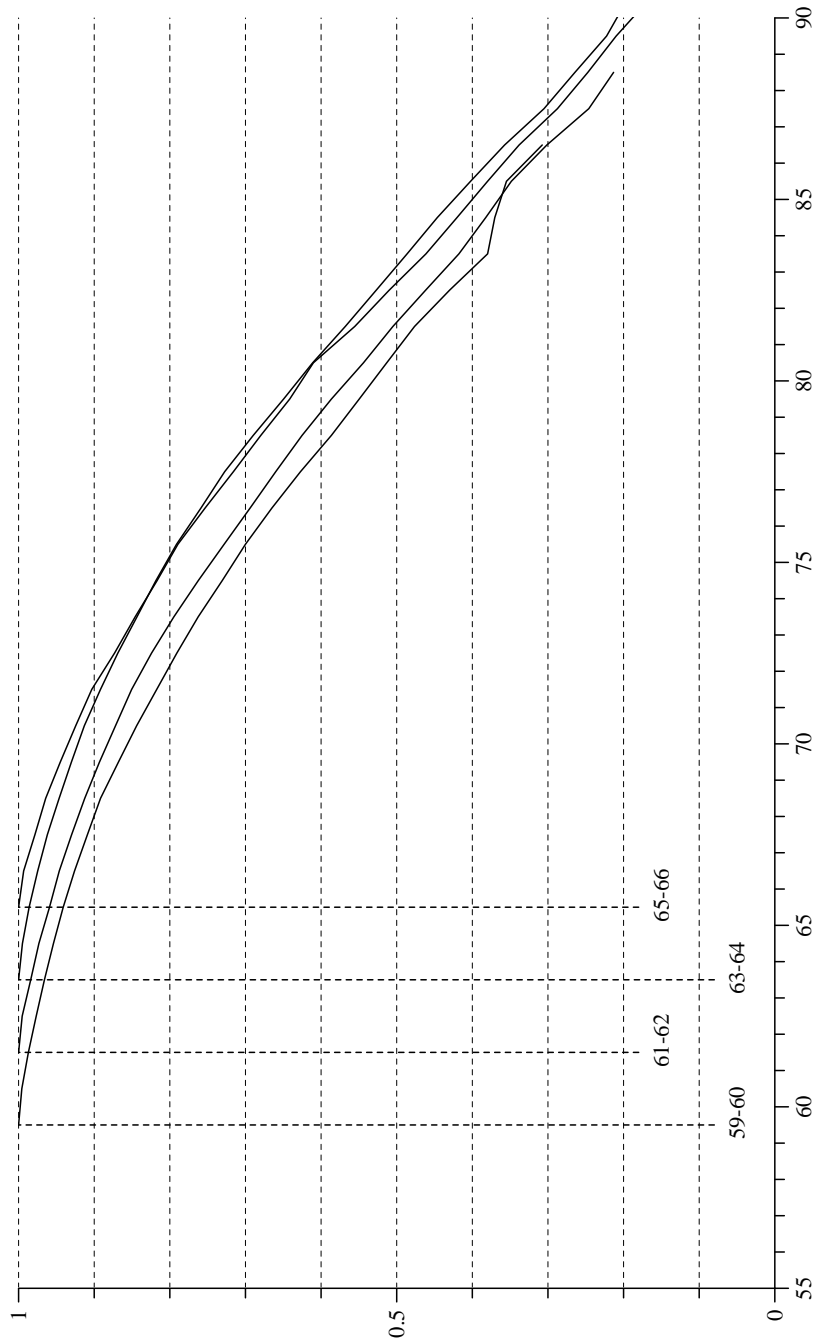


Abb. 3.1-6 Überlebensfunktionen der männlichen GEK-Versicherten, die im Zeitraum von 1975 bis 2001 mit einer Altersrente begonnen haben, differenziert nach dem Rentenzugangsalter.

3.2 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 3.2-1 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 3.2-1 zeigt, in welchen Jahren bei diesen Personen der Rentenbezug begonnen hat. Box 3.2-2 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 3.2-3 zeigt für diese Personen, in welchem Alter sie gestorben sind. Dabei entspricht jede Zeile einem bestimmten Verrentungsalter. Zum Beispiel kann man der Box 3.2-2 entnehmen, dass 17 Personen im Alter 55 verrentet wurden, und dann ersieht

¹⁵Der Ausdruck „Rentenbezug“ bezieht sich hier und im folgenden stets auf Alters-, EU- und BU-Renten.

¹⁶Infratest Sozialforschung, Verbesserung der Datengrundlagen für Mortalitäts- und Mobilitätsanalysen: Verbleibstudie bei Panelausfällen im SOEP, München 2002.

Box 3.2-2 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 3.2-3 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	2	2	0	0	1	0	0	0	1	0	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	1	2	0	4	0	0	1	2	1	2	0	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	3	0	1	1	1	1	1	0	0	0	0	0	1	8
65	0	0	0	0	0	0	0	0	2	2	1	0	0	0	0	2	0	2	1	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 3.2-3, 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 3.2-1 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 3.2-4 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 3.2-5 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,

¹⁷Tatsächlich haben wesentlich mehr Personen einen Rentenbezug angegeben; aber in vielen Fällen gibt es keine gültige Angabe über den Beginn des Rentenbezugs.

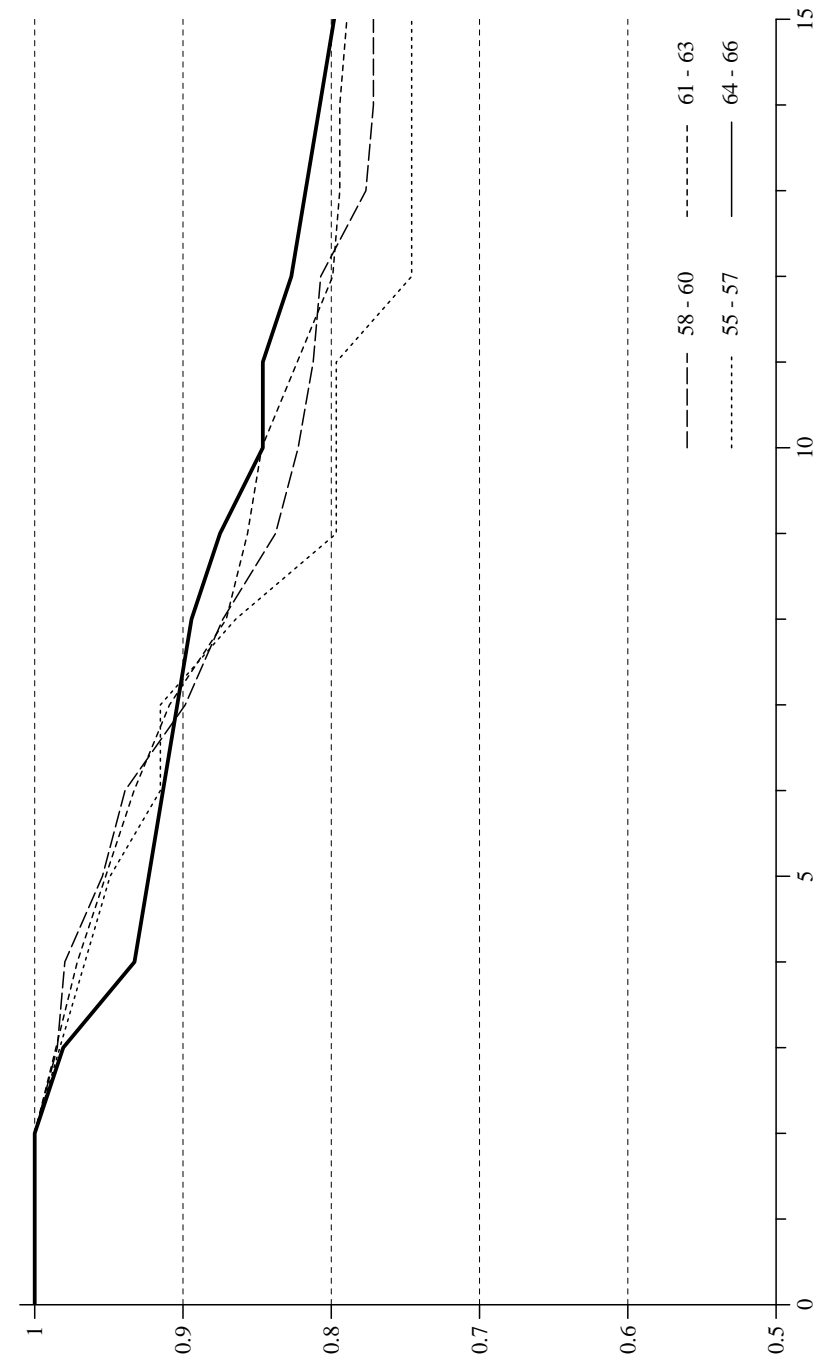


Abb. 3.2-1 Survivorfunktionen für einen Zeitraum von 15 Jahren nach der Verrentung.

Box 3.2-4 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:

- S_i erfasst die Zeitdauer (in Jahren) zwischen dem Rentenbeginn der Person i und dem Jahr 1985.
- D_i ist der Zensierungsindikator. Er bekommt den Wert 1, wenn die Person i bis zum Jahr 2001 gestorben ist, andernfalls den Wert 0.
- 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:

¹⁸Dabei wird ein modifiziertes Kaplan-Meier-Verfahren verwendet; man vgl. Rohwer, Pötter, Grundzüge der sozialwissenschaftlichen Statistik, Weinheim: Juventa 2001, S. 207ff.

Box 3.2-5 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

- Verrentungsalter 55 bis 61 Jahre, dies sind insgesamt 235 Personen, von denen im Beobachtungszeitraum 133 gestorben sind; und
- Verrentungsalter 63 bis 66 Jahre, dies sind insgesamt 279 Personen, von denen im Beobachtungszeitraum 186 gestorben sind.

Abbildung 3.2-2 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 fernere Lebensdauer haben als die Personen in der Gruppe (b).

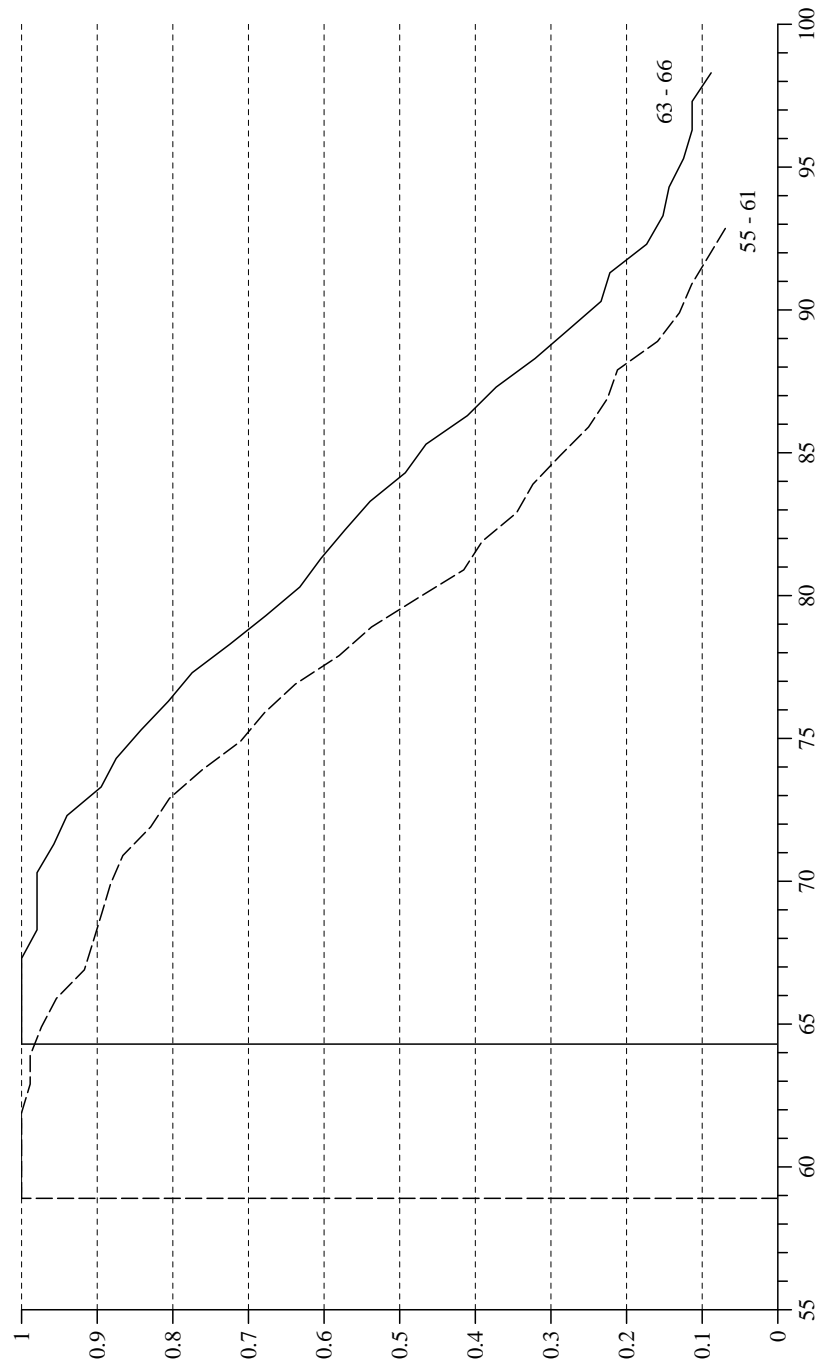


Abb. 3.2-2 Survivorfunktionen für die fernere Lebensdauer nach dem Beginn der Verrentung.

Kapitel 4

Daten der Sozialhilfestatistik

4.1 Berechnung von Bezugsdauern

1. Mikrodaten der Sozialhilfestatistik.
2. Überlegungen zum methodischen Vorgehen.
3. Betrachtung der Jahresraten.
4. Verkettung der Jahresraten.
5. Die Gestalt der Ratenfunktion.
6. Vergleich mit bisherigen Bezugsdauern.
7. Durchschnittliche Bezugsdauern.
8. Alte und neue Bundesländer.

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.

4.1 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 4.1-1 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 4.1-1 (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								

- Eine Information über die „längste bisherige Dauer der Hilfgewährung an die Bedarfsgemeinschaft in der aktuellen Zusammensetzung“, und
- eine Information über die „bisherige Dauer der ununterbrochenen Hilfgewährung 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 4.1-1 zeigt die Daten für die drei Stichtage. Dabei geben S_{97} , S_{98} und S_{99} die Bezugsdauern an und N_{87} , 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

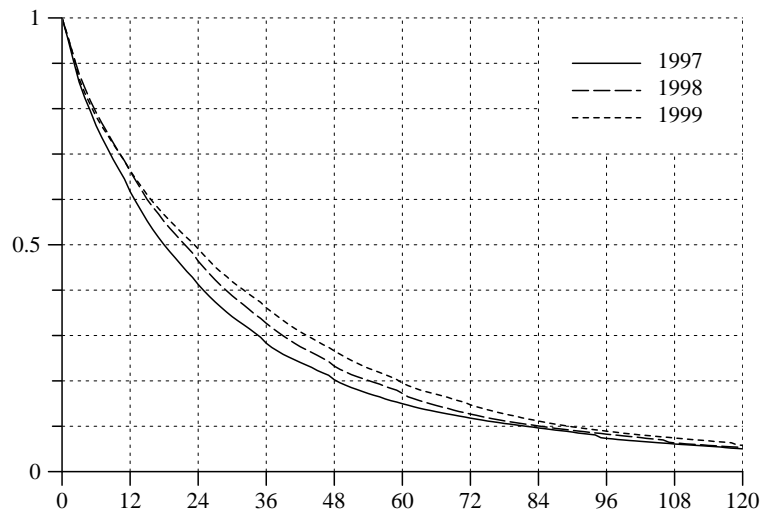


Abb. 4.1-1 Darstellung der Verteilungen der bisherigen Bezugsdauern durch Survivorfunktionen.

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. 4.1-1 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

$$T_j : \Omega_j \longrightarrow \{0, 1, 2, 3, \dots\}$$

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:

¹Nähere 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:

$$n_j(t) := \text{Anzahl der Haushalte, die Ende des Jahres } j \text{ eine} \\ \text{Bezugsdauer von } t \text{ Monaten hatten}$$

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 Auswahlatz 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 Auswahlatz 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.

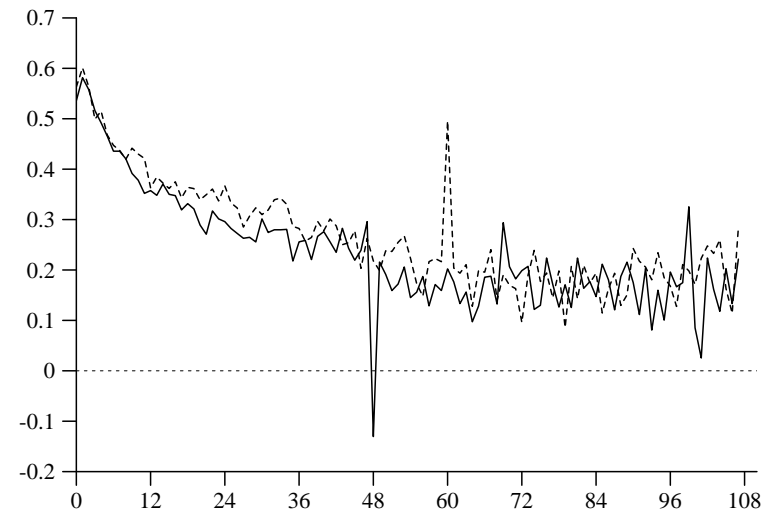


Abb. 4.1-2 Darstellung der Jahresraten $u_{97}(t)$ (durchgezogene Linie) und $u_{98}(t)$ (gestrichelte Linie).

3. Betrachtung der Jahresraten. Bevor überlegt wird, wie man mit den Jahresraten Bezugsdauerverteilungen schätzen kann, sollten sie selbst etwas genauer betrachtet werden. Abb. 4.1-2 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 4.1-1 erkennt man den Grund. Ende 1997 gab es 2500 Haushalte mit einer Bezugsdauer von 48 Monaten, und Ende 1998 gab es *mehr* Haushalte, nämlich 2836, 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 4.1-1 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 4.1-1 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 $u_{98}(t)$. 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 4.1-1 zeigt die Rechenergebnisse.

Abb. 4.1-3 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 4.1-1 und 4.1-3 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.

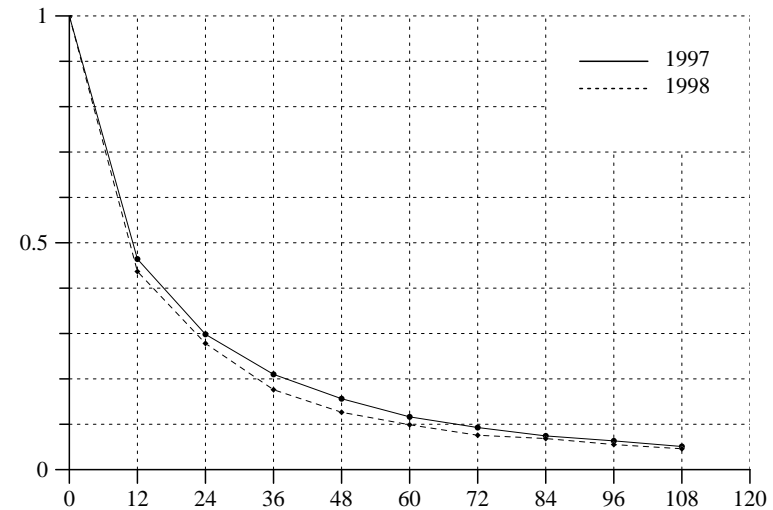


Abb. 4.1-3 Aus Jahresraten berechnete Stützpunkte der Survivorfunktionen G_{97} und G_{98} .

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:

$$f_j(\alpha_{j,1}, \dots, \alpha_{j,119}) := \sum_{t=0}^{107} (u_j^*(t) - u_j(t))^2$$

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:

²Also $\alpha_{j,t} \equiv G_j(t)$. Zusätzlich wird $\alpha_{j,0} = 1$ fest vorgegeben.

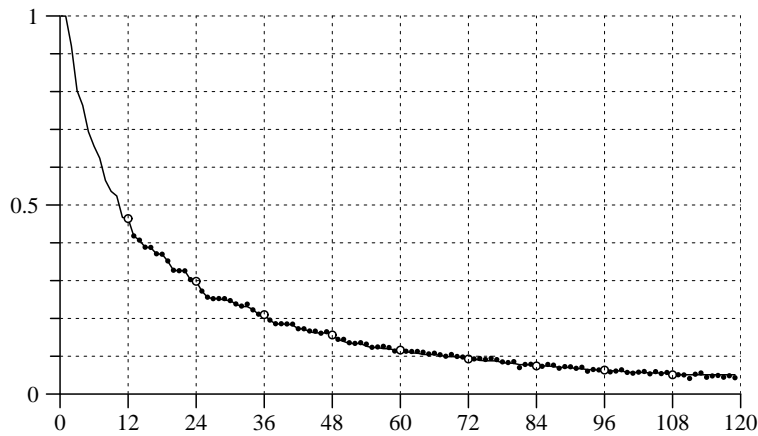


Abb. 4.1-4 Aus der Minimierung der im Text angegebenen Zielfunktion berechnete Survivorfunktion für die Jahresraten $u_{97}(t)$.



Abb. 4.1-5 Aus der Minimierung der im Text angegebenen Zielfunktion berechnete Survivorfunktion für die Jahresraten $u_{98}(t)$.

$$u_j^*(t) = \prod_{k=t}^{t+11} \left(1 - \frac{\alpha_{j,k+1}}{\alpha_{j,k}}\right)$$

Die Zielfunktion ist natürlich unter der Nebenbedingung zu minimieren, dass die Survivorfunktion monoton fällt: $1 \geq \alpha_{j,1} \geq \alpha_{j,2} \geq \dots \geq \alpha_{j,119}$. 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.

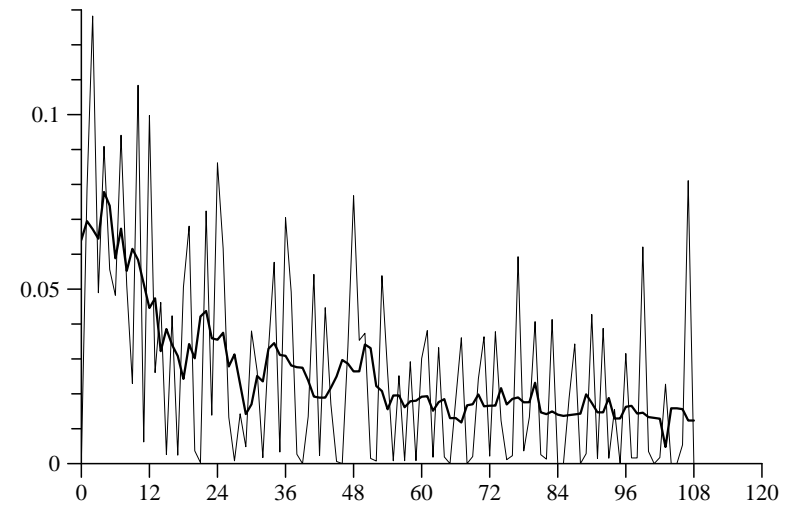


Abb. 4.1-6 Unglättete und durch gleitende Durchschnitte geglättete Raten r_{97}^* , korrespondierend zur Survivorfunktion G_{97}^* in Abbildung 4.1-4.

Die Abbildungen 4.1-4 und 4.1-5 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 4.1-3 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. 4.1-6 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. 4.1-6 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

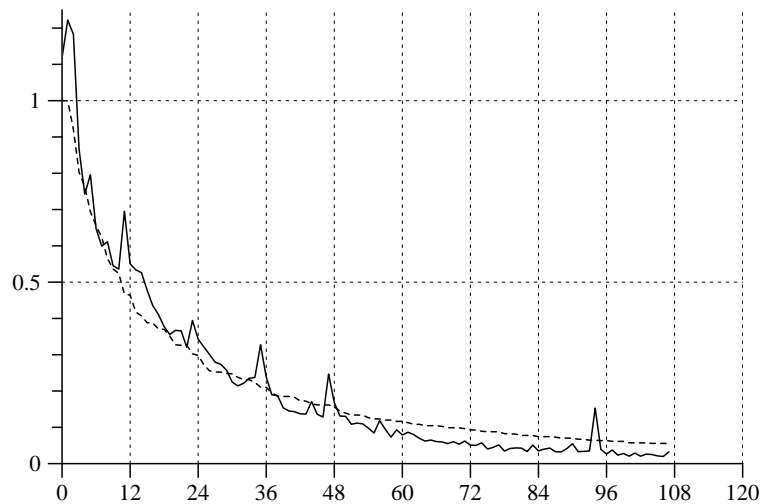


Abb. 4.1-7 Verteilung der bisherigen Bezugsdauern $\alpha n_{97}(t)$ (durchgezogene Linie) und der Survivorfunktion G_{97}^* (gestrichelte Linie).

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 mit der Survivorfunktion G_j^* identisch. 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 4.1-1). Für den Vergleich beschränken wir uns auf die Bezugsdauern $t = 0, \dots, 107$ und berechnen Anteilswerte $\alpha n_{97}(t)$ durch die Normierung

$$\sum_{t=0}^{107} \alpha n_{97}(t) = \sum_{t=0}^{107} G_{97}^*(t)$$

Abb. 4.1-7 vergleicht die beiden Verteilungen bisheriger Bezugsdauern. Es ist bemerkenswert, dass die Verteilungen sehr ähnlich sind. Natürlich kann

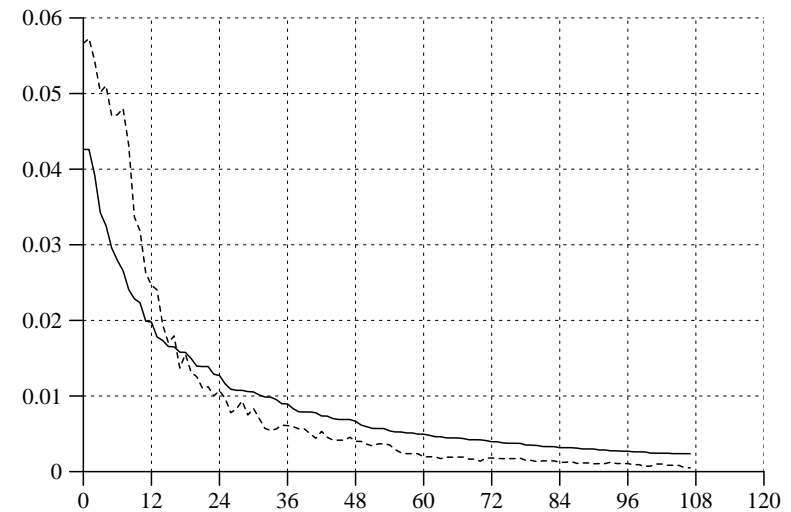


Abb. 4.1-8 Vergleich von G_{97}^* (durchgezogene Linie) mit der geglätteten Verteilung der Häufigkeiten $p_{97}^*(t)$ (gestrichelte Linie).

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 $t = 0, \dots, 107$, so dass $p_{97}^*(0) + \dots + p_{97}^*(107) = 1$ ist. Dementsprechend werden auch die Anteile der bisherigen Bezugsdauern so normiert, dass ihre Summe für die Monate 0 bis 107 gleich 1 ist. Abb. 4.1-8 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,

³Wieder mit gleitenden Durchschnitten, wobei diesmal zu beiden Seiten jeweils 6 Nachbarwerte verwendet worden sind.

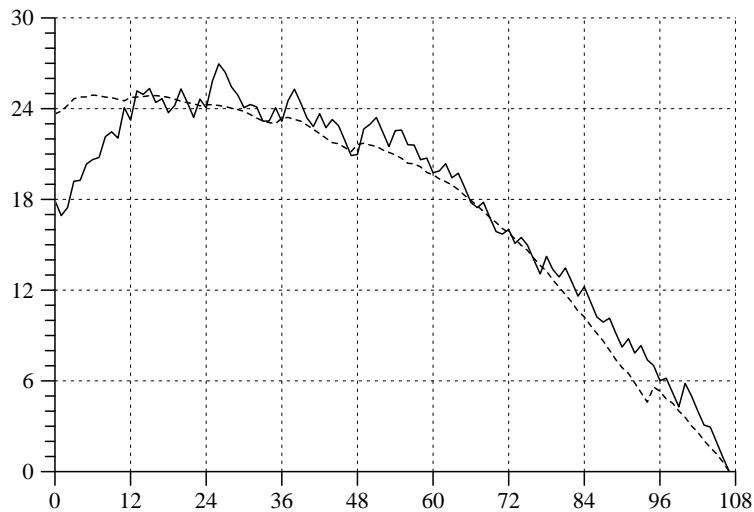


Abb. 4.1-9 Fernere durchschnittliche Bezugsdauern bei tatsächlichen (durchgezogene Linie) und bisherigen Bezugsdauern (gestrichelte Linie).

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. 4.1-1, erhält man Median, 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 4.1-1 verwendet

werden. Dann findet man folgende Mittelwerte:

- Die Mittelwerte der tatsächlichen Bezugsdauern sind von 17.9 Monaten für 1997 auf 17.0 Monaten für 1998 etwas gesunken; dagegen sind
- 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. 4.1-9 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 4.1-2 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						

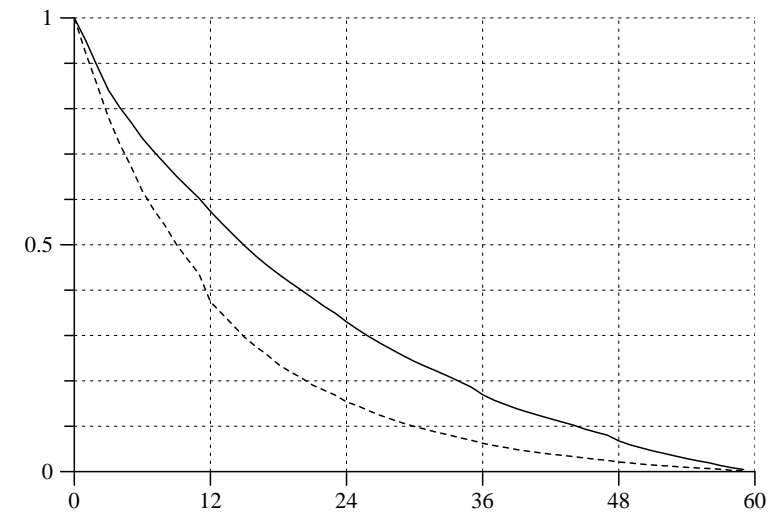


Abb. 4.1-10 Verteilungen der bisherigen Bezugsdauern (maximal 59 Monate) in den alten Bundesländern (durchgezogene Linie) und in den neuen Bundesländern (gestrichelte Linie).

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 4.1-2 zeigt analog zu Box 4.1-1 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.

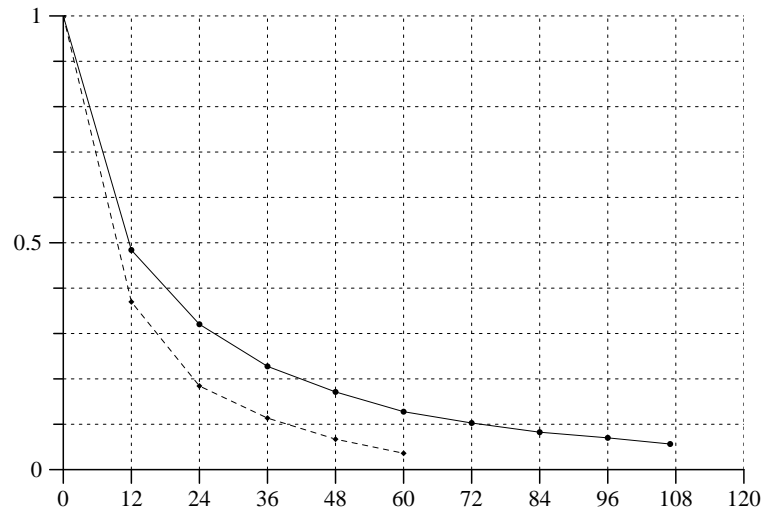


Abb. 4.1-11 Stützpunkte der Survivorfunktionen für tatsächliche Bezugsdauern in den alten Bundesländern (durchgezogene Linie) und in den neuen Bundesländern (gestrichelte Linie).

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. 4.1-10 zeigt die Verteilungen in Gestalt von Survivorfunktionen (analog zu Abb. 4.1-1). 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. 4.1-11 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 Abbildung 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.

⁴Wiederum könnte man, wie in § 4 besprochen wurde, die Gesamtheit der Jahresraten in die Berechnung einbeziehen. Für einen Vergleich ist das jedoch nicht unbedingt erforderlich.

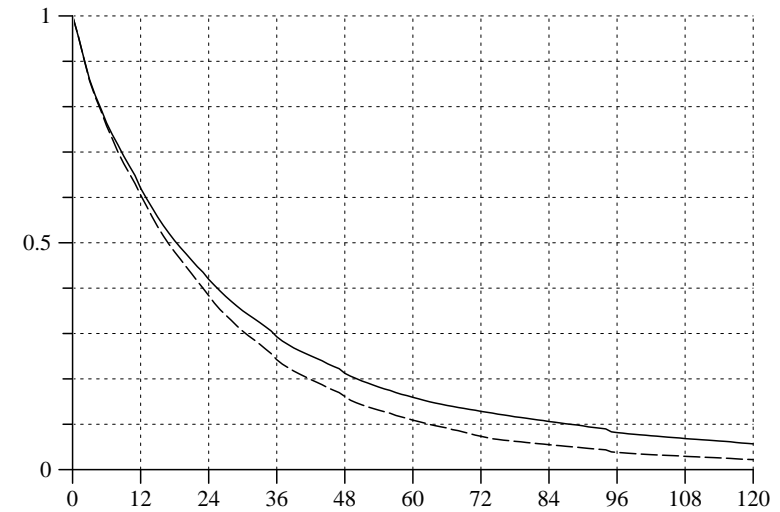


Abb. 4.1-12 Verteilungen der bisherigen Bezugsdauern bei Haushalten mit deutscher (durchgezogene Linie) und ausländischer (gestrichelte Linie) Bezugsperson.

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.

⁵Die 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.

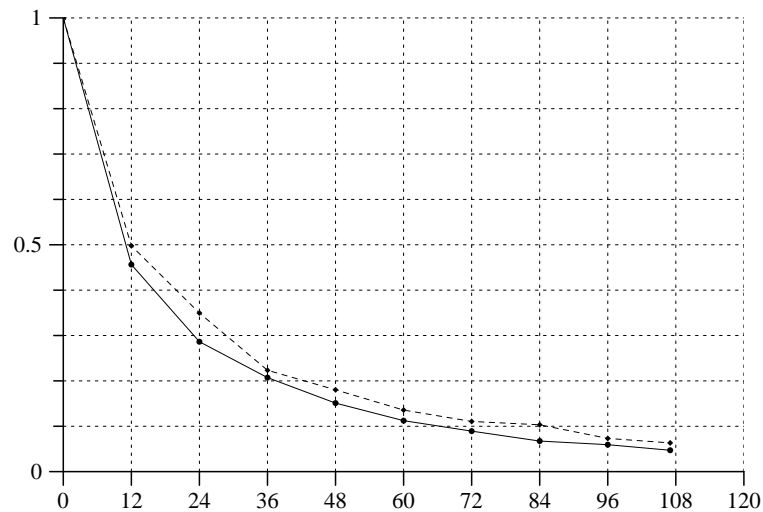


Abb. 4.1-13 Stützpunkte der Survivorfunktionen für tatsächliche Bezugsdauern bei Haushalten mit deutscher (durchgezogene Linie) und ausländischer Bezugsperson (gestrichelte Linie).

Zunächst vergleichen wir die Verteilungen der bisherigen Bezugsdauern Ende des Jahres 1997. Wie Abb. 4.1-12 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. 4.1-13 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 5

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 Todesfälle.

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 Bezeichnung 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

¹Man vgl. etwa Nambodiri 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 Anteilfunktionen vor, die sich auf die Anteile an einer Ausgangsgesamtheit beziehen, die in die unterschiedlichen Folgezustände wechseln. Es wird gezeigt, dass auch solche Anteilfunktionen 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 $t = 0, 1, 2, 3, \dots$ 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 $\Omega := \{\omega_1, \dots, \omega_n\}$. 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 $T : \Omega \rightarrow \{0, 1, 2, 3, \dots\}$ 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 Verweildauernvariablen 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*

$$r^a(t) := \frac{w_t^a}{n_t} = \frac{p^a(t)}{G(t)} \quad \text{und} \quad r^b(t) := \frac{w_t^b}{n_t} = \frac{p^b(t)}{G(t)}$$

²Eine 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.

³Eine 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 $T^a : \Omega^a \rightarrow \{0, 1, 2, 3, \dots\}$ 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*

$$G^a(t) := \sum_{k=t}^{\infty} \frac{w_k^a}{n^a} = 1 - \sum_{k=0}^{t-1} \frac{w_k^a}{n^a} = \prod_{k=0}^{t-1} \left(1 - \frac{w_k^a/n^a}{G^a(k)}\right)$$

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

⁴Es 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

⁵Es 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 5-1 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 zustands-spezifischen 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.

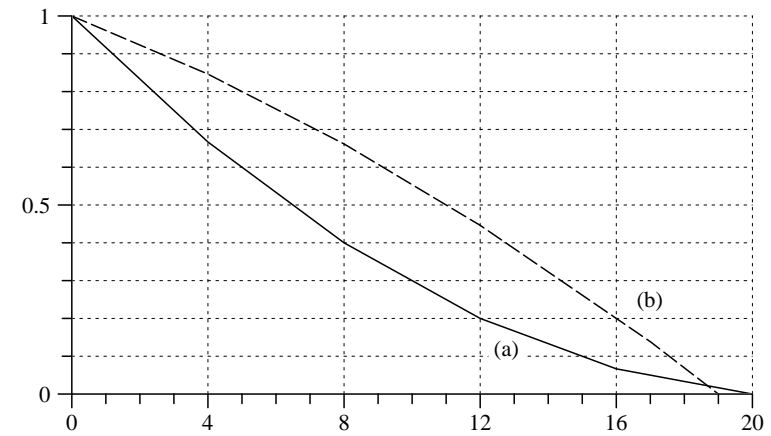
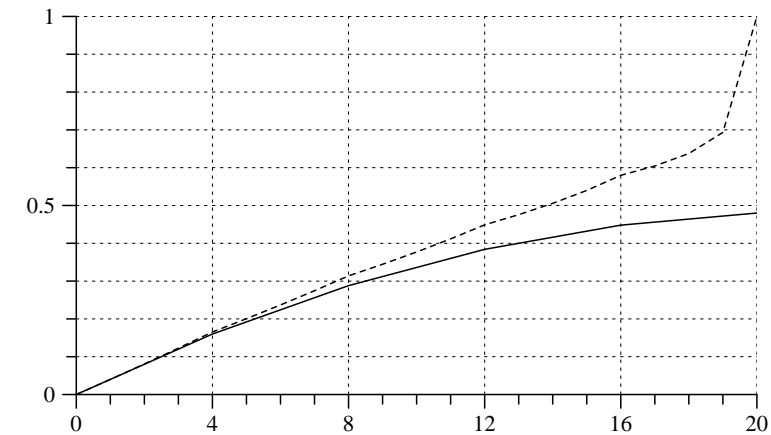
4. *Eine numerische Illustration.* Zur Illustration der Begriffsbildungen und Verdeutlichung des Problems betrachten wir ein Zahlenbeispiel mit fiktiven Daten. Wir gehen von einer Gesamtheit von 2500 arbeitslos gewordenen Personen aus, und es gibt zwei mögliche Folgezustände: (a) Übergang in eine neue Beschäftigung und (b) Verlassen des Arbeitsmarktes (einschließlich Sterbefälle). Tabelle 5-1 zeigt die fiktiven Daten. Die Zeitstellen der Zeitachse $t = 0, \dots, 19$ 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:

- Man findet, dass von den 2500 Personen 1200 (48%) eine neue Beschäftigung annehmen und 1300 (52%) aus dem Arbeitsmarkt ausscheiden.
- Man kann die Survivorfunktionen G^a und G^b für die beiden Folgezustände berechnen (Abb. 5-1).
- 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 5-2 zeigt die Rechenschritte.

⁶Natürlich könnte man berücksichtigen, dass die Prozesszeitachse bei 0 beginnt und jeweils ein halbes Quartal hinzurechnen.

**Abb. 5-1** Survivorfunktionen G^a und G^b für die durch Tabelle 5-1 gegebenen Daten.**Abb. 5-2** Darstellung der Funktionen H^a (durchgezogene Linie) und $1 - \tilde{G}^a$ (gestrichelte Linie) für die durch Tabelle 5-1 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 5-2 Berechnung der Survivorfunktion \tilde{G}^a mit dem Kaplan-Meier-Verfahren aus den Daten in Tabelle 5-1.

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. 5-2 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 Anteilfunktionen zensierte Beobachtungen grundsätzlich nicht verwendet

werden können. Es gilt nämlich folgender Zusammenhang, der bereits in § 3 erwähnt wurde:

$$H^a(t) = \sum_{k=0}^{t-1} p^a(k) = \sum_{k=0}^{t-1} r^a(k) G(k) \quad (5.1)$$

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 Anteilfunktion 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

$$M(T) = \sum_{t=0}^{\infty} t(G(t) - G(t+1))$$

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

$$M(T | T < t_0) = \sum_{t=0}^{t_0-1} t \frac{G(t) - G(t+1)}{1 - G(t_0)}$$

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 Verweildauernvariablen 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 Anteilfunktion H^a berechnet werden:

$$\begin{aligned} M(T^a | T^a < t_0) &= \sum_{t=0}^{t_0-1} t \frac{G^a(t) - G^a(t+1)}{1 - G^a(t_0)} \\ &= \sum_{t=0}^{t_0-1} t \frac{H^a(t+1) - H^a(t)}{H^a(t_0)} = \frac{\sum_{t=0}^{t_0-1} t w_t^a}{\sum_{t=0}^{t_0-1} w_t^a} \end{aligned}$$

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

⁷Alternativ könnten bedingte Mittelwerte der Form $M(T^a | T^a \geq x)$ 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 Anteilfunktion 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.

$w_{j,t}^a$:= Anzahl der Ehen, die im Kalenderjahr j eine Dauer von t Jahren hatten und in diesem Kalenderjahr geschieden wurden.

Ehedauerspezifische Scheidungsraten sind dann durch $q_j^a(t) := w_{j,t}^a/n_{j,t}$ 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

$$\check{C}_j^a(t) := \prod_{k=0}^{t-1} (1 - q_j^a(k))$$

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

$$\tilde{C}_j^a(t) := \prod_{k=0}^{t-1} (1 - q_{j+k}^a(k))$$

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 Scheidungsziffern, die folgendermaßen definiert sind: $p_j^a(t) := w_{j,t}^a/n_{j-t,0}$. 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 5-3 zeigt diese Scheidungsziffern für das Kalenderjahr $j = 1999$ und Ehedauren $t = 0, \dots, 25$. Zum Beispiel wurden in diesem Jahr 11139 Ehen geschieden, die im Jahr 1990 geschlossen wurden. Andererseits kennt man aus früheren Erhebungen die

⁸Diese Definition entspricht derjenigen von I. Esenwein-Rothe (1982: 293).

⁹Da die Kohortenkonstruktion auf einer zeitlichen Verkettung ehedauerspezifischer Periodendaten beruht, sprechen wir von Quasi-Kohorten.

Tabelle 5-3 Ehedauerspezifische Scheidungsziffern für das Jahr 1999, multipliziert mit 1000 (Statistisches Bundesamt 1999: 268).

t	$p_{1999}^a(t)$	t	$p_{1999}^a(t)$	t	$p_{1999}^a(t)$	t	$p_{1999}^a(t)$
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		

Anzahl der Ehen, die 1990 geschlossen wurden, nämlich 516388. So erhält man $p_{1999}^a(9) = 11139/516388 = 0.0216$. 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

$$H_j^a(t) := \frac{1}{n_{j,0}} \sum_{k=0}^{t-1} w_{j+k,k}^a = \sum_{k=0}^{t-1} p_{j+k}^a(k)$$

Zur Berechnung benötigt man natürlich die ehedauerspezifischen Scheidungsziffern für eine Folge von Jahren. Kennt man sie bis zum Jahr 1999, könnte man z.B. $H_{1974}^a(26)$ berechnen, also – unter der genannten Einschränkung – den Anteil der 1974 geschlossenen Ehen, die bis zum Ende des Jahres 1999 geschieden worden sind. Abb. 5-3 zeigt diese Funktion H_{1974}^a als durchgezogene Linie.¹¹ Andererseits könnte man auch eine hypothetische Betrachtung anstellen, bei der man von den ehedauerspezifischen Scheidungsziffern eines bestimmten Kalenderjahrs ausgeht, also

$$\check{H}_j^a(t) := \sum_{k=0}^{t-1} p_j^a(k)$$

berechnet. Mit den Angaben aus Tabelle 5-3 findet man z.B. $\check{H}_{1999}^a(26) = 0.362$. Unter Zugrundelegung der Scheidungsziffern des Jahres 1999 würden also bei einem hypothetischen Eheschließungsjahrgang bis zu einer Ehedauer von einschließlich 25 Jahren 36.2% der Ehen geschieden werden. Abb. 5-3 zeigt die entsprechende Funktion als gestrichelte Linie.

¹⁰Eine weitere Fehlerquelle für Deutschland ist die Gebietsveränderung seit 1990.

¹¹Die Daten entstammen der Tabelle VIII B/181 des Statistischen Bundesamts, die uns freundlicherweise von Herrn H.-P. Bosse zur Verfügung gestellt wurde.

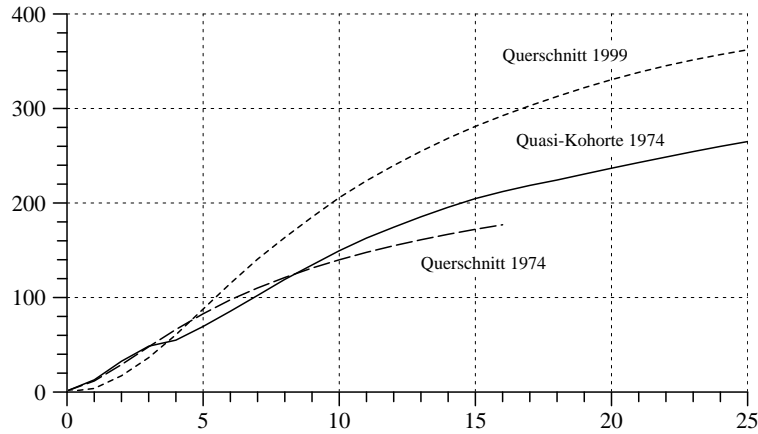


Abb. 5-3 Kumulierte ehedauerspezifische Scheidungsziffern (multipliziert mit 1000).

8. *Scheidungen und Todesfälle.* In § 4 wurde gezeigt, dass es sehr große Differenzen zwischen der Pseudo-Survivorfunktion \bar{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 Hilfe der in Tabelle 5-1 angegebenen Scheidungsziffern für 1999. Um den Simulationszeitraum über 25 Jahre hinaus ausdehnen zu können, wird für $t > 25$ eine konstante Ehescheidungsrate angenommen, die dem Wert für

Tabelle 5-4 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		

$t = 25$ entspricht.¹²

Um zu simulieren, dass Ehen auch durch den Tod eines Ehepartners enden können, verwenden wir die alters- und geschlechtsspezifischen Sterberaten des Jahres 1999:

$$\delta_{\tau}^m := \frac{\text{Anzahl der 1999 gestorbenen Männer im Alter } \tau}{\text{jahresdurchschnittliche Anzahl von Männern im Alter } \tau}$$

$$\delta_{\tau}^f := \frac{\text{Anzahl der 1999 gestorbenen Frauen im Alter } \tau}{\text{jahresdurchschnittliche Anzahl von Frauen im Alter } \tau}$$

¹²Für 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.

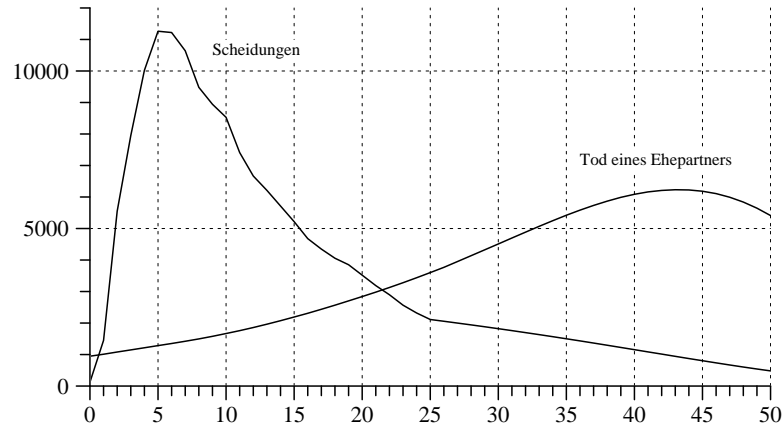


Abb. 5-4 Absolute Häufigkeiten von Ehen, die durch eine Scheidung oder durch den Tod eines Ehepartners beendet wurden, entsprechend den Daten in Tabelle 5-4.

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

$$w_{\tau,\tau',t}^a := n_{\tau,\tau',0} p_{1999}^a(t)$$

Scheidungen, wobei $p_{1999}^a(t)$ die Scheidungsziffer für 1999 bei einer Ehedauer von t Jahren ist; und wenn $t > 25$ ist, gibt es

$$w_{\tau,\tau',t}^a := n_{\tau,\tau',t} \frac{w_{\tau,\tau',25}^a}{n_{\tau,\tau',25}}$$

¹³Die erforderlichen Daten haben wir den Segmenten 685 und 1124-26 der STATIS Datenbank des Statistischen Bundesamts entnommen.

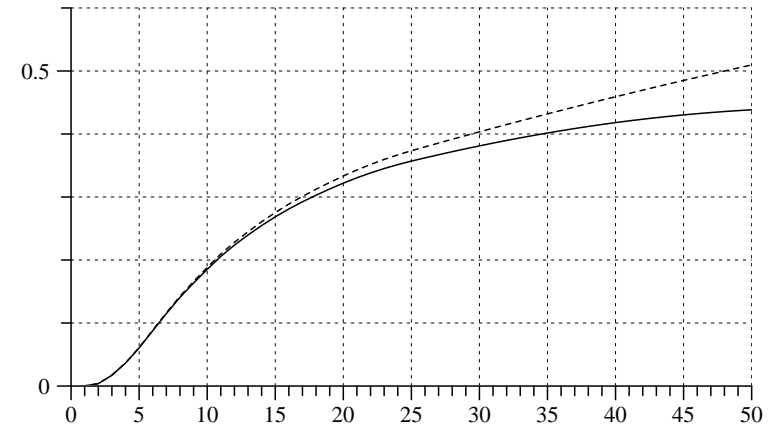


Abb. 5-5 Anteilsfunktion H^a (durchgezogene Linie) und Pseudo-Survivorfunktion $1 - \tilde{G}^a$ (gestrichelte Linie) für die Daten in Tabelle 5-4

Scheidungen. Somit kann man den Bestand der Ehen auf folgende Weise fortschreiben:

$$n_{\tau,\tau',t+1} = n_{\tau,\tau',t} - w_{\tau,\tau',t}^a - w_{\tau,\tau',t}^b$$

Durch Aggregation gewinnt man die in Tabelle 5-4 angegebenen Zahlen. Abb. 5-4 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 5-4 können verwendet werden, um Anteils- und Pseudo-Survivorfunktionen zu berechnen. Abb. 5-5 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.

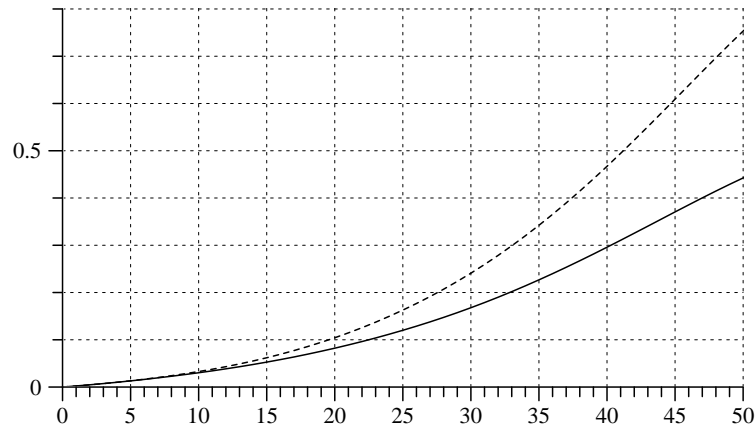


Abb. 5-6 Anteilfunktion H^b (durchgezogene Linie) und Pseudo-Survivorfunktion $1 - \tilde{G}^b$ (gestrichelte Linie) für die Daten in Tabelle 5-4.

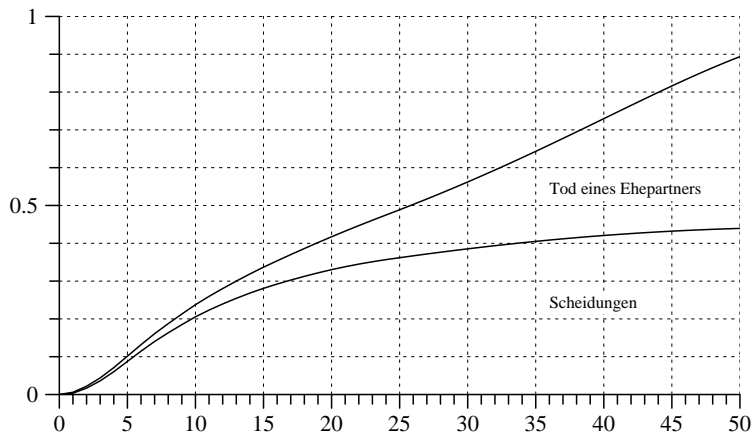


Abb. 5-7 Anteilfunktionen H^a (unterer Linie) und $H^a + H^b$ (obere Linie) für die Daten in Tabelle 5-4.

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. 5-6 zeigt wiederum die Anteilfunktion H^b und die Pseudo-Survivorfunktion $1 - \tilde{G}^b$.

Da Anteilfunktionen im Unterschied zu Pseudo-Survivorfunktionen additiv sind, können sie sinnvoll zusammengefasst werden. Das wird in Abb. 5-7 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.

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.
- 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.
- Blossfeld, H.-P., Hamerle, A., Mayer, K.U. 1986. Ereignisanalyse. *Statistische Theorie und Anwendungen in den Sozialwissenschaften*. Frankfurt: Campus.
- Blossfeld, H.-P., Rohwer, G. 2002. *Techniques of Event History Modeling*, 2nd ed. Mahwah, NJ: Lawrence Erlbaum.
- Diekmann, A., Klein, T. 1991. Bestimmungsgründe des Ehescheidungsrisikos. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 43, 271–290.
- Esenwein-Rothe, I. 1982. *Einführung in die Demographie*. Wiesbaden: Franz Steiner.
- Grusky, D.B. (ed.) 1994. *Social Stratification. Class, Race, and Gender in Sociological Perspective*. Boulder: Westview Press.
- Handl, J. 1985. Mehr Chancengleichheit im Bildungssystem. Erfolg der Bildungsreform oder statistischer Artefakt? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 37, 698–722.
- Hradil, S. 2000. Soziale Ungleichheit. In: G. Reinhold, S. Lamnek, H. Recker (Hg.), *Soziologie-Lexikon*, 589–593. München: Oldenbourg.
- Keyfitz, N. 1977. *Applied Mathematical Demography*, 2nd ed. Berlin: Springer-Verlag.
- 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.
- Kreckel, R. 1992. *Politische Soziologie der sozialen Ungleichheit*. Frankfurt: Campus.
- 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.
- 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*.
- Rousseau, J.-J. 1755. *Abhandlung über den Ursprung und die Grundlagen der Ungleichheit unter den Menschen*. In: *Schriften zur Kulturkritik*, eingel., übers. und hrsg. von K. Weigand. Hamburg: Felix 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.
- 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.
- Swift, A. 2004. *Would Perfect Mobility be Perfect?* *European Sociological Review* 20, 1–11.