# Demography of Germany

## Concepts, Data, and Methods

**G. Rohwer**
**U. Pötter**

**Version 3** **October 2003**

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

goetz.rohwer@ruhr-uni-bochum.de
ulrich.poetter@ruhr-uni-bochum.de

dem.tex October 2003

# Preface

This text is an introduction to concepts and methods of demographic description and analysis. The substantial focus is on the demographic development of Germany, all data refer to this country. The main reason for this focus on a single country is that we want to show how the tools of demography can actually be used for the analysis of demographic problems.

The text consists of two parts. Part I introduces the conceptual framework and explains basic statistical notions. This part also includes a short chapter that explains how we speak of "models" and why we do not make a sharp distinction between "describing" and "modeling" demographic processes. Then follows Part II that deals with data and methods. In the present version of the text, we almost exclusively discuss mortality and fertility data; migration is only mentioned in Chapter 6 and briefly considered in the context of a Leslie model at the end of the text.

In addition to providing a general introduction to concepts of demography, the text also intends to show how to practically work with demographic data. We therefore extensively document all the data used and explain the statistical calculations in detail. In fact, most of these calculations are quite simple; the only exception is the discussion of Leslie models in Chapters 17 and 18 which requires some knowledge of matrix algebra. Except for these chapters, the text has been so written that it may serve as an introduction to elementary statistical methods. The basic approach is identical with the author's *Grundzüge der sozialwissenschaftlichen Statistik* (2001). Virtually no previous knowledge of statistical methods is required for an understanding of the present text. Some notations from set theory that we have used are explained in Appendix A.2.

Most of the data that we have used in this text are taken from publications of official statistics in Germany (Appendix A.1 provides a brief introduction to data sources). We are grateful to Hans-Peter Bosse of the *Statistisches Bundesamt* who provided us with some unpublished materials. We also thank Bernhard Schimpl-Neimanns of ZUMA (Mannheim) who prepared a table with birth data from the 1970 census that we have used for several analyses. In addition, we have used several data files from non-official sources, in particular, data from the *German Life History Study* (Max Planck Institut für Bildungsforschung, Berlin), the *Socio-economic Panel* (Deutsches Institut für Wirtschaftsforschung, Berlin), the *Fertility and Family Survey* (Bundesinstitut für Bevölkerungsforschung, Wiesbaden), the *DJI Family Surveys* (Deutsches Familieninstitut, München), and historical data on mortality prepared by Arthur E. Imhof and his co-workers (1990). All these data sets can be obtained from the *Zentralarchiv* *für Empirische Sozialforschung* in Köln.

The extensive documentation of the data is also intended to allow readers to replicate our calculations. Many calculations can simply be done with paper and pencil. If the amount of data is somewhat larger, one might want to use a computer. Several statistical packages are publicly available. We have used the program TDA which is available from the author's home page: `www.stat.ruhr-uni-bochum/tda.html`. This program was also used to create all of the figures in this text.

For helpful comments and discussions we thank, in particular, Gert Hullen (Bundesinstitut für Bevölkerungsforschung) and Bernhard Schimpl-Neimanns (ZUMA).

Bochum, March 2003
G. Rohwer, U. Pötter

# Contents

# Chapter 1

# Introduction

**1.** The present text is an introduction to concepts and methods of demography, exemplified with data from the demographic development of Germany. The basic idea is to think of a society as a population [Bevölkerung], a set of people. This is the common starting point of almost all demographic investigations and many definitions of demography. As an example, we cite the following definition from a dictionary published by the United Nations (1958, p. 3):[1]

> "Demography is the scientific study of human populations, primarily with respect to their size, their structure and their development."

In a German adaptation of the dictionary by Winkler (1960, p. 17) this definition reads as follows:

> „Die Demographie (Bevölkerungswissenschaft, Bevölkerungslehre) ist die Wissenschaft, die sich hauptsächlich in quantitativer Betrachung mit dem Studium menschlicher Bevölkerungen befaßt: Zahl (Umfang), Gliederung nach allgemeinen Merkmalen (Struktur) und Entwicklung."

Given this understanding, demography is concerned with human populations.

**2.** The focus on populations also provides a view of society. In this view a society simply is a population, a set of people living in some region however demarcated. It might be objected that such a view is greatly incomplete because human societies not only consist of people. It would be difficult, however, to add further characterizations to the definition of a society. All too often the result is no longer a definition but an obscure and dubious statement. The following quotation from Matras (1973, p. 57) can serve as an example:

> "As a working definition, we may say that a *society* is a human population organized, or characterized, by patterns of social relationships for the purpose of collective survival in, and adaptation to, its environment."

This clearly is no longer a definition but an obscure formulation of a dubious assumption. Of course, beginning with the idea that a society is a

---

[1] In this text we distinguish between single and double quotation marks. Single quotation marks are used to refer to linguistic expressions; for example, to say that we are referring to the term 'social structure'. Double quotation marks are used either for citations or to indicate that an expression has no clear meaning or that it is used in a metaphorical way. Within citations, we try to reproduce quotation marks in their original form. If we add something inside a quotation this will be marked by square brackets.

human population, it is possible to describe institutional arrangements and to reflect specific purposes possibly served by such arrangements. But this can only result from an investigation and not anticipated in a definition.

**3.** The demographic view of society is closely linked with a statistical approach. Demography, to a large extent, is the application of statistical methods to study the development of human populations. This is the main idea which accompanied the history of demography from its beginning.[2] Conversely, demography inspired many developments in statistics. The fundamental role played by the word 'population' in the statistical literature is but one indicator. This term has often been used to define statistics; as an example, we refer to Maurice Kendall and Alan Stuart, who begin their "Advanced Theory of Statistics" (1977, p. 1) as follows:

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

As far as demography applies a statistical view to human populations these remarks also contribute to an understanding of demography. The concern is with properties of populations, not with their individual members.

**4.** Since populations do not have properties in an empirical sense of the word, one also needs to understand how demographers construct such properties by using statistical concepts. This will be discussed at length in subsequent chapters. Here we only mention that statistically construed properties of populations are always conceptually derived from properties of their individual members. For example, referring to a human population, each of its members can be assigned a sex and the population can be characterized then by two figures reporting the proportion of male and female members. This also provides a simple example of a statistical distribution: to every individual property is assigned the relative frequency (proportion) of its occurrence in a population.

**5.** Almost always this is also meant when statisticians, including demographers, speak of the "structure" of a population: an account of the frequencies of some individual properties in a population. Here are some examples from the demographic literature:

> "Demography is the discipline that seeks a statistical description of human populations with respect to (1) their demographic structure (the number of the population; its composition by sex, age and marital status; statistics of families, and so on) at a given date, and (2) the demographic events (births, deaths, marriages and terminations of marriages) that take place in them." (Pressat 1972, p. 1) „Unter der demographischen Struktur einer Bevölkerung versteht man ihre Aufgliederung nach demographischen Merkmalen." (Feichtinger 1973, p. 26) „Die Struktur einer bestimmten Bevölkerung wird beschrieben durch die absolute Zahl der Einheiten sowie die Verteilung der jeweils interessierenden Merkmalsausprägungen bei den Einheiten dieser Bevölkerung zu einem bestimmten Zeitpunkt $t$." (Mueller 1993, p. 2)

We mention that sociologists use the word 'structure' often in different meanings. A frequent connotation is that "structure" in some way determines conditions for the behavior of the individual members of a society. It is important, therefore, that this can not be said of a statistical distribution.

**6.** In order to sensibly speak of conditions one would need to think of the individual members of a society as being actors whose possible actions depend in some way on a given environment. The statistical view is quite different. Not only has statistics no conceptual framework for a reference to actors; as shown by the above quotation from Kendall and Stuart, there also is no reference to individuals. Instead, the focus is on populations. This was clearly recognized, for example, by Wilhelm Lexis:

> „Bei der Bildung von Massen für die statistische Beobachtung verschwindet das Individuum als solches, und es erscheint nur noch als eine Einheit in einer Zahl von gleichartigen Gliedern, die gewisse Merkmale gemein haben und von deren sonstigen individuellen Unterschieden abstrahiert wird." (Lexis 1875, p. 1)

The same idea was expressed by another author in the following way:

> „Innerhalb der Demographie interessiert eine individuelle Biographie nur als Element der kollektiven Geschichte der Gruppe, zu welcher das Individuum gehört." (Feichtinger 1979, p. 13)

**7.** The method to characterize populations by statistical distributions (of individual properties) is obviously quite general. Almost all properties which can sensibly be used to characterize individuals can also be used to derive statistical distributions characterizing populations. Statistical methods are therefore used not only in demography but more or less extensively in almost all empirical social research. In fact, there is no clear demarcation between demography and other branches of social research. Some authors have therefore proposed to distinguish between demographic analysis in a narrow sense, also called *formal demography*, and a wider

[2] For an informative overview see Lorimer (1959).

scope, often called *population studies*.[3] Given this distinction, the current text is only concerned with formal demography.[4] The following explanation is taken from a widely known textbook by Shryock and Siegel (1976, p. 1):

"Formal demography is concerned with the size, distribution, structure, and change of populations. *Size* is simply the number of units (persons) in the population. *Distribution* refers to the arrangement of the population in space at a given time, that is, geographically or among various types of residential areas. *Structure*, in its narrowest sense, is the distribution of the population among its sex and age groupings. *Change* is the growth or decline of the total population or of one of its structural units. The components of change in total population are births, deaths, and migrations."

This explanation of formal demography is quite similar to the understanding of *Bevölkerungsstatistik* in the older German literature.[5] It is also similar to the definition of demography cited at the beginning of this chapter.[6] The quotation also shows once more that the term 'structure' is used synonymously with 'statistical distribution'. On the other hand, the word 'distribution' is here not used to refer to a statistical distribution but to "the arrangement of the population in space". — This topic, including internal migration, will not be systematically discussed in the present text. On the other hand, demographic data as provided by official statistics, are always limited to bounded regions, historically defined as "nation states". One therefore cannot avoid to take into account in- and out-migration. This is true, in particular, when dealing with the demographic development in Germany that is the empirical concern of the present text.

---

[3]See Hauser and Duncan (1959, pp. 2-3), and Shryock and Siegel (1976, p. 1). In the older German literature a similar distinction was made between *Bevölkerungsstatistik* and *Bevölkerungslehre*, see v. Bortkiewicz (1919).

[4]This is not to deny the importance of many questions discussed under the heading of population studies. However, we will not try to make this a special sub-discipline of social science but consider a reflection of demographic developments as being an essential part of almost all investigations of social structure.

[5]As an example, we cite L. v. Bortkiewicz (1919, p. 3): „Soll aber eine besondere wissenschaftliche Betrachtung über die Bevölkerung angestellt werden, so kann es sich dabei unmöglich um eine Erörterung alles dessen handeln, was ihr Wohl und Wehe irgendwie angeht. Es gilt hier vielmehr, zunächst die Bevölkerung als unterschiedslose Menschenmasse ins Auge zu fassen, ihre räumliche Verteilung und die zeitlichen Änderungen ihrer Größe zur Darstellung zu bringen, sodann aber auch ihre Gliederung nach gewissen natürlichen Merkmalen, vor allem nach dem Geschlecht und nach dem Alter, klarzulegen und im Anschluß hieran auf die unmittelbaren Ursachen ihres jeweiligen Standes zurückzugehen, als welche sich in erster Linie die Geburten und die Todesfälle und in zweiter Linie die Wanderungen darstellen. Damit ist der Gegenstand der *Bevölkerungsstatistik* im althergebrachten Sinne dieses Wortes angedeutet."

[6]Since there is a common conceptual framework, it seems not necessary, as proposed by the cited dictionary, to distinguish explicitly between formal demography and population statistics.

# Part I

# Conceptual Framework

## Chapter 2

# Temporal References

The chapters in Part I of this text briefly introduce the conceptual framework used to develop a demographic view of society. The main conceptual tool is the notion of a ‘demographic process’. An explicit definition will be given in the next chapter. The present chapter deals with a preliminary question that concerns a suitable temporal framework. Technically, one uses a *time axis* that allows to temporally locate events; but how to represent a time axis? There are two general approaches:

a) One approach treats time as a sequence of temporal locations (e.g., minutes or days) and represents time by integral numbers with an arbitrarily fixed origin. This is called a *discrete time axis*.

b) Another approach treats time as a continuum (a “continuous flow of time”) and represents a time axis by the set of real numbers. This is called a *continuous time axis*.

Since a time axis is used to provide a conceptual framework for the representation of phenomena which occur “in time”, the decision for one or the other of the two approaches should depend on the kind of phenomena that one wants to describe and analyze. In demographic and, more general, social research, the primary phenomena are events, for example, birth and death events. Thinking in terms of a continuous time axis would require to conceive of events as “instantaneous changes”. While this approach is quite widespread in the demographic literature,[1] it conflicts with the simple fact that events always need some time to occur. As we will try to show in the present chapter, this suggests to represent a time axis by real numbers but to think of temporal locations not as “time points” but as temporal intervals. Within such a framework a discrete time axis arises as a special case from an assumption of intervals of equal length. This assumption will be sufficient for most practical purposes and also greatly simplifies the mathematics. Therefore, in later chapters, we most often use a discrete time axis with temporal locations to be understood as temporal intervals having an equal length. In the present chapter we first discuss our notion of events and how thinking in terms of events allows temporal references. We then deal with possibilities of quantifying temporal references.

---

[1]Examples of textbooks that use a continuous time axis for temporal references are, e.g., Keyfitz (1977), and Dinkel (1989).

## 2.1 Events and Temporal Locations

**1.** We all have learned to make temporal references by using clocks and calendars and to think of time as a linearly ordered time axis. But leaving aside for the moment clocks and calendars, what enables us to speak about time? One possible approach begins with events. This notion is extremely general and therefore quite difficult to make precise. However, for the present purpose, it seems possible to neglect philosophical discussions and simply take a common sense view of events.[2] The following four points seem to be essential.

- The occurrence of an event always involves one or more objects.
- Each event has some finite temporal duration.
- For many events one can say that one event occurred earlier than another event.
- Events can be characterized, and classified, by using the linguistic construct of kinds of events.

**2.** Using these assumptions it seems, first of all, important to distinguish between *events* and *kinds of events*. An event is unique; it occurs exactly once. On the other hand, several events can be of the same kind, for example, marriages. Therefore, characterizing an event as being of a certain kind does not give a unique description. Furthermore, an event does not necessarily belong to only a single kind of event. Most often one can characterize an event as an example of several different kinds of events. For example, an event that is a marriage can also be a first marriage.

**3.** While common language clearly distinguishes between objects and events, one might well think of a certain correspondence between, on the one hand, objects and their properties, and on the other hand, events and kinds of events. This has led some authors (e.g., Brand 1982) to think of objects and events as being ontologically similar. Even without defending this position, we will assume that talking of events always implies a reference to objects. The idea is that it should be possible to associate, with each event, *some* objects that are involved in the event. Of course, these objects need not be individuals in the sense of behavioral units.

**4.** Following the common sense view of events it also seems obvious that events occur "in time". The notion of event therefore provides a way to think about time. We assume that one can associate with each event a temporal location. In the following, we will use the letter $e$ to refer to an event and $t(e)$ to denote its temporal location. $t(e)$ will be called the

[2] For related philosophical discussion see Hacker (1982) and Lombard (1986).



*t-location* of the event $e$. While a strict definition cannot be given it seems important to think of $t$-locations *not* as being "time points". Quite to the contrary, one of the most basic facts about events is that each event has a certain temporal duration. This is not only obvious when we think of standard examples of events, but seems logically implied if we think of events in terms of change. Change always needs some amount of time. This also has an important further implication: only when an event *has occurred* and, consequently, when it has become a fact belonging to past history, can we say that the event has, in fact, occurred. We cannot say this *while* the event is occurring.[3]

**5.** That one thinks of events in terms of change is quite essential for the common sense view of events that we try to follow here. Without a change nothing occurs. Fortunately, one need not be very specific about what kinds of changes occur. Also, whether these changes occur "continuously" or "instantaneously" is quite unimportant as long as we require that the event has some temporal duration. The event is defined by what happened during its occurrence and must therefore be taken as a whole. Of course, one might be able to give a description of the event in terms of smaller sub-events; but these will then simply be different events. An event is semantically indivisible. In particular, the beginning of an event is not itself an event, and consequently has no $t$-location.

**6.** Finally, it is important that one can often say of two events that one occurred earlier than the other. Of course, this cannot always be said. One event may occur while another is occurring. However, there are many clear examples where we have no difficulties to say that one event occurred earlier than another one. We therefore assume that the following partial order relations are available when talking about events ($e$ and $e'$ are used to denote events):

$$
\begin{array}{lll}
e \preccurlyeq e' & \text{meaning:} & e' \text{ begins not earlier than } e \\
e \lhd e' & \text{meaning:} & e' \text{ begins not before } e \text{ is finished} \\
e \sqsubseteq e' & \text{meaning:} & e \text{ occurs while } e' \text{ occurs}
\end{array}
$$

We also write $e \sqsubset e'$ if $e \sqsubseteq e'$ and not $e' \sqsubseteq e$. All relations are only partial order relations. Nevertheless, they can be used to define corresponding

[3] Thinking of human actions as particular types of events, this implication has been described by Danto (1985, p. 284) as follows: "Not knowing how our actions will be seen from the vantage point of history, we to that degree lack control over the present. If there is such a thing as inevitability in history, it is not so much due to social processes moving forward under their own steam and in accordance with their own natures, as it is to the fact that by the time it is clear what we have done, it is too late to do anything about it."

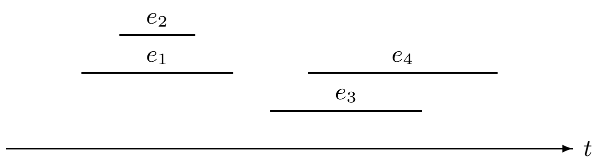



**Fig. 2.1-1** Illustration of order relations between four events on a qualitatively ordered time axis.

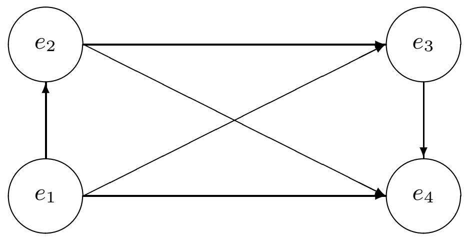



**Fig. 2.1-2** Graph illustration of '$\preccurlyeq$' relation between the four events shown in Figure 2.1-1.

relations between the $t$-locations of events. We use the same symbols:

$$
\begin{aligned}
t(e) \preccurlyeq t(e') &\iff e \preccurlyeq e' \\
t(e) \lhd t(e') &\iff e \lhd e' \\
t(e) \sqsubseteq t(e') &\iff e \sqsubseteq e'
\end{aligned}
$$

We will say that a set of events is equipped with a *qualitatively ordered time axis* if these three relations are available.

**7.** As an illustration consider the four events in Figure 2.1-1 where one can find the following order relations:

$$
\begin{aligned}
&e_1 \preccurlyeq e_2,\ e_1 \preccurlyeq e_3,\ e_1 \preccurlyeq e_4,\ e_2 \preccurlyeq e_3,\ e_2 \preccurlyeq e_4,\ e_3 \preccurlyeq e_4 \\
&e_1 \lhd e_3,\ e_1 \lhd e_4,\ e_2 \lhd e_3,\ e_2 \lhd e_4 \\
&e_2 \sqsubseteq e_1
\end{aligned}
$$

Of course, on a qualitatively ordered time axis, the lengths of the line segments used in Figure 2.1-1 to represent events do not have a quantitative meaning in terms of duration. This becomes clear if one represents the order relations between events by means of a directed graph. This is illustrated in Figure 2.1-2 where the arcs represent the $\preccurlyeq$ relation between the events.



### Composing Events

**8.** Our language is quite flexible to compose two (or more) events into larger events. As an example one can think of clock ticks as elementary events. It seems quite possible to think also of two or more successive clock ticks as events. To capture this idea formally, one can introduce a binary operator, $\sqcup$, that allows to create (linguistically) new events. The rule is: If $e$ and $e'$ are two events then also $e \sqcup e'$ is an event. Events created by using the operator $\sqcup$ will be called *composed events*. When classes of events are considered, one can assume that these are closed with respect to $\sqcup$ by extending the time order relations defined above for composed events in the following way:

$$
\begin{aligned}
e \sqcup e' \preccurlyeq e'' &\iff e \preccurlyeq e'' \text{ or } e' \preccurlyeq e'' \\
e \sqcup e' \lhd e'' &\iff e \lhd e'' \text{ and } e' \lhd e'' \\
e \sqcup e' \sqsubseteq e'' &\iff e \sqsubseteq e'' \text{ and } e' \sqsubseteq e''
\end{aligned}
$$

This also allows to introduce the notion of an elementary event. A possible definition would be that an event, say $e$, is an *elementary event* if there is no other event, $e'$, such that $e' \sqsubset e$. Using this definition, one conceives of elementary events as not being divisible into smaller events with respect to a class of events.

**9.** It might seem questionable whether elementary events do exist. When describing an event it often seems possible to give a description in terms of smaller and smaller sub-events, without definitive limit. However, we are not concerned here with the ontological status of events. Regardless of whether it is possible to give *descriptions* of events in terms of smaller sub-events, when talking about events one cannot avoid to assume *some* "universe of discourse" that provides the necessary linguistic tools. This justifies the assumption that one can single out a finite number of elementary events from any finite collection of events.

**10.** Interestingly, it seems not possible to define a converse operation, $\sqcap$, by using the interpretation that $e \sqcap e'$ occurs *while* both events, $e$ and $e'$, are occurring. The reason is that we should be able to say that an event has, in fact, occurred as soon as the event no longer occurs. But this condition will in general not hold for $e \sqcap e'$ because one can only say that $e$ *and* $e'$ occurred when both are over. There is, therefore, no obvious way to define an algebra of events.

## 2.2 Duration and Calendar Time

**1.** Having introduced the idea of a qualitatively ordered time axis, one can think about possibilities to quantify temporal relations. We begin with an elementary notion of duration. If an event, $e$, occurs while another event,

$e'$, is occurring ($e \sqsubseteq e'$), one can say that the duration of $e$ is not longer than the duration of $e'$. This introduces a partial ordering of events with respect to duration and can be used as a starting point for a quantitative concept of duration.

a) In order to measure the duration of an event $e$ we count the number of pairwise non-overlapping events $e'$ such that $e' \sqsubseteq e$.[4] The maximal number of those events can be used as a discrete measure for the duration of $e$ having $t$-locations as units. As an implication, all elementary events will have a unit duration.

b) In the same way one can measure the duration between two events, say $e$ and $e'$. Again, simply determine the maximal number of pairwise not overlapping events, $e''$, such that $e \lhd e'' \lhd e'$. If such an event cannot be found we say that $e'$ *immediately follows* $e$.[5]

**2.** These definitions make duration dependent on the number of events that can be identified in a given context. An obvious way to cope with this dependency is to enlarge the number of events that can be used to measure duration. This is done by using clocks. Defined in abstract terms, a clock is simply a device that creates sequences of (short) events. Then, if a clock is available when an event occurs, its duration can be measured by counting the clock ticks that occur while the event is occurring. Let $e$ be the event whose duration is to be measured and let $c_n$ denote an event composed of $n$ clock ticks. One might then be able to find a number, $n$, such that

$$t(c_n) \sqsubseteq t(e) \sqsubseteq t(c_{n+1})$$

This will allow to say that the duration of event $e$ is between $n$ and $n+1$ clock ticks.

**3.** Many different kinds of clocks have been invented,[6] and this has led to the difficult question how to compare different clocks with respect to accuracy. Fortunately, we are not concerned here with the problem of how to construct good clocks. We can simply use the clocks that are commonly used in daily life to characterize, and coordinate, events. We are, however, concerned with the problem how to numerically represent durations, independent of the device actually used for measurement. Since

[4] It will be said that two events, $e'$ and $e''$, do not overlap if $e' \lhd e''$ or $e'' \lhd e'$.

[5] In fact, we then do not have any reason to believe in a duration between $e$ and $e'$. Leibniz (1985, p. 7) made this point by saying: "Ein grosser Unterschied zwischen Zeit und Linie: der Zwischenraum zwischen zwei Augenblicken, zwischen denen sich nichts befindet, kann auf keine Weise bestimmt werden und es kann nicht gesagt werden, wieviele Dinge dazwischen gesetzt werden können; [...] In der Zeit berühren sich daher die Momente zwischen denen sich nichts ereignet."

[6] See, e.g., Borst (1990).

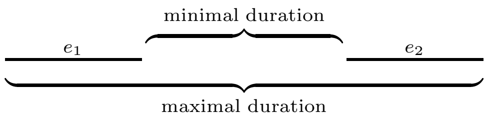


clocks with different accuracies do exist we should find a numerical representation that is independent of any specific clock. This suggests to use intervals of real numbers to represent durations. Since duration is always positive a sensible choice is

$$\mathbf{R}_{]+]} := \{\, ]\,a, b\,] \mid 0 \le a < b, a, b \in \mathbf{R}\}$$

This representation is intended to capture both conceptual and empirical indeterminacy.[7]

**4.** Thinking of events one needs to distinguish between $t$-locations and durations. The duration of an event tells us how long the event lasted while the $t$-location of an event provides information about the location of the event in a set of events equipped with the partial orders, $\preccurlyeq$, $\lhd$, and $\sqsubseteq$. The basic tool for the introduction of quantitative statements about $t$-locations is a *calendar*. Calendars can be defined by specifying a base event and using the concept of duration between events. This allows to locate every event by providing information about the (positive or negative) duration between the event and the base event of the calendar. To make this idea precise one only needs a definition of duration between events. In principle, one can follow the approach already mentioned above. Then, having available a clock, the duration between two events, say $e$ and $e'$, can be measured by counting the number of non-overlapping clock events having a $t$-location between $e$ and $e'$. However, this definition of duration between events is not fully satisfactory because the events also have a duration. This fact obviously creates some conceptual indeterminacy and it seems therefore preferable to proceed in terms of a minimal and maximal duration as follows:




This suggests to use again the set of positive real intervals, $\mathbf{R}_{]+]}$, now for the numerical representation of duration between events.

**5.** The main conclusion is that each event refers to time in two different ways.

a) First, events have an inherent duration. This qualitative notion can be represented numerically by positive real intervals. It will be assumed, therefore, that one can associate with each event, $e$, a positive duration

$$\mathrm{dur}(e) \in \mathbf{R}_{]+]}$$

[7] A fuller exposition of the idea to use intervals for the representation of data having both empirical and conceptual indeterminacies, including a discussion of statistical methods based on this kind of data representation, has been given elsewhere, see Rohwer and Pötter (2001, Part V).

Of course, the interpretation of dur($e$) requires information about the kind of elementary events that have been used to measure duration. If all elementary events are of the same kind, as is normally the case when using clocks, one of these events (or a suitably defined composed event) provides a sensible unit of duration. In any case, it will most often be possible to assume that duration can be measured in some standard units like seconds, days, months, or years.

b) Second, one can associate with each event a $t$-location that provides information about the place of the event in the order of time. Again, this is a purely qualitative notion defined with respect to three partial order relations between events. However, one can introduce a quantitative representation of the duration between events, by using real intervals. Then, for each pair of events, $e$ and $e'$, one can use

$$\mathrm{dur}(e, e') \in \mathbf{R}_{]+]}$$

to represent the duration between the two events.

Finally, one can introduce a calendar as a quantitative representation of $t$-locations. Having specified a base event, $e^\dagger$, one can represent the $t$-location of any other event, say $e$, by the duration between $e$ and $e^\dagger$. Then, if $e^\dagger \preccurlyeq e$, $\mathrm{dur}(e^\dagger, e)$ provides a quantitative representation of the $t$-location of $e$ with respect to the calendar defined by $e^\dagger$.[8] So one finally can use a single numerical representation, $\mathbf{R}_{]+]}$, both for the durations and $t$-locations of events.

## 2.3 Calculations with Calendar Time

**1.** Calendars, like methods of measuring time, changed considerably in the course of history. The choice of a suitable base event and the use of different clocks signify the main differences between the historical calendars used. The idea that nature provides the human experience with many periodic phenomena that, in some sense, should be accommodated by a calendar often provided reasons for calendar reforms.[9] Today, in European countries, the most often used calendar is the *Gregorian calendar* that was introduced by Gregor XII in 1582. A German encyclopedia (Brockhaus, 20th ed. 2001, vol. 11, p. 367) provides the following explanations:

„Der heutige *bürgerliche Kalender* basiert auf dem gregorian. K. Er ist demnach ein Schalt-K. mit einem Gemeinjahr von 365 Tagen. Ein Schalt-Zyklus von 400 K.-Jahren hat 146097 K.-Tage. Ein mittleres K.-Jahr hat somit 365,2425 Tage, ist also um 26 s länger als das trop. Jahr.

[8] If $e \preccurlyeq e^\dagger$, one can use the same approach by allowing for negative real intervals.

[9] The history of calendars is described in several books, see, e.g., Borst (1990) and Richards (1998).



Die K.-Jahre werden ab Christi Geburt gezählt, beginnend mit dem Jahr 1 nach Christus (Abk. n. Chr.). Die K.-Jahre vor dem K.-Jahr 1 werden mit 1 beginnend in die Vergangenheit nummeriert und durch den Zusatz >vor Christus< (Abk. v. Chr.) gekennzeichnet. Ein K.-Jahr 0 gibt es nicht (außer für den Bereich der Astronomie).

Ein K.-Jahr wird in 12 Monate unterteilt, von denen die Monate Januar, März, Mai, Juli, August, Oktober, Dezember 31 Tage haben, die Monate April, Juni, September, November 30 Tage und der Monat Februar 28 oder in einem Schaltjahr 29 Tage. Unabhängig hiervon wird das K.-Jahr in K.-Wochen zu je 7 Wochentagen unterteilt, von denen es 52 oder 53 hat. Als erste K.-Woche eines Jahres zählt diejenige Woche, in die mindestens 4 der ersten 7 Januartage fallen (dabei gilt der Montag als erster Tag der K.-Woche). Ist das nicht der Fall, so zählt diese Woche als letzte K.-Woche des vorausgehenden K.-Jahres.“

Many readers of this text will be familiar with this calender and know how it can be used for temporal references. Some difficulties only arise in the calculation of durations for longer periods. For example, how long is the period beginning June 13, 1911, and ending February 7, 2001, in days, weeks, months?

**2.** To answer this kind of question, an often used method consists in transforming Gregorian dates into numbers defined by an algorithm that simply counts days.[10] The idea is to first fix some day in the Gregorian calendar to become day 0 in the algorithmic calendar, and then to develop an algorithm that allows, for any other day in the Gregorian calendar, to calculate its temporal distance from day 0. As an example, we describe an algorithm proposed by Fliegel and van Flandern (1968) that uses the Gregorian Date November 24, in the year 4714 B.C., as day 0.

**3.** The algorithm consists of two parts. Given a Gregorian date by $d$ (day), $m$ (month) and $y$ (year), one algorithm is used to calculate a corresponding Julian day which we denote by $k$. In a first step, one calculates two auxiliary quantities:

$$a = (m - 14)/12 \quad \text{und} \quad b = y + a + 4800$$

Then the following formula provides the Julian day $k$:

$$k = d - 32075 + 1461\,\frac{b}{4} + 367\,\frac{m - 2 - 12\,a}{12} - 3\,\frac{\frac{b+100}{100}}{4}$$

It should be noticed that all calculations must be done in integer arithmetic. This means that all (intermediate) floating point results must be truncated to the next integer. For example, $25/9 = 2$.

[10] Such an algorithm is often called a *Julian calendar*. The name goes back to Joseph Scaliger who, in the year 1583, first proposed this kind of algorithmical calendar. In fact, it has nothing to do with the calendar, also often called a Julian calendar, that was introduced by Julius Caesar in 46 B.C. [v. Chr.].

**4.** Conversely, a second algorithm is used to calculate the Gregorian day $d$, month $m$, and year $y$, that correspond to a given Julian day $k$. The calculations consist of the following steps:

$$
\begin{aligned}
p &= k + 68569 \\
q &= (4\,p)/146097 \\
r &= p - (146097\,q + 3)/4 \\
s &= 4000\,(r+1)/1461001 \\
t &= r + 31 - (1461\,s)/4 \\
u &= (80\,t)/2447 \\
v &= u/11 \\
d &= t - (2447\,u)/80 \\
m &= u + 2 - 12\,v \\
j &= 100\,(q-49) + v + s
\end{aligned}
$$

The following table shows a few examples.[11]

| $d$ | $m$ | $j$ | $k$ |
|---|---|---|---|
| 1 | 1 | 1 | 1721426 |
| 31 | 12 | 0 | 1721425 |

| $d$ | $m$ | $j$ | $k$ |
|---|---|---|---|
| 1 | 1 | 2001 | 2451911 |
| 31 | 12 | 2000 | 2451910 |

The table also shows that the first year B.C. is given by $y = 0$, not by $y = -1$ as the explanation of the Gregorian calendar cited above might suggest.

## 2.4 Limitations of Accuracy

**1.** Depending on the purpose, temporal references use different units of time: days, weeks, months, years, also smaller units like minutes and seconds. When recording statistical data, a suitable choice of temporal units depends on the kinds of phenomena to be captured by the data. For example, to record the age of a person one can use age in completed years, and there are rarely occasions to use a finer time scale. One exception is the analysis of mortality of newborn children. On the other hand, years are not well suited to record the length of unemployment spells. We would like to distinguish between persons who are unemployed, for example, less than 3 or longer than 6 or 12 months. This suggests to measure unemployment durations not in years, but at least in months. A finer time scale seems

[11] Most statistical packages provide some means to convert between Gregorian dates and Julian days. TDA, for example, provides operators that directly use the algorithms of Fliegel and van Flandern as described above. SPSS uses a similar algorithm but a different base day (October 14, 1582).



to introduce but irrelevant information, since most jobs end at the end of calendar months and start at the beginning of calendar months.

**2.** There are thus no natural temporal units, neither to locate events in historical time nor to measure durations. Moreover, the precision of data recording might be limited. While an observer might be able to measure the duration of a football game in terms of minutes, a demographer can not determine the age of a person by using a clock. He sometimes can rely on records, like birth certificates, but most often needs to ask persons for their age and the dates of other potentially interesting events. The accuracy of demographic data then also depends on person's memory and the temporal framework that is used by them to temporally locate events. While these are empirical limits to the accuracy of demographic data, there also are theoretical limits. One of these limits, already mentioned, derives from the fact that demographic events always have some intrinsic duration. Even if it would be possible to provide a birth date exactly to the hour, or to measure marriage duration in days, the accuracy of the data would be useless because there is no theoretical argument that might justify a distinction. Why should one want to distinguish between two marriages, one of them lasting 5734 and the other one 5735 days? Even if true, it would be misleading to say that the second marriage lasted longer than the first one. As another example suppose that one has a job just in February while someone else has a job just in July. Then the length of employment of the first person is three days shorter than that of the second person, but both get the same renumeration, social security insurances, etc. The point is simply that data should serve to report relevant differences, not just any differences.

# Chapter 3

# Demographic Processes

Since demography is concerned with describing and modeling the development of human populations it is dealing with *Gesamtheiten* embracing many individuals. Their size may vary depending on the spatial or temporal demarcation. However, in most cases already its sheer size makes a direct and complete observation impossible. While it might be possible, at least in principle, to empirically approach each individual member of a population, the same is not true for the population as a whole. For example, we might want to talk about the totality of people who are currently living in Germany. While it is possible to empirically approach any number of individual persons, no one is able to observe the population as a whole. Put somewhat differently, the population as a whole is not an empirical object but a conceptual construction. This is not to deny that all of its members, and consequently also the population, really exists. However, the statement says that one needs some kind of representation of the population in order to have an object that one can think of and talk about. This chapter begins with the introduction of a rudimentary conceptual framework and some notations that allow to make the required representations explicit.

## 3.1 A Rudimentary Framework

**1.** In order to think of a human population one first needs a spatial and temporal context. To specify a temporal context we assume a discrete time axis as discussed in Chapter 2. Such a time axis can be thought of as a sequence of temporal locations which may be days, months, or years. To provide a symbolic representation we use the notation[1]

$$\boldsymbol{\mathcal{T}} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

The elements $t \in \boldsymbol{\mathcal{T}}$ are not just numbers but represent temporal locations. For example, 0 represents a 'day 0', 1 represents a 'day 1', and so on. Since we want to develop a general conceptual framework that can serve both for descriptions and models, the duration of the temporal locations will be left unspecified. We only assume that all temporal locations have the same duration, and the existence of a temporal ordering in the following sense:

[1]In this text we distinguish between '=' and ':='. Preceeding the equality sign by a colon shall mean that the expression on the left-hand side will be defined by the expression on the right-hand side. In contrast, the equality sign without a colon states an equality that requires both sides to be defined beforehand.



temporal location $t$ precedes, and is followed by, the temporal location $t+1$ (for any $t \in \boldsymbol{\mathcal{T}}$). For the moment, we do not require any specific link with historical (calendar) time.

**2.** In a similar way one can introduce a set of spatial locations, in the following denoted by the symbol $\boldsymbol{\mathcal{S}}$. The idea is that the elements of $\boldsymbol{\mathcal{S}}$ provide a spatial context for human individuals. The spatial locations can be defined in many different ways, for example, by referring to geographical or political demarcations. But like temporal locations, spatial locations only need to be specified when it is required by the specific empirical purpose. For the moment, we also do not introduce any kind of topology or metric. Furthermore, we do not make any assumptions about the number of spatial locations in $\boldsymbol{\mathcal{S}}$. In particular, we allow for the limiting case that $\boldsymbol{\mathcal{S}}$ only contains a single spatial location. We only require that our space is complete, in the sense that spatial mobility can only occur across the spatial locations given by $\boldsymbol{\mathcal{S}}$.

**3.** Having introduced a temporal and spatial context, one can think of people who live in this context. The symbol $\Omega_t$ will be used to represent the totality of people who live in the space $\boldsymbol{\mathcal{S}}$ during the temporal location $t \in \boldsymbol{\mathcal{T}}$.[2] The sets $\Omega_t$ are finite, and so one can sensibly speak of the number of people living during the temporal locations $t$. The temporal index $t$ is necessary because the composition of the population sets $\Omega_t$ changes through time. In each temporal location, some people might die and others might be born. Also, if two sets, $\Omega_t$ and $\Omega_{t'}$, contain the same number of people, they might not be identical. Referring to a set of people implies that one is able to identify and distinguish its members. In addition, we assume that, for each individual $\omega \in \Omega_t$, there is exactly one spatial location $s \in \boldsymbol{\mathcal{S}}$ where $\omega$ is currently living.

**4.** One further question needs consideration. Regardless of their specification, temporal locations have some inherent duration. People are born, marry or die *during* a temporal location. One therefore needs a convention about starting and ending times for the membership in the sets $\Omega_t$. Our convention will be as follows: If a child is born in a temporal location $t$, it will be considered as a member of $\Omega_t$ but not of any earlier population set; conversely, if a person dies in a temporal location $t$, she will be regarded as a member of $\Omega_t$ but not of any later population set. How this convention relates to the measurement of age will be discussed in Section 3.4.

**5.** This then is our rudimentary context: a space $\boldsymbol{\mathcal{S}}$ where people live, a time axis $\boldsymbol{\mathcal{T}}$ that allows temporal references, and population sets $\Omega_t$ that contain (fictitious) names of people living in the space $\boldsymbol{\mathcal{S}}$ during the

[2]More precisely, the elements of $\Omega_t$ are not human individuals but (fictitious) names. However, having understood the distinction it should be possible to refer to the elements of $\Omega_t$ as individuals without creating confusion.

temporal locations defined by $\mathcal{T}$. While this context is quite abstract and certainly requires a lot of specifications to become empirically useful, it already allows to formulate the two basic demographic questions: How are the population sets $\Omega_t$ changing across time, and how do these changes depend on births, deaths, and migrations?

### A Fictitious Illustration

**6.** A small fictitious example can serve to illustrate the conceptual framework. Imagine a small island with only a few inhabitants.[3] Sometimes a new child is born or one of the inhabitants dies, and sometimes someone leaves the island or comes from outside as a new member of the island community. How to get more information? This is the task of a chronicler who, more or less systematically, writes down what is happening on the island. His chronicle may contain entries for any kinds of event, but here we are only interested in elementary demographic events. So we assume that the chronicle gets an entry whenever a child is born, one of the inhabitants dies, a person enters the island from outside and becomes a new inhabitant, or one of the inhabitants leaves the island.

**7.** Obviously, the chronicle must begin at some point in time. We assume that the records begin in 1960 and are continued until 1990. In the first year, the chronicler makes a list of all people who are currently living on the island and also records their age and sex. This list might look as follows:[4]

| Name | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_4$ | $\omega_5$ | $\omega_6$ | $\omega_7$ | $\omega_8$ | $\omega_9$ | $\omega_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Age | 40 | 38 | 4 | 16 | 63 | 70 | 25 | 8 | 63 | 11 |
| Sex | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |

This is the stocktaking in the first year, 1960, when the chronicle begins. In the following years the chronicler adds entries whenever a demographic event occurs. The complete chronicle, up to the year 1990, might then look as shown in Table 3.1-1.

**8.** It is quite possible that the chronicler not only records demographic events but adds a lot more information about the life of the people on the island and their living conditions. Since, in this example, the number of people is very small one also can imagine that the chronicler creates his chronicle not simply as a list of records, but uses some literary form and

[3] For example, one may think of Hallig Gröde, a small island at the west coast of Schleswig-Holstein in northern Germany. With currently 16 people living on this island, it is the smallest municipality [Gemeinde] in Germany.

[4] Age is recorded as usual in completed years; sex is represented by numbers, 0 representing 'male' and 1 representing 'female' individuals.



**Table 3.1-1** Chronicle of our fictitious island.

| Year | Name | Age | Sex | Kind of event |
|---|---|---|---|---|
| 1961 | $\omega_4$ | 17 | 0 | leaves the island |
| 1963 | $\omega_6$ | 73 | 0 | dies |
| 1964 | $\omega_{11}$ | 30 | 0 | becomes new inhabitant |
| 1966 | $\omega_{12}$ | 0 | 1 | is born |
| 1970 | $\omega_{13}$ | 0 | 0 | is born |
| 1971 | $\omega_9$ | 74 | 1 | dies |
| 1975 | $\omega_8$ | 23 | 0 | leaves the island |
| 1975 | $\omega_{14}$ | 26 | 1 | becomes new inhabitant |
| 1980 | $\omega_{15}$ | 0 | 0 | is born |
| 1982 | $\omega_{16}$ | 0 | 1 | is born |
| 1985 | $\omega_5$ | 88 | 1 | dies |

really tells a story about the life on the island. It is evident, however, that this is not possible if the number of people becomes very large. But then also the simple list of records becomes larger and larger and difficult to survey; and so it becomes necessary to condense the list into comprehensible information. This is the task of statistical methods. The basic ideas will be discussed in the next chapter.

## 3.2 Representation of Processes

**1.** It is often said that demography, like other social sciences, is concerned with "processes". Taken literally, this only expresses an interest in sequences of events that are assumed to be related in some way. But how does one delineate the events that are part of the process? Observations will not provide an answer because the possibilities to consider objects and events as being part of a process are virtually unlimited. We therefore understand 'process', not as an ontological category (something that exists in addition to objects and events), but as belonging to the ideas and imaginations of humans aiming at an understanding of the occurrences they are observing. Put somewhat differently, we suggest to understand processes as conceptual constructions. This is not do deny that processes can meaningfully be linked to observations of objects and events; but this will then be an indirect link: one can observe objects and events, but not processes. The fictitious chronicle of the previous section can serve as an example. The chronicle can meaningfully be understood as the characterization of a process and, as we have construed the example, it derives from observations. However, what the chronicler actually observes is not a process but the people on the island and a variety of events involving these people. The process only comes into existence by creating the chronicle.

This example also illustrates the abstractions that cannot be avoided in the construction of processes. Only a small number of events can be given an explicit representation. In the example, the chronicler only records some basic demographic events and consequently abstracts from most of what is actually happening on the island.

**2.** In order to explicitly define processes it seems natural to begin with events. For demographic processes, the basic events are births, deaths, and migrations. An explicit representation of these events can be avoided, however, by using the conceptual framework introduced in the previous section.[5] This allows to think of a demographic process simply as a sequence of population sets, $\Omega_t$. Birth and death events are then taken into account by corresponding updates of these population sets; each birth adds a person and each death removes one. This motivates the following notation to represent a *demographic process without external migration*:

$$(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}}^*, \Omega_t)$$

to be understood as a sequence of population sets, $\Omega_t$, which are defined for all temporal locations $t \in \boldsymbol{\mathcal{T}}^*$. In this formulation, $\boldsymbol{\mathcal{T}}^*$ denotes a contiguous subset of the time axis $\boldsymbol{\mathcal{T}}$ that covers the period for which the process shall be considered, and $\boldsymbol{\mathcal{S}}$ provides a representation of the spatial context.[6]

**3.** The assumption on $\boldsymbol{\mathcal{S}}$ introduced in the previous section implies that migration can only occur inside this space. People can move between the spatial locations defined by $\boldsymbol{\mathcal{S}}$, but such events will not change the size of the population and need not be taken into account for a general definition of demographic process. The situation is somewhat different when a demographic process is restricted to a subset of $\boldsymbol{\mathcal{S}}$, say $\boldsymbol{\mathcal{S}}^* \subset \boldsymbol{\mathcal{S}}$, which is often the case in empirical applications, for example, when considering the demographic development in a specific country. People can then migrate between $\boldsymbol{\mathcal{S}}^*$ and $\boldsymbol{\mathcal{S}} \setminus \boldsymbol{\mathcal{S}}^*$. However, if a definition of population sets is restricted to the subspace $\boldsymbol{\mathcal{S}}^*$, such events can formally be treated like births and deaths; in-migration adds a person and out-migration removes a person. One therefore can use an analogous notation,

$$(\boldsymbol{\mathcal{S}}^*, \boldsymbol{\mathcal{T}}^*, \Omega_t)$$

in order to represent a *demographic process with external migration*. As already explained, the notation is meant to imply that $\boldsymbol{\mathcal{S}}^*$ is a proper subset of $\boldsymbol{\mathcal{S}}$ and the population sets $\Omega_t$ are restricted to $\boldsymbol{\mathcal{S}}^*$.

**4.** All further concepts to be introduced in this text, including statistical

[5] An alternative approach that explicitly begins with events is taken, e.g., by Wunsch and Termote (1978, ch. 1).

[6] A fully explicit notation would therefore be: $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}}^*, \{\Omega_t \,|\, t \in \boldsymbol{\mathcal{T}}^*\})$.



variables, will be derived from the notion of a demographic process (with or without external migration). As a first step, one can simply refer to the number of people who are members of the population sets $\Omega_t$. We will use the following notations:

$$\begin{aligned} n_t &:= \text{number of people in temporal location } t \ (n_t = |\Omega_t|) \\ b_t &:= \text{number of children born in temporal location } t \\ d_t &:= \text{number of people dying in temporal location } t \end{aligned}$$

For a demographic process without external migration the relation between population size and birth and death events can then be written as follows:

$$n_{t+1} = n_t + b_{t+1} - d_t \tag{3.2.1}$$

For a demographic process with external migration we use, in addition, the notations:

$$\begin{aligned} m_t^i &:= \text{number of people who enter } \boldsymbol{\mathcal{S}}^* \text{ in temporal location } t \\ m_t^o &:= \text{number of people who leave } \boldsymbol{\mathcal{S}}^* \text{ in temporal location } t \end{aligned}$$

The basic equation then becomes

$$n_{t+1} = n_t + b_{t+1} - d_t + m_{t+1}^i - m_t^o \tag{3.2.2}$$

These equations will be called *accounting equations* of a demographic process (with or without external migration). One should notice that these accounting equations are true by definition. They simply are book-keeping identities about demographic processes and do not have any causal meaning. Illustrations will be given in Chapter 6 with data for the demographic development in Germany.

**5.** Notwithstanding the conventions introduced in the last paragraph of the preceding section, referring to the number of people who live during a temporal location $t$ inevitably involves some conceptual indeterminacies. If temporal locations are short, e.g. days, such indeterminacies might well be ignored. On the other hand, if the temporal index $t$ refers to years, or even longer periods, one might want to distinguish the number of people who live during this period from the number of people who live at the beginning, or end, of the period. This is done, for example, in many publications of population statistics by the *Statistisches Bundesamt*. The distinction is between the number of people at the end of a year, defined as the last day in the year, and a *midyear* population size.[7] When analyzing data

[7] The definitional apparatus of the STATIS data base (see Appendix A.1) provides the following explanations: „Der Bevölkerungsstand gibt die Zahl der Personen an, die zur Bevölkerung gehören, nachgewiesen zu verschiedenen Zeitpunkten. Der Bevöl-

from population statistics it is therefore necessary to distinguish between the different definitions used in the data construction. The notation $n_t$ is always meant to represent some kind of mean population size in the temporal location $t$, most often a year. In addition, we use the following notations:

$$n_t^{\vdash} := \text{population size at the beginning of } t$$

$$n_t^{\dashv} := \text{population size at the end of } t$$

and assume that $n_t^{\dashv} = n_{t+1}^{\vdash}$. The exact meaning of "beginning" and "end" will be left unspecified because possible meanings depend on the application context and availability of data.

**6.** If the beginning and end of a temporal location are explicitly distinguished, the formulation of the accounting equations has to be changed accordingly:

$$n_t^{\dashv} = n_t^{\vdash} + b_t - d_t \tag{3.2.3}$$

for a demographic process without external migration, and

$$n_t^{\dashv} = n_t^{\vdash} + b_t - d_t + m_t^i - m_t^o \tag{3.2.4}$$

for a demographic process with external migration. These versions of the accounting equations will be used in Section 6.4.

## 3.3 Stocks, Flows, and Rates

**1.** Demographers have invented a large number of measures to characterize demographic processes.[8] Some of these measures will be introduced in

kerungsstand im Jahresdurchschnitt insgesamt ist das arithmetische Mittel aus zwölf Monatswerten, die wiederum Durchschnitte aus dem Bevölkerungsstand am Anfang und Ende jeden Monats sind. Zur Berechnung des durchschnittlichen Bevölkerungsstandes nach Altersjahren und Geschlecht wird ein vereinfachtes Verfahren angewendet: Es werden lediglich die arithmetischen Durchschnittswerte aus dem Bevölkerungsstand jeder Gruppe zum Jahresanfang und -ende gebildet und mit einem Korrekturfaktor multipliziert. Dieser Korrekturfaktor ist der Quotient aus dem durchschnittlichen Bevölkerungsstand insgesamt und der Summe aller vereinfacht berechneten Durchschnittswerte des Bevölkerungsstandes in den einzelnen Altersjahren.“ One should also note that these definitions have exceptions and have changed through time. „In den Jahren 1961, 1970 und 1987 wurden keine Durchschnittwerte gebildet, sondern die Ergebnisse der jeweiligen Volks- und Berufszählungen nachgewiesen.“ „Bis 1953 und von 1956 bis 1960 wurde zur Berechnung des Bevölkerungsstandes im Durchschnitt insgesamt das arithmetische Mittel aus jeweils vier Vierteljahreswerten gebildet; dagegen wurde der Bevölkerungsstand von 1953 bis 1955, von 1962 bis 1969 und wird seit 1971 – wie oben beschrieben – als Durchschnitt aus Monatswerten berechnet.“

[8] For a fairly complete compilation of the many measures that are used in the demographic literature see Mueller (1993 and 2000), or Esenwein-Rothe (1982).



Part II of this text when dealing with real data. In the present section we briefly discuss a general idea that has motivated many of these measures and is derived from a distinction between stock and flow quantities. As an example, we refer to equation (3.2.1) in the previous section. The equation connects two kinds of quantity: $n_t$ and $n_{t+1}$ are *stock quantities* [Bestandsgrößen] which record the number of people who live in the respective temporal locations; on the other hand, $b_{t+1}$ and $d_t$ are called *flow quantities* [Stromgrößen] because they record changes (events) which occur during the respective temporal locations. The general idea is: a stock quantity records some state of affairs that is (assumed to be) fixed in a given temporal location, and a flow quantity records changes that occur during some time interval. A flow quantity then counts the number of events of a certain kind which occur during that time interval.

**2.** A further step consists in the definition of *rates* [*Raten*]. The basic idea is to relate the number of events to a number of people who can, in some sense, contribute to the occurrence of the events. A marriage rate can serve as an example:

$$\text{marriage rate} := \frac{\text{number of marriages in year } t}{\text{number of people who might become married in year } t}$$

As shown by this example, a rate is always a ratio where the numerator refers to a flow quantity. The only question is how to define a sensible denominator. In a strict sense, the denominator should refer to the number of people who might experience the events referred to in the numerator. This is possible, for example, when referring to death events. Since any living individual might die at any time, the denominator of a mortality rate can simply refer to all people still alive at a given temporal location. In other cases the definition of a sensible denominator is more difficult. How should one define the number of people who might become married in a certain year? It does not suffice to exclude people who are already married, one should also exclude children below a certain age.

**3.** Actually, the term ‘rate’ is used quite loosely in the demographic literature and other areas of social statistics. While it is most often a ratio where the numerator refers to a flow quantity in the sense of a number of events occurring during a time interval, the denominator might refer to any kind of stock quantity that is assumed to exhibit some sensible relation to the numerator. An example is the *crude birth rate* [*allgemeine Geburtenziffer*[9]] which is defined by

$$\text{crude birth rate} := \frac{\text{number of births in year } t}{\text{mean population size in year } t}$$

[9] This expression that avoids the term ‘rate’ is used by the *Statistisches Bundesamt*. In the literature one also finds other expressions, for example ‘rohe Geburtenrate’.

(often multiplied by 1000).

**4.** Another example is the notion of a *rate of change* [Veränderungsrate], also called a *growth rate* [Wachstumsrate]. This notion can be applied to any sequence of stock quantities. As an example, we refer to a sequence of population sizes, $n_t$. The rate of change, or growth rate, of the population is then defined by[10]

$$\rho_t := \frac{n_{t+1} - n_t}{n_t} \tag{3.3.1}$$

For a demographic process without external migration this can also be written in the form

$$\rho_t = \frac{b_{t+1} - d_t}{n_t} = \frac{b_{t+1}}{n_t} - \frac{d_t}{n_t}$$

This formulation shows that the numerator is a flow quantity and the denominator a stock quantity. However, the calculation only requires a knowledge of the stock quantities. As an example, we use some figures that refer to the demographic development in Germany for the period 1990 to 1996:[11]

| $t$ | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|---|
| $n_t$ | 79365 | 79984 | 80594 | 81179 | 81422 | 81661 | 81896 |
| $\rho_t$ | 0.0078 | 0.0076 | 0.0073 | 0.0030 | 0.0029 | 0.0029 | |

Of course, growth rates can also be expressed in percent.

**5.** The definition of growth rates given above immediately implies the following equation:

$$n_{t+1} = n_t\,(\rho_t + 1)$$

If one considers not just two consecutive temporal locations but some longer time interval, say from $t$ to $t + t'$, one finds the more general relationship

$$n_{t+t'} = n_t\,(1+\rho_t)\cdots(1+\rho_{t+t'-1}) = n_t \prod_{\tau=0}^{t'-1}(1+\rho_{t+\tau})$$

[10]Since a growth rate conceptually relates to two temporal locations, it is an arbitrary convention to index the rate by the first temporal location. Some authors, e.g. Rinne (1996, p. 84), use the second temporal location. We mention that growth rates can also be defined differently. For example, when it seems sensible to distinguish the beginning and end of a period $t$, one might define a growth rate for this period by $(n_t^{\dashv} - n_t^{\vdash})/n_t^{\vdash}$.

[11]The figures refer to the midyear number of people in 1000 and are taken from Statistisches Jahrbuch 1997 für die Bundesrepublik Deutschland (p. 46).



This equation can also be used to define a *mean growth rate* [*durchschnittliche Wachstumsrate*]. The idea is to assume a constant growth rate during the time interval from $t$ to $t + t'$. If this constant growth rate is denoted by $\rho$, one gets the equation

$$n_{t+t'} = n_t\,(1+\rho)^{t'}$$

The mean growth rate for a period from $t$ to $t + t'$, in this text denoted by $\rho_{t,t+t'}$, is defined as the solution of this equation and can be calculated with the following formula:

$$\rho_{t,t+t'} = \left(\frac{n_{t+t'}}{n_t}\right)^{1/t'} - 1$$

As an illustration, using the figures from the above example, one finds

$$\rho_{1990,1996} = \left(\frac{81896}{79365}\right)^{1/6} - 1 \approx 0.00525$$

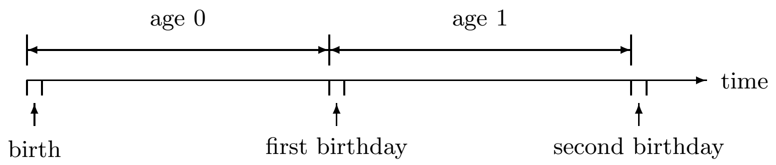
## 3.4 Age and Cohorts

**1.** A substantial part of the demographic literature is concerned with the "structure" of the population sets $\Omega_t$, where 'structure' refers to statistical distributions of individual properties. Two of these properties are of particular importance in demography: sex and age. This is due to the fact that both are important preconditions of many demographic events.[12] For example, only women, during a certain period of their lifes, can give birth to children; and also death events depend in some way on age. So it is often sensible to distinguish people with respect to their sex and age. To distinguish numbers of male and female individuals we use superscripts: $n_t^m$ and $n_t^f$ will denote, respectively, the number of men and women living in temporal location $t$; of course, $n_t = n_t^m + n_t^f$.

**2.** Age refers to the duration between a current temporal location and the date of birth of a person. A commonly used measure is *completed years*. Demographers also use another measure often called *exact age*. The meaning of this term depends on the time axis used to provide a temporal framework. As an example, we assume a discrete time axis, $\boldsymbol{\mathcal{T}}$, with temporal locations defined as days. The exact age of a person is then simply the number of days that passed away since the person was born.

[12]We speak of conditions in order to avoid causal connotations. Both, sex and age, are clearly not "factors" which in some way "produce" demographic events. Thinking of age as a "causal variable" has been called "fallacy of age reification" (Riley 1986, p. 158).

The following graphic illustrates the connection between exact age and age in completed years:




In this example, the exact age of a person is 0 during the day the person is born, it is 1 during the next day, and so on. We will avoid, however, the term ‘exact age’ and speak of age in completed days, or month, or years, whatever unit of time is used for measurement. Furthermore, we simply speak of age if the time unit is identical with the temporal locations of the time axis that provides the temporal framework.

**3.** Given these conventions for the measurement of age, the members of the population sets $\Omega_t$ can be distinguished by their age. We will use the notation $\Omega_{t,\tau}$ to refer to the subset of members of $\Omega_t$ being of age $\tau$ ($\tau = 0, 1, 2, \ldots$). This results in a partition

$$\Omega_t \;=\; \Omega_{t,0} \cup \Omega_{t,1} \cup \Omega_{t,2} \cup \cdots \;=\; \bigcup_{\tau=0}^{\infty} \Omega_{t,\tau}$$

Using the notation $n_{t,\tau}$ to refer to the number of members of $\Omega_{t,\tau}$, a corresponding equation is

$$n_t \;=\; n_{t,0} + n_{t,1} + n_{t,2} + \cdots \;=\; \sum_{\tau=0}^{\infty} n_{t,\tau}$$

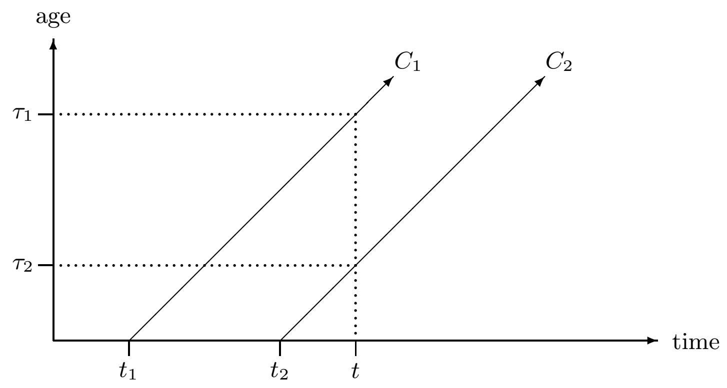
Of course, there is an upper limit to the length of human life; summation to $\infty$ simply avoids the specification of a definite limit.

**4.** If age is measured in the units of a time axis $\boldsymbol{\mathcal{T}}$, the relationship between age and time is quite simple: whenever a person survives a temporal location $t \in \boldsymbol{\mathcal{T}}$ this adds one unit to the person’s age.[13] This relationship can be graphically illustrated by an *age-period diagram*, also called a *Lexis diagram*, as follows:[14]

[13]This statement assumes that age is measured in the same units as used for the definition of $\boldsymbol{\mathcal{T}}$; this also implies that, if an individual is of age $\tau$ in temporal location $t$, then it was born in temporal location $t - \tau$.

[14]Beginning with G. F. Knapp and W. Lexis, demographers have developed many variations of this basic age-period diagram; for a discussion see Esenwein-Rothe (1992, pp. 16-30).






The horizontal axis represents time, the vertical axis represents age. The diagonal arrows depict the life courses of individuals born in the same temporal location. For example, the arrow denoted by $C_1$ refers to individuals born in temporal location $t_1$. When time goes on they grow older and their age can be read from the vertical scale. A vertical line that begins at any temporal location, say $t$, intersects the diagonal life course lines at the corresponding ages.

**5.** We use the notation $\mathcal{C}_t$ to refer to the set of people born in the temporal location $t$, and $\mathcal{C}_{[t,t')}$ to refer to the set of people born in the time interval from $t$ to $t'$. Such sets are called *birth cohorts*.[15] As shown by the Lexis diagram, a demographic process can be considered as a temporal sequence of birth cohorts, and this is one reason why this concept plays a prominent role in demography. It is also a basic concept for much of the research that deals with human life courses.[16] The approach bases an understanding of long-term changes in the development of societies on a comparison of the life courses of members of successive birth cohorts.[17] This research has also led to a more general definition of the term ‘cohort’:

> “A cohort is an aggregate of individual elements, each of which experienced a significant event in its life history during the same chronological interval.” (Ryder 1968, p. 546)

A similar definition was given by Glenn (1977, p. 8):

[15]Some demographers also use the word ‘generation’ or use both terms synonymously. However, except when referring to relationships between parents and children, we will avoid to speak of “generations” because this word has many different and often unclear meanings; see, e.g., the discussion by Mannheim (1952), Pfeil (1967), and Kertzer (1983).

[16]For an overview, see Wagner (2001).

[17]An early exposition of this view was given by Ryder (1965).

"a cohort is defined as those people within a geographically or otherwise delineated population who experienced the same significant life event within a given period of time."

Not just birth events, also most other kinds of event can then be used to define cohorts; for example, one can speak of a marriage cohort that comprises all people who married in the same time period.

**6.** In the present text cohorts will always be birth cohorts for which we have introduced the notation $\mathcal{C}_t$. Similar to the population sets $\Omega_t$, we think of cohorts as *sets* of people, not as some kind of social group.[18] We shall also avoid any kind of analogy with individuals. A cohort is a conceptual construction, not some object that exists apart and independent of its members. In particular, we shall not think of cohorts as some kind of "agents" that "drive" social change.[19] This is not to deny that one may find some similarities among the members of the same cohort. However, whatever similarities one may find, they result from the life courses of the individual cohort members.[20] It is, therefore, conceptually senseless to think of such similarities as resulting from being members of the same cohort, that is, from being born in the same year. The argument becomes not better if one refers, not just to the fact of being born in the same year, but to events and social conditions that were experienced by members of the same cohort at the same age. Such considerations might well be used for a retrospective interpretation of similarities among members of the same cohort. It would be a mistake, however, to use such considerations for implicitly changing the meaning of the term 'cohort'. The term is defined by referring to people born in the same year, or historical period, possibly adding some spatial demarcation. So whatever happens to be the case afterwards is irrelevant for the meaning of the term 'cohort'.[21] The only fact derivable from the definition is that members of the same cohort are always of (approximately) the same age during their life courses. But being of the same age can not be used to explain facts or events. In our view, therefore, cohort is not an explanatory concept, but a conceptual tool that allows to think in terms of life courses and their development. As we have mentioned, this also might help in retrospective interpretations. But the explanatory value of such interpretations derives from locating individual life courses within the historical periods in which they develop.

[18] This distinctions has already been stressed by Mannheim (1952, pp. 288-9). Unfortunately, Mannheim's notions of a "generation as an actuality" and a "generation unit" eventually obscure this distinction.

[19] This view has been suggested, more or less explicitly, by Ryder (1965). In another paper, he wrote: "Some reservations to this discussion are necessary to obviate the implication that cohorts are the exclusive agents of social change." (Ryder 1968, p. 548) The point, however, is not that there are also other agents, but that cohorts never are agents. Cohorts do not bring anything about.

[20] This has been explicitly recognized by Mayer and Huinink (1990, p. 213): "the characteristics of a cohort are aggregated outcomes of the individual behavior of cohort members in the social context, indicated crudely by calender time."

[21] An additional argument is simply that the members of a cohort experience quite different events, and live under quite different social conditions, during their life courses; see, e.g., Rosow (1978).



**7.** It should also be stressed that 'cohort' is essentially a retrospective concept. The definition given above suggests that cohorts, like $\Omega_t$, are sets of people. While this is formally correct, there is an important ontological distinction. The population sets $\Omega_t$ consist of people who live in the historical time $t$. Therefore, if we know that a person is a member of $\Omega_t$, we can infer that this person is alive in $t$. But now think of a cohort of people born in some year $t_0$. Knowing that a person is member of the cohort $\mathcal{C}_{t_0}$ only allows to infer that this person lived during $t_0$. For all later periods, nothing definite can be inferred. Of course, most members of $\mathcal{C}_{t_0}$ will live for some period following $t_0$. But life beyond $t_0$ is a property of an individual member of the cohort, not of the cohort itself. It is thus dubious how to speak of the temporal existence of a cohort. One might say that $\mathcal{C}_{t_0}$ appears in the year $t_0$ and remains existent until the death of its last survivor. But this implies that the cohort is no longer a definite population set but part of a demographic process, formally

$$\mathcal{C}_{t_0} \equiv (\mathcal{C}_{t_0,0}, \mathcal{C}_{t_0,1}, \mathcal{C}_{t_0,2}, \ldots)$$

where $\mathcal{C}_{t_0,\tau}$ denotes the set of members of $\mathcal{C}_{t_0}$ still alive at age $\tau$. In the framework of a demographic process without migration one can then formally identify the sets $\mathcal{C}_{t_0,\tau}$ with the population sets $\Omega_{t_0+\tau,\tau}$.

**8.** This temporal view is quite sensible and is used, for example, in the construction of cohort life tables (Chapter 8). However, it is no longer possible, then, to identify a cohort with a definite set of people. The problem is somewhat obscured by the fact that most empirical cohort studies actually condition on survivorship. They are concerned with people born in $t_0$ and having survived until some temporal location $t > t_0$. So one can speak of a definite set of people, formally identical with $\Omega_{t,t-t_0}$; but clearly, this set is not identical with $\mathcal{C}_{t_0}$. Of course, there is nothing wrong in referring to a population set $\Omega_{t,t-t_0}$. It implies, however, a retrospective point of view. The population set $\Omega_{t,t-t_0}$ only comes into existence when history has passed the temporal location $t$. Therefore, thinking of cohorts as definite population sets presupposes a retrospective point of view.

# Chapter 4

# Variables and Distributions

The last chapter introduced the notion of a demographic process, formally denoted by $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}}^*, \Omega_t)$. This suffices to represent the population size, that is, the number of people in the population sets $\Omega_t$, and to record its development through time. Additional questions concern properties of the members of $\Omega_t$. Two such properties, sex and age, were already part of the example discussed in Section 3.1, but many other properties can also be considered. If the size of $\Omega_t$ is large, how can one sensibly represent the properties of all its members? This is the task of statistical methods as has been expressed by a famous statistician, Ronald A. Fisher, in the following way:

> "Briefly, and in its most concrete form, the object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data." (Fisher 1922, p. 311)

The present chapter introduces two basic notions, statistical variables and statistical distributions (additional concepts will be added in later chapters). Definitions and notations mainly follow the author's "Grundzüge der sozialwissenschaftlichen Statistik" (2001). Some notational simplifications will be discussed in Section 4.3.

## 4.1 Statistical Variables

**1.** Unfortunately, the word 'variable' is easily misleading because it suggests something that "varies" or being a "variable quantity". In order to get an appropriate understanding it is first of all necessary to distinguish statistical from logical variables. Consider the expression '$x \leq 5$'. In this expression, $x$ is a *logical variable* that can be replaced by a name. Obviously, without substituting a specific name, the expression '$x \leq 5$' has no definite meaning and, in particular, is neither a true nor a false statement. The expression is actually no statement at all but a *sentential function* [*Aussageform*]. A statement that is true or false or meaningless only results when a name is substituted for $x$. For example, if the symbol 1 is substituted for $x$, the result is a true statement ($1 \leq 5$); if the symbol 9 is substituted for $x$, the result is a false statement ($9 \leq 5$); and if some name not referring to a number is substituted for $x$, the result is neither true nor false but meaningless. As the reader will remember from his or her mathematical education such logical variables are heavily used in mathematics,



often to formulate general statements, for example:

For all numbers $x$: if $0 \leq x \leq 5$, then $0 \leq x^2 \leq 25$

This example also shows that logical variables are in no way "variable quantities".[1]

**2.** Statistical variables serve a quite different purpose. They are used to represent the data for statistical calculations which refer to properties of objects. The basic idea is that one can characterize objects by properties. Since this is essentially an assignment of properties to objects, statistical variables are defined as functions:[2]

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

$X$ is the name of the function, $\Omega$ is its domain, and $\tilde{\mathcal{X}}$ is a set of possible values. To each element $\omega \in \Omega$, the statistical variable $X$ assigns exactly one element of $\tilde{\mathcal{X}}$ denoted by $X(\omega)$. In this sense, a statistical variable is simply a function.[3] What distinguishes statistical variables from other functions is a specific purpose: statistical variables serve to characterize objects. Therefore, in order to call $X$ a statistical variable (and not just a function), its domain, $\Omega$, should be a set of objects and the set of its possible values, $\tilde{\mathcal{X}}$, should be a set of properties that can be meaningfully used to characterize the elements of $\Omega$. To remind of this purpose, the set of possible values of a statistical variable will be called its *property space* [*Merkmalsraum*] and its elements will be called *property values* [*Merkmalswerte*].[4]

**3.** As was mentioned in the Introduction, in the statistical literature domains of statistical variables are often called populations. This is unfortunate because a statistical variable can refer to any kind of object. We therefore use the term 'population' only if one is actually referring to sets

[1] This has been stressed by many logicians, see, e.g., Frege (1990, p. 142); and already in 1903, B. Russell (1996, p. 90) wrote: "Originally, no doubt, the variable was conceived dynamically as something which changed with the lapse of time, or, as is said, as something which successively assumed all values of a certain class. This view cannot be too soon dismissed."

[2] Since the notion of a 'function' is used throughout the whole text we have added a section in Appendix A.2 providing basic definitions.

[3] It follows that logical and statistical variables are completely different things. Moreover, the term 'variable' is misleading in both cases. For a more extensive discussion that also shows how both notions, logical and statistical variables, can be linked by using sentential functions, see Rohwer and Pötter (2002b, ch. 9). — When there is no danger of confusion, we will drop the attribute 'statistical' and simply speak of variables.

[4] We generally denote statistical variables by upper case letters $(A, B, C, \ldots, X, Y, Z)$ and their property spaces by corresponding calligraphic letters that are marked by a tilde $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}, \ldots, \tilde{\mathcal{X}}, \tilde{\mathcal{Y}}, \tilde{\mathcal{Z}})$.

of people. In the general case we speak of the *domain* or, equivalently, of the *reference set* of a statistical variable.

**4.** As an illustration we use the chronicle introduced in Section 3.1. Referring to the year 1960, the reference set consists of the names of all individuals who, in 1960, were inhabitants of the island:

$$\Omega_{1960} := \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10}\}$$

Given this reference set, one can think of characterizing its elements by properties. Two property spaces have been used in the example, age and sex. The latter property space consists of two elements, 'male' and 'female', and may be written explicitly as follows:

$$\tilde{\mathcal{S}} := \{\text{male}, \text{female}\}$$

This then allows to define a statistical variable, say $S$, that assigns to each individual $\omega \in \Omega$ a property value $S(\omega) \in \tilde{\mathcal{S}}$. Of course, if the statistical variable is intended to represent data, the assignment should not be made arbitrarily but reflect actual properties of the individuals. In our example, the assignment should be made in the following way:

| $\omega$ | $S(\omega)$ |
|---|---|
| $\omega_1$ | male |
| $\omega_2$ | female |
| $\omega_3$ | female |
| $\omega_4$ | male |
| $\omega_5$ | female |
| $\omega_6$ | male |
| $\omega_7$ | female |
| $\omega_8$ | male |
| $\omega_9$ | female |
| $\omega_{10}$ | male |

This example also demonstrates that, in contrast to most functions that are used in mathematics, statistical variables cannot be defined by referring to some kind of rule. There is no rule that would allow to infer the sex, or any other property, of an individual by knowing its name. In order to make a statistical variable explicitly known one almost always needs a tabulation of its values.

**5.** As the example also shows, the elements of a property space are not numbers but properties that can be used to characterize objects. However, it is general practice in statistics to *represent* properties by numbers. This was already done in Section 3.1 where we have represented the properties 'male' and 'female' by the numbers 0 and 1, respectively. One reason for doing so is the resulting simplification in the tabulation of statistical data.



The main reason is, however, another one: numerical representations allow to perform statistical calculations. As an example, we introduce the *mean value* of a statistical variable, say $X$, denoted by $\mathrm{M}(X)$. The definition is

$$\mathrm{M}(X) := \frac{1}{|\Omega|} \sum_{\omega \in \Omega} X(\omega)$$

The calculation consists in summing up the values of the variable for all elements in the reference set and then dividing by the number of elements. Obviously, the calculation requires a numerical representation for the values of the variable, that is, for the elements of its property space. But as soon as one has introduced a numerical representation one can do anything that can be done with numbers also with the values of a variable. To be sure, this does not guarantee a result with an immediate and sensible interpretation. This might, or might not, be the case and can never be guaranteed from a statistical calculation alone.[5] However, in our example we get a sensible result. Performing the calculations of a mean value for the variable $S$, we get the value

$$\mathrm{M}(S) = \frac{1}{10}(0+0+0+1+0+1+0+1+0+1+1) = 0.5$$

providing the proportion of female individuals in the set of inhabitants of our fictitious island.

**6.** The chronicle also provides information about the age of the inhabitants of the island. Referring again to the set of people who lived on the island in 1960, denoted by $\Omega_{1960}$, we can define another statistical variable that assigns to each individual $\omega \in \Omega_{1960}$ its age in 1960. We will call this variable $A$, and denote its property space by $\tilde{\mathcal{A}}$, defined by

$$\tilde{\mathcal{A}} := \{0, 1, 2, 3, \ldots\}$$

In this case we do not need to explicitly introduce a numerical representation because age, given its usual meaning in terms of completed years, already has a numerical expression. Of course, an age of 40, say, is not identical with the number 40. It is a number which is given a specific meaning. Consequently, $\tilde{\mathcal{A}}$, while formally identical with the set of natural numbers, has an additional meaning which is not a part of the definition of natural numbers, namely that its elements are agreed upon to denote ages in completed years. Many other methods to provide information about age could equally well be used, for example, measuring age in months or days, or simply distinguishing between children and older people. These

[5] One cannot rely on any general rules but needs to consider each statistical calculation in its specific context. As an example, think of household income and rent. Subtracting rent from household income provides a meaningful result, but simply to add both quantities does not.

are considerations that precede the definition of a property space and, consequently, a statistical variable. Here we follow the chronicler who has used completed years to record ages allowing to define a statistical variable

$$A : \Omega_{1960} \longrightarrow \tilde{\mathcal{A}}$$

that provides for each inhabitant of the island his or her age in 1960. Of course, the definition does not suffice to record ages. As in the first example one needs to distinguish between merely assuming the existence of a variable and actually knowing its values. To provide such knowledge the values of a statistical variable must be explicitly recorded in some way. In this example we can use again a simple table as follows:

| $\omega$ | $A(\omega)$ |
|---|---|
| $\omega_1$ | 40 |
| $\omega_2$ | 38 |
| $\omega_3$ | 4 |
| $\omega_4$ | 16 |
| $\omega_5$ | 63 |
| $\omega_6$ | 70 |
| $\omega_7$ | 25 |
| $\omega_8$ | 8 |
| $\omega_9$ | 63 |
| $\omega_{10}$ | 11 |

Such a table, often called a *data matrix*, provides the values of a variable and can be used as a starting point for further calculations. For example, we can calculate the mean value of $A$ which is

$$\mathrm{M}(A) \;=\; \frac{1}{10}\,(40+38+4+16+60+70+25+8+65+11) \;=\; 33.7$$

In contrast to the first example, this is not a proportion but the value of the mean age of the inhabitants of the island in 1960.

**7.** Our second example also shows that, in general, one needs to distinguish between a property space as it is used to define a statistical variable and the set of property values that are actually realized in some given reference set $\Omega$. The former will be called a *conceptual property space* and the latter a *realized property space*.[6] In our example, the realized property space is

$$A(\Omega) \;=\; \{A(\omega)\,|\,\omega\in\Omega\} \;=\; \{4,8,11,16,25,38,40,63,70\} \subset \tilde{\mathcal{A}}$$

and is obviously not identical with $\tilde{\mathcal{A}}$.

[6]The realized property space of a statistical variable can also be called its *range*. This term is commonly used to denote, for an arbitrary function $f : B \longrightarrow C$, the image $f(B) \subseteq C$; see also Appendix A.2.



**8.** A further point is worth attention. The same property value can have several realizations in a reference set. In our example, there are two individuals, $\omega_5$ and $\omega_9$, having the same age in 1960 (and, of course, during their whole lifes). Also in the first example, several people share the properties 'male' and 'female', respectively. Using terminology from Appendix A.2, one can say that statistical variables are, in general, not injective functions. Inverse functions must therefore be defined in terms of sets. For example,

$$A^{-1}(\{4\}) \;=\; \{\omega_3\},\;\; A^{-1}(\{5\}) \;=\; \emptyset,\;\; A^{-1}(\{63\}) \;=\; \{\omega_5,\omega_9\}$$

The interpretation is straightforward: If $\tilde{A}$ is a subset of the property space $\tilde{\mathcal{A}}$, then $A^{-1}(\tilde{A})$ is the subset of members of $\Omega$ having a property value in $\tilde{A}$. The same notation is used with any other statistical variable. For example,

$$S^{-1}(\{1\}) \;=\; \{\omega_2,\omega_3,\omega_5,\omega_7,\omega_9\}$$

is the set of female members of $\Omega_{1960}$.

**9.** A final consideration concerns the reference sets that are used as domains to define statistical variables. In both previous examples, $\Omega_{1960}$ was taken to be a set of people, the inhabitants living on our fictitious island in 1960. In general, $\Omega$ can be any set of objects. The formal notion of a statistical variable only requires that the elements of $\Omega$ can be identified and distinguished, and to which values of a property space can be meaningfully assigned. This generality of statistical variables should be taken with some caution, however, depending on the purpose to be served. In this text, statistical variables will be used as means for demographic descriptions and models. The basic conceptual framework is a demographic process that consists of a space $\boldsymbol{\mathcal{S}}$, a time axis $\boldsymbol{\mathcal{T}}$, and population sets $\Omega_t$ comprising the people who are living in $\boldsymbol{\mathcal{S}}$ during the temporal location $t$. As our examples have shown, statistical variables can be used to represent information about the members of the sets $\Omega_t$. It also seems possible to use the space, $\boldsymbol{\mathcal{S}}$, as a domain for statistical variables which then take the form

$$L : \boldsymbol{\mathcal{S}} \longrightarrow \tilde{\mathcal{L}}$$

Such variables will be called *spatial variables*. They are always statistical variables. Examples would be the characterization of spatial locations by their size (e.g., in square kilometers), or by the number of people who currently live in these locations. These examples show that spatial locations, as understood in this text, are similar to objects. Both have some kind of physical existence and can sensibly be characterized by properties. This ontological status is not shared, however, by temporal locations. Temporal locations are in no way similar to objects and do not have a physical existence. So it seems not possible to sensibly characterize temporal locations by properties, and we therefore shall avoid to use a time axis as a domain for the definition of statistical variables in a proper sense.

## 4.2 Statistical Distributions

**1.** Statistical variables provide the starting point for all further statistical concepts. These concepts, directly or indirectly, always relate to reference sets of statistical variables and not to their individual members,[7] more specifically: they relate to *distributions* of properties in a reference set. In order to introduce this notion explicitly, consider a statistical variable

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

If all values of this variable were known it would be possible to use that knowledge to characterize all individual members of $\Omega$. However, statistical concepts and methods have a quite different purpose. Statistical questions do not concern individual members of $\Omega$ but frequencies of property values in the reference set $\Omega$. This was stated in a *Declaration of Professional Ethics*, published by the *International Statistical Institute* (1986, p. 238), as follows:[8]

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

Accordingly, the basic idea is that statistical concepts and methods are concerned, not with individuals, but with frequencies of properties.[9]

[7]See the quotations from Lexis and Feichtinger cited in the Introduction.

[8]See also Bürgin and Schnorr-Bäcker (1986).

[9]It should be mentioned, however, that also the complementary idea, to use statistical data for the characterization of individuals, accompanies the history of statistics. The following quotation from one of its founders, Francis Galton (1889, p. 35–37), provides an example. Galton begins: "We require no more than a fairly just and comprehensive method of expressing the way in which each measurable quality is distributed among the members of any group, whether the group consists of brothers or of members of any particular social, local, or other body of persons, or whether it is co-extensive with an entire nation or race." Then follows, however, a quite different reasoning: "A knowledge of the distribution of any quality enables us to ascertain the Rank that each man holds among his fellows, in respect to that quality. This is a valuable piece of knowledge in this struggling and competitive world, where success is to the foremost, and failure to the hindmost, irrespective of absolute efficiency. [...] When the distribution of any faculty has been ascertained, we can tell from the measurement, say of our child, how he ranks among other children in respect to that faculty, whether it be a physical gift, or one of health, or of intellect, or of morals. As the years go by, we may learn by the same means whether he is making his way towards the front, whether he just holds his place, or whether he is falling back towards the rear. Similarly as regards the position of our class, or of our nation, among other classes and other nations."



**2.** Following this idea, one no longer refers to individual members of $\Omega$ but to elements, or subsets, of a variable's property space. So let $\tilde{X}$ be any subset of $\tilde{\mathcal{X}}$; such subsets will be called *property sets*. The question concerns the proportion of members in $\Omega$ who were assigned a value in the property set $\tilde{X}$ via the function $X$. A general answer is provided by the *frequency function* of $X$, that is, a function

$$\mathrm{P}[X] : \mathcal{P}(\tilde{\mathcal{X}}) \longrightarrow \mathbf{R}$$

In this formulation, $\mathcal{P}(\tilde{\mathcal{X}})$ is the power set of $\tilde{\mathcal{X}}$, that is, the set of all subsets of $\tilde{\mathcal{X}}$. So the domain of the frequency function $\mathrm{P}[X]$ consists of all property sets that can be created from the property space $\tilde{\mathcal{X}}$. The assignment is defined by

$$\mathrm{P}[X](\tilde{X}) := \frac{1}{|\Omega|} \left| \{\omega \in \Omega \,|\, X(\omega) \in \tilde{X}\} \right|$$

Thus, for every property set $\tilde{X} \in \mathcal{P}(\tilde{\mathcal{X}})$, $\mathrm{P}[X](\tilde{X})$ is the relative frequency of $\tilde{X}$ in $\Omega$.

**3.** Frequency functions always refer to *relative* frequencies (proportions). We therefore adopt the terminological convention that the word 'frequency', without a qualifying attribute, also always means relative frequency. Since also absolute frequencies are often used to characterize populations, we introduce the complementary notion of an absolute frequency function defined by

$$\mathrm{P}^*[X](\tilde{X}) := \left| \{\omega \in \Omega \,|\, X(\omega) \in \tilde{X}\} \right| = \mathrm{P}[X](\tilde{X})\,|\Omega|$$

**4.** It is fairly obvious how to derive a frequency function from the values of a statistical variable. Given a property set, say $\tilde{X}$, one simply counts the number of elements of $\Omega$ having a property value in $\tilde{X}$, and then divides the resulting count by the number of elements of $\Omega$. Less obvious is that, from a statistical or demographic point of view, all relevant information about a statistical variable is contained in its frequency distribution. In fact, all proper statistical concepts are derived from frequency distributions. To illustrate this, we refer to the first example, variable $S$, discussed in the previous section. The property space is $\tilde{S} = \{0, 1\}$, and so it suffices to consider the property sets $\{0\}$ and $\{1\}$.[10] Using the data which are tabulated on page 40, one immediately finds:

$$\mathrm{P}[S](\{0\}) = \mathrm{P}[S](\{1\}) = 0.5$$

[10]The power set of a property space $\tilde{\mathcal{X}}$ also contains $\tilde{\mathcal{X}}$ and the empty set, $\emptyset$. However, since $\mathrm{P}[X](\tilde{\mathcal{X}}) = 1$ and $\mathrm{P}[X](\emptyset) = 0$ for any variable $X$, these property sets can be neglected.

Now, this information also suffices to calculate the mean value of $S$. This is possible because, for any statistical variable, say $X$, its mean value can also be expressed by the formula

$$\mathrm{M}(X) = \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \tilde{x}\, \mathrm{P}[X](\{\tilde{x}\})$$

Therefore, in our example, knowing the frequencies of $\{0\}$ and $\{1\}$, one immediately finds $\mathrm{M}(S) = 0.5$.

**5.** The fact that all relevant information about a statistical variable is contained in its distribution can be used to stress again the specific statistical abstraction that derives from the consideration of frequency distributions. In general, knowing the frequency distribution of a statistical variable, it is no longer possible to infer property values for the individual members of the reference set that was used to define the variable. In this sense, as said by Lexis, "verschwindet das Individuum als solches" (see the quotation in the Introduction).

**6.** Although the domain of the frequency function of a statistical variable, say $X$, is defined as the power set of the variable's property space, $\tilde{\mathcal{X}}$, it is not necessary to explicitly tabulate the frequencies of all possible property sets. This is due to the fact that frequency functions are additive: if $\tilde{X}$ and $\tilde{X}'$ are any two *disjoint* property sets, then

$$\mathrm{P}[X](\tilde{X} \cup \tilde{X}') = \mathrm{P}[X](\tilde{X}) + \mathrm{P}[X](\tilde{X}')$$

One can express, therefore, the frequency of any property set in the following way:

$$\mathrm{P}[X](\tilde{X}) = \sum_{\tilde{x} \in \tilde{X}} \mathrm{P}[X](\{\tilde{x}\})$$

This shows that it suffices to know the frequencies of the one-element property sets $\{\tilde{x}\}$, corresponding to the property values $\tilde{x} \in \tilde{\mathcal{X}}$, in order to have complete knowledge of the frequency distribution of $X$. This also makes clear how the consideration of frequency distributions serves the main goal of statistical methods, namely to make an often large number of values of a statistical variable comprehensible. Instead of a separate entry for each individual member of the reference set $\Omega$, the representation of a frequency distribution only requires a separate entry for each property value in the realized property space of a statistical variable. Of course, many further statistical concepts can then be used to describe, analyze, and compare frequency distributions of one or more statistical variables. We will introduce some of these concepts in later chapters when they can help in a discussion of substantial questions.



## 4.3 Remarks about Notations

**1.** The notations introduced in the foregoing sections are somewhat more involved than those often found in introductory textbooks. This is done in order to make as clear as possible the logical structure of the corresponding concepts, in particular, two basic ideas:

- A statistical variable is not any kind of "variable quantity", but an assignment of properties to objects. So it is formally a function in the mathematical sense of this term. This also implies that statistical variables can not sensibly be thought of as "factors" which, in any dubious sense, can "influence", or "cause", the behavior of objects.
- Statistical notions refer, not to individual objects, but to frequency distributions of properties in sets of objects. These frequency distributions which contain all statistically relevant information are, again, functions in the mathematical sense of the word.

**2.** However, having recognized these conceptual foundations, we will reduce the notational burden and introduce the following abbreviations:

a) If $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ is a statistical variable and $\tilde{x} \in \tilde{\mathcal{X}}$ a property value, its frequency must correctly be written as $\mathrm{P}[X](\{\tilde{x}\})$ because the domain of the frequency function, $\mathrm{P}[X]$, is the power set of $\tilde{\mathcal{X}}$. However, it will save notational overhead to omit, in this case, the curly brackets around $\tilde{x}$ and simply write $\mathrm{P}[X](\tilde{x})$. For example, to refer to the proportion of male individuals in $\Omega_{1960}$ we might simply write $\mathrm{P}[S](0)$.

b) One often refers to the frequency, not of a single property values $\tilde{x} \in \tilde{\mathcal{X}}$, but of a set of property values, which we have called a property set, $\tilde{X} \subset \tilde{\mathcal{X}}$. The correct formulation is then $\mathrm{P}[X](\tilde{X})$, because the function is $\mathrm{P}[X]$ and its argument is $\tilde{X}$. While formally not correct, an often used alternative formulation is $\mathrm{P}(X \in \tilde{X})$. However, this alternative formulation is sometimes practical. For example, we might want to refer to the frequency of people being of age 65 or above. The property set is then $\{\tilde{a} \in \tilde{\mathcal{A}} \,|\, \tilde{a} \geq 65\}$, and its frequency would need to be written as $\mathrm{P}[A](\{\tilde{a} \in \tilde{\mathcal{A}} \,|\, \tilde{a} \geq 65\})$. But obviously, it saves notational overhead to simply write $\mathrm{P}(A \geq 65)$.

# Chapter 5

# Modal Questions and Models

**1.** We conclude the discussion of the conceptual framework with a few remarks concerning the term 'model'. This will also make clear why we do not make a sharp distinction between population statistics and the construction of demographic models. — The basic idea is *not* to follow a widespread conception that thinks of models as being "simplified descriptions" of some part of reality. For example:

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

Quite similar is the view that models are in some way "mappings" [Abbildungen] of parts of reality. For example:

„Modelle können wir uns in erster Näherung denken als begriffliche Konstrukte zur 'Abbildung' realer Systeme oder zum Umgang mit solchen." (Balzer 1997, p. 16) „Ein Modell ist wohl immer aufzufassen als eine Abbildung. Die Frage ist nur, was abgebildet wird, und wie die Abbildungsfunktion aussieht." (Frey 1961, p. 89)

The main objection is that models as used in scientific discourse almost never serve the purpose of *describing* something. While descriptions certainly play an important role in scientific work, this is done by documenting observations. In contrast, most models serve a quite different purpose, namely to provide a framework for thinking about *modal* questions. For example, What *might* have caused the decline of birth rates in Germany following the baby boom of the 1960s? or, To which extend *might* the proportion of old people increase during the next 20 years?

**2.** This is a basic distinction: statements may relate to facts or to possibilities. Descriptions are intended to provide facts, but most human reasoning concerns possibilities. This is also true of most scientific reasoning.[1] Reasoning concerned with possibilities will be called *modal reasoning*. There is a wide variety of different forms.[2] Often modal reasoning concerns

[1]An opposite view was expressed, e.g., by Samuelson (1952, p. 61): "All sciences have the common task of describing and summarizing empirical reality." However, if the task really consists in describing something one would not need a model but could simply report observations. Therefore, given that models do not have the task to describe something, it would also be misleading to say that models can only provide "simplified", or even "distorted descriptions" (see, e.g., Baumol 1966, p. 90). After all, why should somebody be interested in "distorted descriptions" of reality?

[2]For a good introduction see White (1975).



future possibilities, for example, the future demographic development in Germany. However, also with respect to the past one often is not able to simply state facts. As a consequence one needs to think about a modal question: What *might* have been the case? The question indicates that one can only speculate about possible facts. Of course, how this can be done depends on the available knowledge. For example, there are no reliable recordings of the number of births in Germany during World War II. Nevertheless, some information is available and can be used to provide reasons for the belief that birth rates declined. A researcher might then say: for these reasons it seems highly probable that birth rates declined during the years of war.

**3.** We propose to use the term 'model' in the following way: *Models are explicitly formulated means intended to serve modal reasoning.* Since modal reasoning comes in many different forms, the same is true of models. Some distinctions will be suggested below. However, one should also recognize that the construction and use of models is not automatically implied by modal reasoning. In fact, most modal reasoning goes without employing a model. A model only comes into existence when it is explicitly formulated as a means intended to provide a conceptual framework for reasoning about specific questions. For example, using information from a weather report to support speculations about tomorrow's weather is not using a model; but possibly the people preparing the forecast have used a model. As a condition of its existence, a model needs some kind of representation independent of its actual use. This is not to require any specific conceptual tools for the formulation of a model. There is again a wide variety of possibilities. The formulation of a model can be purely verbal or, in addition, employ symbols, graphs, figures, even physical devices. But in whatever form a model is presented, it must be possible to think about the model in its own right.

**4.** Models do not provide answers to modal questions; they serve to think about, and evaluate, possible answers.[3] The main service consists in providing a framework for explicit reasoning. One has to state explicitly the available knowledge, additional assumptions, and how both are used to draw inferences. This is particularly important with regard to assumptions because possible answers to modal questions often heavily depend on assumptions. One might also say that assumptions are required by the very nature of modal questions because, by definition, the available

[3]We therefore do not agree with Baumol (1966, p. 90 - 91) who conceived of "predictive models" in the following way: "A predictive model need require relatively little comprehension on the part of its users or even its designers. It is a machine which grinds out its forecasts more or less mechanically, and for such tasks, unreasoning, purely extrapolative techniques frequently still turn out the best results." Like oracles, such machines should not be called models because, as Baumol rightly says, they do not serve any kind of reasoning.

evidence is not sufficient for an answer. However, this statement needs a qualification. Modal reasoning does not, by itself, require assumptions. For example, being interested in tomorrow's weather, one can ask a weather forecast. In order to believe in the information one does not need the assumption that the forecaster actually knows the weather of tomorrow. In fact, one can simply use the information without making any assumptions whatsoever. Assumptions are no prerequisite for reasoning, nor for the formation of beliefs. It is easily misleading to say that assumptions are required to allow reasoning in situations where the available evidence is incomplete. Assumptions are only required if one is interested in making reasoning explicit, that is, in making reasoning an object of critique and evaluation. This also is the main task of models. Their job is to show explicitly how one might, or might not, arrive at certain conclusions with regard to a modal question that motivates the reasoning.

**5.** This also allows to explain the affinity between thinking in terms of models and what might be called rule-based reasoning. The idea is: given certain assumptions one can draw some inferences while others are ruled out because they would violate established rules of reasoning. Of course, not only possible mistakes in applying rules, but the rules themselves, can become a matter of dispute. Furthermore, human reasoning cannot be reduced to a mechanical application of given rules. As a limiting case, one can think of mathematical proofs; but mathematics is not concerned with modal questions and, consequently, not with the construction of models. At least in the dominating understanding, mathematics is basically interested in the *formal* implications that can be derived from assumptions according to given rules. This allows to make use of mathematical results in many areas of rule-based reasoning. However, when thinking about modal questions, the interest is not in the formal implications of assumptions and rules but concerns possible answers. It is, therefore, not only important that the reasoning is formally correct, but of even greater importance is that assumptions are reasonable. We therefore avoid to speak of 'formal models'. Irrespective of the conceptual tools used to formulate a model, which often are borrowed from mathematics, its task is not to allow formal inferences but to support modal reasoning.[4]

**6.** Since there is a great variety in modal questions and conceptual tools, also many different kinds of models do exist. Thinking of models used in the social sciences, the following aspects provide hints to introduce some broad distinctions: the conceptual framework that provides the model's ontology, the kind of modal question that the model is intended to serve, and the linguistic and technical tools used to formulate the model and derive possible implications.

[4] One should notice, however, that the word 'formal' is often used in two different meanings. As the word is used above, it refers to arguments which are true only because of their form. In a different meaning, 'formal' is often understood as the opposite of 'informal'. Its meaning then becomes similar to what we tried to express with the word 'explicit'.



**7.** We begin with the first aspect and broadly distinguish two types of model:

a) *statistical models* which conceptually relate to distributions of statistical variables, and

b) *behavioral models* which explicitly refer to the behavior of individuals.

Almost all demographic models belong to the first type. Assumptions concern, for example, the development of birth and death rates, or the number and age distribution of immigrants. Despite the possibility to metaphorically link such assumptions to the behavior of people, they conceptually relate to demographic processes formulated in terms of population sets. All further concepts used for the model formulation are derived thereof and do not relate to individual behavior. Such models are therefore concerned, not with individual behavior, but with the development, and relations between, quantities derived from statistical distributions. In this sense, all models discussed in subsequent chapters are statistical models. In contrast, we propose to speak of behavioral models only if a model explicitly refers to individuals and allows to reason about individual behavior.

**8.** The second aspect concerns the modal reasoning that a model is intended to serve. We broadly distinguish three groups of models:

a) *representational models* whose purpose is to provide a view of something,

b) *analytical models* which are used to provide a conceptual framework for reasoning about relationships and rules, and

c) *technical and political models* that have the purpose to support people in the design and implementation of technical artefacts and institutions.

**9.** There is a great variety of representational models. For example, a map can be called a representational model as it is intended to provide a specific view of some area. Another example would be the model of a building that an architect plans to build. The model is then intended to provide a view of a building that might come into existence in the future. Of course, the model is not a description because there is no "future building" which could be described. The model is rather used to support reasoning about modal questions concerning the possible features of a building that might become realized in the future. The examples discussed in the present text mainly relate to statistical distributions. One of the techniques widely used in statistics to construct representational models is smoothing. An example would be the construction of trends by smoothing a time series. As the purpose is to provide a specific view of a process one can rightly speak of a representational model. More involved examples concern the construction of representational models in situations where the available information is incomplete. One can think, e.g., of the construction of world maps in early times when substantial parts of the earth were not known by the map makers. A similar situation often occurs in the construction of statistical models. Examples which will be discussed in later chapters concern the estimation of statistical distributions when part of the available data is incomplete. Estimation procedures are then based on assumptions which might be wrong and cannot be tested with the available data. So one is actually concerned with a modal question concerning an unknown distribution. A further and somewhat more complicated example of this type will be discussed in Chapter 9 where we deal with data from surveys in which respondents provide information about birth and, possibly, death dates of their parents and the question concerns whether such data can be used to construct cohort life tables. As will be seen, this requires several assumptions which should be considered explicitly. We therefore continue, in Section 9.2, with the discussion of a simulation model to find out in which way conclusions depend on alternative assumptions.

**10.** Since every model requires in some way a representation of its subject matter there is no sharp distinction between representational and analytical models. The distinction is mainly a question of emphasis. Nevertheless, there is a specific concern, not normally present in the construction of representational models, that justifies a distinction. Analytical models, as we propose to understand this term, are intended to support speculations about relationships and rules. How this can be done depends on the conceptual framework. In the construction of a behavioral model one would need to think about relationships between individuals and rules of their behavior. This will not be further discussed because, in the present text, we only deal with statistical models. Relationships then concern statistical distributions or quantities derived thereof. Furthermore, because statistical models do not explicitly refer to individuals, it is not necessary to speculate about "behavioral rules" for the model's objects. Instead, rules only concern the argumentation used to establish the model. As an example, we discuss in Section 13.3 a rudimentary model of the baby boom that occurred in Germany in the period 1955–1965. The modal question motivating the model concerns "timing effects" on the development of the number of newborn children. The model tries to add to an understanding of the baby boom by showing that, without certain "timing effects", a quite different development might have occurred. This will not be a causal explanation. In fact, we use a statistical model without any reference to individuals who can bring about changes by their activities. Consequently, also the rules used in the argumentation do not refer to the behavior of individuals. Instead, they solely concern logical implications of hypothetical assumptions which, in turn, relate to the distribution of birth events as represented by cumulated cohort birth rates. — More abstract versions of analytical models will be discussed in later chapters. The model introduced in Chapter 17 is not related to historical events, like the baby boom of the 1960s, but at the beginning only provides a general conceptual framework for reasoning about possible demographic processes. As will be seen, this framework can nevertheless be used to gain insights into how such processes "work". Moreover, as we try to show in another chapter, the model also provides a starting point for a discussion of modal questions concerning the effects of immigration on the demographic development in Germany.



**11.** A third group of models will be called technical and political models. Again, the distinction is to some extent a question of emphasis. As an example, one can think of the model created by the architect mentioned above. The model can be understood as representational because its immediate purpose is to provide a view of the building planned by the architect. However, given that the plan becomes realized, new modal questions arise: How can/should the work be done? As a consequence, also the model, or variants derived thereof, has to serve reasoning about these additional questions. It must be transformed into a technical model that can actually be used as a guide to realize the initial idea. Of course, depending on the kind of artefacts, or systems, in the original sense of this word, which are planned and possibly realized, there is a great variety in the details of corresponding models. An important distinction concerns whether such systems also contain human individuals who can act in their own right. If this is not the case, we propose to speak of *technical models*, otherwise of *political models*. However, this distinction is mainly important only for behavioral models. Statistical models and, in particular, the demographic models discussed in the present text are only concerned with assumptions about statistical distributions or quantities derived thereof. Distinctions concerning the behavior of objects, whether they should be considered as building materials or as actors, are therefore not directly relevant. Political questions only come into play when models are also used in political discussions and decision-making. Prominent examples would be models used for population projections.

**12.** Finally, models can be distinguished with regard to the linguistic tools used in their formulation. While most models use symbolic, mainly mathematical, notations, this is not an essential feature. As an example, one can think of Keynes' "General Theory" that was originally developed without any usage of symbolic notation. However, the same example also shows that using symbolic notations can help in the understanding of a model and its potential use for reasoning about modal questions. Furthermore, using symbolic notations often allows an easier understanding of the assumptions built into a model, and supports the discussion of their implications. In particular, almost all statistical models employ symbolic notations as these are already available by the initial introduction of the conceptual framework. A further distinction concerns the technical means used to derive implications of the assumptions put into a model. The classical way is reasoning, supported by paper and pencil and, if available, further instruments. New methods became available by the modern computer. In particular, the computer allows to implement so-called simulation models. An example will be discussed in Section 9.2.1.

# Part II

# Data and Methods

## Chapter 6

# Basic Demographic Data

Based on the conceptual framework introduced in Part I, we now begin with a description and analysis of the demographic development in Germany. The present chapter provides a brief presentation of some basic figures concerning the number of people and then discusses age and sex distributions.

## 6.1 Data Sources

**1.** We begin with a few remarks about data sources.[1] Most of the basic data are provided by official statistics [amtliche Statistik], its central office in Germany being the *Statistisches Bundesamt*.[2] In this and the following chapters we mainly rely on such data from official statistics. Supplementary data from retrospective surveys will be used in later chapters.

**2.** Most demographic data published by official statistics are based on two sources, censuses and population registers, corresponding loosely to the distinction between stock and flow quantities (see Section 3.3). A census [Volkszählung] is intended to provide information about the number of people who live in a certain region at a specific date, so it is a kind of stock-taking.[3] In contrast, population registers record events, in particular, births, deaths, marriages and migrations. Official statistics uses both data sources. Since censuses only take place at greater temporal intervals, data from population registers are used to provide estimates of the population size in years between censuses.

**3.** Like the political history of Germany, also its history of censuses is quite irregular. A publication of the *Statistisches Bundesamt* that deals with the historical development of official statistics in Germany provides the following information:

„Nach der territorialen Neuordnung der Nachfolgestaaten des Heiligen Römischen Reichs Deutscher Nation auf dem Wiener Kongreß wurde 1816 erstmals in Preußen innerhalb der neuen Grenzen eine Volkszählung durchgeführt. Die anderen Länder des Deutschen Bundes führten in der Folgezeit Volkszählungen durch,

[1] For an extensive survey of sources of demographic data see the reports by Carola Schmid (1993 and 2000).

[2] For references to publications see Appendix A.1.

[3] Most often, a census not only counts people but also records some of their properties, like age, sex, marital status and citizenship. For information about the questionnaire that was used in the latest census of 1987 see Würzberger, Störtzbach and Stürmer (1986).

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

Since then, a further census in the territory of the former FRG took place on May 25, 1987. Censuses in the territory of the former GDR were perfomed in 1950 (August 31), 1964 (December 31), and 1981 (December 31).[4]

**4.** A further question concerns the demarcation of people who are counted in a census. The just cited document provides the following information:

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

[4]For additional information see Schmid (1993, pp. 55-57).



It might be added that the concept of population which is used by official statistics in Germany has again slightly changed with the introduction of a new registration law [Meldegesetz] in 1983.

**5.** The second main source of demographic data are population registers. In Germany, different kinds of such registers exist. To provide basic demographic data, the *Statistisches Bundesamt* mainly uses information from two such registers:

a) Registers of births, deaths, and marriages, which are kept by offices of local authorities, called *Standesamt*.[5]

b) Registers of residences, also kept by offices of local authorities, called *Einwohnermeldeamt*. In addition, there is a central register for persons without a German citizenship, called *Ausländerzentralregister*. Data from these registers are used by the *Statistisches Bundesamt* for its statistics about internal and external migration.[6]

## 6.2 Number of People

**1.** Our first question concerns the number of people who lived in Germany during its history. Data from official statistics begin with the first census in Preußen in 1816. A difficulty results from the fact that the political boundaries of Germany have often changed during its history; the latest change occurred in October 1990 through unification with the former GDR (Deutsche Demokratische Republik). Since we are mainly interested in the development after World War II, it suffices to distinguish two territories: (a) the *territory of the former FRG* (Bundesrepublik Deutschland),[7] and

[5]These registers were introduced in 1875. For a history of corresponding laws and institutions see Schütz (1977). Registration forms as used by the *Statistisches Bundesamt* have been published in Fachserie 1, Reihe 1, 1990 (pp. 312-323).

[6]Fachserie 1, Reihe 1, 1999 (pp. 13-14).

[7]The *Saarland* became part of the former FRG only in 1957. However, many time series from official statistics include the *Saarland* also for the period 1950 – 1956.

**Table 6.2-1** Number of people (in 1000) in the territory of the former FRG. Source: Statistisches Jahrbuch 2001 (p. 44).

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

(b) the *territory of the former GDR* (including the eastern part of Berlin). We simply speak of *Germany* when referring to both territories.

**2.** The data in Table 6.2-1 are taken from the Statistisches Jahrbuch 2001 (p. 44) and refer to the territory of the former FRG. The yearbook provides the following hints about sources.

a) The figures for 1961, 1970, and 1987 are based on census data and relate to their target dates (June 6, 1961, May 27, 1970, and May 25, 1987). The remaining figures for the period since 1946 are estimates of the midyear population size and are derived from register data in connection with census data and data from the Wohnungsstatistik.[8]

[8] Actually, the figures result from backward projections. The Statistisches Jahrbuch 2001 (p. 41) provides the following remarks: „Bei den [. . .] für die Jahre 1950 bis 1970 nachgewiesenen Fortschreibungszahlen handelt es sich um rückgerechnete Einwohnerzahlen aufgrund der Ergebnisse der Wohnungsstatistik vom 25.9.1956 (1950 bis 1955), der Volkszählung vom 6.6.1961 (1957 bis 1960) und der Volkszählung vom 27.5.1970 (1962 bis 1969). Die für die Jahre ab 1970 bis einschl. 1986 nachgewiesenen Bevölkerungszahlen sind Fortschreibungsdaten, die von den Ergebnissen der Volkszählung 1970 ausgehen. Die ab 30.6.1987 nachgewiesenen Bevölkerungszahlen beruhen auf den Ergebnissen der Volkszählung 1987.“



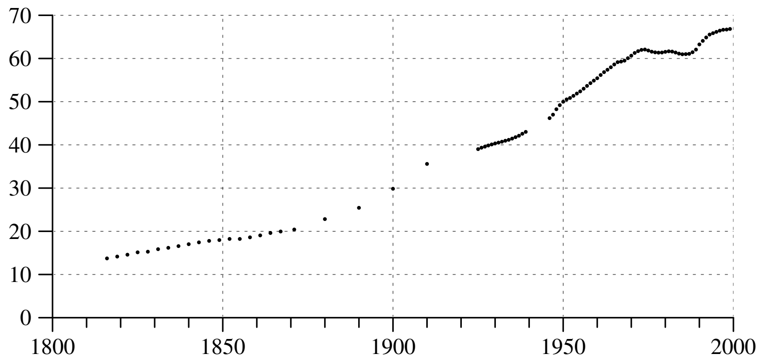



**Fig. 6.2-1** Graphical presentation of the data from Table 6.2-1. The scale of the ordinate is in million.

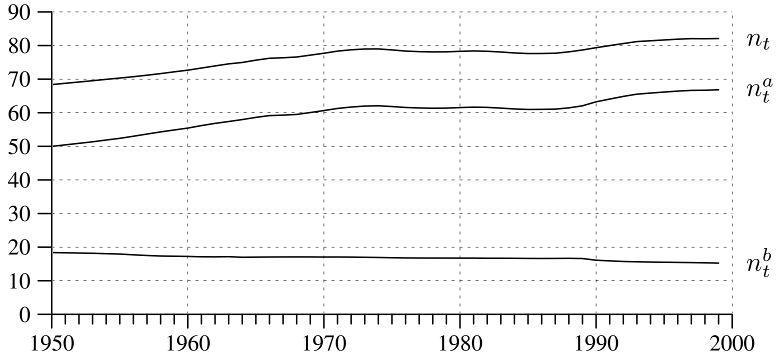



**Fig. 6.2-2** Graphical presentation of the data from Table 6.2-2. The scale of the ordinate is in million.

b) The sources of the figures for earlier periods are not explicitly documented. It can be supposed that the primary sources for the period since 1871 are censuses that have taken place in the years 1871, 1880, 1890, 1900, 1910, 1925, 1933, and 1939, and that the figures for years in between are estimates, possibly also based on additional register data.

c) It might further be supposed that the figures for the period before 1871 are based on censuses that began 1816 in Preußen and were then periodically continued in 3-year intervals, with some delay also in other parts of the Zollverein. It is not clear, however, in which way the figures were adjusted to the territory of the former FRG.

**3.** In order to get a first impression of the long-term development of the number of people in Germany the data of Table 6.2-1 are plotted in Figure 6.2-1. Since the data are unevenly spaced, we have represented each number

**Table 6.2-2** Number of people (in 1000) on the territory of the former FRG ($n_t^a$) and the former GDR ($n_t^b$). Source: Statistisches Jahrbuch 2001 (p. 44).

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

from Table 6.2-1 by a separate dot. Of course, this is just a first impression and we need to analyze more carefully the components, births, deaths, and migrations, that contributed to the overall picture. Here we only add some information about the development in the two parts of Germany after World War II. Table 6.2-2 presents the basic figures taken again from Statistisches Jahrbuch 2001 (p. 44), Figure 6.2-2 gives a graphical presentation. $n_t$, the number of people in both territories, is calculated by adding $n_t^a$ and $n_t^b$. Notice that the demarcation of the two territories has not been changed after October 1990, the eastern part of Berlin is considered a part of the former territory of the GDR.

## 6.3 Births and Deaths

**1.** The number of people living in some territory changes with births, deaths, and migration. We therefore should consider these components in order to get a better understanding of the demographic development in Germany. We begin with a consideration of births and deaths in the post-World War II period. The basic figures as published by the *Statistisches Bundesamt* are shown in Table 6.3-1. Following our general convention, we denote the number of births and deaths that occurred during a year $t$ by $b_t$ and $d_t$, respectively. Additional indices are used to distinguish between (a) the territory of the former FRG and (b) the territory of the former GDR. Figure 6.3-1 provides a graphical view of these data.

**2.** Comparing Figures 6.2-2 and 6.3-1, one observes that the turning points in the development of the number of people roughly corresponds to peri-



**Table 6.3-1** Births ($b_t$) and deaths ($d_t$) in the territory of the former FRG (first four columns) and in the territory of the former GDR (last four columns); all counts in 1000. Source: Fachserie 1. Reihe 1, 1999 (pp. 43-44).

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

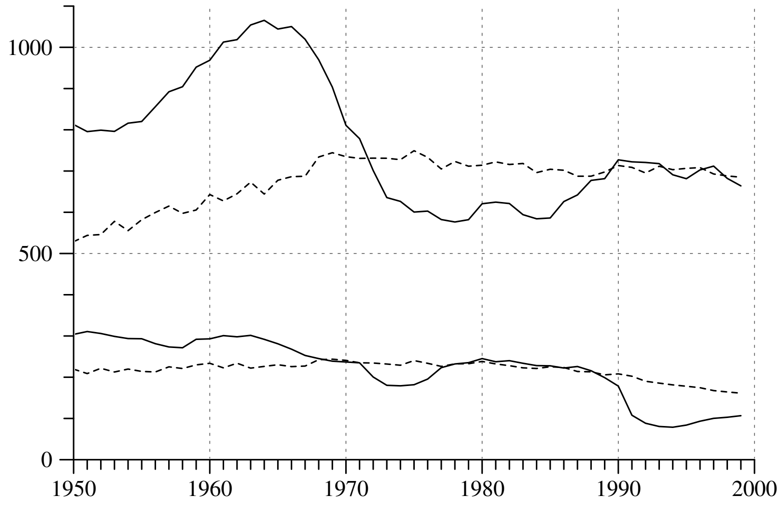



**Fig. 6.3-1** Births (solid line) and deaths (dotted line) in the territories of the former FRG (upper part) and the former GDR (lower part). The scale of the ordinate is in 1000. The data are taken from Table 6.3-1

ods where the number of births is above, or below, the number of deaths (differences are due to migration). Most remarkable is the high amount of variation in the number of births. In the territory of the former FRG the "baby boom" of the sixties is followed by a substantial decline in number of births. In the territory of the former GDR occurred a huge decline in the number of births during the years following the unification in 1989.

**3.** A closer analysis will be given in later chapters. In fact, since simple time series data do not take into account changes in the age distribution of the population, possible conclusions are quite limited. For example, it is not possible to derive any safe conclusions about changes in the length of life and conditions of mortality from the time series shown in Figure 6.3-1. It might well be that a growing number of deaths as shown in this figure for the first two decades is accompanied by an increase in the mean life length. This will be further discussed in Chapter 7.

**4.** The development of birth and death rates in the recent German history should also be related to a broader historical and international context. While we are not able to provide an adequate discussion in the present text, we only present some data that roughly indicate some long-term changes.[9]

[9] There are several studies which provide extensive discussions of the long-term demographic development in Germany. For an introduction, see Marschalck (1984). A thorough discussion of the fertility decline in the period 1871–1939 was given by Knodel (1974).



**Table 6.3-2** Crude birth rates (CBR) and crude death rates (CDR) in Germany. Data for the period 1841 – 1943 refer to the territory of the *Deutsches Reich* with varying boundaries (see footnote 11); data for the period 1946 – 1999 refer to the territory of the former FRG. Sources: Statistisches Bundesamt, Bevölkerung und Wirtschaft 1872 – 1972 (pp. 101-103), and Fachserie 1, Reihe 1, 1999 (p. 50).

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

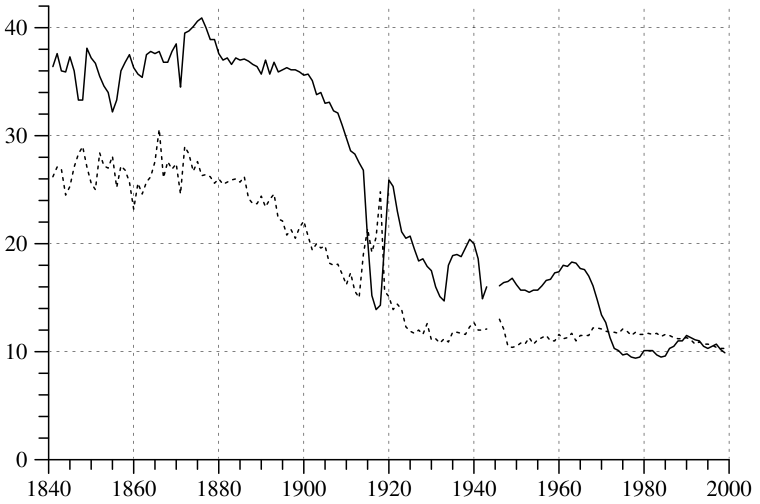



**Fig. 6.3-2** Crude birth rates (solid line) and crude death rates (dotted line) for the period 1841–1999 in Germany, based on the data shown in Table 6.3-2

We simply use crude birth and death rates which have been published by the *Statistisches Bundesamt* for most years since 1841.[10] Table 6.3-2 shows the data.[11] The crude birth rates and the crude death rates are calculated per 1000 of the midyear population. Figure 6.3-2 provides a visual impression. The plot impressively shows the long-term decline of both, the crude birth rates and the crude death rates. The plot also shows that, until about 1970, birth rates were most often quite higher than death rates (a dramatic exception is only during the years of World War I).

## 6.4 Accounting Equations

**1.** Changes in the size of a population are a result of births, deaths, and migration. The basic relationships can be expressed by accounting equations. As was discussed in Section 3.3, there are two variants. Since the *Statistisches Bundesamt* has also published population size data which

[10] The crude birth and death rates are defined, respectively, as $b_t/n_t$ and $d_t/n_t$, multiplied by 1000.

[11] For the period 1841 – 1943, the source uses the term 'Reichsgebiet' and, for the years 1938 to 1943, provides the remark: "Gebietsstand 31.12.1937." (Statistisches Bundesamt 1972, p. 103) For the years 1871–1918 (and presumably also for previous years, see Statistisches Jahrbuch für das Deutsche Reich 1919, p. 2), the data refer to the territory of the *Deutsches Reich*. Notice that, for the years before 1871, other sources sometimes provide different figures which refer to the territory of the *Deutscher Zollverein*.



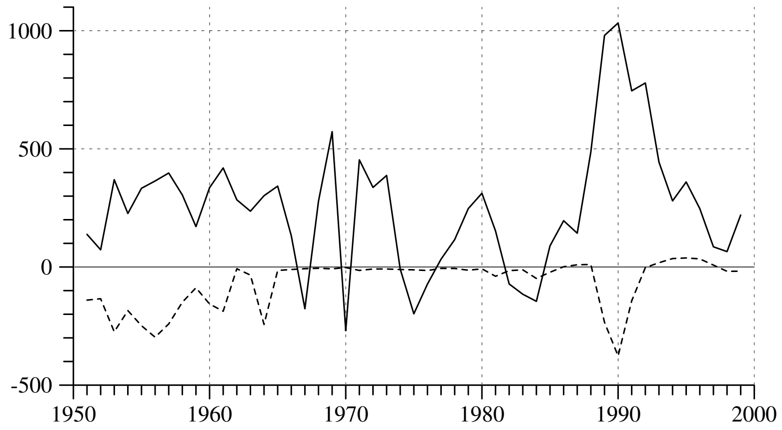



**Fig. 6.4-1** Balance of migration in the territory of the former FRG (solid line) and the territory of the former GDR (dotted line). The scale of the ordinate is in 1000. Data are taken from Tables 6.4-1 and 6.4-2.

refer to the end of each year, we can use the second variant:

$$n^{\vdash}_{t+1} = n^{\dashv}_t = n^{\vdash}_t + b_t - d_t + m^i_t - m^o_t \qquad (6.4.1)$$

In this equation, $t$ refers to calendar years, and $n^{\vdash}_t$ and $n^{\dashv}_t$ denote, respectively, the population size at the beginning and end of the year $t$. These are the stock quantities. The other symbols represent flow quantities, that is, number of events which occurred during the year: births ($b_t$), deaths ($d_t$), in-migration ($m^i_t$), and out-migration ($m^o_t$).

**2.** Values for $n^{\vdash}_t$ are shown in the second column of Tables 6.4-1 and 6.4-1; the first table refers to the territory of the former FRG, the second table to the territory of the former GDR.[12] Both tables also show the number of births ($b_t$) and deaths ($d_t$) which are identical with the entries in Table 6.3-1. These data are already sufficient to calculate the balance of migration. As implied by the accounting equation (6.4.1), one gets

$$(m^i_t - m^o_t) = (n^{\dashv}_t - n^{\vdash}_t) - (b_t - d_t) \qquad (6.4.2)$$

This equation has been used to calculate the entries in the last column of Tables 6.4-1 and 6.4-2.

**3.** Figure 6.4-1 shows the development of the balance of migration as given in the last column of Tables 6.4-1 and 6.4-2, respectively. For the territory of the former FRG, the in-migration exceeded the out-migration for most years. Until 1961 when the GDR closed its borders, the excess

[12] These data are published in Fachserie 1, Reihe 1, 1999 (p. 30).

**Table 6.4-1** Population changes in the territory of the former FRG. All counts in 1000. Source: Fachserie 1, Reihe 1, 1999 (p. 30 and p. 43).

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



**Table 6.4-2** Population changes in the territory of the former GDR. All counts in 1000. Source: Fachserie 1, Reihe 1, 1999 (p. 30 and p. 44).

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

of in-migration mainly resulted from people who came from the GDR into the FRG. A similar movement has taken place in the years immediately following the unification in 1989. One should note that our data refer separately to the territories of the former FRG and GDR and also after 1989 include migrations between both territories.

**4.** Table 6.4-1 shows that, beginning in 1972, in most of the following years the number of deaths exceeded the number of births. Thus, without an excess of in-migration, the population size would have declined in this period. It is difficult, however, to answer the modal question, Which development of population sizes would have taken place in the absence of migration? Simply subtracting $(m_t^i - m_t^o)$ from $n_{t+1}$ will not give a convincing answer because, in the absence of migration, the development of births and deaths would also have been different. So we will postpone a discussion of such modal questions to a later chapter.

## 6.5 Age and Sex Distributions

From a demographic point of view, the two most important characteristics of people are their age and their sex. We therefore discuss in the remaining sections of this chapter how to construct, and graphically present, statistical distributions of these characteristics in a population.

### 6.5.1 Age Distributions

**1.** Both, age and sex, can be represented by statistical variables. We begin with age. Referring to a population $\Omega_t$, one can define a statistical variable

$$A_t : \Omega_t \longrightarrow \tilde{\mathcal{A}} := \{0, 1, 2, 3, \ldots\}$$

that provides, for each person $\omega \in \Omega_t$, a value $A_t(\omega)$ which is the age of the person in the temporal location $t$. We will assume that age is measured in completed years, so the elements of the property space $\tilde{\mathcal{A}}$ are to be interpreted as completed years. Given this conceptual approach, corresponding data could be provided in a table as follows:

| $\omega$ | $A_t(\omega)$ |
|---|---|
| $\omega_1$ | $A_t(\omega_1)$ |
| $\vdots$ | $\vdots$ |
| $\omega_{n_t}$ | $A_t(\omega_{n_t})$ |

(6.5.1)

where $n_t$ denotes the number of persons in $\Omega_t$. Each person is identified by a (fictitious) name, often some arbitrary identification number, given in the first column. The second column contains the corresponding value of the variable $A_t$, in this example, the person's age.



**2.** Since the number of persons is often very large it would be senseless to print the full table. So the next step is to extract relevant information. The statistical approach, as discussed in Section 4.2, always begins with a calculation of frequencies. The first step is to find the realized property space

$$\tilde{\mathcal{A}}_t^* := A_t(\Omega_t) = \{A_t(\omega) \,|\, \omega \in \Omega_t\}$$

that is, the set of all elements of $\tilde{\mathcal{A}}$ which occur at least once in the population $\Omega_t$. This already provides a first piece of information about the range of the variable. In a second step, one can calculate for each element of $\tilde{\mathcal{A}}_t^*$ the frequency of its occurrence in the population $\Omega_t$. Since we are concerned here with ages, we will refer to the elements of $\tilde{\mathcal{A}}_t^*$ by the letter $\tau$, corresponding to our convention to generally denote ages by $\tau$. For the calculation of frequencies one needs to distinguish between absolute and relative frequencies. The absolute frequency of some age value $\tau \in \tilde{\mathcal{A}}_t^*$ is simply the number of persons in $\Omega_t$ of age $\tau$. Using the notation introduced in Section 4.2, they can be written as

$$\mathrm{P}^*[A_t](\tau) = |\{\omega \in \Omega_t \,|\, A_t(\omega) = \tau\}|$$

The corresponding relative frequencies are

$$\mathrm{P}[A_t](\tau) = \frac{\mathrm{P}^*[A_t](\tau)}{n_t}$$

Of course, it will always be true by definition that

$$\sum_{\tau \in \tilde{\mathcal{A}}_t^*} \mathrm{P}^*[A_t](\tau) = n_t \quad \text{and} \quad \sum_{\tau \in \tilde{\mathcal{A}}_t^*} \mathrm{P}[A_t](\tau) = 1$$

Having performed the calculations for all values in $\tilde{\mathcal{A}}_t^*$, one gets a statistical distribution, in this example, an *age distribution* for the population $\Omega_t$. This distribution provides the required statistical information and can be tabulated or graphically presented.

**3.** Actually, most data from official statistics are already in the form of statistical distributions. To continue with our example, the latest data currently available for the age distribution in Germany are published by the *Statistisches Bundesamt* in Fachserie 1, Reihe 1, 1999 (pp. 64-65).[13] These data refer to the midyear population size in the year 1999 and

[13] After having written this section, an update was published in the STATIS database of the *Statistisches Bundesamt*. Some of these data will be used in a later chapter for the construction of life tables.

**Table 6.5-1** Midyear population size (in 1000) in the year 1999 in Germany by age and sex; age in completed years. Source: Fachserie 1, Reihe 1, 1999 (pp. 64-65).

| $\tau$ | $n_{t,\tau}$ | $n^m_{t,\tau}$ | $n^f_{t,\tau}$ | $\tau$ | $n_{t,\tau}$ | $n^m_{t,\tau}$ | $n^f_{t,\tau}$ |
|---|---|---|---|---|---|---|---|
| 0 | 777.9 | 399.6 | 378.3 | 46 | 1132.0 | 570.3 | 561.7 |
| 1 | 800.2 | 410.8 | 389.4 | 47 | 1124.4 | 566.1 | 558.4 |
| 2 | 805.7 | 413.8 | 391.9 | 48 | 1120.0 | 563.9 | 556.0 |
| 3 | 785.1 | 403.1 | 382.0 | 49 | 1107.1 | 558.6 | 548.5 |
| 4 | 776.6 | 398.8 | 377.7 | 50 | 1047.1 | 529.9 | 517.2 |
| 5 | 797.6 | 409.8 | 387.9 | 51 | 977.2 | 494.2 | 483.0 |
| 6 | 823.0 | 422.1 | 400.9 | 52 | 892.9 | 450.6 | 442.2 |
| 7 | 848.6 | 435.2 | 413.4 | 53 | 790.2 | 397.3 | 393.0 |
| 8 | 908.6 | 466.4 | 442.1 | 54 | 867.9 | 434.9 | 433.0 |
| 9 | 947.0 | 486.0 | 461.0 | 55 | 1000.4 | 502.2 | 498.2 |
| 10 | 954.9 | 490.1 | 464.8 | 56 | 997.3 | 500.9 | 496.4 |
| 11 | 957.9 | 492.5 | 465.4 | 57 | 1089.7 | 545.7 | 544.1 |
| 12 | 939.3 | 482.6 | 456.7 | 58 | 1228.5 | 612.5 | 616.0 |
| 13 | 915.4 | 469.6 | 445.8 | 59 | 1252.0 | 621.7 | 630.3 |
| 14 | 898.5 | 461.2 | 437.3 | 60 | 1199.8 | 593.3 | 606.6 |
| 15 | 901.6 | 463.2 | 438.4 | 61 | 1121.1 | 551.2 | 569.9 |
| 16 | 919.5 | 472.8 | 446.7 | 62 | 1069.6 | 522.7 | 546.9 |
| 17 | 933.2 | 479.9 | 453.2 | 63 | 1037.5 | 502.8 | 534.7 |
| 18 | 938.6 | 481.4 | 457.2 | 64 | 983.4 | 472.3 | 511.1 |
| 19 | 924.6 | 473.3 | 451.3 | 65 | 854.9 | 409.8 | 445.1 |
| 20 | 903.8 | 462.2 | 441.6 | 66 | 760.9 | 361.2 | 399.7 |
| 21 | 902.7 | 461.0 | 441.7 | 67 | 764.5 | 359.3 | 405.2 |
| 22 | 902.2 | 460.3 | 442.0 | 68 | 789.4 | 365.7 | 423.6 |
| 23 | 893.1 | 456.3 | 436.7 | 69 | 793.7 | 362.4 | 431.2 |
| 24 | 897.5 | 458.8 | 438.6 | 70 | 773.4 | 348.0 | 425.4 |
| 25 | 918.5 | 469.2 | 449.3 | 71 | 736.0 | 319.3 | 416.7 |
| 26 | 977.7 | 500.6 | 477.1 | 72 | 694.2 | 283.8 | 410.4 |
| 27 | 1085.3 | 557.1 | 528.2 | 73 | 674.7 | 258.7 | 416.0 |
| 28 | 1172.2 | 603.4 | 568.8 | 74 | 635.6 | 228.3 | 407.4 |
| 29 | 1248.5 | 644.1 | 604.4 | 75 | 593.0 | 202.3 | 390.7 |
| 30 | 1328.6 | 686.0 | 642.6 | 76 | 583.6 | 196.3 | 387.4 |
| 31 | 1380.5 | 712.4 | 668.1 | 77 | 588.2 | 193.2 | 395.0 |
| 32 | 1419.5 | 732.7 | 686.8 | 78 | 571.0 | 180.1 | 391.0 |
| 33 | 1445.6 | 748.1 | 697.5 | 79 | 472.7 | 144.3 | 328.4 |
| 34 | 1464.1 | 758.2 | 705.9 | 80 | 317.8 | 96.6 | 221.2 |
| 35 | 1473.5 | 761.7 | 711.7 | 81 | 230.7 | 69.0 | 161.8 |
| 36 | 1447.0 | 746.6 | 700.4 | 82 | 221.5 | 65.1 | 156.4 |
| 37 | 1414.9 | 727.4 | 687.5 | 83 | 247.9 | 70.4 | 177.5 |
| 38 | 1386.3 | 711.2 | 675.1 | 84 | 292.7 | 79.5 | 213.2 |
| 39 | 1347.1 | 690.9 | 656.2 | 85 | 297.7 | 78.1 | 219.6 |
| 40 | 1293.6 | 663.6 | 630.1 | 86 | 265.3 | 67.9 | 197.4 |
| 41 | 1250.0 | 641.1 | 608.9 | 87 | 224.2 | 55.5 | 168.7 |
| 42 | 1225.0 | 627.2 | 597.8 | 88 | 186.3 | 44.3 | 142.0 |
| 43 | 1194.7 | 610.0 | 584.7 | 89 | 155.8 | 35.8 | 120.0 |
| 44 | 1170.0 | 594.4 | 575.7 | 90* | 481.4 | 106.6 | 374.8 |
| 45 | 1145.7 | 578.8 | 566.9 | Total | 82086.6 | 40048.0 | 42038.6 |



record age in completed years. We will use the following abbreviations:

$$n_{t,\tau} := \text{number of persons in } t \text{ being of age } \tau$$

$$n^m_{t,\tau} := \text{number of men in } t \text{ being of age } \tau$$

$$n^f_{t,\tau} := \text{number of women in } t \text{ being of age } \tau$$

We also define

$$n^m_t := \sum_{\tau=0}^{\infty} n^m_{t,\tau} \quad \text{and} \quad n^f_t := \sum_{\tau=0}^{\infty} n^f_{t,\tau}$$

to denote the total number of men and women, respectively. The total population size is then given by $n_t = n^m_t + n^f_t$.

**4.** Table 6.5-1 shows values (in 1000) for the year $t = 1999$. This table is actually a cross-tabulation with respect to age and sex, but for the moment we are only interested in a simple classification by age, that is, in the values of $n_{t,\tau}$. These are absolute frequencies, and the relationship with our previous notation is therefore given by $n_{t,\tau} = \mathrm{P}^*[A_t](\tau)$. Of course, one immediately also gets relative frequencies:

$$\mathrm{P}[A_t](\tau) = \frac{n_{t,\tau}}{n_t}$$

The last age category in Table 6.5-1, denoted by $90^*$, comprises age 90 and all higher ages. The frequency for this age category is therefore not directly comparable with the other frequencies. However, one can safely assume that

$$n_{t,90^*} = \sum_{\tau=90}^{\infty} n_{t,\tau}$$

since this equation directly follows from the additivity of frequencies.

**5.** Summing up the values for $n_{t,\tau}$ shows that $n_t = 82086600$ which is the midyear number of people living in Germany in 1999. Compared with an original raw data file, Table 6.5-1 is obviously much smaller. However, even the condensed frequency table is difficult to survey. How can one extract the information in the table in a more comprehensible form? One possibility is to aggregate the property space; this will be discussed in Section 6.5.4. Here we use a graphical display, called *frequency curves*, which does not require aggregation. The basic idea is simple: One employs a two-dimensional coordinate system and uses the horizontal axis (abscissa) to represent the elements of the property space and the vertical axis (ordinate) to represent the (absolute or relative) frequencies. As an example, Figure 6.5-1 shows the age distribution, in terms of absolute frequencies


**Fig. 6.5-1** Plot of a frequency curve that represents the age distribution in Germany in the year 1999. Data are taken from Table 6.5-1, omitting category 90*. The scale of the ordinate is in 1000.

$n_{t,\tau}$, in the form of a frequency curve. This plot may then be used as a starting point for further interpretation.

**6.** Note that the age distribution in a given year is the result of demographic events which occurred over a long period of time, beginning about 100 years ago when the oldest persons still living were born. In this sense an age distribution is to be viewed as a transitory result of a long-term demographic process. In general, the number of people being of age $\tau$ in year $t$ results from the number born $\tau$ years before $t$ and having survived until age $\tau$.[14] This relates the age categories to historically earlier periods. For example, the low frequency of people of age 53 can be traced back to the year 1946 and related to the low number of births in that year.

## 6.5.2 Decomposition by Sex

**1.** Instead of age, one can use any other property space to construct a statistical distribution. As a second example, and in order to explain the idea of a two-dimensional distribution, we refer to people's sex. So we set up another statistical variable

$$S_t : \Omega_t \longrightarrow \tilde{\mathcal{S}} := \{0, 1\}$$

which assigns to each member $\omega \in \Omega_t$ a value $S_t(\omega) \in \tilde{\mathcal{S}}$ which is 0 for men and 1 for women. Assuming that data are given in the form of a data

[14] Of course, people may have been born anywhere and become members of a population set $\Omega_t$ by immigration.



matrix

$$\begin{array}{c|c} \omega & S_t(\omega) \\ \hline \omega_1 & S_t(\omega_1) \\ \vdots & \vdots \\ \omega_{n_t} & S_t(\omega_{n_t}) \end{array} \tag{6.5.2}$$

one can calculate absolute and relative frequencies. Using the data from Table 6.5-1, and the abbreviations introduced in the previous section, one finds for $t = 1999$ the values

$$\mathrm{P}^*[S_t](0) = n_t^m = 40048.0 \qquad \mathrm{P}[S_t](0) = \frac{n_t^m}{n_t} = 0.488$$

$$\mathrm{P}^*[S_t](1) = n_t^f = 42038.6 \qquad \mathrm{P}[S_t](1) = \frac{n_t^f}{n_t} = 0.512$$

which show that there are somewhat more female than male persons who are currently living in Germany.

**2.** We now have two frequency distributions, one for age and another one for sex, but this will not allow, for example, to compare the age distribution of men and women. Also knowing the data in the form of (6.5.1) and (6.5.2) will not suffice because one cannot know whether the individual members of $\Omega_t$ have been given the same names (identification numbers) in both tables. So one needs a different starting point, namely a statistical variable that provides, for each person in $\Omega_t$, *simultaneously* both an age and a sex. This is formally expressed by a *two-dimensional variable*

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

This variable is called two-dimensional because it consists of two components and correspondingly has a two-dimensional property space being the cartesian product of $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{S}}$. Written explicitly:

$$\tilde{\mathcal{A}} \times \tilde{\mathcal{S}} = \{(0,0), (0,1), (1,0), (1,1), (2,0), (2,1), \ldots\}$$

Each element of the property space is now a pair of values where the first value indicates the age and the second value indicates the sex. For example, $(A, S)_t(\omega) = (27, 1)$ would mean that $\omega$ is a male individual of age 27. Of course, given a two-dimensional variable one can derive two one-dimensional variables, in our example, one variable for age and another one for sex. The important point, however, is that given two one-dimensional variables, it will normally not be possible to reconstruct a two-dimensional variable.

**3.** This also means that the values of a two-dimensional variable must be tabulated simultaneously. Instead of two separate forms, like (6.5.1) and

(6.5.2), one needs a table that provides, for each individual $\omega \in \Omega_t$, a value of the two-dimensional variable $(A, S)_t$. Of course, the organization of the table is not important as long as one can identify for each person both its age and sex. So we might formally identify the two-dimensional variable $(A, S)_t$ with a pair of two one-dimensional variables, $(A_t, S_t)$, and organize the table as follows:

$$\begin{array}{c|cc} \omega & A_t(\omega) & S_t(\omega) \\ \hline \omega_1 & A_t(\omega_1) & S_t(\omega_1) \\ \vdots & \vdots & \vdots \\ \omega_{n_t} & A_t(\omega_{n_t}) & S_t(\omega_{n_t}) \end{array} \tag{6.5.3}$$

This is often called a *cross-tabulation* of the variables $A_t$ and $S_t$. However, the table should not be viewed as providing values for two variables separately. The important point is that we want to construct a two-dimensional distribution which provides a frequency for each element in the combined property space $\tilde{\mathcal{A}} \times \tilde{\mathcal{S}}$. Given any element $(\tau, s) \in \tilde{\mathcal{A}} \times \tilde{\mathcal{S}}$, we want to calculate the number of individuals in $\Omega_t$ who are of age $\tau$ *and* sex $s$, that is, the frequency

$$\mathrm{P}^*[A_t, S_t](\tau, s) := \left|\{\omega \in \Omega_t \,|\, (A, S)_t(\omega) = (\tau, s)\}\right|$$

The corresponding relative frequencies are then given by

$$\mathrm{P}[A_t, S_t](\tau, s) := \frac{\mathrm{P}^*[A_t, S_t](\tau, s)}{n_t}$$

These frequencies can finally be documented in a frequency table. For our example, such a table might be organized as follows:

$$\begin{array}{cc|cc} \tau & s & \mathrm{P}^*[A_t, S_t](\tau, s) & \mathrm{P}[A_t, S_t](\tau, s) \\ \hline 0 & 0 & \mathrm{P}^*[A_t, S_t](0,0) & \mathrm{P}[A_t, S_t](0,0) \\ 0 & 1 & \mathrm{P}^*[A_t, S_t](0,1) & \mathrm{P}[A_t, S_t](0,1) \\ 1 & 0 & \mathrm{P}^*[A_t, S_t](1,0) & \mathrm{P}[A_t, S_t](1,0) \\ 1 & 1 & \mathrm{P}^*[A_t, S_t](1,1) & \mathrm{P}[A_t, S_t](1,1) \\ \vdots & \vdots & \vdots & \vdots \end{array} \tag{6.5.4}$$

Each row shows the absolute and relative frequencies for one specific element in the combined property space $\tilde{\mathcal{A}} \times \tilde{\mathcal{S}}$. Notice that this way of organizing a frequency table can also be used for distributions having three or more dimensions. For two-dimensional distributions, another possibility would be to associate the elements of the two components of the property space with, respectively, the rows and columns of a frequency table. In our example, the table would then look as follows:



$$\begin{array}{c|cc} \tau & s = 0 & s = 1 \\ \hline 0 & \mathrm{P}^*[A_t, S_t](0,0) & \mathrm{P}^*[A_t, S_t](0,1) \\ 1 & \mathrm{P}^*[A_t, S_t](1,0) & \mathrm{P}^*[A_t, S_t](1,1) \\ 2 & \mathrm{P}^*[A_t, S_t](2,0) & \mathrm{P}^*[A_t, S_t](2,1) \\ \vdots & \vdots & \vdots \end{array} \tag{6.5.5}$$

In this example we have used absolute frequencies; it should be obvious, however, that the same kind of table can also be used to represent relative frequencies.

**4.** It is now easily seen that Table 6.5-1 that was used in Section 6.5.1 to report the data as published by the *Statistisches Bundesamt* is actually a combination of two frequency tables. The first one, already discussed in the previous section, consists of the first two columns. This part of the table refers to a one-dimensional age distribution, that is, tabulates the values of a function

$$\tau \longrightarrow n_{t,\tau} = \mathrm{P}^*[A_t](\tau)$$

In addition, the first, third, and fourth columns document a two-dimensional frequency distribution which refers simultaneously to age and sex. Obviously, this part of the table is organized in the same way as shown schematically in (6.5.5) and may be explicitly written as a function in the following way:

$$(\tau, s) \longrightarrow \mathrm{P}^*[A_t, S_t](\tau, s) = \begin{cases} n^m_{t,\tau} & \text{if } s = 0 \\ n^f_{t,\tau} & \text{if } s = 1 \end{cases}$$

**5.** Again, the question arises how to represent the data in a more accessible way. A general approach is to consider *conditional distributions*. For a two-dimensional distribution, the distribution of one of its components in sub-populations defined by specific values of the other component are considered. To illustrate, in our example one can use the elements of $\tilde{\mathcal{S}}$ to define two sub-populations

$$\Omega^m_t := \{\omega \in \Omega_t \,|\, S_t(\omega) = 0\} \quad \text{and} \quad \Omega^f_t := \{\omega \in \Omega_t \,|\, S_t(\omega) = 1\}$$

the first one comprising all male individuals and the second one all female individuals in $\Omega_t$. This then allows to define separate distributions of $A_t$ for the two sub-populations:

$$\mathrm{P}[A_t | S_t = s](\tau) := \frac{\mathrm{P}[A_t, S_t](\tau, s)}{\mathrm{P}[S_t](s)}$$

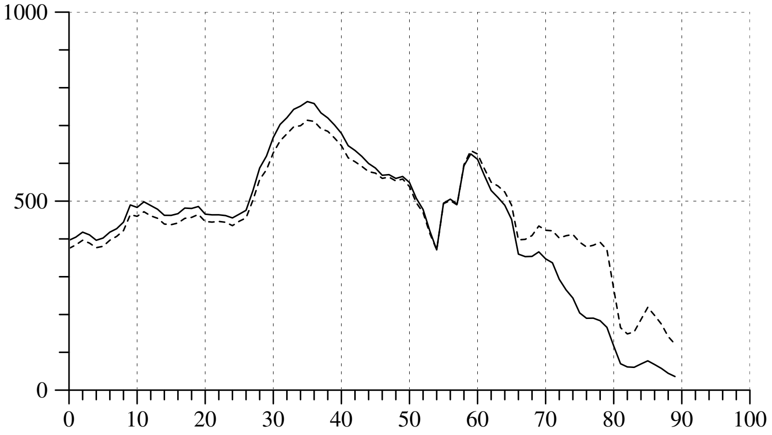



**Fig. 6.5-2** Plot of two frequency curves that represent the age distribution in Germany at the end of 1999 for men (solid line) and women (dotted line). Data are taken from Table 6.5-1, omitting category 90*. The scale of the ordinate is in 1000.

Specifically, in our example, $\mathrm{P}[A_t|S_t = 0](\tau) = n^m_{t,\tau}/n^m_t$ is the age distribution in sub-population $\Omega^m_t$, and $\mathrm{P}[A_t|S_t = 1](\tau) = n^f_{t,\tau}/n^f_t$ is the age distribution in sub-population $\Omega^f_t$.

**6.** Conditional distributions are most often expressed in terms of relative frequencies. It is easy, however, to derive corresponding absolute frequencies. One simply has to multiply $\mathrm{P}[A_t|S_t = y](\tau)$ by the number of people in the sub-population defined by $S_t = y$, explicitly written:

$$\mathrm{P}^*[A_t|S_t = y](\tau) := \mathrm{P}[A_t|S_t = y](\tau)\,\mathrm{P}^*[S_t](y)$$

The third column of 6.5-1 provides values for $\mathrm{P}^*[A_t|S_t = 0](\tau) = n^m_{t,\tau}$ and the fourth column provides values for $\mathrm{P}^*[A_t|S_t = 1](\tau) = n^f_{t,\tau}$.

**7.** In the same way as was done in the previous section, one can use frequency curves to get a visual impression of the distributions. In order to allow for a comparison, one can plot both curves in the same coordinate system as shown by Figure 6.5-2.[15] There are obviously some remarkable differences in the age distribution of men and women, especially in the older ages.

[15]In much of the demographic literature, one often finds a slightly different graphical presentation, called a *population pyramid*, which results from drawing the age distributions of men and women in opposite directions. To allow for an easy comparison, we prefer to plot the frequency curves in the same coordinate system.



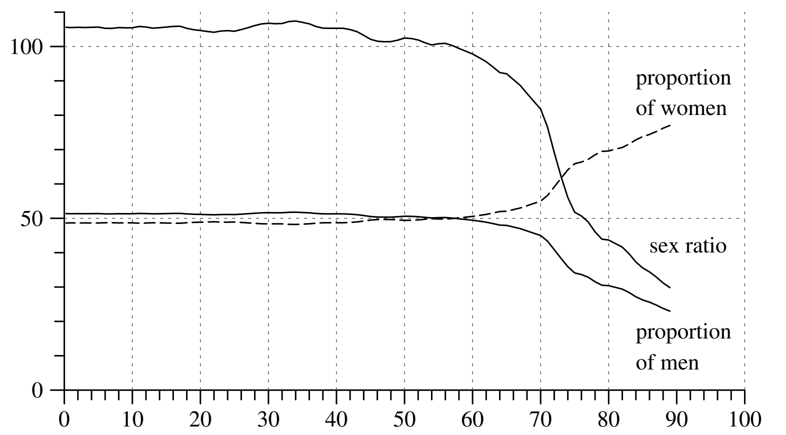



**Fig. 6.5-3** Dependency of sex ratios, and male-female proportions (in percent), on age in Germany, 1999; calculated from Table 6.5-1.

### 6.5.3 Male-Female Proportions

**1.** Figure 6.5-2 shows that the proportion of male and female individuals in a society depends on age. As a natural starting point one can consider the number of male births per 100 female births. This is called the *sex ratio at birth*. For an arbitrary age the definition is:

$$\text{Sex ratio at age } \tau := \frac{\text{number of men at age } \tau}{\text{number of women at age } \tau} \quad \text{(multiplied by 100)}$$

It has often been found that the sex ratio at birth is about 105 or 106. For example, in 1999 in Germany the number of male births was 396296 and the number of female births was 374448,[16] so one finds a sex ratio of 105.8.

**2.** Another possibility is to use proportions. We will use the notations

$$\sigma_{t,m} := \text{proportion of male birth in year } t$$
$$\sigma_{t,f} := \text{proportion of female birth in year } t$$

Referring again to the births in Germany in 1999, one finds $\sigma_{1999,m} = 0.514$ and $\sigma_{1999,f} = 0.486$. Given the sex ratio, one can calculate the male proportion by the sex ratio divided by 100 plus the sex ratio, and correspondingly for females.

**3.** Male and female proportions also depend on age. This can be seen by comparing age distributions for men and women as in Figure 6.5-2.

[16]Fachserie 1, Reihe 1, 1999 (p. 42).

Alternatively, one can calculate sex ratios, or, equivalently, proportions as functions of age. Based on the data in Table 6.5-1, this is shown in Figure 6.5-3. The changes in higher ages mainly result from different mortality rates of men and women; this will be further discussed in the next chapter. In Germany, an additional source of variation is due to in-migration.

### 6.5.4 Aggregating Age Values

**1.** A statistical distribution shows, for each value in a property space, its frequency in a population. The problem of comprehensible representation of statistical distributions will therefore depend on the number of elements of a property space. If the number is small, as in the case of sex, one can simply report the frequencies in the form of a small table. If, on the other hand, the number of values is large as, for example, in the case of age, frequency tables are not easily surveyed and one needs some additional means for the presentation of a distribution. In the foregoing sections we have used frequency curves to provide graphical displays. In this section we discuss an approach that relies on the aggregation of the elements in a property space.

**2.** The problems which arise from a large number of different elements of a property space are solved by merging several of them into classes of properties. Formally, this means that a property space is partitioned into classes and these classes are then considered as elements of a new property space. To illustrate, we use the property space $\tilde{\mathcal{A}}$ for age in completed years as introduced in Section 6.5.1. Its values can be partitioned into classes, for example:

$$\begin{array}{llll} \tilde{a}_1^* := \{0,\ldots,5\} & & \tilde{a}_4^* := \{31,\ldots,64\} \\ \tilde{a}_2^* := \{6,\ldots,18\} & & \tilde{a}_5^* := \{65,\ldots,79\} \\ \tilde{a}_3^* := \{19,\ldots,30\} & & \tilde{a}_6^* := \{80,\ldots\} \end{array}$$

These age classes can then be considered as elements of a new property space

$$\tilde{\mathcal{A}}^* := \{\tilde{a}_1^*, \tilde{a}_2^*, \tilde{a}_3^*, \tilde{a}_4^*, \tilde{a}_5^*, \tilde{a}_6^*\}$$

which, in turn, can be used to define a new variable $A_t^* : \Omega_t \longrightarrow \tilde{\mathcal{A}}^*$. It should be obvious how its values are derived from the values of the original variable $A_t$. For any $\omega \in \Omega_t$, if $A_t(\omega) = \tau$ and $\tau \in \tilde{a}_j^*$, then $A_t^*(\omega) = \tilde{a}_j^*$. Less formally, if an individual is of age $\tau$, it is assigned the age class that contains $\tau$.

**3.** The distribution of $A_t^*$ is easily derived from the distribution of $A_t$ because frequencies are additive. For each $\tilde{a}_j^* \in \tilde{\mathcal{A}}^*$, its absolute and relative



frequencies are given by

$$\mathrm{P}^*[A_t^*](\tilde{a}_j^*) = \sum_{\tau \in \tilde{a}_j^*} \mathrm{P}^*[A_t](\tau) \quad \text{and} \quad \mathrm{P}[A_t^*](\tilde{a}_j^*) = \sum_{\tau \in \tilde{a}_j^*} \mathrm{P}[A_t](\tau)$$

respectively. To illustrate, using the data from Table 6.5-1, we find the following frequencies:

| $\tilde{a}^*$ | | $\mathrm{P}^*[A_t^*](\tilde{a}^*)$ | $\mathrm{P}[A_t^*](\tilde{a}^*)$ |
|---|---|---|---|
| $\tilde{a}_1^*$ | $\{0,\ldots,5\}$ | 4743.1 | 0.0578 |
| $\tilde{a}_2^*$ | $\{6,\ldots,18\}$ | 11886.1 | 0.1448 |
| $\tilde{a}_3^*$ | $\{19,\ldots,30\}$ | 12154.7 | 0.1481 |
| $\tilde{a}_4^*$ | $\{31,\ldots,64\}$ | 40095.6 | 0.4885 |
| $\tilde{a}_5^*$ | $\{65,\ldots,79\}$ | 10285.8 | 0.1253 |
| $\tilde{a}_6^*$ | $\{80,\ldots\}$ | 2921.3 | 0.0356 |

(6.5.6)

This will be called an *aggregated frequency table*. Of course, the expression 'aggregated' is to be understood as referring to the original frequency table from which the aggregated table is derived.

### 6.5.5 Age Distributions since 1952

**1.** Any age distribution refers to a certain historical time $t$. How do age distributions change with time? We base our description of the German development through the last 50 years on data available from the STATIS data base of the *Statistisches Bundesamt* (see Appendix A.1). The data set refers to the territory of the former FRG and covers the years 1952 to 1998.[17] For each of these years, there is an age distribution (in completed years) both for men and women. Using previously introduced notations, the absolute frequencies are given as values for $n_{t,\tau}^m$ and $n_{t,\tau}^f$. In order to simplify the presentation, we disregard sex and use $n_{t,\tau} = n_{t,\tau}^m + n_{t,\tau}^f$. Schematically, the data file then looks as follows:

| $\tau$ | 1952 | $\cdots$ | 1998 |
|---|---|---|---|
| 0 | $n_{1952,0}$ | $\cdots$ | $n_{1998,0}$ |
| 1 | $n_{1952,1}$ | $\cdots$ | $n_{1998,1}$ |
| 2 | $n_{1952,2}$ | $\cdots$ | $n_{1998,2}$ |
| $\vdots$ | $\vdots$ | | $\vdots$ |
| 90* | $n_{1952,90^*}$ | $\cdots$ | $n_{1998,90^*}$ |

[17]Segment 36, as updated in June 2000.

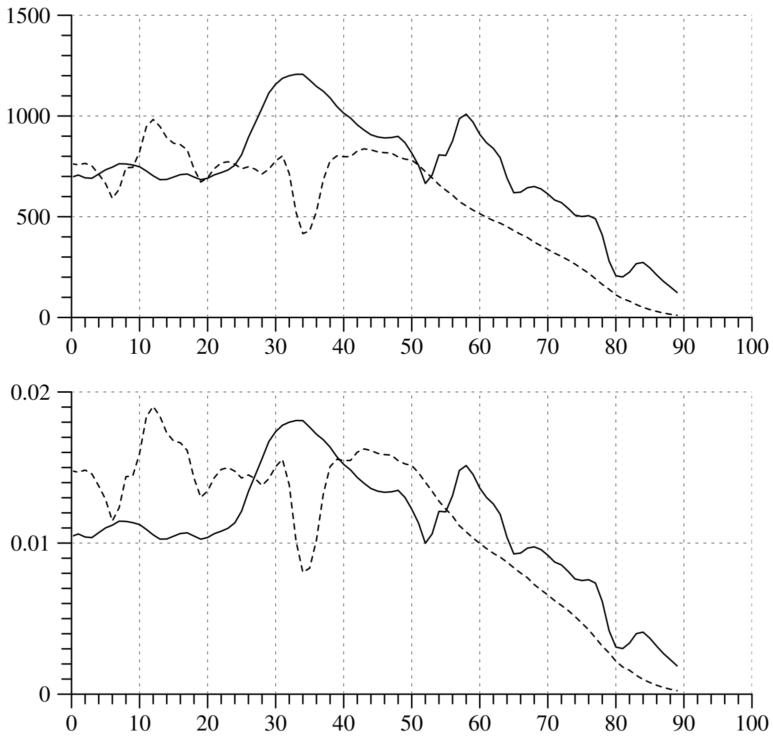



**Fig. 6.5-4** Age distribution in the territory of the former FRG in 1998 (solid line) and 1952 (dotted line). In the upper plot, the frequency curves refer to absolute frequencies (in 1000), in the lower plot they refer to relative frequencies.

The property space for age is the same as in Table 6.5-1; the last value, 90*, represents an open-ended class that comprises ages of 90 and above.

**2.** To begin with, we compare the age distribution in the years 1952 and 1998 with the help of frequency curves. The result is shown in Figure 6.5-4. Since population size has changed during this period,[18] the figure provides curves both for absolute frequencies (in the upper plot) and relative frequencies (in the lower plot). Obviously, there is a remarkable increase in the number of people in higher ages, both in terms of absolute and relative frequencies. The comparison also shows that the marked irregularities in the age distributions are mainly due to the development of the demographic process since World War I.

**3.** It remains the question of how to get an impression of the year-to-year changes in the age distribution for the whole period from 1952 to 1998. It

[18]Based on this data file, the population size (in 1000) is 51620 in 1952 and 66697 in 1998.



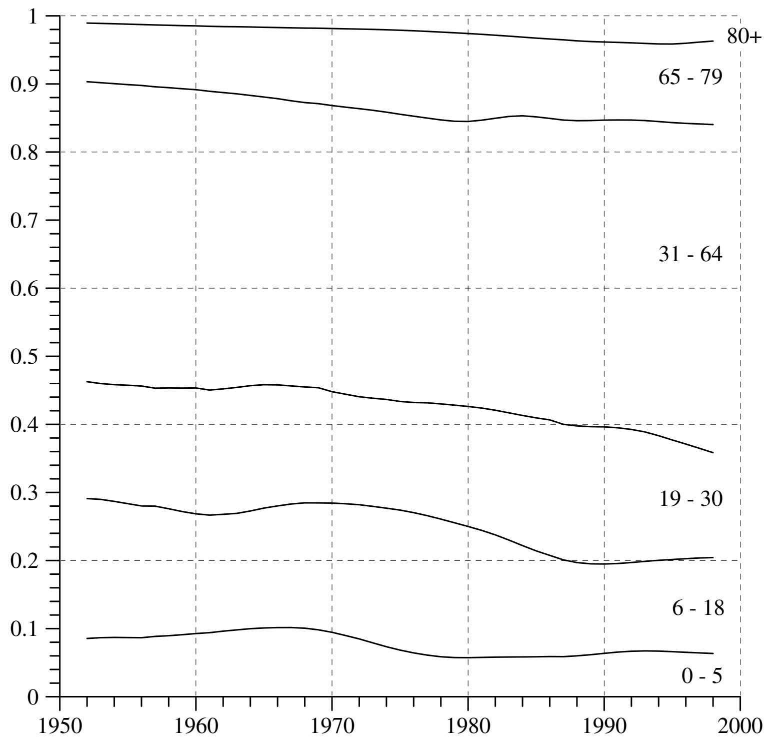



**Fig. 6.5-5** Development of the age distribution in the territory of the former FRG, from 1952 to 1998. Shown are proportions in six age classes as indicated on the right-hand side of the graphic.

is obviously not sensible to provide frequency curves separately for each of these years. The general question is how to describe a time series of frequency distributions. A simple approach would be to characterize each distribution by its mean value and then to plot the development of these means. This would show how the mean age of the population has changed over the period. In our application we find that this mean age continuously increased from 38 years in 1952 to 44 years in 1998.[19]

**4.** However, just to report the development of mean ages provides only limited information. To provide additional information, we use the method of aggregation discussed in the previous section. To illustrate, the same partition into 6 age classes is used. Calculating relative frequencies for these age classes, results in proportions

$$\mathrm{P}[A_t^*](\tilde{a}_1^*), \ldots, \mathrm{P}[A_t^*](\tilde{a}_6^*) \qquad (t = 1952, \ldots, 1998)$$

[19]For these calculations we have assumed an age of 90 for all people in the age class 90*. Other assumptions would result in slightly higher mean age values.

For each year, these proportions sum to unity and can be displayed in a *plot of proportions* as shown in Figure 6.5-5. The plot impressively shows the rise of the proportion of elderly people during the period from 1952 to 1998.

# Chapter 7

# Mortality and Life Tables

The development of a demographic process primarily depends on birth and death events. In this chapter, we discuss methods that have been proposed to quantify mortality and illustrate these methods with data from German official statistics. We begin with death events because birth events require more complicated methods. The reason is simply that each person must eventually die, and will only die once, while a women can give birth to several children.

## 7.1 Mortality Rates

**1.** A simple way to quantify mortality is to count the number of deaths which occur in a year and relate this to the midyear size of the population in that year. This is called a *crude death rate* or *crude mortality rate*.[1] Using previously introduced notations, the definition is

$$\text{Crude death rate} := \frac{d_t}{n_t} \quad \text{(multiplied by 1000)}$$

where the temporal index $t$ most often refers to calendar years. How these crude death rates have developed in Germany has already been shown in Section 6.3.

**2.** An obvious problem with crude death rates is that they do not take into account changes in the age distribution of a population, or differences between the age distributions of two or more populations that are to be compared with respect to mortality. Since mortality is highly dependent on age, a better starting point is to calculate *age-specific death rates* which will be denoted by

$$\delta_{t,\tau} := \frac{d_{t,\tau}}{n_{t,\tau}}$$

In this definition, $d_{t,\tau}$ denotes the number of people who died in year $t$ at the age of $\tau$, and $n_{t,\tau}$ is the midyear estimate of the number of people in year $t$ being of age $\tau$. Furthermore, since mortality is also different for men and women, we use the notations

$$\delta^m_{t,\tau} := \frac{d^m_{t,\tau}}{n^m_{t,\tau}} \qquad \text{and} \qquad \delta^f_{t,\tau} := \frac{d^f_{t,\tau}}{n^f_{t,\tau}}$$

[1] In publications of the *Statistisches Bundesamt* this is called *allgemeine Sterbeziffer*. Some authors also use the term ‘rohe Sterblichkeitsrate’, or ‘rohe Mortalitätsrate’.

**Table 7.1-1** Midyear population and number of deaths in Germany 1999, subdivided by age ($\tau$); 95* comprises all ages $\tau \geq 95$. Source: Segments 685 and 1124-26 of the STATIS data base of the *Statistisches Bundesamt*.

| $\tau$ | $n^m_{t,\tau}$ | $d^m_{t,\tau}$ | $n^f_{t,\tau}$ | $d^f_{t,\tau}$ | $\tau$ | $n^m_{t,\tau}$ | $d^m_{t,\tau}$ | $n^f_{t,\tau}$ | $d^f_{t,\tau}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 399633 | 1979 | 378251 | 1517 | 48 | 563942 | 2422 | 556038 | 1222 |
| 1 | 410782 | 173 | 389437 | 137 | 49 | 558611 | 2555 | 548502 | 1320 |
| 2 | 413836 | 126 | 391872 | 83 | 50 | 529900 | 2745 | 517217 | 1354 |
| 3 | 403107 | 90 | 381972 | 63 | 51 | 494176 | 2691 | 482985 | 1414 |
| 4 | 398813 | 79 | 377741 | 54 | 52 | 450618 | 2873 | 442241 | 1421 |
| 5 | 409761 | 52 | 387869 | 41 | 53 | 397251 | 2432 | 392979 | 1287 |
| 6 | 422128 | 70 | 400905 | 43 | 54 | 434885 | 3182 | 432973 | 1618 |
| 7 | 435168 | 66 | 413392 | 52 | 55 | 502194 | 4012 | 498201 | 2023 |
| 8 | 466447 | 76 | 442107 | 44 | 56 | 500889 | 4244 | 496426 | 2010 |
| 9 | 485976 | 56 | 461036 | 54 | 57 | 545671 | 5189 | 544075 | 2493 |
| 10 | 490076 | 67 | 464833 | 42 | 58 | 612495 | 5906 | 615985 | 2831 |
| 11 | 492537 | 71 | 465361 | 42 | 59 | 621650 | 6970 | 630333 | 3345 |
| 12 | 482637 | 76 | 456705 | 54 | 60 | 593275 | 7303 | 606556 | 3514 |
| 13 | 469636 | 79 | 445784 | 50 | 61 | 551237 | 7359 | 569910 | 3497 |
| 14 | 461216 | 114 | 437317 | 65 | 62 | 522738 | 7819 | 546881 | 3692 |
| 15 | 463159 | 145 | 438441 | 84 | 63 | 502833 | 8422 | 534670 | 3985 |
| 16 | 472798 | 194 | 446710 | 108 | 64 | 472336 | 8835 | 511058 | 4451 |
| 17 | 479914 | 314 | 453245 | 148 | 65 | 409837 | 8207 | 445060 | 4173 |
| 18 | 481413 | 487 | 457174 | 159 | 66 | 361225 | 7990 | 399667 | 4074 |
| 19 | 473334 | 454 | 451301 | 168 | 67 | 359314 | 8984 | 405178 | 4766 |
| 20 | 462189 | 435 | 441633 | 137 | 68 | 365748 | 10288 | 423645 | 5572 |
| 21 | 460967 | 468 | 441747 | 156 | 69 | 362427 | 11122 | 431229 | 6279 |
| 22 | 460272 | 412 | 441953 | 125 | 70 | 347956 | 11690 | 425406 | 6858 |
| 23 | 456346 | 401 | 436715 | 118 | 71 | 319299 | 11439 | 416670 | 7489 |
| 24 | 458828 | 441 | 438622 | 135 | 72 | 283797 | 10934 | 410392 | 8366 |
| 25 | 469223 | 377 | 449325 | 155 | 73 | 258704 | 11017 | 416006 | 9404 |
| 26 | 500605 | 443 | 477112 | 143 | 74 | 228253 | 10847 | 407353 | 10147 |
| 27 | 557051 | 473 | 528212 | 169 | 75 | 202335 | 10404 | 390714 | 11172 |
| 28 | 603444 | 533 | 568797 | 200 | 76 | 196282 | 11027 | 387353 | 12516 |
| 29 | 644108 | 528 | 604417 | 214 | 77 | 193182 | 12292 | 395011 | 14616 |
| 30 | 686013 | 614 | 642576 | 235 | 78 | 180053 | 12610 | 390951 | 16287 |
| 31 | 712393 | 627 | 668147 | 263 | 79 | 144258 | 12452 | 328396 | 17041 |
| 32 | 732651 | 666 | 686810 | 308 | 80 | 96563 | 7484 | 221190 | 10678 |
| 33 | 748125 | 740 | 697478 | 345 | 81 | 68960 | 6478 | 161774 | 9951 |
| 34 | 758218 | 809 | 705906 | 365 | 82 | 65068 | 6763 | 156443 | 11125 |
| 35 | 761731 | 872 | 711736 | 431 | 83 | 70399 | 7840 | 177521 | 13397 |
| 36 | 746559 | 1066 | 700436 | 433 | 84 | 79463 | 10704 | 213228 | 20226 |
| 37 | 727433 | 1076 | 687469 | 539 | 85 | 78125 | 11001 | 219593 | 22124 |
| 38 | 711239 | 1126 | 675052 | 559 | 86 | 67852 | 10565 | 197433 | 22470 |
| 39 | 690941 | 1300 | 656171 | 629 | 87 | 55469 | 9517 | 168730 | 21892 |
| 40 | 663561 | 1334 | 630073 | 667 | 88 | 44330 | 8191 | 141995 | 20816 |
| 41 | 641087 | 1418 | 608883 | 717 | 89 | 35770 | 7471 | 120032 | 19758 |
| 42 | 627205 | 1572 | 597762 | 786 | 90 | 27775 | 6466 | 97204 | 18053 |
| 43 | 610005 | 1719 | 584657 | 858 | 91 | 20817 | 5174 | 75582 | 15700 |
| 44 | 594357 | 1788 | 575662 | 892 | 92 | 15772 | 4072 | 58386 | 13384 |
| 45 | 578806 | 1983 | 566937 | 980 | 93 | 11724 | 3170 | 42726 | 10696 |
| 46 | 570259 | 2055 | 561714 | 1097 | 94 | 8658 | 2446 | 30973 | 8332 |
| 47 | 566062 | 2223 | 558353 | 1136 | 95* | 21807 | 4871 | 69931 | 20949 |



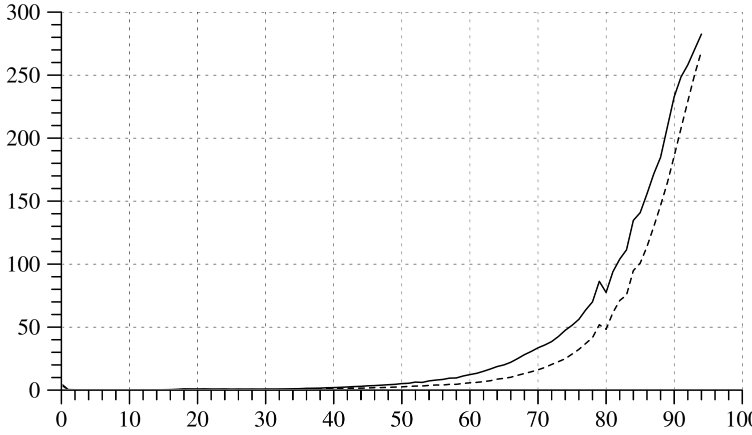



**Fig. 7.1-1** Age-specific death rates (per 1000) for men (solid line) and women (dotted line) in Germany 1999, calculated from Table 7.1-1. The plot is restricted to ages less than 95.

to refer to age-specific death rates for men and women, respectively. Like crude death rates, also age-specific death rates are often multiplied by 1000, this is marked by adding a tilde:

$$\tilde{\delta}^m_{t,\tau} := 1000\,\delta^m_{t,\tau}, \quad \tilde{\delta}^f_{t,\tau} := 1000\,\delta^f_{t,\tau}, \quad \tilde{\delta}_{t,\tau} := 1000\,\delta_{t,\tau}$$

**3.** To illustrate the calculations, we use data for the year 1999 shown in Table 7.1-1.[2] For an age of 60, one finds

$$\tilde{\delta}^m_{1999,60} = \frac{7303}{593.275} = 12.31 \quad \text{and} \quad \tilde{\delta}^f_{1999,60} = \frac{3514}{606.556} = 5.79$$

Performing these calculations for all ages, the resulting death rates can be visualized as shown in Figure 7.1-1. The figure clearly shows how mortality varies with age, and also shows that, in older ages, death rates of men are higher than death rates of women. So the figure also confirms that, when investigating mortality, one should take into account age and sex.

**4.** The higher mortality of male individuals already begins in an early age. This is hidden in Figure 7.1-1 because death rates are generally very low until an age of about 50. The rates for the range of ages from 0 to 50 are shown in Figure 7.1-2. The figure shows that significant differences in the death rates already begin at an age of about 15. It seems plausible that at least some part of the higher mortality of male individuals is also due to different behavior and/or socio-economic conditions.

---

[2] Values for the midyear population correspond to those given in Table 6.5-1; of course, the seemingly exact values in Table 7.1-1 result from population projections and should be understood as estimates.

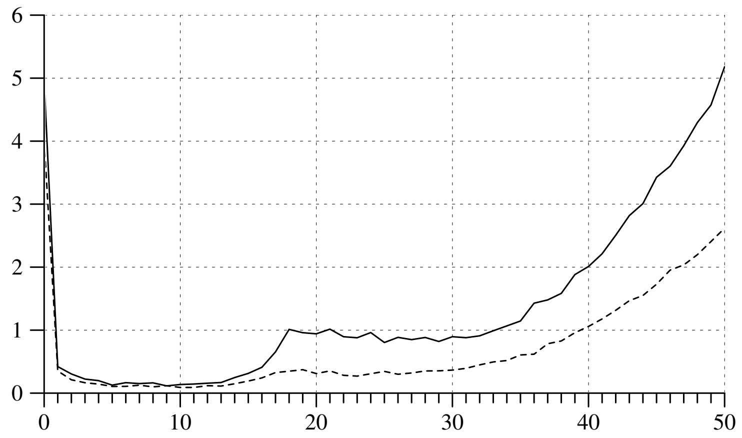



**Fig. 7.1-2** Age-specific death rates (per 1000) for men (solid line) and women (dotted line) in Germany 1999, restricted to ages from 0 to 50. Calculated from Table 7.1-1.

**5.** A consideration of age-specific death rates also shows why crude death rates may suggest misleading conclusions. While the crude death rates, as shown in Figure 6.3-1, have increased until about 1970, the age-specific death rates actually declined during this period. This is shown by the figures in Table 7.1-2. For example, for men, the crude death rate was 12.0 in 1952 and 13.1 in 1970, but in almost all age groups the age-specific death rates were lower in 1970 than in 1952.

**6.** The crude death rate can be written as a weighted mean of age-specific death rates. Denoting by $a_{t,\tau} := n_{t,\tau}/n_t$ the proportion of persons of age $\tau$, in year $t$, the relationship is as follows:

$$\frac{d_t}{n_t} = \frac{\sum_\tau d_{t,\tau}}{n_t} = \sum\nolimits_\tau \frac{d_{t,\tau}}{n_t} = \sum\nolimits_\tau \frac{d_{t,\tau}}{n_{t,\tau}} \frac{n_{t,\tau}}{n_t} = \sum\nolimits_\tau \delta_{t,\tau}\, a_{t,\tau}$$

The equation shows that the crude death rate depends on both, the age-specific death rates $\delta_{t,\tau}$, and the age distribution given by $a_{t,\tau}$. This suggests that, in order to compare death rates for different years, or among different territories, one can use *standardized death rates* which refer to a common age distribution. As an example, Table 7.1-3 shows crude and standardized death rates for Germany, as calculated by the *Statistisches Bundesamt*. The standardization is based on the age distribution in 1995; the standardized death rates shown in this table are therefore calculated with the following formula:

$$\text{Standardized death rate for year } t = \sum\nolimits_\tau \delta_{t,\tau}\, a_{1995,\tau}$$

In contrast to the crude death rates, the standardized death rates declined



**Table 7.1-2** Death rates for specified age groups in Germany. For all years, the figures refer to the territories of both the former FRG and the former GDR. Source: Fachserie 1, Reihe 1, 1999 (p. 230).

| | Age | 1952 | 1970 | 1990 | 1999 |
|---|---|---|---|---|---|
| Men | 0 | 59.661 | 25.242 | 8.223 | 4.952 |
| | 1 – 4 | 2.242 | 1.058 | 0.469 | 0.288 |
| | 5 – 9 | 0.818 | 0.600 | 0.243 | 0.144 |
| | 10 – 14 | 0.686 | 0.493 | 0.243 | 0.170 |
| | 15 – 19 | 1.266 | 1.395 | 0.819 | 0.672 |
| | 20 – 24 | 1.883 | 1.731 | 1.097 | 0.938 |
| | 25 – 29 | 1.862 | 1.616 | 1.133 | 0.848 |
| | 30 – 34 | 2.062 | 1.814 | 1.467 | 0.950 |
| | 35 – 39 | 2.701 | 2.443 | 2.026 | 1.495 |
| | 40 – 44 | 3.785 | 3.725 | 2.816 | 2.497 |
| | 45 – 49 | 5.951 | 5.736 | 4.797 | 3.960 |
| | 50 – 54 | 10.068 | 9.155 | 7.539 | 6.036 |
| | 55 – 59 | 15.428 | 15.240 | 12.526 | 9.458 |
| | 60 – 64 | 23.805 | 26.353 | 19.486 | 15.038 |
| | 65 – 69 | 36.701 | 44.226 | 30.838 | 25.068 |
| | 70 – 74 | 58.846 | 69.332 | 47.465 | 38.892 |
| | 75 – 79 | 96.879 | 102.786 | 80.038 | 64.168 |
| | 80 – 84 | 158.387 | 153.799 | 129.312 | 103.216 |
| | 85 – 89 | 242.165 | 232.039 | 197.655 | 166.030 |
| | 90 – | 352.626 | 341.384 | 309.298 | 245.878 |
| | Age | 1952 | 1970 | 1990 | 1999 |
| Women | 0 | 47.075 | 19.089 | 6.209 | 4.011 |
| | 1 – 4 | 1.773 | 0.836 | 0.377 | 0.219 |
| | 5 – 9 | 0.533 | 0.389 | 0.196 | 0.111 |
| | 10 – 14 | 0.398 | 0.299 | 0.165 | 0.111 |
| | 15 – 19 | 0.720 | 0.561 | 0.325 | 0.297 |
| | 20 – 24 | 1.095 | 0.616 | 0.405 | 0.305 |
| | 25 – 29 | 1.282 | 0.703 | 0.434 | 0.335 |
| | 30 – 34 | 1.550 | 0.903 | 0.626 | 0.446 |
| | 35 – 39 | 2.153 | 1.425 | 1.035 | 0.755 |
| | 40 – 44 | 2.864 | 2.237 | 1.515 | 1.308 |
| | 45 – 49 | 4.225 | 3.606 | 2.440 | 2.062 |
| | 50 – 54 | 6.311 | 5.328 | 3.497 | 3.127 |
| | 55 – 59 | 9.494 | 7.869 | 5.697 | 4.561 |
| | 60 – 64 | 15.436 | 13.050 | 9.196 | 6.912 |
| | 65 – 69 | 27.302 | 23.033 | 15.301 | 11.813 |
| | 70 – 74 | 48.909 | 41.437 | 25.386 | 20.360 |
| | 75 – 79 | 86.029 | 73.536 | 48.149 | 37.852 |
| | 80 – 84 | 142.401 | 126.783 | 87.257 | 70.286 |
| | 85 – 89 | 220.005 | 204.513 | 152.491 | 126.282 |
| | 90 – | 328.131 | 317.971 | 268.712 | 232.427 |

**Table 7.1-3** Crude and standardized death rates in Germany. The standardized death rates are based on the age distribution in 1995. Source: Fachserie 1, Reihe 1, 1999 (p. 55).

| | Crude death rates | | | Standardized death rates | | |
|---|---|---|---|---|---|---|
| Year | Male | Female | Both | Male | Female | Both |
| 1952 | 12.0 | 10.1 | 11.1 | 14.8 | 20.9 | 17.9 |
| 1960 | 13.2 | 11.0 | 12.0 | 15.2 | 20.0 | 17.6 |
| 1970 | 13.1 | 12.0 | 12.6 | 14.9 | 18.2 | 16.6 |
| 1980 | 12.2 | 12.1 | 12.1 | 13.3 | 15.1 | 14.2 |
| 1990 | 11.1 | 12.1 | 11.6 | 11.5 | 12.7 | 12.1 |
| 1999 | 9.8 | 10.8 | 10.3 | 9.2 | 10.3 | 9.8 |

over the whole period from 1952 to 1999.[3] These rates, therefore, summarize the development of the age-specific death rates.

## 7.2 Mean Age at Death

**1.** Another approach to summarize information about mortality uses either mean life length or mean age at death. In order to define these concepts one needs to refer to a population. Mean age at death refers to all people who died in a specific year, while mean life length refers to birth cohorts, that is, to sets of people born in the same year. Actually, many calculations of "life expectations" neither follow the first nor the second of these two approaches. In fact, they do not refer to any population at all but construct a fictitious distribution for the length of life with the help of a period life table. This will be discussed in Section 7.3. In the present section we briefly discuss the calculation of mean age at death. The discussion of mean life length will be postponed because most often one then needs to take into account incomplete observations.

**2.** We will denote the set of people who died in year $t$ by $\Omega_t^\dagger$. Implied by the general framework introduced in Chapter 3, this is a subset of $\Omega_t$. We can then formally define a variable

$$A_t^\dagger : \Omega_t^\dagger \longrightarrow \tilde{\mathcal{A}} := \{0, 1, 2, \ldots\}$$

which provides, for each person $\omega \in \Omega_t^\dagger$, an age at death, $A_t^\dagger(\omega)$. This then implies a statistical distribution for the variable $A_t^\dagger$, and its mean value,

$$\mathrm{M}(A_t^\dagger) = \sum_{\omega \in \Omega_t^\dagger} A_t^\dagger(\omega) \,/\, |\Omega_t^\dagger|$$

[3]One exception is male mortality in the period following World War II.



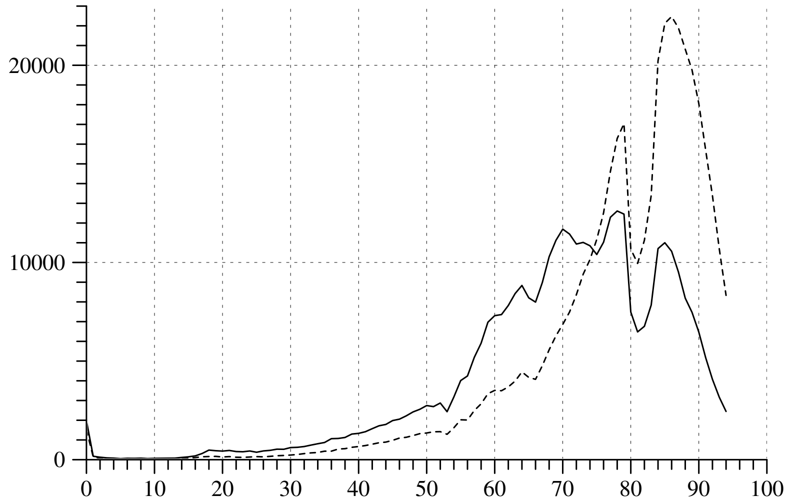



**Fig. 7.2-1** Frequency curves showing the distribution of age at death for men (solid line) and women (dotted line) who died in Germany in 1999; curves are restricted to ages less than 95. Calculated from Table 7.1-1.

is the mean life length of the people in $\Omega_t^\dagger$. In the following, we use this conceptual framework but additionally distinguish men and women: $\Omega_t^\dagger = \Omega_t^{\dagger,m} \cup \Omega_t^{\dagger,f}$. The corresponding variables will be denoted by $A_t^{\dagger,m}$ and $A_t^{\dagger,f}$, respectively.

**3.** Table 7.1-1 provides information on $\Omega_{1999}^{\dagger,m}$ and $\Omega_{1999}^{\dagger,f}$, the sets of, respectively, men and women who died in Germany in 1999. Absolute frequencies for the variables $A_t^{\dagger,m}$ and $A_t^{\dagger,f}$ are given by the entries $d_{t,\tau}^m$ and $d_{t,\tau}^f$. Summing up these values, one finds

$$d_t^m := |\,\Omega_t^{\dagger,m}\,| = 390742 \quad \text{and} \quad d_t^f := |\,\Omega_t^{\dagger,f}\,| = 455588$$

and, by dividing the entries by these numbers, one immediately also finds the relative frequencies:

$$\mathrm{P}[A_t^{\dagger,m}](\tau) = \frac{d_{t,\tau}^m}{d_t^m} \quad \text{and} \quad \mathrm{P}[A_t^{\dagger,f}](\tau) = \frac{d_{t,\tau}^f}{d_t^f}$$

**4.** In a next step, one can visualize the distributions of $A_t^{\dagger,m}$ and $A_t^{\dagger,f}$ by frequency curves. Using absolute frequencies, the curves are shown in Figure 7.2-1. It is seen that the curves are remarkably different for men and women and also depend on the age distribution in 1999.[4] Since age is

[4]This is true, in particular, for ages around 80; see the age distributions in Figure 6.5-2.

recorded in completed years, it seems sensible to use the formula

$$\sum_{\tau=0}^{\infty} (\tau + 0.5)\, \mathrm{P}[A_t^\dagger](\tau) \;=\; \mathrm{M}(A_t^\dagger) + 0.5$$

suitably modified for men and women, to calculate the mean age at death. However, an obvious problem concerns the open-ended age class beginning at age 95 in Table 7.1-1. One needs some estimate, say $a$, for the mean age of the people who died at an age equal to, or greater than, 95. This would then allow to rewrite the formula as

$$\sum_{\tau=0}^{94} (\tau + 0.5)\, \mathrm{P}[A_t^\dagger](\tau) + a\, \mathrm{P}[A_t^\dagger](95^*)$$

Using $a = 95.5$ would result in a minimal mean life length. It might be more reasonable to use a somewhat higher estimate. To see the dependency on $a$, we calculate the first term which is 69.46 for men and 74.48 for women. Using the proportions of men and women in the $95^*$ age class, one gets the formulas

$$69.46 + a\,0.0125 \quad \text{and} \quad 74.48 + a\,0.0460$$

Assuming $a = 95.5$, the mean age at death would be 70.7 for men and 78.9 for women. However, also a much higher value of $a$ would only slightly increase the estimates. Thus it seems safe to believe that the mean age at death in Germany in 1999 is about 71 years for men and 80 years for women.

**5.** We briefly mention another possibility to report some kind of mean value for a distribution which is called its *median* and refers to the distribution function of a statistical variable. We first introduce the notion of a distribution function. If $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ is any statistical variable, its *distribution function* is defined as a function $F[X] : \mathbf{R} \longrightarrow \mathbf{R}$ which associates with each number $x \in \mathbf{R}$ the proportion of elements of $\Omega$ having a value of the variable $X$ which is less than, or equal to, $x$. In a formal notation:

$$F[X](x) \;:=\; \frac{\left|\{\omega \in \Omega \,|\, X(\omega) \leq x\}\right|}{|\,\Omega\,|}$$

which also shows that the values of a distribution function are always between 0 and 1. If the number of elements in the property space $\tilde{\mathcal{X}}$ is finite (which is always the case in the examples of this text), there is a simple relationship between a distribution function and relative frequencies:

$$F[X](x) \;=\; \sum_{\tilde{x} \leq x} \mathrm{P}[X](\tilde{x})$$



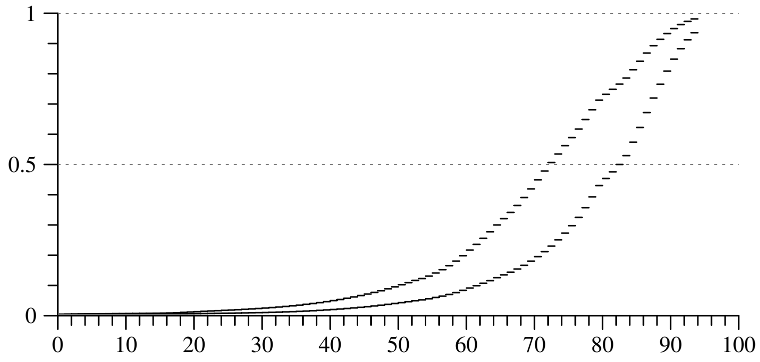



**Fig. 7.2-2** Distribution functions for the variables $A_t^{\dagger,m}$ and $A_t^{\dagger,f}$ which record the age at death in Germany 1999. The upper curve refers to men, the lower curve to women, both curves are restricted to ages less than 95. Calculated from Table 7.1-1.

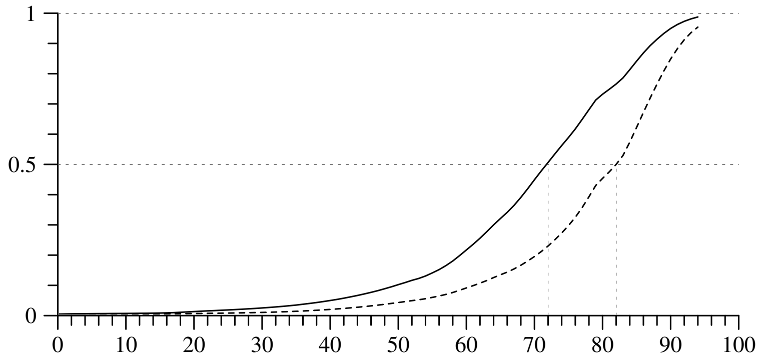



**Fig. 7.2-3** Distribution functions for the variables $A_t^{\dagger,m}$ (solid line) and $A_t^{\dagger,f}$ (dotted line) which record the age at death in Germany 1999; both curves are restricted to ages less than 95. Calculated from Table 7.1-1.

Therefore, a distribution function is sometimes also called a *cumulated frequency function*.

**6.** Using the data from Table 7.1-1, distribution functions of the variables $A_t^{\dagger,m}$ and $A_t^{\dagger,f}$ are plotted in Figure 7.2-2. As seen from this figure, distribution functions are step functions with jumps at the elements of a variable's realized property space.[5] Of course, if the realized property space contains many values which are near together, it might be sensible, for better visibility, to connect the function values by linear line segments.

[5]The figure shows the distribution function only for age values less than 95 since, from the data in Table 7.1-1, we do not know where the function approaches unity. By the definition of a distribution function, one only knows that $F(\infty) = 1$.

For our example, this is shown in Figure 7.2-3. This figure also illustrates the notion of median: If $F[X]$ denotes the distribution function of a statistical variable $X$, its *median* is defined as a number, say $m_x$, such that $F[X](m_x) \approx 0.5$.[6] Using this definition, one finds from Figure 7.2-3 that the median life length is about 72 years for men and 82 years for women. These are somewhat higher than the mean values calculated above since most of the frequencies occur in the upper right part of the distribution.

**7.** The median of a distribution can be interpreted in the following way: about half of the population has property values below and another half has property values above that number. In our example, about half of the men who died in 1999 died at ages below 72 years. One might notice that the calculation of a median does not require complete knowledge about a distribution. Contrary to the calculation of mean values discussed above, the median life length is quite independent of the form of the distribution function below and above its median. In particular, in our example, one does not need any assumptions about the mean age of the people who died in ages higher than 90 years.

## 7.3 Life Tables

Mean age at death refers to people who died in a specific year. Another approach is to think in terms of life length of people born in the same year or period. This leads to the idea of life tables. As will be discussed later one has to distinguish cohort and period life tables. In order to prepare this discussion we first introduce the notion of duration variables.

### 7.3.1 Duration Variables

**1.** Life length is just one example of duration data. In general, duration data can refer to almost any kind of duration, for example, job durations and marriage durations. In this section, before continuing with a discussion of mortality, we introduce definitions and notations which are helpful to deal not only with life length but with other kinds of duration data as well. The starting point is a general *duration variable*

$$T : \Omega \longrightarrow \tilde{\mathcal{T}} := \{0, 1, 2, 3, \ldots\}$$

which is defined for some population $\Omega$. For each individual $\omega \in \Omega$, the variable $T$ records a duration $T(\omega) \in \tilde{\mathcal{T}}$. As mentioned, $\tilde{\mathcal{T}}$ can refer to life length, job duration, marriage duration, or any other kind of duration. In

[6]Since distribution functions are step functions, there normally is no unique number $m_x$ such that $F[X](m_x)$ exactly equals 0.5. For practical computations, an often used approach is to sort the values of a variable in ascending order and then to choose the mid-value, if the number of data is uneven, or otherwise the mean of two neighboring mid-values.



any case, $\tilde{\mathcal{T}}$ will be considered as a discrete time axis representing temporal locations $0, 1, 2, \ldots$ which might be days, months, or years. Therefore, if $T(\omega) = t$, this means that the event terminating $\omega$'s duration occurs somewhere in the temporal location $t$, and the duration amounts to $t$ completed time units.

**2.** Since $T$ is a statistical variable, it has a statistical distribution defined by a frequency function

$$\mathrm{P}[T](t) \;=\; \frac{\big|\{\omega \in \Omega \,|\, T(\omega) = t\}\big|}{|\,\Omega\,|}$$

For each $t \in \tilde{\mathcal{T}}$, $\mathrm{P}[T](t)$ is the proportion of individuals in $\Omega$ for whom the variable $T$ has the value $t$. For example, if $\tilde{\mathcal{T}}$ refers to life length, $\mathrm{P}[T](t)$ would be the proportion of individuals whose life length is $t$.

**3.** As already discussed in Section 7.2, the distribution of a statistical variable can also be described by a distribution function. Applied to the duration variable $T$, values of the distribution function are given by

$$F[T](t) \;=\; \frac{\big|\{\omega \in \Omega \,|\, T(\omega) \leq t\}\big|}{|\,\Omega\,|}$$

where now $t \in \mathbf{R}$ is any real number. One may notice that both functions, $\mathrm{P}[T]$ and $F[T]$, provide the same information because one can be derived from the other. If the frequency function is given, then

$$F[T](t) = \sum_{t' \leq t} \mathrm{P}[T](t')$$

On the other hand, if $t$ is any value in $\tilde{\mathcal{T}}$, then

$$\mathrm{P}[T](t) \;=\; F[T](t) - F[T](t-1)$$

if $t > 0$, and $\mathrm{P}[T](0) = F[T](0)$.

**4.** A further concept often used in a discussion of duration variables is called a *survivor function* and denoted by $G[T]$. We will use the following definition:

$$G[T](t) \;:=\; \frac{\big|\{\omega \in \Omega \,|\, T(\omega) \geq t\}\big|}{|\,\Omega\,|}$$

where $t$ can be any real number.[7] Again, $F[T]$ and $G[T]$ provide the same

[7]In the literature one also finds a slightly different definition:

$$G[T](t) \;:=\; \frac{\big|\{\omega \in \Omega \,|\, T(\omega) > t\}\big|}{|\,\Omega\,|} \;=\; 1 - F[T](t)$$

This definition was used, for example, by Rohwer and Pötter (2001, p. 198). The definition given above is preferred for the present text because it better suits a discrete time axis.

information. $F[T](t)$ is the proportion of individuals whose duration is less than, or equal to, $t$; and $G[T](t)$ is the proportion of individuals whose duration is greater than, or equal to, $t$. For example, if $\tilde{\mathcal{T}}$ refers to life length, $G[T](70)$ would be the proportion of people still alive at age 70.

**5.** Finally, one can characterize the distribution of a duration variable by a *rate function*. A rate function

$$r[T] : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$$

associates to each duration $t \in \tilde{\mathcal{X}}$ a number

$$r[T](t) := \frac{\left|\{\omega \in \Omega \,|\, T(\omega) = t\}\right|}{\left|\{\omega \in \Omega \,|\, T(\omega) \geq t\}\right|}$$

The numerator is the number of individuals in $\Omega$ whose duration is $t$, and the denominator is the number of individuals with a duration not less than $t$. For example, assuming that $T$ refers to life length, if the number of individuals still alive at age 90 is 1000 and, of these people, 100 die at age 90, then the rate for $t = 90$ would be

$$r[T](90) = 100/1000 = 0.1$$

**6.** Another way to interprete rates is in terms of events, in this example, in terms of death events. One can define a *risk set*

$$\mathcal{R}(t) := \{\omega \in \Omega \,|\, T(\omega) \geq t\}$$

containing all individuals who still might experience the event (which, in turn, defines the duration) in $t$; and also an *event set*

$$\mathcal{E}(t) := \{\omega \in \Omega \,|\, T(\omega) = t\}$$

containing the members of $\mathcal{R}(t)$ who actually experienced the event in $t$. The definition of a rate as given above is then equivalent to

$$r[T](t) = \frac{|\,\mathcal{E}(t)\,|}{|\,\mathcal{R}(t)\,|}$$

**7.** We mention that a rate function provides the same information about the distribution of $T$ as the frequency function $\mathrm{P}[T]$, the distribution function $F[T]$, and the survivor function $G[T]$. First, since

$$|\,\mathcal{E}(t)\,| = \mathrm{P}[T](t)\,|\Omega| \quad \text{and} \quad |\,\mathcal{R}(t)\,| = G[T](t)\,|\Omega|$$

one directly finds that

$$r[T](t) = \frac{\mathrm{P}[T](t)}{G[T](t)}$$



On the other hand, assume that the rate function is given. Since always $G[T](0) = 1$, the survivor function may be written in the form

$$G[T](t) = \frac{G[T](t)}{G[T](t-1)}\,\frac{G[T](t-1)}{G[T](t-2)} \cdots \frac{G[T](1)}{G[T](0)}$$

However, since the factors can also be written as

$$\frac{G[T](t)}{G[T](t-1)} = \frac{G[T](t-1) - \mathrm{P}[T](t-1)}{G[T](t-1)} = 1 - r[T](t-1)$$

it follows that

$$\begin{aligned} G[T](t) &= (1 - r[T](t-1))\,(1 - r[T](t-2)) \cdots (1 - r[T](0)) \\ &= \prod_{j=0}^{t-1} (1 - r[T](j)) \qquad (7.3.1) \end{aligned}$$

Therefore, given the rate function, one can derive the survivor function, and consequently also the frequency and distribution functions.

### 7.3.2 Cohort and Period Life Tables

**1.** An often used method to record mortality data is the construction of a *life table* [Sterbetafel]. There are two variants:

a) A *cohort life table* records the mortality of a birth cohort and refers to the historical period during which members of the birth cohort lived.

b) A *period life table* is derived from the age-specific mortality rates of one or more consecutive years and, consequently, reflects the mortality conditions of these years.

**2.** The construction of a cohort life table refers to a birth cohort, say $\mathcal{C}_{t_0}$, whose members are born in the year $t_0$. One can think, then, of a duration variable

$$T_{t_0} : \mathcal{C}_{t_0} \longrightarrow \tilde{\mathcal{T}} = \{0, 1, 2, 3, \ldots\}$$

that records, for each individual $\omega \in \mathcal{C}_{t_0}$, its life length $T_{t_0}(\omega)$. A life table is then simply a table that describes the distribution of $T_{t_0}$, most often in terms of a survivor function or a rate function.

**3.** Actually, most life tables, and in particular life tables published by official statistics, are period life tables. One reason is that period life tables are better suited to keep track of mortality conditions as they are changing from year to year. In contrast, a cohort life table would refer to a relatively long historical period. For example, a life table for persons born in 1900 would be the result of all changes in mortality conditions that occurred during the whole last century. A second reason is that it is more difficult to find suitable data for cohort life tables. In the remainder of the present section we therefore concentrate on period life tables. Some approaches to construct cohort life tables will be discussed in Chapter 8.

**4.** A period life table refers to a population of people who live during a period $t$. For the moment, we will assume that $t$ refers to a specific year and denote the population by $\Omega_t$. Most of the members of $\Omega_t$ will be still alive in the next year, $t+1$, but some will die during the year $t$. This can be represented by a two-dimensional statistical variable

$$(A_t, D_t) : \Omega_t \longrightarrow \tilde{\mathcal{A}} \times \tilde{\mathcal{D}}$$

$\tilde{\mathcal{A}}$ is a property space for age in completed years, so $A_t(\omega)$ is the age of $\omega$ in the year $t$, measured in completed years; and $\tilde{\mathcal{D}} := \{0,1\}$ is the property space for variable $D_t$ which is used to record whether a person dies during the year $t$ or survives to the next year:

$$D_t(\omega) := \begin{cases} 1 & \text{if } \omega \text{ dies during the year } t \\ 0 & \text{otherwise} \end{cases}$$

For example, $(A_t, D_t)(\omega) = (50, 1)$ would mean that $\omega$ died at age 50 during the year $t$; and $(A_t, D_t)(\omega) = (50, 0)$ would mean that $\omega$ is of age 50 in year $t$ but survived to the following year. Given this two-dimensional variable, one can define age-specific death rates. If $n_{t,\tau} = |\{\omega \in \Omega_t \,|\, X_t(\omega) = \tau\}|$ is the number of persons in $\Omega_t$ who are of age $\tau$ in the year $t$, and $d_{t,\tau} = |\{\omega \in \Omega_t \,|\, X_t(\omega) = \tau, D_t(\omega) = 1\}|$ is the number of persons in $\Omega_t$ who died during the year $t$ at the age $\tau$, the age-specific death rates are given by

$$\delta_{t,\tau} = \frac{d_{t,\tau}}{n_{t,\tau}}$$

Obviously, this is identical with the definition of age-specific death rates given in Section 7.1.[8]

**5.** These age-specific mortality rates can now be used to construct a kind of fictitious distribution. To motivate the construction, authors often refer to a fictitious cohort in the following way: Think of a set of $l_{t,0}$ people, all born at the same time, day 0. Then assume that, for each year $\tau$,

[8]These age-specific death rates are often called “death probabilities” [Sterbewahrscheinlichkeiten]. This is misleading because these rates refer to frequencies, not to probabilities. Unfortunately, there is a general tendency in the statistical literature to confuse probabilities and frequencies. For a discussion, and critique, see Rohwer and Pötter (2002b).



beginning in day 0, the proportion of people dying during the year $\tau$ is given by $\delta_{t,\tau}$. This implies:

$$\begin{aligned} l_{t,1} &= l_{t,0}\,(1 - \delta_{t,0}) \\ l_{t,2} &= l_{t,1}\,(1 - \delta_{t,1}) \\ l_{t,3} &= l_{t,2}\,(1 - \delta_{t,2}) \end{aligned}$$

and, in general,

$$l_{t,\tau} = l_{t,\tau-1}\,(1 - \delta_{t,\tau-1}) = l_{t,0} \prod_{j=0}^{\tau-1} (1 - \delta_{t,j})$$

until, eventually, all members of the fictitious cohort are dead.[9] The construction of a period life table basically consists in performing these calculations and presenting the results in a table where the essential columns are: the age $\tau$, the age-specific death rates $\delta_{t,\tau}$, and the number of people still alive at age $\tau$.

**6.** Alternatively, one can think in terms of a fictitious duration variable, $T_t$, that has a distribution defined by the rate function

$$r[T_t](\tau) := \delta_{t,\tau}$$

This rate function implies a survivor function

$$G[T_t](\tau) = \prod_{j=0}^{\tau-1} (1 - r[T_t](j)) = \prod_{j=0}^{\tau-1} (1 - \delta_{t,j})$$

and it follows that $G[T_t](\tau) = l_{t,\tau}/l_{t,0}$. The sequence $l_{t,0}, l_{t,1}, l_{t,2}, \ldots$ can therefore be interpreted as the values of a survivor function for the fictitious duration variable $T_t$.

**7.** To illustrate the calculations, we use data for Germany in 1999 as shown in Table 7.1-1. The result of the calculations, separately for men and women, is shown in Table 7.3-1. The initial size of the fictitious cohorts is $l^m_{t,0} = 100000$ and $l^f_{t,0} = 100000$. Further values of $l^m_{t,\tau}$ and $l^f_{t,\tau}$ can then be calculated recursively as described above. For example,

$$l^m_{t,1} = l^m_{t,0}\,(1 - \delta^m_{t,0}) = 100000 \cdot \left(1 - \frac{4.95}{1000}\right) = 99505$$

From the 100000 men assumed to be alive at the beginning, 99505 survive their first birth day. Figure 7.3-1 shows the corresponding survivor functions for men and women. These functions are only shown up to an age

[9]Obviously, in order to provide sensible results, it is required that all death rates are strictly less than 1 until, at the maximal age (or open-ended age class) the death rate gets the value 1.

**Table 7.3-1** Period life table for Germany in 1999, calculated from the data in Table 7.1-1.

| $\tau$ | $\tilde{\delta}^m_{t,\tau}$ | $l^m_{t,\tau}$ | $\tilde{\delta}^f_{t,\tau}$ | $l^f_{t,\tau}$ | $\tau$ | $\tilde{\delta}^m_{t,\tau}$ | $l^m_{t,\tau}$ | $\tilde{\delta}^f_{t,\tau}$ | $l^f_{t,\tau}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 4.95 | 100000 | 4.01 | 100000 | 48 | 4.29 | 94574 | 2.20 | 97143 |
| 1 | 0.42 | 99505 | 0.35 | 99599 | 49 | 4.57 | 94168 | 2.41 | 96929 |
| 2 | 0.30 | 99463 | 0.21 | 99564 | 50 | 5.18 | 93737 | 2.62 | 96696 |
| 3 | 0.22 | 99433 | 0.16 | 99543 | 51 | 5.45 | 93252 | 2.93 | 96443 |
| 4 | 0.20 | 99410 | 0.14 | 99526 | 52 | 6.38 | 92744 | 3.21 | 96160 |
| 5 | 0.13 | 99391 | 0.11 | 99512 | 53 | 6.12 | 92153 | 3.27 | 95851 |
| 6 | 0.17 | 99378 | 0.11 | 99502 | 54 | 7.32 | 91588 | 3.74 | 95537 |
| 7 | 0.15 | 99362 | 0.13 | 99491 | 55 | 7.99 | 90918 | 4.06 | 95180 |
| 8 | 0.16 | 99347 | 0.10 | 99478 | 56 | 8.47 | 90192 | 4.05 | 94794 |
| 9 | 0.12 | 99330 | 0.12 | 99469 | 57 | 9.51 | 89428 | 4.58 | 94410 |
| 10 | 0.14 | 99319 | 0.09 | 99457 | 58 | 9.64 | 88577 | 4.60 | 93978 |
| 11 | 0.14 | 99305 | 0.09 | 99448 | 59 | 11.21 | 87723 | 5.31 | 93546 |
| 12 | 0.16 | 99291 | 0.12 | 99439 | 60 | 12.31 | 86740 | 5.79 | 93049 |
| 13 | 0.17 | 99275 | 0.11 | 99427 | 61 | 13.35 | 85672 | 6.14 | 92510 |
| 14 | 0.25 | 99259 | 0.15 | 99416 | 62 | 14.96 | 84528 | 6.75 | 91942 |
| 15 | 0.31 | 99234 | 0.19 | 99401 | 63 | 16.75 | 83264 | 7.45 | 91322 |
| 16 | 0.41 | 99203 | 0.24 | 99382 | 64 | 18.70 | 81869 | 8.71 | 90641 |
| 17 | 0.65 | 99162 | 0.33 | 99358 | 65 | 20.03 | 80338 | 9.38 | 89852 |
| 18 | 1.01 | 99098 | 0.35 | 99326 | 66 | 22.12 | 78729 | 10.19 | 89009 |
| 19 | 0.96 | 98997 | 0.37 | 99291 | 67 | 25.00 | 76988 | 11.76 | 88102 |
| 20 | 0.94 | 98902 | 0.31 | 99254 | 68 | 28.13 | 75063 | 13.15 | 87066 |
| 21 | 1.02 | 98809 | 0.35 | 99223 | 69 | 30.69 | 72951 | 14.56 | 85920 |
| 22 | 0.90 | 98709 | 0.28 | 99188 | 70 | 33.60 | 70713 | 16.12 | 84669 |
| 23 | 0.88 | 98621 | 0.27 | 99160 | 71 | 35.83 | 68337 | 17.97 | 83304 |
| 24 | 0.96 | 98534 | 0.31 | 99134 | 72 | 38.53 | 65889 | 20.39 | 81807 |
| 25 | 0.80 | 98439 | 0.34 | 99103 | 73 | 42.59 | 63350 | 22.61 | 80139 |
| 26 | 0.88 | 98360 | 0.30 | 99069 | 74 | 47.52 | 60652 | 24.91 | 78328 |
| 27 | 0.85 | 98273 | 0.32 | 99039 | 75 | 51.42 | 57770 | 28.59 | 76377 |
| 28 | 0.88 | 98190 | 0.35 | 99008 | 76 | 56.18 | 54800 | 32.31 | 74193 |
| 29 | 0.82 | 98103 | 0.35 | 98973 | 77 | 63.63 | 51721 | 37.00 | 71796 |
| 30 | 0.90 | 98022 | 0.37 | 98938 | 78 | 70.03 | 48430 | 41.66 | 69139 |
| 31 | 0.88 | 97935 | 0.39 | 98901 | 79 | 86.32 | 45038 | 51.89 | 66259 |
| 32 | 0.91 | 97849 | 0.45 | 98863 | 80 | 77.50 | 41151 | 48.28 | 62820 |
| 33 | 0.99 | 97760 | 0.49 | 98818 | 81 | 93.94 | 37961 | 61.51 | 59788 |
| 34 | 1.07 | 97663 | 0.52 | 98769 | 82 | 103.94 | 34395 | 71.11 | 56110 |
| 35 | 1.14 | 97559 | 0.61 | 98718 | 83 | 111.37 | 30820 | 75.47 | 52120 |
| 36 | 1.43 | 97447 | 0.62 | 98658 | 84 | 134.70 | 27388 | 94.86 | 48187 |
| 37 | 1.48 | 97308 | 0.78 | 98597 | 85 | 140.81 | 23699 | 100.75 | 43616 |
| 38 | 1.58 | 97164 | 0.83 | 98520 | 86 | 155.71 | 20362 | 113.81 | 39222 |
| 39 | 1.88 | 97010 | 0.96 | 98439 | 87 | 171.57 | 17191 | 129.75 | 34758 |
| 40 | 2.01 | 96828 | 1.06 | 98344 | 88 | 184.77 | 14242 | 146.60 | 30248 |
| 41 | 2.21 | 96633 | 1.18 | 98240 | 89 | 208.86 | 11610 | 164.61 | 25814 |
| 42 | 2.51 | 96419 | 1.31 | 98124 | 90 | 232.80 | 9185 | 185.72 | 21565 |
| 43 | 2.82 | 96178 | 1.47 | 97995 | 91 | 248.55 | 7047 | 207.72 | 17560 |
| 44 | 3.01 | 95906 | 1.55 | 97852 | 92 | 258.18 | 5295 | 229.23 | 13912 |
| 45 | 3.43 | 95618 | 1.73 | 97700 | 93 | 270.39 | 3928 | 250.34 | 10723 |
| 46 | 3.60 | 95290 | 1.95 | 97531 | 94 | 282.51 | 2866 | 269.01 | 8039 |
| 47 | 3.93 | 94947 | 2.03 | 97341 | 95 | | 2056 | | 5876 |



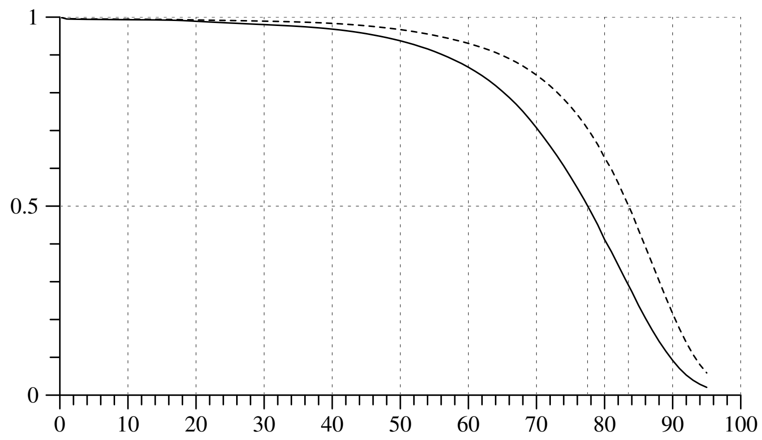



**Fig. 7.3-1** Plot of the survivor functions calculated in Table 7.3-1. For men: $l^m_{t,\tau}/100000$ (solid line), for women: $l^f_{t,\tau}/100000$ (dotted line).

of 94 because our data group all higher ages into a single age class (95*). For the same reason, Table 7.3-1 does not provide death rates for $\tau = 95$. If one would refer to the age class 95*, the death rate would simply be 1 since any person in this age class must eventually die.

**8.** We mention that the survivor functions shown in Figure 7.3-1 are different from the survivor functions that correspond to the distribution functions shown in Figure 7.2-2. While these distribution functions, and the corresponding survivor functions, refer to a definite population, namely all people who died in Germany in 1999, the survivor functions shown in Figure 7.3-1 do not refer to any identifiable population but are conceptual constructions derived from the age-specific death rates in 1999. The difference also becomes visible when calculating median life lengths. Based on Figure 7.3-1, one finds about 77.5 years for men and 83.5 years for women. This is significantly higher than the median life length of those men and women who actually died in 1999, as calculated in Section 7.2, namely 72 years for men and 82 years for women. Of course, these values are lower because they reflect the mortality conditions during the life courses of these people, and not just in 1999.

### 7.3.3 Conditional Life Length

**1.** A period life table can be thought of as the representation of the distribution of a fictitious duration variable $T_t$. The corresponding mean value of $T_t$, $\mathrm{M}(T_t)$, might be interpreted as the mean life length corresponding to the mortality conditions in $t$. In a further step, one can condition the calculation on the assumption that people have already reached a certain age, say $\tau_0$. One might then ask for the mean life length of these people.

**2.** The formal framework is provided by the notion of *conditional mean value*. We first introduce this notion for a general duration variable $T : \Omega \longrightarrow \tilde{\mathcal{T}}$. Given any value $t_0 \in \tilde{\mathcal{T}}$, the risk set

$$\mathcal{R}(t_0) := \{\omega \in \Omega \,|\, T(\omega) \geq t_0\}$$

consists of those people in $\Omega$ whose values of $T$ are not less than $t_0$. The conditional mean value of $T$, given $T \geq t_0$, is then simply the mean value of $T$ in the subpopulation $\mathcal{R}(t_0)$. We use the following notation:

$$\mathrm{M}[T|T \geq t_0] := \frac{\sum_{\omega \in \mathcal{R}(t_0)} T(\omega)}{|\mathcal{R}(t_0)|}$$

Since $T$ can only assume non-negative values, the unconditional mean value is a special case: $\mathrm{M}(T) = \mathrm{M}[T|T \geq 0]$. It is also easy to see that

$$\text{if } t_0 \leq t_1, \text{ then } \mathrm{M}[T|T \geq t_0] \leq \mathrm{M}[T|T \geq t_1]$$

In any case, the calculation of conditional mean values only requires a knowledge of the distribution of $T$ beginning at $t_0$, as shown by the following equation:

$$\mathrm{M}[T|T \geq t_0] = \frac{\sum_{t=t_0}^{\infty} t\,\mathrm{P}[T](t)}{\sum_{t=t_0}^{\infty} \mathrm{P}[T](t)} = \frac{\sum_{t=t_0}^{\infty} t\,\mathrm{P}^*[T](t)}{\sum_{t=t_0}^{\infty} \mathrm{P}^*[T](t)}$$

**3.** The notion of a conditional mean can also be applied to a fictitious duration variable $T_t$ defined by a period life table for the period $t$. Using notations from the previous section, and omitting indices which distinguish male and female quantities, one may write:

$$\mathrm{P}[T_t](\tau) = \frac{l_{t,\tau} - l_{t,\tau+1}}{100000} = \frac{l_{t,\tau}\,\delta_{t,\tau}}{100000}$$

This then allows, for any age $\tau_0$, to calculate a conditional mean value by

$$\mathrm{M}[T_t|T_t \geq \tau_0] = \frac{\sum_{\tau=\tau_0}^{\infty} \tau\, l_{t,\tau}\,\delta_{t,\tau}}{\sum_{\tau=\tau_0}^{\infty} l_{t,\tau}\,\delta_{t,\tau}} \tag{7.3.2}$$

**4.** To illustrate the calculations we use the data from Table 7.3-1. The only difficulty concerns the age class 95*. As was already discussed in Section 7.2, one needs an assumption about the mean age at death in this age class, that is, about the conditional life length for $\tau_0 = 95$. Assuming without further justification

$$\mathrm{M}(T_t^m|T_t^m \geq 90] = 97 \quad \text{and} \quad \mathrm{M}(T_t^f|T_t^f \geq 90] = 99$$



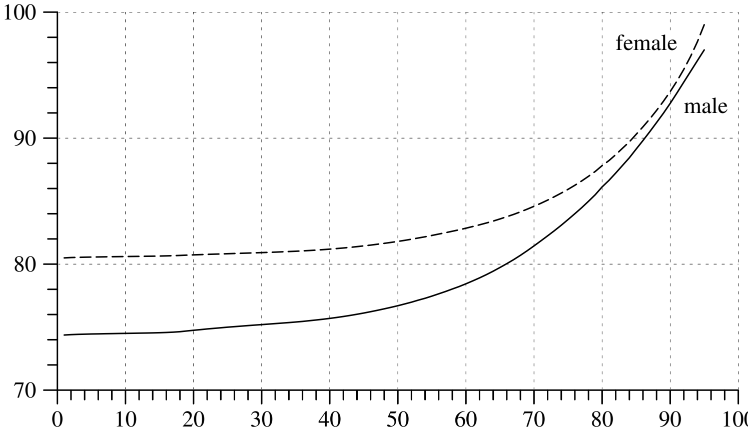



**Fig. 7.3-2** Conditional life length in Germany, 1999, derived from the data in Table 7.3-1.

one can use (7.3.2) to calculate conditional life lengths for all $\tau_0 \geq 0$. The result is shown in Figure 7.3-2 where the abscissa refers to age values $\tau_0$ and the ordinate records the conditional life length. The unconditional mean values, corresponding to $\tau_0 = 0$, are about 74.5 years for men and 80.5 years for women. Obviously, if $\tau_0$ increases, also the conditional life length increases. One can also derive a *mean residual life function* [fernere Lebenserwartung] defined by

$$\mathrm{M}(T_t|T_t \geq \tau_0] - \tau_0$$

For example, given that people have already reached an age of 70, our period life table would estimate a mean residual life length of about 11.5 years for men and 14.5 years for women.

## 7.4 Official Life Tables in Germany

In the present section we discuss life tables by official statistics in Germany. We begin with a brief overview and then consider changes in mortality as reflected by a series of period life tables beginning in the 1871–81 period.

### 7.4.1 Introductory Remarks

**1.** In Section 7.3 we used age-specific death rates of the year 1999 in order to construct a life table for that year. This rather straightforward method is often modified.[10] Differences mainly concern the following points:

[10] Much of the discussion of different methods to construct life tables occurred already during the 19th century. For a fairly complete report see v. Bortkiewicz (1911).

a) How to calculate age-specific death rates? While the numerator simply counts the number of deaths that occurred in a given age and year, there are several possible choices for the definition of the denominator. Instead of using the midyear population, $n_{t,\tau}$, as we have done in Section 7.3, one might want to take into account also the temporal distribution of death events during the year.[11]

b) Another point concerns the calculation of age-specific death rates for very old ages. In the example presented in Section 7.3 this was not done because the data did not provide any information about ages greater than, or equal to, 90. If further data would be available one might be able to estimate age-specific death rates also for these higher ages. Alternatively, one might apply some interpolation procedure.[12]

c) A further point concerns the use of data for a single year. An alternative would be to combine the data of several years. The latter approach is often used as a kind of smoothing procedure. For the same reason, one might apply analytical smoothing procedures to single-year data.

We will not here discuss the many different methods that have been proposed for the construction of life tables. Instead, we describe the methods used by the *Statistisches Bundesamt*.

**2.** Official statistics in Germany distinguishes between general life tables [Allgemeine Sterbetafeln] and abridged life tables [abgekürzte Sterbetafeln]. General life tables refer to periods which are centered around the year of a census. The most recent general life table is based on the census in 1987 and refers to the three-year period 1986 – 88. The methods to calculate general life tables changed several times.[13] For the last two tables (1970 – 72 and 1986 – 88), calculations begin with age-specific death rates which, in our standard notation, are defined as

$$\delta_{t,\tau} = \frac{d_{t,\tau}}{n_{t,\tau}}$$

[11] A discussion of alternative methods can be found, e.g., in Namboodiri and Suchindran (1987, pp. 12-19), or Flaskämper (1962, pp. 351-366).

[12] Namboodiri and Suchindran (1987, p. 19-20) write: "Population and death data at the very old ages, when they are available, are generally disregarded in computing a life table, mainly because they are considered inaccurate. It has therefore been a common practice to use arbitrary methods for computing $q_x$ [the death rate for age $x$] at the very old ages (usually 85 and above). For practical purposes, any reasonable method is satisfactory, since the arbitrariness involved in the method has only a small effect on the life table as a whole. The major requirement that is usually kept in mind when choosing a procedure in this connection is that the procedure should produce a smooth junction with the $q_x$ values already computed and a smooth upward progression of $q_x$ with advancing age." The authors then briefly describe four different methods.

[13] Detailed explanations are available in Fachserie 1, Reihe 1 S.2, Allgemeine Sterbetafel für die Bundesrepublik Deutschland 1986/88.



where $\tau$ refers to age in completed years and $t$ refers to a calendar year. These rates are then modified to

$$q_{t,\tau} := \frac{d_{t,\tau}}{n_{t,\tau} + \frac{d_{t,\tau}}{2}}$$

The reason behind this modification is that about half of the people who die at age $\tau$ during the year $t$ are not counted in $n_{t,\tau}$. Therefore, in order to get an estimate of the number of people who are actually at risk of dying in year $t$, $d_{t,\tau}/2$ is added to $n_{t,\tau}$.[14] These modified age-specific death rates are then calculated for a three-year period as follows:

$$q_{(t),\tau} := \frac{d_{t-1,\tau} + d_{t,\tau} + d_{t+1,\tau}}{n_{t-1,\tau} + n_{t,\tau} + n_{t+1,\tau} + \frac{d_{t-1,\tau}+d_{t,\tau}+d_{t+1,\tau}}{2}}$$

where $t = 1987$ for the life table which refers to the period 1986 – 88.

**3.** Abridged life tables, like general life tables, are based on three-year intervals and use the same method of calculating modified age-specific death rates.[15] The main differences are as follows:

a) For the construction of general life tables, additional calculations are performed to provide more detailed information about death rates during the first year after birth. The abridged life tables simply use $q_{(t),0}$ without further subdivisions.

b) While the calculation of abridged life tables is directly derived from the death rates $q_{(t),\tau}$, these rates are smoothed before they are used in the calculation of general life tables.[16]

[14] In the literature, these modified rates, $q_{t,\tau}$, are often called "age-specific death probabilities". However, for reasons already mentioned, we avoid the term 'probability' and simply speak of modified age-specific death rates.

[15] Abridged life tables have been calculated regularly for each year since 1957; results are published in Fachserie 1, Reihe 1.

[16] Fachserie 1, Reihe 1, S.2 (p. 13) provides the following reasons: „Um einen möglichst wirklichkeitsgetreuen Verlauf der Sterbewahrscheinlichkeiten in Abhängigkeit vom Alter $x$ zu erreichen, ist es notwendig, die rohen Sterbewahrscheinlichkeiten $\bar{q}_x$ auszugleichen, das heißt, von zufallsbedingten Schwankungen und solchen systematischen Sprüngen zu bereinigen, die an bestimmte Geburtsjahrgänge gebunden sind. An das Ausgleichungsverfahren sind damit die folgenden Anforderungen zu stellen:
– Der Verlauf der ausgeglichenen Sterbewahrscheinlichkeiten $q_x$ in Abhängigkeit vom Alter $x$ soll möglichst „glatt" sein, das heißt hier, möglichst kleine Krümmungen haben und keine Sprungstellen und keine Knicke aufweisen.
– Zufallsbedingte Schwankungen sollen ausgeglichen werden.
– Typische altersspezifische Besonderheiten im Sterblichkeitsverlauf sollen bewahrt bleiben, zum Beispiel das relative Maximum bei den 20jährigen.
– Besonderheiten im Sterblichkeitsverlauf, die an bestimmte Geburtsjahrgänge gebunden sind (Kohorteneffekte), zum Beispiel die relative hohe Sterbewahrscheinlichkeit bei den „Kriegsjahrgängen" des Ersten Weltkriegs, müssen eliminiert werden."

One should notice that the term ‘abridged life table’ is used differently in the demographic literature. In contrast to ‘abgekürzte Sterbetafel’, an abridge life table most often refers to a calculation based on 5-year or 10-year age intervals.[17]

### 7.4.2 General Life Tables 1871 – 1988

**1.** In Germany, general life tables have been constructed by official statistics for the following periods:

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

All tables are period life tables. Until 1932 – 34, they refer to the territory of the former *Deutsches Reich*; all other tables refer to the territory of the former FRG. As has been mentioned in Section 7.4.1, methods of table construction have slightly changed throughout the years.

**2.** Separately for men and women, the survivor functions of all life tables are reproduced in Tables 7.4-1- 4. Following general conventions in the presentation of life tables, beginning with initial values $l_0^m = 100000$ men and

[17]See, e.g., Namboodiri and Suchindran (1987, pp. 21-26).



**Table 7.4-1** Male survivor functions in German life tables from official statistics. Source: see text.

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

**Table 7.4-2** Male survivor functions in German life tables from official statistics. Source: see text.

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



**Table 7.4-3** Female survivor functions in German life tables from official statistics. Source: see text.

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

**Table 7.4-4** Female survivor functions in German life tables from official statistics. Source: see text.

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



$l_0^f = 100000$ women, subsequent values show how many of these men and women have survived until an age $\tau$ (completed age in years). Therefore,

$$l_\tau^m/100000 \quad \text{and} \quad l_\tau^f/100000$$

directly provide values of the life table survivor functions. All further quantities commonly presented in publications of life tables can be derived:

a) Omitting the superscript indicating sex, the number of individuals (per 100000) who died at age $\tau$ is

$$d_\tau := l_\tau - l_{\tau+1}$$

For example, referring to the life table for 1910–11, $d_{10}^m = 166$ and $d_{10}^f = 174$. Of course, calculating these quantities for the last age class, $\tau = 100$, requires additional assumptions.

b) Conditional death frequencies[18] can be calculated by

$$q_\tau := \frac{d_\tau}{l_\tau} = \frac{l_\tau - l_{\tau+1}}{l_\tau}$$

For example, referring again to the life table for 1910–11, one can calculate $q_{10}^m = 0.00218$ and $q_{10}^f = 0.00221$.

c) Calculation of conditional mean life lengths requires an assumption about mortality in the last age class, $\tau = 100$. Assuming that 100 is the oldest possible age, calculation can be done with the formula

$$e_\tau := \frac{\sum_{j=\tau}^{100} (j + 0.5)\, d_j}{\sum_{j=\tau}^{100} d_j}$$

This is the mean life duration of individuals who reached age $\tau$. For example, referring again to the life table for 1910–11, one finds $e_{10}^m = 62.08$ and $e_{10}^f = 63.99$. We mention that, in the presentation of life tables, one also finds figures for $e_\tau - \tau$, often called *mean residual life length* [fernere Lebenserwartung].

### 7.4.3 Increases in Mean Life Length

**1.** The data in Tables 7.4-1-4 can be used to investigate changes in (fictitious) mean life length.[19] We begin with directly plotting the survivor functions as given in the tables. This is shown in Figure 7.4-1. For women

[18] Often called "death probabilities" [Sterbewahrscheinlichkeiten]. However, since the quantities refer to frequencies, we avoid to speak of "probabilities".

[19] See also Proebsting (1984) who has discussed all these data sets, except the one for 1986 – 88.

and also for younger men, the functions follow the chronological order from bottom to top. For example, in the period 1871–81, about 41 % of the men and about 38 % of the women died before age 20, while in the period 1986–88 these proportions have declined to about 1–2 %. A substantial decline in mortality occurred, in particular, for newborn children. This can also be calculated directly from Tables 7.4-1 and 7.4-3. The following table shows the proportion (in %) of male and female babies who died during their first year of life:

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

**2.** Instead of directly comparing survivor functions, one can compare age-dependent mean life lengths, $e_\tau$, as defined in the previous section. They are shown in Figure 7.4-2. Again, the graphs follow the chronological order from bottom to top. It is seen that the greatest increases in mean life length occurred in young ages. To keep the plot easy to survey, the graphs begin at age 1. However, changes in the mean life length of newborn children can be calculated directly from the data. Comparing the periods 1871–81 and 1986–88, these mean life lengths have increased from about 36 to 72 years for male, and from about 38 to 79 years for female children.

**3.** These changes can also be visualized in historical time by locating the values roughly at the center years of the life table periods. This is done in Figure 7.4-3. Shown are the historical changes in values of $e_\tau$ for a selected number of ages ($\tau = 0, 10, 20, 30, 40, 50, 60, 70$). One sees, again, that the most significant increases in the mean life length occurred in younger ages, in particular for newborn children.

### 7.4.4 Life Table Age Distributions

**1.** In general, changes in the age distribution of a population not only depend on death rates but also on the development of births and migration. So it is difficult to isolate the contribution of changes in mortality. Nevertheless, some ideas can be gained from a hypothetical consideration. Assume that for a longer period, say 100 years, each year the same number



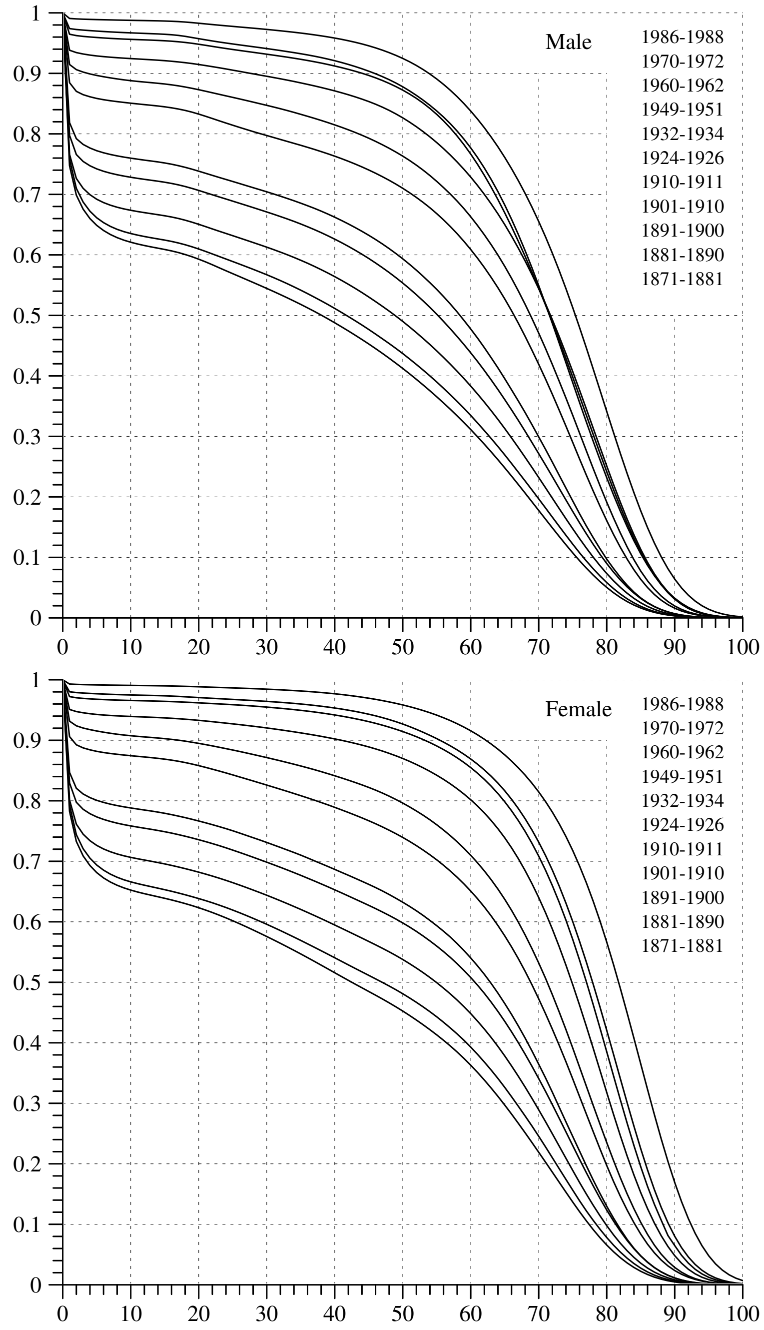



**Fig. 7.4-1** Male and female survivor functions in Germany, 1871 – 1988. At an age of 10 years the functions are in chronological order.

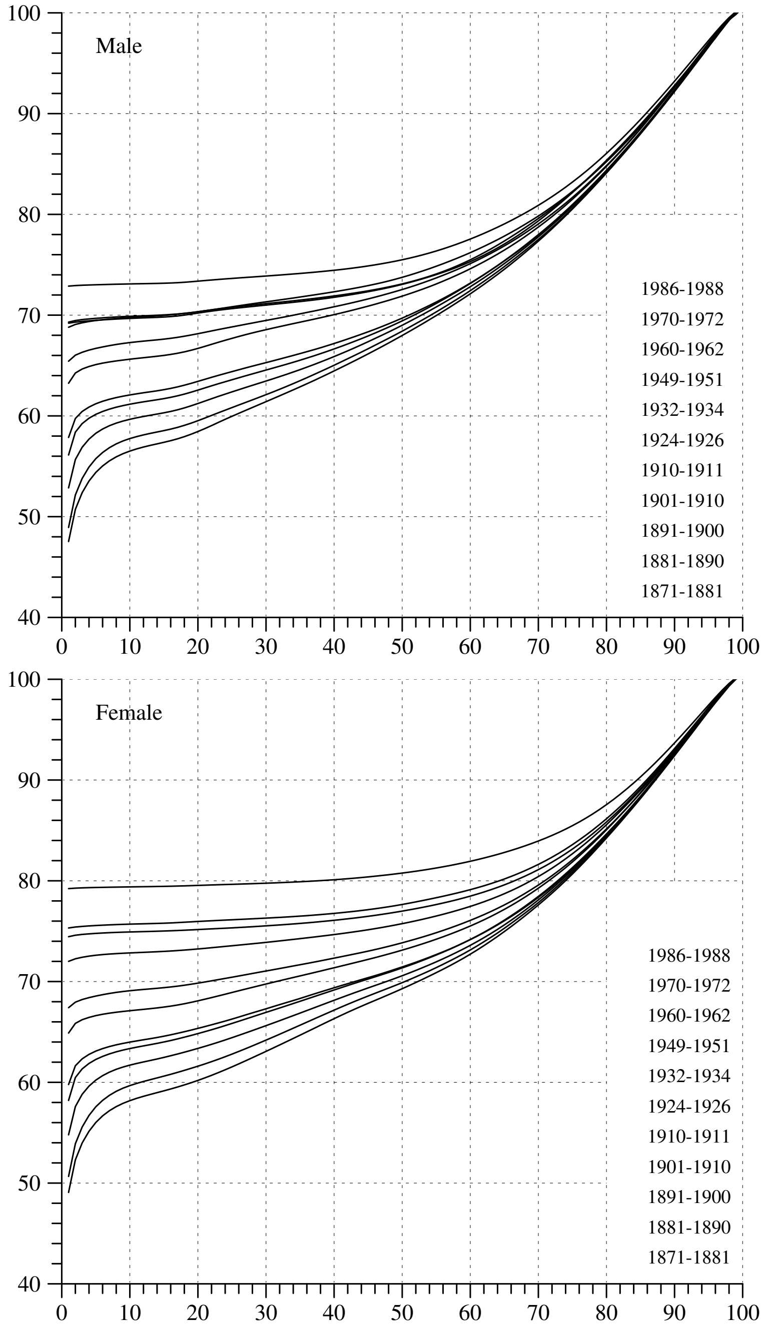



**Fig. 7.4-2** Male and female mean life durations $e_\tau$ in Germany, 1871 – 1988, conditional on age $\tau$ as specified on the abscissa.



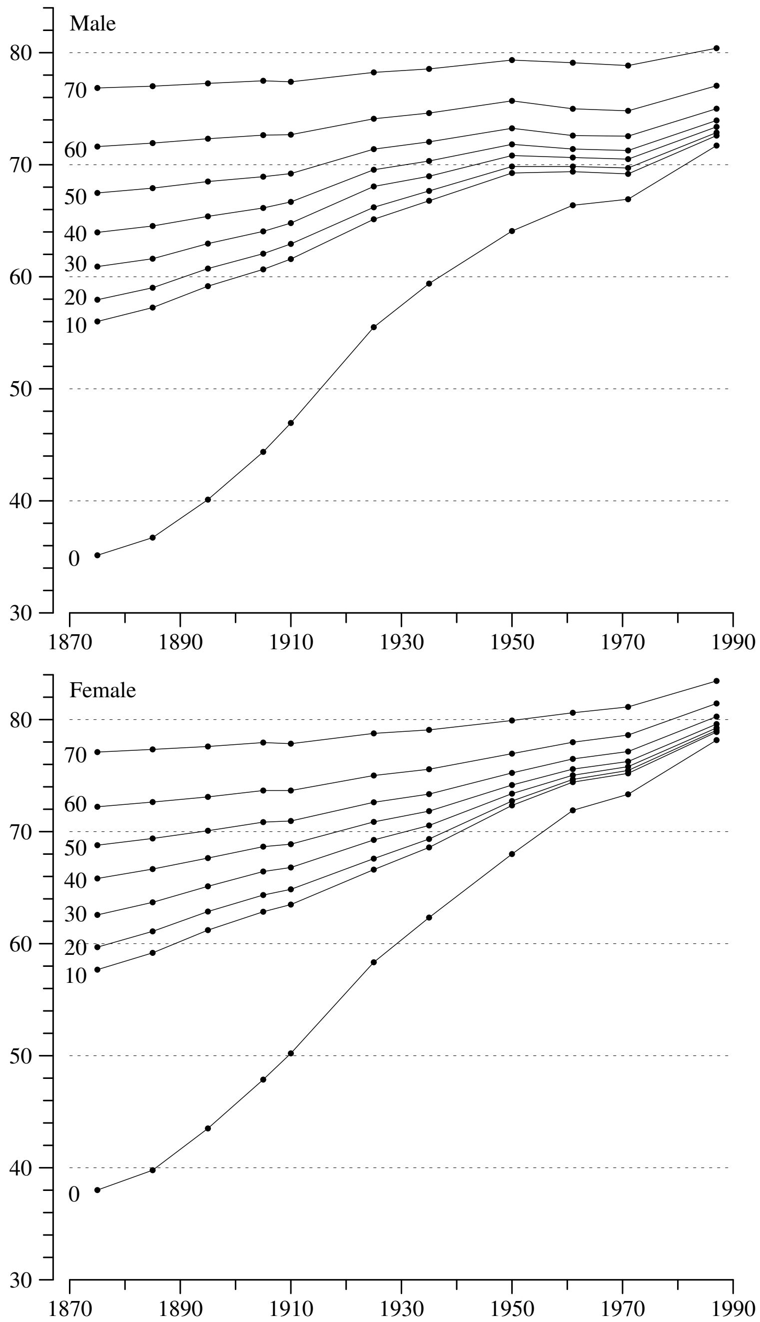



**Fig. 7.4-3** Changes in male and female mean life durations $e_\tau$ in Germany, 1871 – 1988, conditional on ages $\tau = 0, 10, 20, 30, 40, 50, 60, 70$.

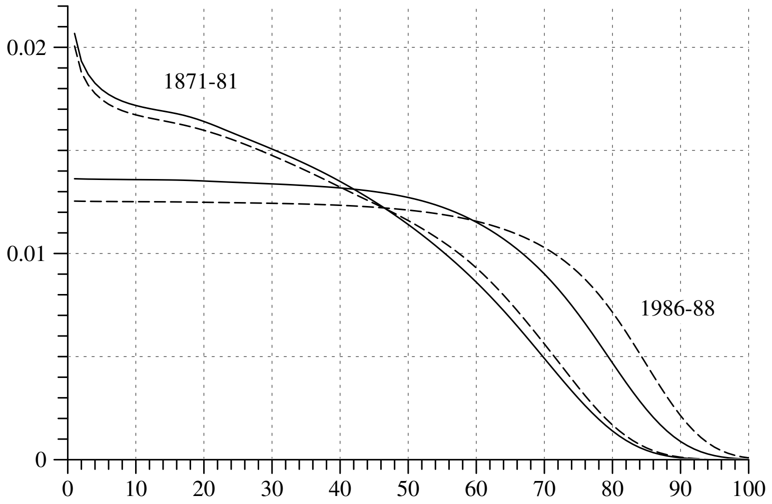



**Fig. 7.4-4** Hypothetical male (solid line) and female (dotted line) age frequencies calculated from period life tables 1871 – 81 and 1986 – 88.

of children is born and that they survive according to a given period life table. This then implies a stable age distribution completely determined by the given life table. In fact, this age distribution is simply proportional to the life table's survivor function. Of course, since death rates are different for men and women, also the corresponding age distributions are different. In our hypothetical population generated by a constant number of 100000 births per year and constant mortality conditions given by some period life table, the number of women of age $\tau$ in a given year is just $l_\tau^f$. Thus, the total number of women alive in that year is $\sum_\tau l_\tau^f$. The relative frequency of women of age $\tau$ is therefore

$$l_\tau^m / \sum_{j=1}^{100} l_j^m$$

and analoguously for men. Figure 7.4-4 directly compares the sex-specific age frequency curves implied by the 1871 – 81 and 1986 – 88 period life tables.

**2.** Another possibility is to aggregate ages into age classes and then to calculate frequencies for each age class. This allows to visualize how the hypothetical age distributions that can be associated with each period life table have changed in historical time. Using 10-year age classes, this is shown, separately for men and women, in Figure 7.4-5.



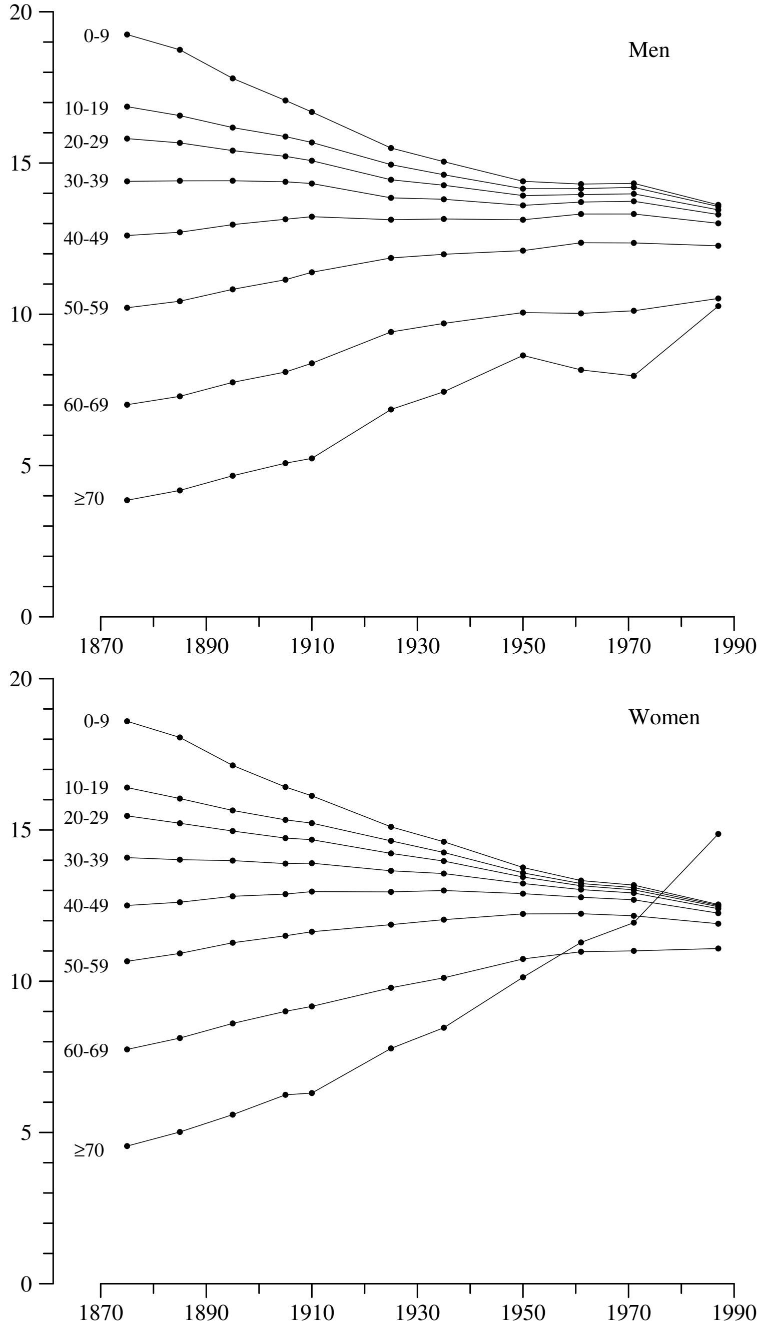



**Fig. 7.4-5** Development of hypothetical age distributions of men and women, calculated from the period life tables from 1871 – 81 to 1986 – 88. The ordinate is in percent for specified age classes.

# Chapter 8

# Mortality of Cohorts

While period life tables reflect the mortality conditions of a given period, cohort life tables try to reconstruct mortality conditions as they developed during the life time of birth cohorts. The latter are much more difficult to produce, mainly due to insufficient data. There are basically two approaches: one can either try to reconstruct cohort life tables from period data, or one can try to actually follow the members of a birth cohort through their life.[1] In the present chapter, we begin with a discussion of the first approach, an attempted reconstruction from period data from official statistics.[2] The second approach is more difficult. Assuming that one wants to construct a cohort life table for a birth cohort $\mathcal{C}_{t_0}$, one would need to actually follow all of its members beginning in the year $t_0$. Such historical data are very scarce. One example will be discussed in Section 8.3. An alternative is to follow the members of $\mathcal{C}_{t_0}$ only through part of their life histories, say, beginning in some year $t > t_0$. This is possible with surveys organized as panels, that is, surveys repeated year by year and targeted at the same people. Such data will allow to construct incomplete life tables which condition on survivorship until an age of $\tau = t - t_0$. Based on data from the German Socio-economic Panel, this approach will be discussed in Section 8.4.

## 8.1 Cohort Death Rates

**1.** We begin with a few definitions. As already introduced in Section 3.4, the symbol $\mathcal{C}_{t_0}$ will be used to denote a birth cohort, that is, a set of people all born during the year $t_0$. Correspondingly, $\mathcal{C}^m_{t_0}$ and $\mathcal{C}^f_{t_0}$ denote, respectively, the male and female members of $\mathcal{C}_{t_0}$. Furthermore, $\mathcal{C}_{t_0,\tau}$ is the set of members of $\mathcal{C}_{t_0}$ being of age $\tau$, that is, who survived at least until age $\tau$. Again, we use $\mathcal{C}^m_{t_0,\tau}$ and $\mathcal{C}^f_{t_0,\tau}$ to distinguish male and female members.

**2.** A cohort view of mortality can be depicted in the following way:

$$\mathcal{C}_{t_0} = \mathcal{C}_{t_0,0} \supseteq \mathcal{C}_{t_0,1} \supseteq \mathcal{C}_{t_0,2} \supseteq \cdots$$

In general, $\mathcal{C}_{t_0,\tau} \setminus \mathcal{C}_{t_0,\tau+1}$ is the set of members of $\mathcal{C}_{t_0}$ who died at age $\tau$

[1]A further possibility is based on data from surveys in which respondents are asked to provide information about dates of birth and, possibly, also dates of death of their parents. Such data will be discussed in Chapter 9.

[2]Historically, this was the main approach to the construction of cohort life tables, see the historical survey by Young (1978). For an early example see Merrell (1947).



(equivalently, in the year $t_0 + \tau$). So we can introduce *age-specific cohort death rates* referring to the proportion of members of $\mathcal{C}_{t_0,\tau}$ who died at age $\tau$. We will use the notation

$$\eta_{t_0,\tau} := \frac{|\,\mathcal{C}_{t_0,\tau} \setminus \mathcal{C}_{t_0,\tau+1}\,|}{|\,\mathcal{C}_{t_0,\tau}\,|}$$

The numerator refers to the number of members of $\mathcal{C}_{t_0}$ who died at age $\tau$, and the denominator refers to the number of members of $\mathcal{C}_{t_0}$ who survived age $\tau - 1$ and might die at age $\tau$. In order to distinguish male and female mortality we also use the notations

$$\eta^m_{t_0,\tau} := \frac{|\,\mathcal{C}^m_{t_0,\tau} \setminus \mathcal{C}^m_{t_0,\tau+1}\,|}{|\,\mathcal{C}^m_{t_0,\tau}\,|} \quad \text{and} \quad \eta^f_{t_0,\tau} := \frac{|\,\mathcal{C}^f_{t_0,\tau} \setminus \mathcal{C}^f_{t_0,\tau+1}\,|}{|\,\mathcal{C}^f_{t_0,\tau}\,|}$$

**3.** One also can think in terms of a duration variable

$$T_{t_0} : \mathcal{C}_{t_0} \longrightarrow \tilde{\mathcal{T}} := \{0, 1, 2, 3, \ldots\}$$

that records the life length of the members of $\mathcal{C}_{t_0}$. As was introduced in Section 7.3, a cohort life table is simply a description of the distribution of $T_{t_0}$. This can be done in terms of a frequency function $\mathrm{P}[T_{t_0}]$, a distribution function $F[T_{t_0}]$, a survivor function $G[T_{t_0}]$, or a rate function $r[T_{t_0}]$. All descriptions are equivalent in that each one can be derived from any other one. Particularly useful is the rate function that records the age-specific cohort death rates:

$$\tau \longrightarrow r[T_{t_0}](\tau) = \eta_{t_0,\tau}$$

## 8.2 Reconstruction from Period Data

**1.** The idea is to reconstruct cohort life tables from age-specific death rates of consecutive years. As an example we consider the birth cohort $t_0 = 1910$. When referring to the territory of the former *Deutsches Reich* in 1910, this birth cohort has 1924778 members.[3] Of course, nobody knows the true age-specific cohort death rates

$$\eta_{1910,\tau} = \frac{\text{members of } \mathcal{C}_{1910} \text{ who died at age } \tau}{\text{members of } \mathcal{C}_{1910} \text{ who survived age } \tau - 1}$$

Data from official statistics can be used, however, to calculate age-specific period death rates:

$$\delta_{1910,0},\ \delta_{1911,1},\ \delta_{1912,2},\ \delta_{1913,3},\ \ldots$$

[3]Statistisches Jahrbuch für das Deutsche Reich 1919 (p. 41).

If we ignore in- and out-migration, and if we also ignore changing political borders, we might assume that these death rates provide sensible estimates for the cohort death rates $\eta_{1910,\tau}$. The reconstruction of a cohort life table then simply consists of using the death rates $\delta_{1910+\tau,\tau}$ instead of the rates $\eta_{1910,\tau}$.[4]

**2.** Unfortunately, the data available in the STATIS data base of the *Statistisches Bundesamt* only allow to calculate the death rates beginning at age 42, corresponding to the year 1952. We therefore calculate a *conditional survivor function*. Let $T$ denote the statistical variable that would record the life length as implied by a complete knowledge of the death rates $\delta_{1910+\tau,\tau}$. We can then define, for each age $\tau_0$, a conditional survivor function

$$G[T|T \geq \tau_0](\tau) := \prod_{j=\tau_0}^{\tau-1} (1 - \delta_{1910+j,j})$$

defined for all $\tau > \tau_0$. As a convention, we also define $G[T|T \geq \tau_0](\tau_0) = 1$. If $\tau_0 = 0$, one gets the unconditional survivor function $G[T|T \geq 0](\tau) = G[T](\tau)$. In general, the relationship is

$$G[T](\tau) = G[T](\tau_0)\, G[T|T \geq \tau_0](\tau)$$

As a special case, if $\tau_0 = 42$, we get the formulation

$$G[T](\tau) = G[T](42)\, G[T|T \geq 42](\tau) \qquad (\text{for } \tau \geq 42) \tag{8.2.1}$$

Given our data, we can only calculate the second term on the right-hand side. But, since the first term on the right-hand side is a constant, the second term provides a function proportional to the survivor function for all ages $\tau \geq 42$.

**3.** Table 8.2-1 provides the respective data which refer to the territory of Germany since 1990: values for the midyear population size, $n^m_{t,\tau}$ and $n^f_{t,\tau}$, are taken from Segment 685 of the STATIS data base; values of the number of deaths, $d^m_{t,\tau}$ and $d^f_{t,\tau}$, are taken from Fachserie 1, Reihe 1.S.3 (Gestorbene nach Alters- und Geburtsjahren sowie Familienstand, 1948 – 1989) and from Segments 1124-26 of the same data base.[5] Death rates (per 1000) are calculated as

$$\tilde{\delta}^m_{t,\tau} = \frac{1000\, d^m_{t,\tau}}{n^m_{t,\tau}} \quad \text{and} \quad \tilde{\delta}^f_{t,\tau} = \frac{1000\, d^f_{t,\tau}}{n^f_{t,\tau}}$$

[4] There are several proposals that do not start with age-specific death rates but try to concatenate information from period life tables; see Höhn (1984), Dinkel (1984). Dinkel (1992) has also suggested to combine both methods. For further discussion see also the contributions in Dinkel, Höhn and Scholz (1996).

[5] For the year 1990, we have used additional data from Fachserie 1, Reihe 1, 1990 (p. 136).



**Table 8.2-1** Age-specific midyear population (in 1000), number of deaths, and age-specific death rates in Germany, 1952–1999. Source: STATIS data base and Fachserie 1, Reihe 1 (see text).

| $t$ | $\tau$ | $n^m_{t,\tau}$ | $d^m_{t,\tau}$ | $\tilde{\delta}^m_{t,\tau}$ | $n^f_{t,\tau}$ | $d^f_{t,\tau}$ | $\tilde{\delta}^f_{t,\tau}$ |
|---|---|---|---|---|---|---|---|
| 1952 | 42 | 477.7 | 1775 | 3.72 | 633.5 | 1776 | 2.80 |
| 1953 | 43 | 476.7 | 1797 | 3.77 | 632.2 | 1806 | 2.86 |
| 1954 | 44 | 486.9 | 1896 | 3.89 | 636.3 | 1830 | 2.88 |
| 1955 | 45 | 473.9 | 2189 | 4.62 | 629.3 | 2024 | 3.22 |
| 1956 | 46 | 465.2 | 2390 | 5.14 | 622.7 | 2117 | 3.40 |
| 1957 | 47 | 467.6 | 2650 | 5.67 | 623.9 | 2275 | 3.65 |
| 1958 | 48 | 465.8 | 2671 | 5.73 | 622.6 | 2407 | 3.87 |
| 1959 | 49 | 463.8 | 3036 | 6.55 | 621.1 | 2552 | 4.11 |
| 1960 | 50 | 461.0 | 3577 | 7.76 | 618.6 | 2771 | 4.48 |
| 1961 | 51 | 456.2 | 3640 | 7.98 | 613.9 | 2955 | 4.81 |
| 1962 | 52 | 453.2 | 3974 | 8.77 | 612.6 | 3239 | 5.29 |
| 1963 | 53 | 449.9 | 4518 | 10.04 | 609.8 | 3517 | 5.77 |
| 1964 | 54 | 444.8 | 4918 | 11.06 | 604.8 | 3575 | 5.91 |
| 1965 | 55 | 440.4 | 5522 | 12.54 | 601.4 | 3948 | 6.57 |
| 1966 | 56 | 435.2 | 5907 | 13.57 | 597.9 | 4281 | 7.16 |
| 1967 | 57 | 428.2 | 6302 | 14.72 | 593.6 | 4507 | 7.59 |
| 1968 | 58 | 420.9 | 7154 | 17.00 | 589.0 | 5175 | 8.79 |
| 1969 | 59 | 414.0 | 8013 | 19.36 | 584.1 | 5549 | 9.50 |
| 1970 | 60 | 404.9 | 8497 | 20.99 | 574.9 | 5889 | 10.24 |
| 1971 | 61 | 396.5 | 8999 | 22.69 | 570.2 | 6428 | 11.27 |
| 1972 | 62 | 387.3 | 9622 | 24.85 | 564.1 | 6983 | 12.38 |
| 1973 | 63 | 377.7 | 10048 | 26.60 | 557.3 | 7391 | 13.26 |
| 1974 | 64 | 367.2 | 10792 | 29.39 | 549.9 | 8031 | 14.60 |
| 1975 | 65 | 355.9 | 11663 | 32.77 | 541.8 | 8674 | 16.01 |
| 1976 | 66 | 343.8 | 12042 | 35.03 | 532.7 | 9453 | 17.75 |
| 1977 | 67 | 331.7 | 12254 | 36.95 | 523.4 | 9758 | 18.64 |
| 1978 | 68 | 319.0 | 13266 | 41.59 | 513.7 | 10458 | 20.36 |
| 1979 | 69 | 305.6 | 13727 | 44.92 | 502.8 | 11507 | 22.88 |
| 1980 | 70 | 291.9 | 14433 | 49.44 | 491.3 | 12510 | 25.46 |
| 1981 | 71 | 277.3 | 14920 | 53.81 | 478.6 | 13443 | 28.09 |
| 1982 | 72 | 262.2 | 15171 | 57.87 | 464.9 | 14252 | 30.66 |
| 1983 | 73 | 247.0 | 15460 | 62.60 | 450.1 | 15460 | 34.35 |
| 1984 | 74 | 231.4 | 15237 | 65.84 | 434.3 | 16296 | 37.52 |
| 1985 | 75 | 215.6 | 16083 | 74.58 | 417.5 | 17400 | 41.68 |
| 1986 | 76 | 199.7 | 15833 | 79.29 | 399.6 | 18398 | 46.04 |
| 1987 | 77 | 185.4 | 15433 | 83.23 | 378.3 | 18922 | 50.01 |
| 1988 | 78 | 170.0 | 15488 | 91.09 | 360.2 | 19881 | 55.19 |
| 1989 | 79 | 154.6 | 14997 | 97.00 | 339.9 | 20881 | 61.44 |
| 1990 | 80 | 139.7 | 15290 | 109.47 | 318.9 | 22230 | 69.71 |
| 1991 | 81 | 125.1 | 14452 | 115.49 | 297.2 | 22433 | 75.49 |
| 1992 | 82 | 111.0 | 13570 | 122.21 | 274.9 | 22154 | 80.59 |
| 1993 | 83 | 97.9 | 13152 | 134.34 | 252.6 | 22977 | 90.96 |
| 1994 | 84 | 85.2 | 12123 | 142.35 | 229.8 | 22895 | 99.63 |
| 1995 | 85 | 73.5 | 11248 | 153.12 | 206.8 | 23085 | 111.63 |
| 1996 | 86 | 62.5 | 10489 | 167.71 | 184.1 | 22810 | 123.92 |
| 1997 | 87 | 52.7 | 9183 | 174.40 | 162.0 | 21713 | 134.03 |
| 1998 | 88 | 43.7 | 8371 | 191.35 | 140.6 | 20896 | 148.59 |
| 1999 | 89 | 35.8 | 7471 | 208.69 | 120.0 | 19758 | 164.65 |
| 2000 | 90 | 28.4 | 6406 | 225.56 | 100.7 | 18054 | 179.29 |

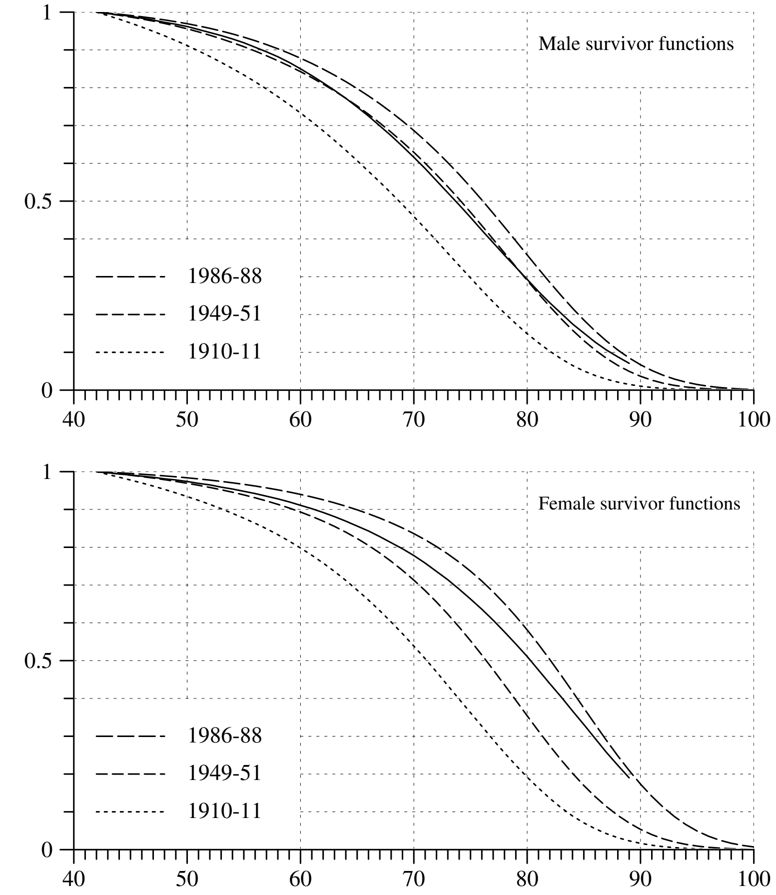



**Fig. 8.2-1** Conditional survivor functions ($\tau \geq 42$) for the members of the birth cohort 1910 (solid line), compared with conditional survivor functions from period life tables (dotted lines).

Death rates for the reconstruction of a cohort life table can be derived by defining $\delta^m_\tau := \tilde{\delta}^m_{t,\tau}/1000$ and $\delta^f_\tau := \tilde{\delta}^f_{t,\tau}/1000$. So one can finally calculate conditional survivor functions

$$G[T^m | T^m \geq 42](\tau) = \prod_{j=42}^{\tau-1} (1 - \delta^m_\tau)$$

$$G[T^f | T^f \geq 42](\tau) = \prod_{j=42}^{\tau-1} (1 - \delta^f_\tau)$$

**4.** Results of this calculation are shown as solid lines in Figure 8.2-1. In



order to provide a context for an interpretation we have added conditional survivor functions from period life tables.[6] It is seen how period life tables systematically underestimate reductions in death rates that occurred in historical time. An remarkable exception is the 1949-51 life table for men. The conditional survivor function from this table seems to be mainly identical with the conditional survivor function of the 1910 quasi-cohort.[7]

[6] The data are taken from Tables 7.4-1-4 in Section 7.4.2.

[7] It is known, however, that the 1949-51 period life table, in particular for men, is based on underestimated death rates, see Dinkel and Meinl (1991, p. 117).

## 8.3 Historical Data

As an alternative to the reconstruction of cohort life tables from period data, one can try to actually follow the life courses of people born in the past. As an example, we discuss a data set that was made available by Arthur E. Imhof and his co-workers (see Imhof et al. 1990).

### 8.3.1 Data Description

**1.** The data set is available from the *Zentralarchiv für empirische Sozialforschung* (Köln). A basic description can be found in Imhof et al. (1990). The data result from so-called "Ortssippenbücher" (see Imhof et al. 1990, pp. 57-66, also Knodel 1975) and refer to several different local areas. Here we only use the data file from Ostfriesland (a region between Aurich and Leer). This data file file contains information about 24971 persons belonging to 3882 families. All families consist of a married women, in 3756 cases also information about the husband is available; the remaining 17333 persons are children. The following tables shows the distribution of family sizes.

| Number of children | Number of families | Number of children | Number of families |
|---|---|---|---|
| 0 | 321 | 8 | 280 |
| 1 | 472 | 9 | 173 |
| 2 | 374 | 10 | 114 |
| 3 | 422 | 11 | 61 |
| 4 | 429 | 12 | 17 |
| 5 | 434 | 13 | 6 |
| 6 | 411 | 14 | 3 |
| 7 | 364 | 15 | 1 |

For 584 persons no valid information about the birth year is available, we therefore consider only the remaining 24387 persons, 7145 parents and 17242 children. The birth years of parents range from 1616 until 1835,[8] birth years of children range from 1636 to 1871. Figure 8.3-1 shows the frequency distributions on a historical time axis.

**2.** In many cases additional information about death years is available. In order to use this information for the estimation of survivor functions it is important to distinguish between parents and children. For parents we need to take into account that they have already survived until the

[8] This is due to the selection of families: „Es wurden nur Daten von Kindern aus solchen Ehen erhoben, bei denen das Todesdatum beider Elternteile (bei unehelichen Kindern das der Mutter) bekannt war und der Todesfall im Untersuchungsgebiet eintrat. Zudem mußte die Ehe vor 1850 geschlossen oder die erste uneheliche Geburt vor 1860 erfolgt sein.“ (Imhof et al. 1990, pp. 62-63)



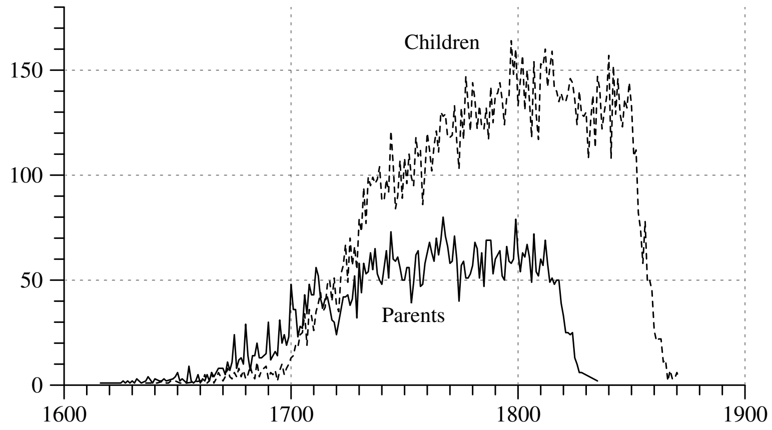



**Fig. 8.3-1** Absolute frequencies of birth years of 7145 parents and 17242 children in the Ostfriesland data set.

age of marriage and/or giving birth to children. We therefore proceed in two steps: We begin with calculating survivor functions for parents, this is easy because there is complete information about their death dates; we then try to estimate survivor functions for the children. Finally we compare parent's and children's life length.

### 8.3.2 Parent's Survivor Functions

**1.** Since the birth years range over a very long period (Fig. 8.3-1), we distinguish four broad birth cohorts (all parents are born before 1850):

| Birth years | Mothers | Fathers |
|---|---|---|
| 1616 – 1699 | 260 | 322 |
| 1700 – 1749 | 1111 | 1176 |
| 1750 – 1799 | 1524 | 1455 |
| 1800 – 1849 | 729 | 567 |

The total number of mothers is 3624 and the total number of fathers is 3520. These are the cases where the birth year is known. Fortunately, for all these cases also the death year is known.[9] So one can directly calculate all life lengths as shown in Table 8.3-1.

**2.** The data in Table 8.3-1 provide values of statistical variables $\hat{T}_c^{f_1}$ and $\hat{T}_c^{m_1}$ that record, respectively, the life length of mothers and fathers belonging to birth cohort $c$.[10] The only problem concerns the fact that mothers

[9] As already mentioned, this is implied by the selection of families for the data set.

[10] The superscripts are meant to indicate female ($f_1$) and male ($m_1$) persons of the first generation.

**Table 8.3-1** Number of mothers and fathers in the Ostfriesland data set who died in the specified age.

| | Mothers | | | | Fathers | | | |
|---|---|---|---|---|---|---|---|---|
| Age | 1616/ 1699 | 1700/ 1749 | 1750/ 1799 | 1800/ 1849 | 1616/ 1699 | 1700/ 1749 | 1750/ 1799 | 1800/ 1849 |
| 17 | | 2 | | | | | | |
| 18 | | | | | | | | |
| 19 | | | 1 | | | | | |
| 20 | | 4 | 1 | | | | | |
| 21 | | 2 | 1 | 3 | | | | |
| 22 | | 2 | 1 | 2 | | | | |
| 23 | 1 | 3 | 3 | 4 | | | 2 | |
| 24 | 1 | 4 | 8 | 3 | | | | |
| 25 | | 10 | 11 | 4 | 1 | 2 | 4 | |
| 26 | 1 | 8 | 10 | | | 3 | | 1 |
| 27 | 1 | 5 | 11 | 5 | 1 | 1 | 4 | 1 |
| 28 | 1 | 13 | 3 | 8 | 1 | 3 | 2 | 4 |
| 29 | 2 | 5 | 5 | 5 | | 5 | 3 | 2 |
| 30 | 2 | 3 | 14 | 10 | 1 | 6 | 5 | 3 |
| 31 | | 5 | 9 | 5 | | 2 | 6 | 2 |
| 32 | 4 | 11 | 12 | 11 | 2 | 6 | 9 | 4 |
| 33 | 2 | 9 | 9 | 12 | 2 | 12 | 10 | 3 |
| 34 | 4 | 7 | 14 | 5 | 2 | 8 | 8 | 2 |
| 35 | 1 | 8 | 8 | 10 | 3 | 5 | 9 | 6 |
| 36 | 7 | 7 | 24 | 7 | 1 | 8 | 15 | 5 |
| 37 | 3 | 11 | 9 | 8 | | 3 | 14 | 8 |
| 38 | 6 | 5 | 12 | 6 | 1 | 7 | 11 | 5 |
| 39 | 2 | 11 | 20 | 6 | 2 | 9 | 7 | 2 |
| 40 | 2 | 11 | 16 | 9 | 2 | 16 | 13 | 3 |
| 41 | 5 | 8 | 18 | 7 | 2 | 11 | 17 | 6 |
| 42 | 6 | 15 | 11 | 6 | 1 | 10 | 17 | 5 |
| 43 | 2 | 20 | 10 | 10 | 3 | 12 | 20 | 8 |
| 44 | 2 | 14 | 8 | 16 | | 9 | 11 | 4 |
| 45 | 2 | 12 | 27 | 13 | 2 | 11 | 11 | 10 |
| 46 | 1 | 3 | 13 | 7 | 4 | 10 | 9 | 4 |
| 47 | 1 | 9 | 16 | 7 | 5 | 12 | 17 | 5 |
| 48 | 2 | 10 | 30 | 8 | 2 | 10 | 19 | 6 |
| 49 | 4 | 7 | 10 | 7 | 2 | 14 | 12 | 7 |
| 50 | 4 | 17 | 17 | 4 | 5 | 16 | 15 | 6 |
| 51 | 3 | 10 | 17 | 5 | 6 | 8 | 14 | 7 |
| 52 | 5 | 11 | 18 | 4 | 7 | 15 | 23 | 9 |
| 53 | 4 | 9 | 12 | 7 | 3 | 14 | 26 | 9 |
| 54 | 3 | 11 | 23 | 6 | 3 | 20 | 25 | 9 |
| 55 | 3 | 15 | 14 | 8 | 7 | 16 | 26 | 12 |
| 56 | 3 | 16 | 23 | 8 | 6 | 26 | 26 | 10 |
| 57 | 4 | 18 | 22 | 8 | 3 | 16 | 26 | 9 |
| 58 | 6 | 17 | 18 | 14 | 5 | 19 | 26 | 10 |
| 59 | 3 | 20 | 21 | 7 | 5 | 21 | 31 | 10 |
| 60 | 2 | 29 | 19 | 11 | 5 | 26 | 30 | 12 |
| 61 | 2 | 18 | 23 | 13 | 1 | 15 | 39 | 10 |
| 62 | 1 | 18 | 30 | 17 | 6 | 26 | 30 | 10 |
| 63 | 8 | 24 | 30 | 8 | 3 | 21 | 30 | 9 |
| 64 | 4 | 25 | 30 | 16 | 5 | 20 | 43 | 7 |
| 65 | 2 | 27 | 25 | 14 | 18 | 26 | 30 | 14 |



**Table 8.3-1** (continued) Number of mothers and fathers in the Ostfriesland data set who died in the specified age.

| | Mothers | | | | Fathers | | | |
|---|---|---|---|---|---|---|---|---|
| Age | 1616/ 1699 | 1700/ 1749 | 1750/ 1799 | 1800/ 1849 | 1616/ 1699 | 1700/ 1749 | 1750/ 1799 | 1800/ 1849 |
| 66 | 8 | 21 | 28 | 16 | 10 | 21 | 32 | 9 |
| 67 | 4 | 23 | 25 | 15 | 11 | 37 | 32 | 11 |
| 68 | 7 | 21 | 31 | 21 | 13 | 34 | 32 | 7 |
| 69 | 7 | 23 | 35 | 13 | 6 | 19 | 37 | 14 |
| 70 | 7 | 35 | 39 | 8 | 13 | 36 | 56 | 14 |
| 71 | 5 | 24 | 41 | 19 | 12 | 32 | 36 | 11 |
| 72 | 6 | 46 | 41 | 17 | 9 | 37 | 35 | 14 |
| 73 | 11 | 36 | 45 | 21 | 12 | 38 | 40 | 18 |
| 74 | 6 | 22 | 42 | 18 | 4 | 41 | 51 | 9 |
| 75 | 9 | 33 | 45 | 27 | 5 | 39 | 44 | 19 |
| 76 | 5 | 25 | 29 | 19 | 7 | 25 | 55 | 17 |
| 77 | 8 | 25 | 38 | 19 | 10 | 35 | 38 | 19 |
| 78 | 10 | 22 | 46 | 15 | 8 | 28 | 33 | 5 |
| 79 | 5 | 22 | 58 | 16 | 8 | 24 | 39 | 12 |
| 80 | 5 | 31 | 49 | 14 | 15 | 46 | 23 | 18 |
| 81 | 1 | 22 | 39 | 20 | 4 | 26 | 22 | 19 |
| 82 | 2 | 26 | 30 | 15 | 7 | 35 | 23 | 11 |
| 83 | 5 | 20 | 40 | 18 | 3 | 21 | 27 | 17 |
| 84 | 5 | 18 | 31 | 15 | 8 | 20 | 26 | 18 |
| 85 | 7 | 15 | 16 | 10 | 4 | 15 | 22 | 13 |
| 86 | 2 | 18 | 15 | 10 | 5 | 16 | 15 | 7 |
| 87 | 4 | 14 | 16 | 7 | 3 | 15 | 9 | 4 |
| 88 | 1 | 9 | 18 | 4 | 5 | 7 | 15 | 5 |
| 89 | 1 | 5 | 10 | 3 | 5 | 4 | 5 | 3 |
| 90 | 1 | 12 | 10 | 3 | 1 | 9 | 5 | 4 |
| 91 | 1 | 3 | 6 | 3 | 1 | 2 | 5 | 2 |
| 92 | | 6 | 1 | 1 | 1 | 5 | 2 | 2 |
| 93 | 2 | 3 | 8 | 3 | 1 | 3 | 1 | 3 |
| 94 | 1 | | 2 | 2 | | 2 | | 1 |
| 95 | | 1 | 1 | 2 | | 1 | | 1 |
| 96 | 1 | | 1 | | | 1 | | |
| 97 | | 1 | | | 2 | | | 1 |
| 98 | 1 | | | 1 | | 1 | | |
| 99 | | | | | | | | |
| 100 | | | 1 | | 1 | 1 | | |
| Total | 260 | 1111 | 1524 | 729 | 322 | 1176 | 1455 | 567 |

and fathers already survived until some age. We therefore calculate conditional survivor functions. For mothers we begin at age 25, and for fathers at age 30. So we use the formula

$$G[\hat{T}_c^{f_1} | \hat{T}_c^{f_1} \geq 25](\tau) = \frac{\sum_{k=\tau}^{\infty} d_{c,\tau}^{f_1}}{\sum_{k=25}^{\infty} d_{c,\tau}^{f_1}}$$

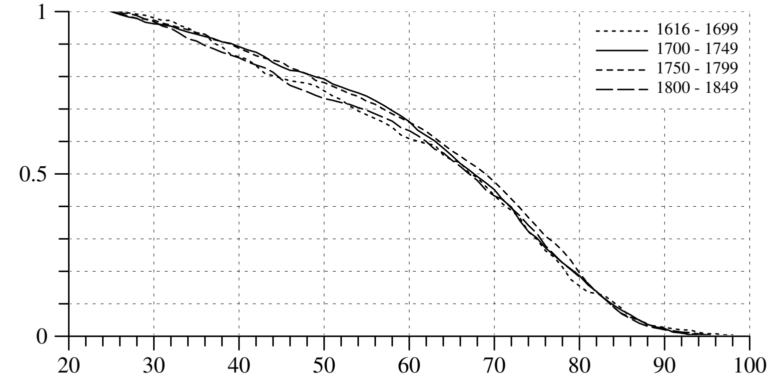



**Fig. 8.3-2** Conditional survivor functions $G[\hat{T}_c^f|\hat{T}_c^f \geq 25]$ for mothers in the Ostfriesland data set.

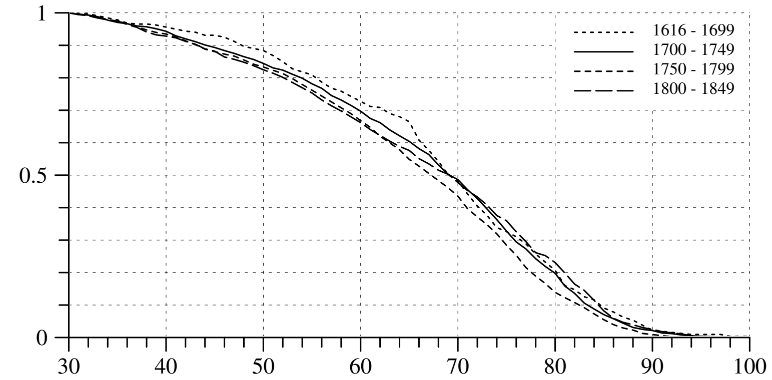



**Fig. 8.3-3** Conditional survivor functions $G[\hat{T}_c^m|\hat{T}_c^m \geq 30]$ for fathers in the Ostfriesland data set.

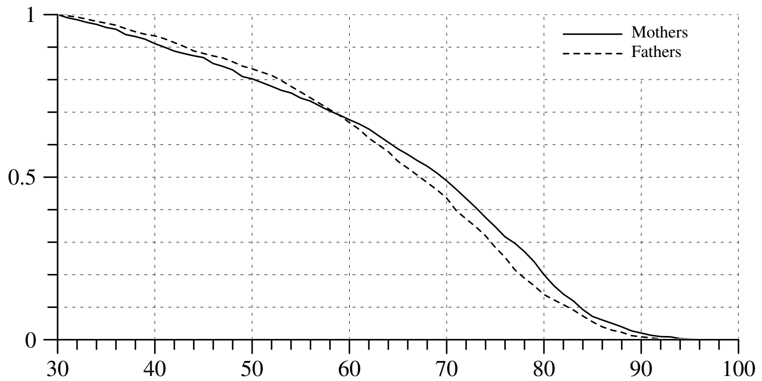



**Fig. 8.3-4** Comparison of conditional survivor functions $G[\hat{T}_c^f|\hat{T}_c^f \geq 30]$ and $G[\hat{T}_c^m|\hat{T}_c^m \geq 30]$ for birth cohort 1750 – 1799.



for mothers, and the formula

$$G[\hat{T}_c^{m_1}|\hat{T}_c^{m_1} \geq 25](\tau) = \frac{\sum_{k=\tau}^{\infty} d_{c,\tau}^{m_1}}{\sum_{k=30}^{\infty} d_{c,\tau}^{m_1}}$$

for fathers; $d_{c,\tau}^{f_1}$ and $d_{c,\tau}^{m+1}$ are, respectively, the number of mothers and fathers belonging to birth cohort $c$ who died at the age of $\tau$ (see Table 8.3-1). Figures 8.3-2 and 8.3-3 show these conditional survivor functions. Interestingly, there are only small variations across the different birth cohorts. As seen in Figure 8.3-4, these survivor functions are also very similar for mothers and fathers. Of course, one needs to recognize that we have used conditional survivor functions which only refer to parents.

## 8.3.3 Children's Survivor Functions

**1.** The calculation of survivor functions for children is more complicated because the observations are incomplete. In about 2 % of all cases we do not know the child's sex, and in order to distinguish female and male children, these cases cannot be used. Furthermore, the data set does not provide a valid birth year for all the remaining 8295 female and 8723 male children. However, as shown in the following table, this only concerns the first birth cohort.

| Birth years | Female children | (a) | (b) | (c) | Male children | (a) | (b) | (c) |
|---|---|---|---|---|---|---|---|---|
| 1616 – 1699 | 154 | 46 | 108 | 59 | 141 | 36 | 105 | 68 |
| 1700 – 1749 | 1382 | 0 | 1382 | 946 | 1588 | 0 | 1588 | 1141 |
| 1750 – 1799 | 2932 | 0 | 2932 | 2029 | 3117 | 0 | 3117 | 2304 |
| 1800 – 1849 | 3362 | 0 | 3362 | 1939 | 3421 | 0 | 3421 | 2174 |
| 1850 – 1881 | 465 | 0 | 465 | 245 | 456 | 0 | 456 | 242 |
| Total | 8295 | 46 | 8249 | 5218 | 8723 | 36 | 8687 | 5929 |

Columns labeled (a) show the number of cases without a valid birth year, columns (b) and (c) show, respectively, the number of cases with a valid birth year and a valid death year. So the question is how to use this incomplete information in order to estimate survivor functions.

**2.** As an example we consider male children born in the years 1750 – 1799. A first possibility would be to use only the 2304 complete observations. It would be possible then to immediately calculate a survivor function in the same way as was done in the previous section for parents. However, would the result be trustworthy? Assuming that we do not have any idea about the selection process that created the incomplete observations, no answer can be given. Nevertheless, we can at least calculate lower and upper bounds for a range of possible survivor functions. In order to calculate a lower bound we can simply assume that all children with an unknown death year died at age $\tau = 0$, and in order to calculate an upper bound we can assume that all these children survived the highest observed age which

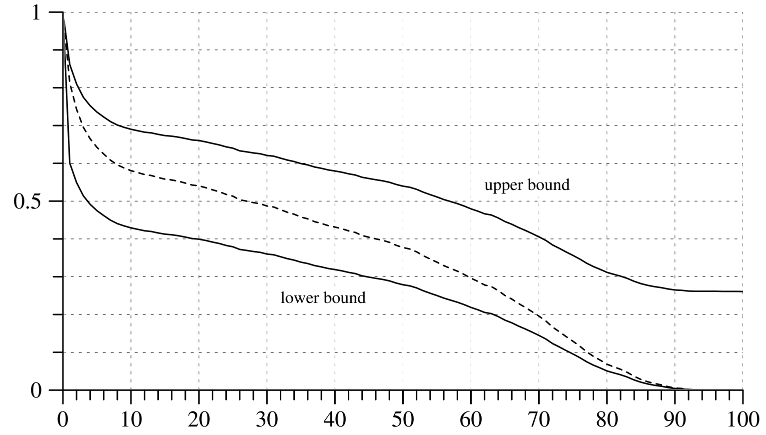



**Fig. 8.3-5** Lower and upper bounds for the survivor function of male children born between 1750 and 1799 in the Ostfriesland data set. The dotted line shows a survivor function calculated from only the complete observations.

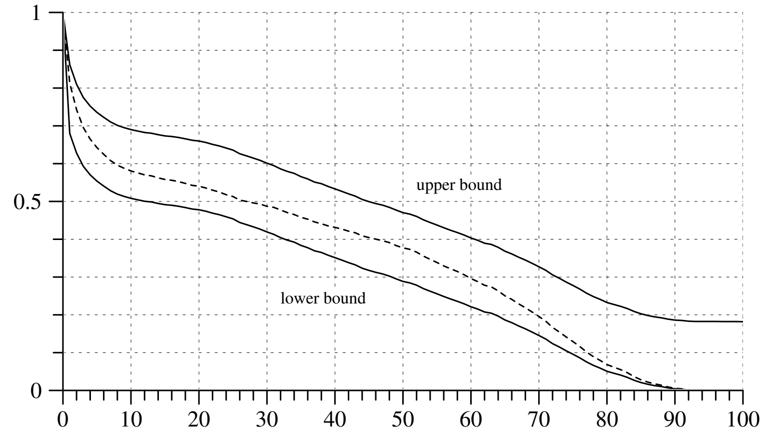



**Fig. 8.3-6** Lower and upper bounds for the survivor function of male children born between 1750 and 1799 in the Ostfriesland data set calculated by using additional information about latest observation. The dotted line shows a survivor function calculated from only the complete observations.

is 99 in this example. Figure 8.3-5 shows these bounds and also a survivor function calculated from only the complete observations. Obviously, there is a broad range for possible survivor functions.

**3.** The question therefore arises whether one can find additional information that can be used to get more narrow bounds. For this purpose any



kind of information can be used that allows to conclude that a person has survived some known age. For example, there might be information about a date of marriage or child-bearing (see the discussion in Imhof et al. 1990, p. 68 and p. 71). For some of those persons without a valid death year such information about a latest observation is provided in the data set, in most cases this is the marriage year. The following table shows the availability of this information.

| | Female children | | Male children | |
|---|---|---|---|---|
| Birth years | (a) | (b) | (a) | (b) |
| 1616 – 1699 | 49 | 11 | 37 | 7 |
| 1700 – 1749 | 436 | 142 | 447 | 119 |
| 1750 – 1799 | 903 | 393 | 813 | 246 |
| 1800 – 1849 | 1423 | 1021 | 1247 | 696 |
| 1850 – 1881 | 220 | 124 | 214 | 80 |

In our example, there are 813 male children without a valid death year, but in 246 cases we know a date of latest observation and can use this additional information to get better bounds for the survivor function. This is shown in Figure 8.3-6. Obviously, compared with Figure 8.3-5, the bounds are somewhat narrower.

**4.** Without the introduction of additional assumptions, the calculation of bounds to include the unknown survivor function is the best one can do. Of course, depending on the proportion of incomplete observations and the possibilities to use additional information, the range of possible survivor functions might become very broad and then looses almost all informational content. An alternative approach which is often followed in statistical practice would be to make assumptions about the process that leads to incomplete observations. The simplest assumption would be that the durations which are incompletely observed "randomly" result from the same distribution as the completely observed durations. Of course, this assumption might be wrong and there are almost no possibilities for checking the assumption with the given data set. The most often used estimation method for survivor functions which is based on this assumption is the *Kaplan-Meier procedure* (Kaplan and Meier 1958) and will be discussed in the next section.

## 8.3.4 The Kaplan-Meier Procedure

**1.** In order to explain the Kaplan-Meier procedure we refer to a general duration variable

$$\hat{T} : \Omega \longrightarrow \tilde{\mathcal{T}} := \{0, 1, 2, 3, \ldots\}$$

which is defined for some population $\Omega$. For each individual $\omega \in \Omega$, the variable $\hat{T}$ records a duration $\hat{T}(\omega) \in \tilde{\mathcal{T}}$ (see Section 7.3.1). This is the variable of theoretical interest. Observations are given by a two-dimensional variable

$$(T, D) : \Omega \longrightarrow \tilde{\mathcal{T}} \times \{0, 1\}$$

If $D(\omega) = 1$ the observation is complete and we can conclude that $\hat{T}(\omega) = T(\omega)$. On the other hand, if $D(\omega) = 0$ the observation is *right censored* and we can only conclude that $\hat{T}(\omega) \geq T(\omega)$.[11] The question then is how to estimate the distribution of $\hat{T}$ by using the information provided by $(T, D)$.

**2.** The Kaplan-Meier procedure is intended to provide one kind of answer. One possibility to explain this method is by referring to rates. Let $r[\hat{T}]$ denote the rate function corresponding the distribution of $\hat{T}$. As was shown in Section 7.3.1, the survivor function of $\hat{T}$ can then be calculated as follows:

$$G[\hat{T}](t) = \prod_{j=0}^{t-1} (1 - r[\hat{T}](j))$$

Of course, with partially censored data we do not know the rate function $r[\hat{T}]$, but we can use the observations provided by $(T, D)$ to get estimates and then use the above formula to calculate an estimate of the survivor function. Estimates of values of the rate function can be calculated in the following way:

$$r[\hat{T}](t) \approx_e r^*(t) := \frac{|\{\omega \in \Omega \,|\, T(\omega) = t, D(\omega) = 1\}|}{|\{\omega \in \Omega \,|\, T(\omega) \geq t\}|}$$

$r^*(t)$ might be called the *observed rate* at $t$, as derived from the observations which might, and actually have, ended their duration in this temporal location.[12] Of course, whether this observed rate is approximately equal to the value of the rate function $r[\hat{T}]$ at $t$ is not known and, as already remarked at the end of the previous section, can also not be checked with incomplete data. We therefore use the notation '$\approx_e$' to indicate that the right-hand side is assumed to be a reasonable estimate of the left-hand quantity. Given this assumption one immediately derives an estimate of the survivor function, namely

$$G[\hat{T}](t) \approx_e G^*(t) := \prod_{j=0}^{t-1} (1 - r^*(j))$$

[11] Also a strict inequality sign might here be used. However, with broadly defined units of the time axis, it is often plausible that an episode might end in the same temporal location where the observation ends.

[12] The set referred to in the denominator is sometimes called the observed *risk set*, and the set referred to in the numerator is called the observed *event set*.



$G^*$ is then called the *Kaplan-Meier estimate* of the unknown survivor function $G[\hat{T}]$.

**3.** As an illustration we continue with the example from the previous section and consider the male children born in the years 1750 to 1799. The variable $(T, D)$ will be defined as follows:

a) If $\omega$ refers to a male child in this birth cohort and we know the death year, then $D(\omega) = 1$ and $T(\omega)$ records the life length of $\omega$.

b) If we do not know the death year but have information about a latest observation, then $D(\omega) = 0$ and $T(\omega)$ is the age at latest observation.

c) If we know neither a death year nor a year of latest observation, then $D(\omega) = 0$ and $T(\omega) = 0$, that is, the observation is right censored already at the beginning.

Table 8.3-2 shows the data for male and female children born between 1750 and 1799. The numbers of male and female children who died at age $\tau$ are denoted, respectively, by $d_\tau^{m_2}$ and $d_\tau^{f_2}$, and the number of censored observations are denoted by $c_\tau^{m_2}$ and $c_\tau^{f_2}$.

**4.** Table 8.3-3 illustrates the calculations. The column labeled $n_\tau^{m_2}$ shows the number of cases in the "risk set", that is, the number of persons who are still at risk to die at age $\tau$. Then follow the number of persons who actually died ($d_\tau^{m_2}$) and the number of censored observations ($c_\tau^{m_2}$) at the current age. This then allows to calculate the observed rate $r^*(\tau)$ and to update the survivor function $G^*(\tau)$. The resulting survivor function is shown in Figure 8.3-7. Also shown is a survivor function calculated from only the complete observations which is located below the Kaplan-Meier survivor function. This follows from the fact that, in the calculation of observed rates, the Kaplan-Meier procedure takes into account also the censored observations. However, this does not make the Kaplan-Meier estimate always superior to the other one that only uses complete observations. As shown in the figure, both survivor functions are in the range that is indicated by the lower and upper bounds.

**5.** Using the data from Table 8.3-2, one can compare survivor functions for male and female children born between 1750 and 1799. The Kaplan-Meier estimates shown in Figure 8.3-8 suggest somewhat higher death rates for male children.

**6.** Survivor functions of parents and children cannot be compared directly because parents already survived until some age. But we can compare conditional survivor functions. As was done in Figure 8.3-4, we condition on having survived age 30. Conditional survivor functions for children can directly be derived from the Kaplan-Meier estimates: if $G^*$ is an estimate

**Table 8.3-2** Information about male and female children belonging to birth cohort 1750 – 1799 in the Ostfriesland data set.

| $\tau$ | $d_\tau^{m_2}$ | $c_\tau^{m_2}$ | $d_\tau^{f_2}$ | $c_\tau^{f_2}$ | $\tau$ | $d_\tau^{m_2}$ | $c_\tau^{m_2}$ | $d_\tau^{f_2}$ | $c_\tau^{f_2}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 431 | 567 | 403 | 510 | 51 | 17 | 2 | 12 | 5 |
| 1 | 161 | | 134 | | 52 | 24 | 4 | 13 | |
| 2 | 110 | | 95 | | 53 | 20 | 2 | 8 | 2 |
| 3 | 70 | | 54 | | 54 | 19 | 2 | 13 | 3 |
| 4 | 52 | | 46 | | 55 | 21 | 1 | 11 | |
| 5 | 41 | | 42 | | 56 | 17 | 4 | 15 | |
| 6 | 38 | | 32 | | 57 | 17 | 2 | 16 | |
| 7 | 28 | | 23 | | 58 | 20 | 1 | 13 | 1 |
| 8 | 19 | | 9 | | 59 | 23 | 1 | 18 | |
| 9 | 16 | | 13 | | 60 | 19 | | 12 | |
| 10 | 12 | | 9 | | 61 | 22 | 2 | 13 | 1 |
| 11 | 12 | | 5 | | 62 | 10 | | 19 | 1 |
| 12 | 6 | | 8 | | 63 | 24 | | 22 | |
| 13 | 12 | | 8 | | 64 | 30 | 2 | 24 | |
| 14 | 10 | | 7 | | 65 | 22 | | 20 | |
| 15 | 5 | | 4 | | 66 | 27 | | 17 | 1 |
| 16 | 8 | | 7 | | 67 | 24 | | 18 | |
| 17 | 11 | | 11 | | 68 | 26 | 1 | 22 | |
| 18 | 12 | | 7 | 2 | 69 | 27 | 1 | 20 | |
| 19 | 5 | 1 | 10 | 4 | 70 | 29 | | 23 | |
| 20 | 11 | 1 | 4 | 6 | 71 | 38 | | 30 | |
| 21 | 13 | 3 | 8 | 9 | 72 | 28 | | 36 | |
| 22 | 12 | 1 | 10 | 12 | 73 | 30 | 1 | 35 | |
| 23 | 16 | 1 | 15 | 23 | 74 | 28 | | 28 | |
| 24 | 11 | 6 | 12 | 21 | 75 | 30 | | 28 | 1 |
| 25 | 22 | 8 | 8 | 19 | 76 | 34 | | 23 | |
| 26 | 8 | 10 | 13 | 30 | 77 | 28 | | 31 | |
| 27 | 9 | 9 | 13 | 20 | 78 | 23 | | 28 | |
| 28 | 7 | 12 | 6 | 19 | 79 | 25 | 1 | 26 | |
| 29 | 14 | 9 | 8 | 20 | 80 | 16 | | 31 | |
| 30 | 6 | 14 | 13 | 11 | 81 | 15 | | 28 | 2 |
| 31 | 16 | 11 | 11 | 12 | 82 | 19 | | 19 | |
| 32 | 16 | 5 | 7 | 8 | 83 | 24 | | 22 | |
| 33 | 12 | 4 | 11 | 21 | 84 | 19 | | 13 | |
| 34 | 17 | 10 | 12 | 13 | 85 | 14 | | 8 | |
| 35 | 11 | 8 | 7 | 10 | 86 | 11 | | 10 | |
| 36 | 17 | 9 | 17 | 15 | 87 | 8 | | 16 | |
| 37 | 11 | 2 | 9 | 15 | 88 | 11 | | 9 | |
| 38 | 13 | 10 | 15 | 12 | 89 | 8 | | 3 | |
| 39 | 9 | 11 | 11 | 12 | 90 | 3 | | 5 | |
| 40 | 10 | 10 | 16 | 11 | 91 | 6 | | 5 | |
| 41 | 13 | 8 | 12 | 9 | 92 | 2 | | | |
| 42 | 10 | 9 | 7 | 14 | 93 | | | 5 | |
| 43 | 19 | 8 | 6 | 6 | 94 | | | 1 | |
| 44 | 10 | 8 | 9 | 12 | 95 | | | 1 | |
| 45 | 10 | 6 | 16 | 4 | 96 | 1 | | 1 | |
| 46 | 9 | 4 | 14 | 4 | 97 | | | | |
| 47 | 12 | 6 | 6 | | 98 | | | | |
| 48 | 18 | 5 | 15 | 1 | 99 | 1 | | | |
| 49 | 14 | 7 | 6 | 1 | 100 | | | 1 | |
| 50 | 9 | 3 | 13 | | | | | | |



**Table 8.3-3** Application of the Kaplan-Meier procedure to the data for male children in Table 8.3-2.

| $\tau$ | $n_\tau^{m_2}$ | $d_\tau^{m_2}$ | $c_\tau^{m_2}$ | $r^*(\tau)$ | $G^*(\tau)$ | $\tau$ | $n_\tau^{m_2}$ | $d_\tau^{m_2}$ | $c_\tau^{m_2}$ | $r^*(\tau)$ | $G^*(\tau)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3117 | 431 | 567 | 0.1383 | 1.0000 | 50 | 899 | 9 | 3 | 0.0100 | 0.4404 |
| 1 | 2119 | 161 | 0 | 0.0760 | 0.8617 | 51 | 887 | 17 | 2 | 0.0192 | 0.4360 |
| 2 | 1958 | 110 | 0 | 0.0562 | 0.7963 | 52 | 868 | 24 | 4 | 0.0276 | 0.4276 |
| 3 | 1848 | 70 | 0 | 0.0379 | 0.7515 | 53 | 840 | 20 | 2 | 0.0238 | 0.4158 |
| 4 | 1778 | 52 | 0 | 0.0292 | 0.7231 | 54 | 818 | 19 | 2 | 0.0232 | 0.4059 |
| 5 | 1726 | 41 | 0 | 0.0238 | 0.7019 | 55 | 797 | 21 | 1 | 0.0263 | 0.3965 |
| 6 | 1685 | 38 | 0 | 0.0226 | 0.6852 | 56 | 775 | 17 | 4 | 0.0219 | 0.3860 |
| 7 | 1647 | 28 | 0 | 0.0170 | 0.6698 | 57 | 754 | 17 | 2 | 0.0225 | 0.3776 |
| 8 | 1619 | 19 | 0 | 0.0117 | 0.6584 | 58 | 735 | 20 | 1 | 0.0272 | 0.3691 |
| 9 | 1600 | 16 | 0 | 0.0100 | 0.6507 | 59 | 714 | 23 | 1 | 0.0322 | 0.3590 |
| 10 | 1584 | 12 | 0 | 0.0076 | 0.6442 | 60 | 690 | 19 | 0 | 0.0275 | 0.3475 |
| 11 | 1572 | 12 | 0 | 0.0076 | 0.6393 | 61 | 671 | 22 | 2 | 0.0328 | 0.3379 |
| 12 | 1560 | 6 | 0 | 0.0038 | 0.6344 | 62 | 647 | 10 | 0 | 0.0155 | 0.3268 |
| 13 | 1554 | 12 | 0 | 0.0077 | 0.6320 | 63 | 637 | 24 | 0 | 0.0377 | 0.3218 |
| 14 | 1542 | 10 | 0 | 0.0065 | 0.6271 | 64 | 613 | 30 | 2 | 0.0489 | 0.3096 |
| 15 | 1532 | 5 | 0 | 0.0033 | 0.6230 | 65 | 581 | 22 | 0 | 0.0379 | 0.2945 |
| 16 | 1527 | 8 | 0 | 0.0052 | 0.6210 | 66 | 559 | 27 | 0 | 0.0483 | 0.2833 |
| 17 | 1519 | 11 | 0 | 0.0072 | 0.6177 | 67 | 532 | 24 | 0 | 0.0451 | 0.2696 |
| 18 | 1508 | 12 | 0 | 0.0080 | 0.6133 | 68 | 508 | 26 | 1 | 0.0512 | 0.2575 |
| 19 | 1496 | 5 | 1 | 0.0033 | 0.6084 | 69 | 481 | 27 | 1 | 0.0561 | 0.2443 |
| 20 | 1490 | 11 | 1 | 0.0074 | 0.6063 | 70 | 453 | 29 | 0 | 0.0640 | 0.2306 |
| 21 | 1478 | 13 | 3 | 0.0088 | 0.6019 | 71 | 424 | 38 | 0 | 0.0896 | 0.2158 |
| 22 | 1462 | 12 | 1 | 0.0082 | 0.5966 | 72 | 386 | 28 | 0 | 0.0725 | 0.1965 |
| 23 | 1449 | 16 | 1 | 0.0110 | 0.5917 | 73 | 358 | 30 | 1 | 0.0838 | 0.1822 |
| 24 | 1432 | 11 | 6 | 0.0077 | 0.5851 | 74 | 327 | 28 | 0 | 0.0856 | 0.1670 |
| 25 | 1415 | 22 | 8 | 0.0155 | 0.5806 | 75 | 299 | 30 | 0 | 0.1003 | 0.1527 |
| 26 | 1385 | 8 | 10 | 0.0058 | 0.5716 | 76 | 269 | 34 | 0 | 0.1264 | 0.1373 |
| 27 | 1367 | 9 | 9 | 0.0066 | 0.5683 | 77 | 235 | 28 | 0 | 0.1191 | 0.1200 |
| 28 | 1349 | 7 | 12 | 0.0052 | 0.5646 | 78 | 207 | 23 | 0 | 0.1111 | 0.1057 |
| 29 | 1330 | 14 | 9 | 0.0105 | 0.5616 | 79 | 184 | 25 | 1 | 0.1359 | 0.0939 |
| 30 | 1307 | 6 | 14 | 0.0046 | 0.5557 | 80 | 158 | 16 | 0 | 0.1013 | 0.0812 |
| 31 | 1287 | 16 | 11 | 0.0124 | 0.5532 | 81 | 142 | 15 | 0 | 0.1056 | 0.0730 |
| 32 | 1260 | 16 | 5 | 0.0127 | 0.5463 | 82 | 127 | 19 | 0 | 0.1496 | 0.0653 |
| 33 | 1239 | 12 | 4 | 0.0097 | 0.5394 | 83 | 108 | 24 | 0 | 0.2222 | 0.0555 |
| 34 | 1223 | 17 | 10 | 0.0139 | 0.5341 | 84 | 84 | 19 | 0 | 0.2262 | 0.0432 |
| 35 | 1196 | 11 | 8 | 0.0092 | 0.5267 | 85 | 65 | 14 | 0 | 0.2154 | 0.0334 |
| 36 | 1177 | 17 | 9 | 0.0144 | 0.5219 | 86 | 51 | 11 | 0 | 0.2157 | 0.0262 |
| 37 | 1151 | 11 | 2 | 0.0096 | 0.5143 | 87 | 40 | 8 | 0 | 0.2000 | 0.0206 |
| 38 | 1138 | 13 | 10 | 0.0114 | 0.5094 | 88 | 32 | 11 | 0 | 0.3438 | 0.0164 |
| 39 | 1115 | 9 | 11 | 0.0081 | 0.5036 | 89 | 21 | 8 | 0 | 0.3810 | 0.0108 |
| 40 | 1095 | 10 | 10 | 0.0091 | 0.4995 | 90 | 13 | 3 | 0 | 0.2308 | 0.0067 |
| 41 | 1075 | 13 | 8 | 0.0121 | 0.4950 | 91 | 10 | 6 | 0 | 0.6000 | 0.0051 |
| 42 | 1054 | 10 | 9 | 0.0095 | 0.4890 | 92 | 4 | 2 | 0 | 0.5000 | 0.0021 |
| 43 | 1035 | 19 | 8 | 0.0184 | 0.4843 | 93 | 2 | 0 | 0 | 0.0000 | 0.0010 |
| 44 | 1008 | 10 | 8 | 0.0099 | 0.4755 | 94 | 2 | 0 | 0 | 0.0000 | 0.0010 |
| 45 | 990 | 10 | 6 | 0.0101 | 0.4707 | 95 | 2 | 0 | 0 | 0.0000 | 0.0010 |
| 46 | 974 | 9 | 4 | 0.0092 | 0.4660 | 96 | 2 | 1 | 0 | 0.5000 | 0.0010 |
| 47 | 961 | 12 | 6 | 0.0125 | 0.4617 | 97 | 1 | 0 | 0 | 0.0000 | 0.0005 |
| 48 | 943 | 18 | 5 | 0.0191 | 0.4559 | 98 | 1 | 0 | 0 | 0.0000 | 0.0005 |
| 49 | 920 | 14 | 7 | 0.0152 | 0.4472 | 99 | 1 | 1 | 0 | 1.0000 | 0.0005 |

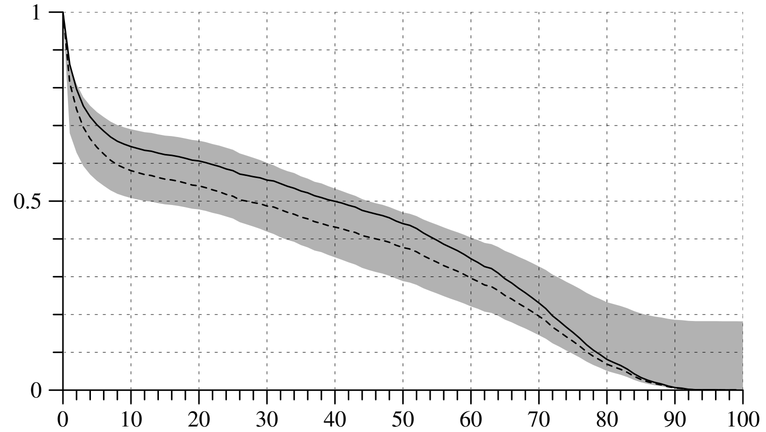



**Fig. 8.3-7** Kaplan-Meier survivor function for male children born between 1750 and 1799 in the Ostfriesland data set calculated in Table 8.3-3 (solid line). The dotted line and the grey-scaled bounds are taken from Figure 8.3-6.

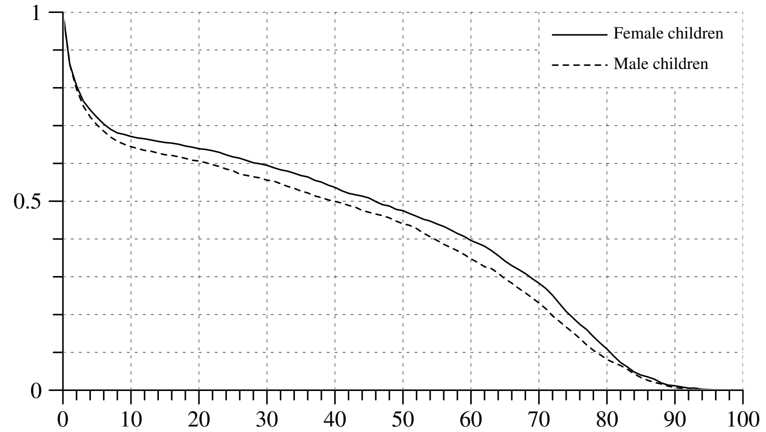



**Fig. 8.3-8** Kaplan-Meier survivor functions for female and male children born between 1750 and 1799 in the Ostfriesland data set (Table 8.3-2).

of $G[\hat{T}]$, then

$$G[\hat{T}|\hat{T} \geq t_0] \approx_e \frac{G^*(t)}{G^*(t_0)} = \prod_{j=t_0}^{t-1} (1 - r^*(j))$$

Figure 8.3-9 compares mothers and female children, both born between



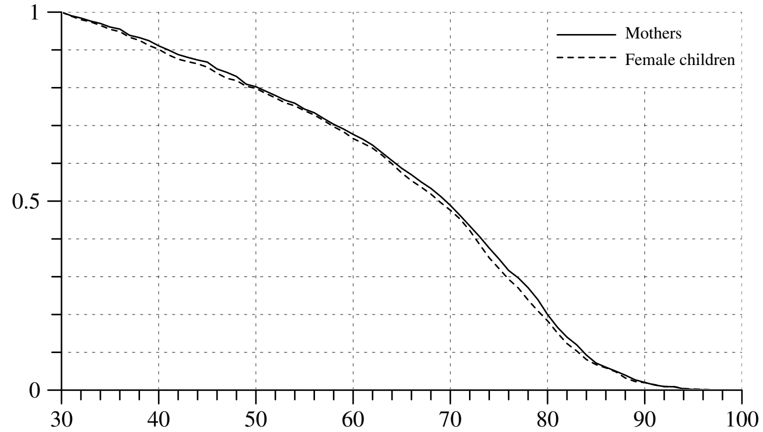



**Fig. 8.3-9** Kaplan-Meier survivor function for mothers and female children born between 1750 and 1799 in the Ostfriesland data set.

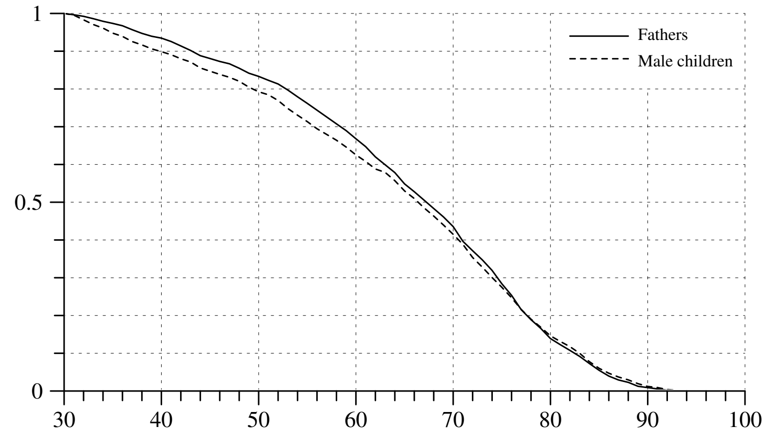



**Fig. 8.3-10** Kaplan-Meier survivor function for fathers and male children born between 1750 and 1799 in the Ostfriesland data set.

1750 and 1799, Figure 8.3-11 provides the corresponding curves for fathers and male children. In both cases, the conditional survivor functions are more or less similar, providing some confidence into the Kaplan-Meier estimates of the survivor functions for children, at least for higher ages.

## 8.4 Mortality Data from Panel Studies

This section is not finished.

# Chapter 9

# Parent's Length of Life

Additional information about the survival of people in historical time is available from the German Life History Study (GLHS) and the Socio-economic Panel (SOEP). In both surveys respondents were asked to provide information about their parents, in particular about their parent's birth years, whether they were still alive at the interview date, and, if not, about their respective years of death. We can try to use this information to enlarge the knowledge about mortality conditions in earlier periods.[1] However, we need first to consider the specific features of the data generating process because, in this case, the information about the parents results from a sample of their children. Information is therefore only available for persons who became a parent of at least one child, and this information also depends on the child's survival to the interview dates. We first introduce the notion of left truncated data, and then use a simulation model to study possible complications. The insights gained by this study will finally be used to draw some inferences from the GLHS and SOEP data.

## 9.1 Left Truncated Data

**1.** The first problem obviously concerns the fact that the available data contain information only about those persons who became mother or father of at least one child. One possibility would be to restrict any inferences to those persons. This would allow to directly apply the standard Kaplan-Meier procedure to estimate survivor functions with partially censored data (see Section 8.3.4). On the other hand, one may also assume that mortality is independent of whether or not persons became parents of a child. This assumption would open the possibility to draw at least some inferences about the whole population. Of course, it will not be possible to estimate complete survivor functions because no information is available about death events occurring at early ages. But given the independence assumption, it might be possible to estimate survivor functions conditional on having survived to the age at which children are born.

**2.** In order to discuss this question we consider a simple model where we are given a population set $\Omega$ and a two-dimensional variable:

$$(T, C) : \Omega \longrightarrow \tilde{\mathcal{T}} \times \tilde{\mathcal{T}} \cup \{-1\}$$

$\tilde{\mathcal{T}} := \{0, 1, 2, \ldots\}$ is a property space for age. For each $\omega \in \Omega$, $T(\omega)$ is $\omega$'s

[1]For previous analyzes of the SOEP data about the life lengths of parents see Schepers and Wagner (1989), and Klein (1993).



life length, and $C(\omega)$ records the age at which $\omega$ became, for the first time, mother of a child, or is -1 if this did not happen during $\omega$'s lifetime.[2] The problem can now be stated as follows: The available data only refer to a subset of $\Omega$, namely

$$\Omega^* := \{\omega \in \Omega \,|\, C(\omega) \geq 0\}$$

consisting of women who became a mother of at least one child. We start from the assumption that information from all children is available, ignoring their mortality up to the interview date. The question then is, how, and to what extent, can these data be used to assess the distribution of $T$ in $\Omega$?

**3.** We follow the basic idea of the Kaplan-Meier procedure to assess the distribution of $T$ via rates (see Section 8.3.4). Assume complete observations. It would be possible, then, to create a risk set

$$\Omega_\tau := \{\omega \in \Omega \,|\, T(\omega) \geq \tau\}$$

containing all members of $\Omega$ who might die at age $\tau$, and an event set

$$\{\omega \in \Omega_\tau \,|\, T(\omega) = \tau\}$$

containing those members of $\Omega_\tau$ who actually died at age $\tau$. From these sets one can calculate rates

$$r(\tau) := \frac{|\{\omega \in \Omega_\tau \,|\, T(\omega) = \tau\}|}{|\Omega_\tau|}$$

which can be used to find the survivor function

$$G[T](\tau) = \prod_{j=0}^{\tau-1} (1 - r(j))$$

Now, since our data only refer to $\Omega^*$, we cannot create these sets and consequently cannot calculate the rates $r(\tau)$. One can only try to estimate these rates, but this will then require an assumption. Our assumption will be that mortality does not depend on whether, and when, people became mothers and fathers. In terms of the model, the assumption is[3]

$$r(\tau) \approx_e \tilde{r}^*(\tau) := \frac{|\{\omega \in \Omega^*_\tau \,|\, T(\omega) = \tau\}|}{|\Omega^*_\tau|}$$

[2]For the present discussion we assume that $\Omega$ refers to women only. The same reasoning, however, applies to men with minor modifications.

[3]As in Section 8.3.4, we use the notation '$\approx_e$' to indicate that the right-hand side is assumed to be a reasonable estimate of the left-hand quantity.

where the risk set on the right-hand side is now defined by

$$\Omega^*_\tau := \{\omega \in \Omega^* \,|\, T(\omega) \geq \tau, 0 \leq C(\omega) \leq \tau\}$$

Since both this risk set and the corresponding event set can be calculated from data restricted to $\Omega^*$, one gets estimates of the rates $r(\tau)$. Of course, this will be possible only for ages

$$\tau \geq a^+ := \min\{C(\omega) \,|\, \omega \in \Omega^*\}$$

which implies that only the conditional survivor function $G[T|T \geq a^+]$ can be estimated:

$$G[T|T \geq a^+](\tau) \approx_e \prod_{j=0}^{\tau-1} (1 - \tilde{r}^*(j)) \qquad (9.1.1)$$

Notice also that in general

$$\Omega^*_\tau \neq \{\omega \in \Omega^* \,|\, T(\omega) \geq \tau\}$$

because a women in $\Omega^*$ might get her first child later than $\tau$. In order to create suitable risk sets $\Omega^*_\tau$ one has to apply the same conditioning as used for the event sets to meet the assumption that mortality does not depend on whether, and when, women become mothers.

**4.** An example can serve to illustrate the reasoning. We assume that $\Omega$ contains 1000 women and consider, in turn, five age classes:

0 In the age class $\tau = 0$ all 1000 women are at risk of dying, and we assume that 100 women actually die.

1 There remain 900 women who might die in the age class $\tau = 1$. We assume that 100 of these women actually die. However, some of these women will also become mothers of children. We assume that this is true of 200 women. Implied by the assumption that mortality does not depend on becoming a mother, about

$$\frac{100}{900}\, 200 \approx 22$$

will also die; of course, they belong to the 100 persons who die in this age class.

2 There remain 800 women who might die in the age class $\tau = 2$. We assume that 200 of these women actually die. Furthermore, we assume that 300 women become mothers of a child. The assumption of equal mortality implies that about 300 / 4 = 75 of these women also die. In addition, there are 178 women who became mothers in age class $\tau = 1$, and of these about $178/4 \approx 45$ will die.



3 The remaining number of women is $800 - 200 = 600$, and we assume that 200 of these women die in the age class $\tau = 3$. Furthermore we assume that again 200 women become mothers of a child. Consequently, about $200/3 \approx 67$ of these women also will die. Furthermore, there are 358 women who became mothers before $\tau = 3$, and of these 119 will die.

4 Finally, there remain 400 women and all will die because $\tau = 4$ is the last and open-ended age class.

**5.** Given this situation, we can first assume that complete data are available. This would allow to calculate the survivor function in the following way:

| $\tau$ | $\lvert\Omega_\tau\rvert$ | $\lvert\{\omega \in \Omega_\tau \,\lvert\, T(\omega) = \tau\}\rvert$ | $r(\tau)$ | $G[T](\tau)$ |
|---|---|---|---|---|
| 0 | 1000 | 100 | 1/10 | 1.00 |
| 1 | 900 | 100 | 1/9 | 0.90 |
| 2 | 800 | 200 | 1/4 | 0.80 |
| 3 | 600 | 200 | 1/3 | 0.60 |
| 4 | 400 | 400 | 1 | 0.40 |

Obviously, the survivor function is simply proportional to the number of persons in the risk set. In a next step, we assume that data are only available for $\Omega^*$, that is, women who gave birth to at least one child. In our example, there are altogether $200 + 300 + 200 = 700$ women. We now perform the same calculations for these women using the risk and event sets as defined above. This can be summarized in the following table:

| $\tau$ | $\lvert\Omega^*_\tau\rvert$ | $\lvert\{\omega \in \Omega^*_\tau \,\lvert\, T(\omega) = \tau\}\rvert$ | $\tilde{r}^*(\tau)$ | $G^*(\tau)$ |
|---|---|---|---|---|
| 0 | | | | 1 |
| 1 | 200 | 22 | 0.110 | $g$ |
| 2 | 478 | 120 | 0.251 | $g \cdot 0.890$ |
| 3 | 558 | 186 | 0.333 | $g \cdot 0.667$ |
| 4 | 372 | 372 | 1.000 | $g \cdot 0.445$ |

For $\tau = 0$, the risk set is empty and we cannot calculate a death rate. Consequently, we also cannot estimate the value of the survivor function for $\tau = 1$ which, in the table, is substituted by the unknown value $g$. For $\tau > 0$ it is possible, however, to create risk and event sets and calculate corresponding rates $\tilde{r}^*(\tau)$. And these rates can finally be used to derive the values of the conditional survivor function

$$G[T|T \geq a^+](\tau) = \frac{G[T](\tau)}{G[T](a^*)} \approx_e \frac{1}{g}\, G^*(\tau)$$

where $a^+ = 1$ in the present example.

**6.** The important point is to recognize the difference between the unconditional rates

$$r^*(\tau) := \frac{|\{\omega \in \Omega^* \,|\, T(\omega) = \tau\}|}{|\{\omega \in \Omega^* \,|\, T(\omega) \geq \tau\}|}$$

and the rates $\tilde{r}^*(\tau)$ defined above. Using the rates $r^*(\tau)$ would result in a survivor function for the variable $T^*$ defined for the reference set $\Omega^*$. But this survivor function will not, in general, be proportional to a conditional survivor function for the variable of interest, $T$, which is defined for the reference set $\Omega$. In order to calculate a conditional survivor function for $T$ one needs the rates $\tilde{r}^*(\tau)$, see formula (9.1.1). The risk sets $\Omega^*_\tau$ from which the rates $\tilde{r}^*(\tau)$ are derived take into account the temporal nature of becoming a member of $\Omega^*$. In our example, a woman becomes a member of $\Omega^*$ after the birth of her first child. Corresponding observations are therefore called *left truncated*, in this example, left truncated at the age of first child-bearing. Since our observations of death events only relate to members of $\Omega^*$, the risk of an observed death at the age $\tau$ only relates to persons who became members of $\Omega^*$ until $\tau$. This argument will again be used in Section 9.2.2.

## 9.2 Selection by Survival

Before applying the method discussed in the previous section to the GLHS and SOEP data we need to discuss the additional complications that result from a retrospective survey of children who are asked about their parents. In order to understand some of the problems that might result from this specific data generating process we use a simulation model.

### 9.2.1 The Simulation Model

**1.** The basic idea is to simulate data for a set of women according to a known survivor function and then to compare this known function with estimates based on information from the women's children who survived until some fixed interview date. In the first version of the model we refer to a set of $N = 10000$ women all born in the year $t_0 := 1900$; this set will be denoted by $\Omega$. We assume that these women survive according to the 1891–1900 period life table for Germany (see Table 7.4-3 in Section 7.4.2); the corresponding age-specific death rates will be denoted by $\delta^f_\tau$. Additional assumptions concern the birth of children.[4] We assume age- and parity-specific birth rates

$$\beta_{\tau,k} := \frac{\text{Number of women giving birth to a further child at age } \tau}{\text{Number of women aged } \tau \text{ and having } k \text{ children}}$$

[4]Women, as well as men, can become parents in different ways. In the model we only consider women who might become mothers by giving birth to children.



**Box 9.2-1** Skeleton of the simulation model.

```
For each ω ∈ Ω do:
    n(ω) := 0;                    # counter for ω's children
    For (τ = 0, ..., 100 ) {
        Get a random number ε;
        If (ε ≤ β_{τ,n(ω)})
          add one child to n(ω), create a new entry
          in Ω^c, and record the mother's identification
          number and age;
        Get another random number ε;
        If (ε ≤ δ^f_τ)
          goto L1;
    }
L1:
    Record that ω died at age τ and has given birth to c(ω)
    children, also record for all children the mother's age at death;
For each ω ∈ Ω^c do:
    For (τ = 0, ..., 100 ) {
        Get a random number ε;
        If (ε ≤ δ^c_τ)
          goto L2;
    }
L2:
    Record that ω died at age τ;
```

In order to arrive at a simulation model that roughly corresponds to the historical situation these rates are calculated from a subsample of the census that took place in Germany in the year 1970 (see Section 12.2.1). For each women who survived until 1970, this sample contains information about the birth years of up to 12 children. For the calculation of age- and parity-specific birth rates we have used all of these women who were born between 1870 and 1925. These rates are only used to set up our simulation model, we therefore do not pay attention to historical accuracy.

**2.** We can thus think of a second reference set, $\Omega^c$, containing identification numbers of all children born of the women in $\Omega$. Of course, the number of members of $\Omega^c$ is not known in advance but depends on the death rates $\delta^f_\tau$ and the birth rates $\beta_{\tau,k}$. But given these rates, we can finally create two lists. One list containing, for each women in $\Omega$, her identification number, her death year, and her number of children. And another list that contains, for each child in $\Omega^c$, an identification number, the birth year, and the identification number of the mother. In addition, in order

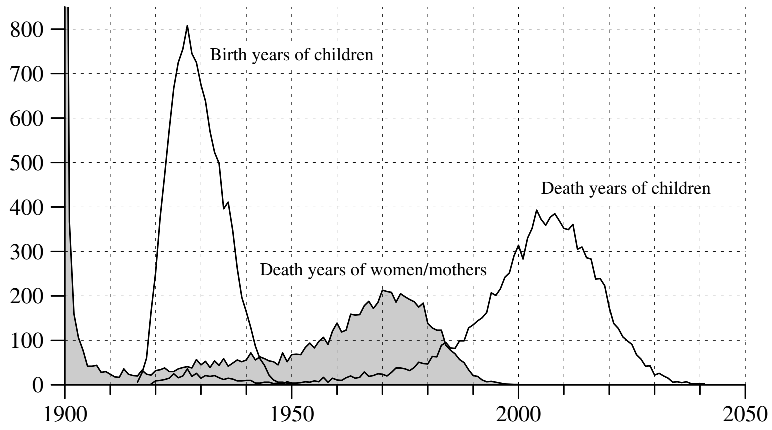



**Fig. 9.2-1** Frequency distributions of birth and death years in the simulated data set.

to simulate a retrospective survey, we assume that the children survive according to the 1960–1962 period life table for women (see Table 7.4-3 in Section 7.4.2); the corresponding age-specific death rates will be denoted by $\delta^c_\tau$. So we can add, for each child in the second list, also a death year.

**3.** Box 9.2-1 depicts the algorithm that we have used to generate the data for the simulation model. In this description, $\epsilon$ refers to a draw from random numbers which are equally distributed in the interval from 0 to 1. Using this algorithm we get the first list with $N = 10000$ entries that record the identification numbers of the women in $\Omega$, their age at death, and their number of children. Of these women, 4776 have at least one child.[5] We also get the second list which, in our implementation of the model, contains entries for 11407 children. Figure 9.2-1 shows a frequency distribution of the years in which the women in $\Omega$ died on a historical time axis. Also shown are frequency distributions of the birth and death years of the children. Note that the algorithm is based on the assumption that women's survival is independent of their giving birth to children. Problems that might result from a violation of this assumption can therefore not be checked within this model.

### 9.2.2 Considering Left Truncation

**1.** Before using the model to discuss the question whether we might be able to recover the survivor function of the members of $\Omega$ based on information resulting from a retrospective survey of their children, we illustrate the

[5]One should note that, based on the 1891–1900 period life table, only 68 % of the women survived age 20, and only 60 % survived age 40.



importance of correctly taking into account left truncated observations. What we want to recover is some part of the distribution of the variable

$$T : \Omega \longrightarrow \tilde{\mathcal{T}} := \{0, 1, 2, 3, \ldots\}$$

that records the life length of the members of $\Omega$. Of course, our observations refer at best to a subset of $\Omega$ that consists of those members of $\Omega$ who gave birth to at least one child. This subset will be denoted by $\Omega^*$. We can now again define a variable

$$T^* : \Omega^* \longrightarrow \tilde{\mathcal{T}} := \{0, 1, 2, 3, \ldots\}$$

that records the life length of the members of $\Omega^*$. However, as already mentioned in Section 9.1, it is important to recognize that the distribution of $T^*$ will not, in general, be identical with a conditional distribution of $T$.

**2.** Referring to the simulation model of the previous section, we have defined the distribution of $T$ by the death rates $\delta^f_\tau$. The survivor function of $T$ is therefore given by

$$G[T](\tau) = \prod_{j=0}^{\tau-1} (1 - \delta^f_j)$$

Obviously, in order to recover (some part of) this survivor function we need estimates of the death rates $\delta^f_\tau$. However, these death rates are systematically different from the death rates

$$r^*(\tau) := \frac{|\{\omega \in \Omega^* \,|\, T^*(\omega) = \tau\}|}{|\{\omega \in \Omega^* \,|\, T^*(\omega) \geq \tau\}|}$$

which correspond to the variable $T^*$ and might be used to calculate its survivor function. In order to find estimates of $\delta^f_\tau$, we need to take into account that women only become members of $\Omega^*$ when they have given birth to a first child. We therefore consider a two-dimensional variable

$$(T^*, C^*) : \Omega^* \longrightarrow \tilde{\mathcal{T}} \times \tilde{\mathcal{T}}$$

where $T^*$ is defined as before and $C^*$ records the age at which members of $\Omega^*$ gave, for the first time, birth to a child. This then allows to define a rate function

$$\tilde{r}^*(\tau) := \frac{|\{\omega \in \Omega^* \,|\, T^*(\omega) = \tau\, C^*(\omega) \leq \tau\}|}{|\{\omega \in \Omega^* \,|\, T^*(\omega) \geq \tau, C^*(\omega) \leq \tau\}|}$$

The denominator counts the number of members of the risk set at $\tau$, defined as

$$\Omega^*_\tau := \{\omega \in \Omega^* \,|\, T^*(\omega) \geq \tau, C^*(\omega) \leq \tau\}$$

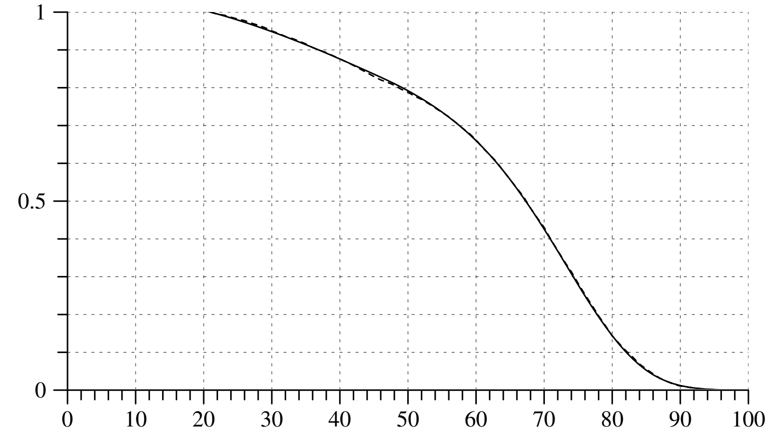



**Fig. 9.2-2** Survivor functions, conditional on $\tau \geq 21$, from the 1891–1900 period life table (solid line) and from women with at least one child in the simulated data set (dotted line).

that is, members of $\Omega^*$ who actually have given birth to a first child not later than $\tau$; and the numerator counts the number of members of the risk set who actually died at the age $\tau$. Obviously, $\tilde{r}^*(\tau) \neq r^*(\tau)$, but $\tilde{r}^*(\tau)$ is the death rate of women who actually are members of $\Omega^*$ *at the age of* $\tau$ and, as we have construed the model, is a reasonable estimate of $\delta^f_\tau$. We therefore should use $\tilde{r}^*(\tau)$ to estimate a conditional version of the survivor function $G[T]$.

**3.** In principle, it would be possible to obtain estimates of $\tilde{r}^*$ from the age at the first birth onward. Since this rate is zero up to the age of the first observed death in $\Omega^*$, one might as well start at this age, say $a^*$,[6] so that the conditional survivor function is then

$$\tilde{G}^*_{a^*}(\tau) := \prod_{j=a^*}^{\tau-1} (1 - \tilde{r}^*(j))$$

It might be taken as an estimate of the conditional survivor function $G[T|T \geq a^*]$. To illustrate, we use the simulated data set from our model. Assuming complete knowledge about all women in $\Omega^*$, we find that the earliest death occurs at age 19. However, this occurs only once, and at $\tau = 20$ there is no death at all. We therefore define $a^* := 21$ and, given complete knowledge, can directly calculate $\tilde{G}^*_{a^*}$. This is shown in Figure 9.2-2 as a dotted line. Also shown as a solid line is $G[T|T \geq a^*]$ calculated

[6] Of course, due to the small number of cases in a sample of observations, $\tilde{r}^*(a^*)$ might not be a good estimate of $\delta^f(a^*)$ and one should condition on some later age. In fact, it might happen that $\tilde{r}^*(a^*) = 1$ so that one cannot find a reasonable estimate of a survivor function beginning at $a^*$.



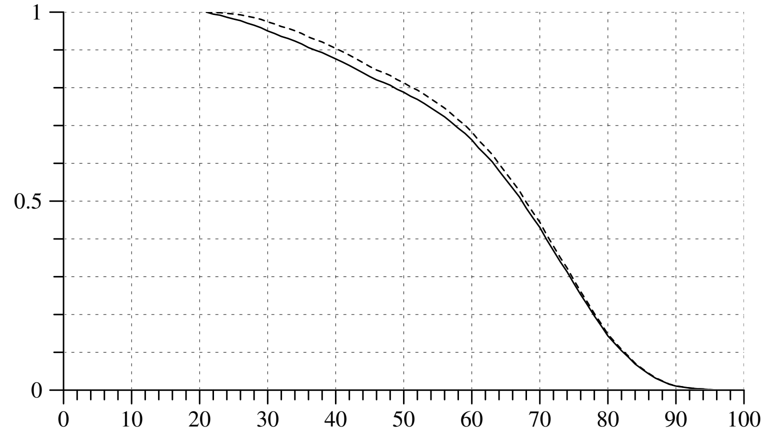



**Fig. 9.2-3** Conditional survivor functions $\tilde{G}^*_{a^*}$ (solid line) and $G[T^*|T^* \geq a^*]$ calculated from the simulated data set with $a^* = 21$.

from the 1891-1900 period life table for women. Obviously, both curves agree quite well. On the other hand, if we had not taken into account the fact that women become members of $\Omega^*$ only after having given birth to a first child, but estimated the survivor function $G[T^*|T^* \geq a^*]$, the result would be systematically biased as a consequence of the inequality $r^*(\tau) \leq \tilde{r}^*(\tau)$. This is illustrated by Figure 9.2-3 where the solid line shows $\tilde{G}^*_{a^*}$ and the dotted line shows $G[T^*|T^* \geq a^*]$.

**4.** The fact that women become members of $\Omega^*$ only after the birth of a child is formally equivalent to treating the observations as left truncated at the age at first birth. Of course, nothing is wrong with estimating the survivor function of $T^*$ instead of $\tilde{G}^*_{a^*}$. The argument has only shown that one should use the latter one if the interest is in recovering part of the distribution of $T$. One might also notice that, while $G[T^*]$ refers to a well-defined statistical variable, this cannot be said of $\tilde{G}^*_{a^*}$. This function actually results from a mixture of rate functions. This is seen by a partition of $\Omega^*$ into subsets $\Omega^*_{[a]} := \{\omega \in \Omega^* \,|\, C^*(\omega) = a\}$, consisting of those members of $\Omega^*$ who had a first birth at the age $a$. Defining rate functions for these subsets by

$$\tilde{r}^*_a(\tau) := \frac{|\{\omega \in \Omega^*_{[a]} \,|\, T^*(\omega) = \tau\}|}{|\{\omega \in \Omega^*_{[a]} \,|\, T^*(\omega) \geq \tau\}|}$$

one can express $\tilde{r}^*(\tau)$ as a mixture

$$\tilde{r}^*(\tau) = \sum_{a \leq \tau} \tilde{r}^*_a(\tau)\, w_a(\tau)$$

where the weights, defined as

$$w_a(\tau) := \frac{|\{\omega \in \Omega^*_{[a]} \,|\, T^*(\omega) \geq \tau\}|}{\sum_{a' \leq \tau} |\{\omega \in \Omega^*_{[a']} \,|\, T^*(\omega) \geq \tau\}|}$$

reflect the composition of the risk set at $\tau$.

### 9.2.3 Using Information from Children

**1.** We now turn to the question of how to estimate conditional survivor functions for the members of $\Omega$ when we only have information from the children, that is, members of $\Omega^c$. So we need to take into account the relationship between $\Omega^c$ and $\Omega^*$. To make this explicit, we introduce a variable (function)

$$m : \Omega^c \longrightarrow \Omega^*$$

such that for each child $\omega \in \Omega^c$, $m(\omega)$ refers to the mother of $\omega$ in $\Omega^*$. Conversely, for each women $\omega \in \Omega^*$, $m^{-1}(\{\omega\})$ is the set of her children in $\Omega^c$. Now let $\bar{\Omega}^c$ denote a simple random sample from $\Omega^c$. This induces a random sample from $\Omega^*$, namely

$$\bar{\Omega}^* := \{\omega \in \Omega^* \,|\, \text{there is an } \omega' \in \bar{\Omega}^c \text{ with } m(\omega') = \omega\}$$

But $\bar{\Omega}^*$ is not a simple random sample from $\Omega^*$ because women with more children are more frequent in $\bar{\Omega}^*$ than in $\Omega^*$. This should be taken into account when estimating $\tilde{r}^*(\tau)$ from information provided by the children in the sample $\bar{\Omega}^c$.

**2.** A further problem concerns the temporal nature of the membership of women in $\Omega^*$. As has been discussed in the previous section, given the data generating process assumed in our simulation model, a women belongs to $\Omega^*$ as soon as she has given birth to her first child. The definition of the rates $\tilde{r}^*$ makes the condition explicit by including the variable $C^*$ referring to the age at the first birth. Therefore, if $\omega$ is any member of the sample $\bar{\Omega}^c$, one should not condition on the mother's age when giving birth to $\omega$, but on the age of her first child-bearing. To illustrate, we use the data from the simulation model and compare two fictitious samples: $\bar{\Omega}^c_1$ contains all first-born children from $\Omega^c$, and $\bar{\Omega}^c_2$ contains all last-born children from $\Omega^c$. Of course, both samples provide the same information about the life length of women in $\Omega^*$. But there are now different ways to select truncation times. If we condition on the age of the mothers when giving birth to the children in the samples, we get the results shown in Figure 9.2-4. Obviously, conditioning on the mother's age when giving birth to her last child would result in an extremely biased estimate.



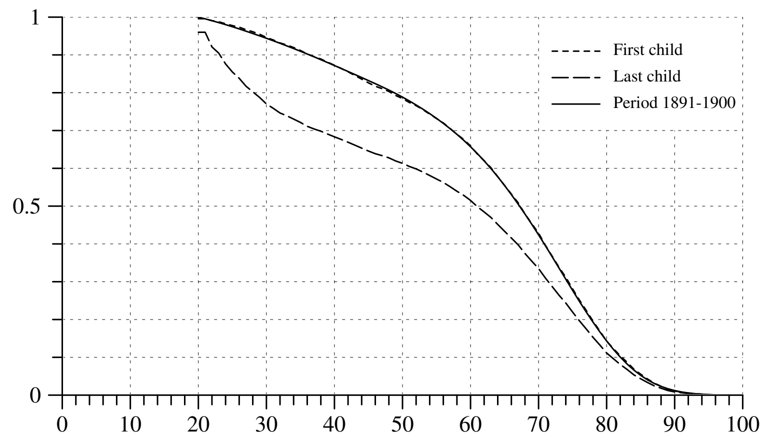



**Fig. 9.2-4** Comparison of conditional survivor functions calculated from two different samples from the simulated data set.

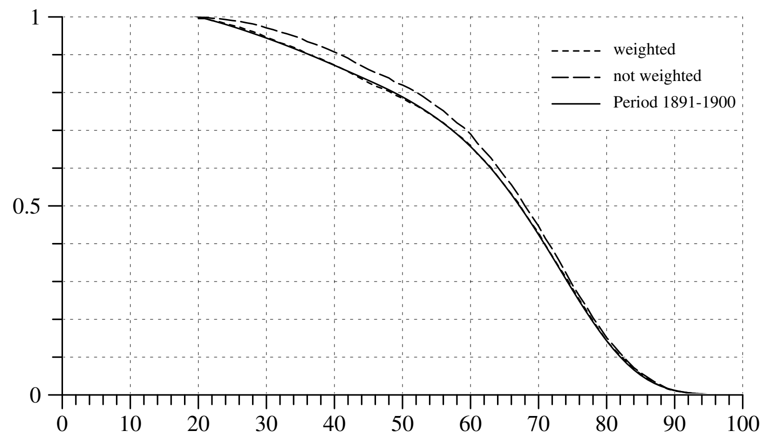



**Fig. 9.2-5** Comparison of conditional survivor functions estimated with, and without, weights from the simulated data set.

**3.** In order to avoid this mistake, we should, ideally, have values of the following variable:

$$(T^*_c, C^*_c, N^*_c) : \Omega^c \longrightarrow \tilde{\mathcal{T}} \times \tilde{\mathcal{T}} \times \{1, 2, 3, \ldots\}$$

where $T^*_c(\omega)$ provides information about the (possibly censored) life length of $\omega$'s mother, $C^*_c(\omega)$ provides information about the mother's age at *first* child-bearing, and $N^*_c(\omega)$ counts the mother's number of children. Since $N^*_c$ will be used to provide weights for the observations in the sample $\bar{\Omega}^c$, this should be the number of children surviving up to the time when the sample is drawn. Now, assuming that this information is available from a simple random sample $\bar{\Omega}^c$, the rates $\tilde{r}^*$ can be estimated in the following way:[7]

$$\tilde{r}^*(\tau) \approx_e \tilde{r}^*_w(\tau) := \frac{\sum_{\omega \in \bar{\Omega}^c} \frac{1}{N^*_c(\omega)} I[T^*_c = \tau, C^*_c \leq \tau](\omega)}{\sum_{\omega \in \bar{\Omega}^c} \frac{1}{N^*_c(\omega)} I[T^*_c \geq \tau, C^*_c \leq \tau](\omega)}$$

To illustrate, we use again data from the simulation model. Figure 9.2-5 compares conditional survivor functions calculated from estimated rates $\tilde{r}^*_w(\tau)$ and from analogously defined rates where the weights are dropped.[8] The figure clearly indicates that one should use the weights $1/N^*_c$ if this information is available.

**4.** However, this information might not be available and it is important, therefore, that there is also another and simpler way to arrive at reasonable estimates. In order to explain this possibility consider the risk set

$$\Omega^*_\tau = \{\omega \in \Omega^* \,|\, T^*(\omega) \geq \tau, C^*(\omega) \leq \tau\}$$

at $\tau$. The death rates to be estimated can then be written as

$$\tilde{r}^*(\tau) = \frac{|\{\omega \in \Omega^*_\tau \,|\, T^*(\omega) = \tau\}|}{|\Omega^*_\tau|}$$

By assumption, these rates do not depend on the number of children born of members of $\Omega^*_\tau$ until $\tau$, and also do not depend on the children's birth dates. To make this explicit, we may partition the risk sets into subsets according to the number of children born until $\tau$. Let $K^*_\tau(\omega)$ denote the number of children born of $\omega$ until $\tau$. Each risk set $\Omega^*_\tau$ may then be written as a union of subsets

$$\Omega^*_{\tau,k} := \{\omega \in \Omega^*_\tau \,|\, K^*_\tau(\omega) = k\}$$

taken over all possible values of $k$. Furthermore, we can define death rates for these subsets,

$$\tilde{r}^*_k(\tau) := \frac{|\{\omega \in \Omega^*_{\tau,k} \,|\, T^*(\omega) = \tau\}|}{|\Omega^*_{\tau,k}|}$$

[7] The notation uses *indicator variables*. If $X$ is any statistical variable with a possible value $\tilde{x}$, then

$$I[X = \tilde{x}](\omega) := \begin{cases} 1 & \text{if } X(\omega) = \tilde{x} \\ 0 & \text{otherwise} \end{cases}$$

[8] In the calculation we have used all observations from $\Omega^c$, but basically the same differences would result from a simple random sample from $\Omega^c$.



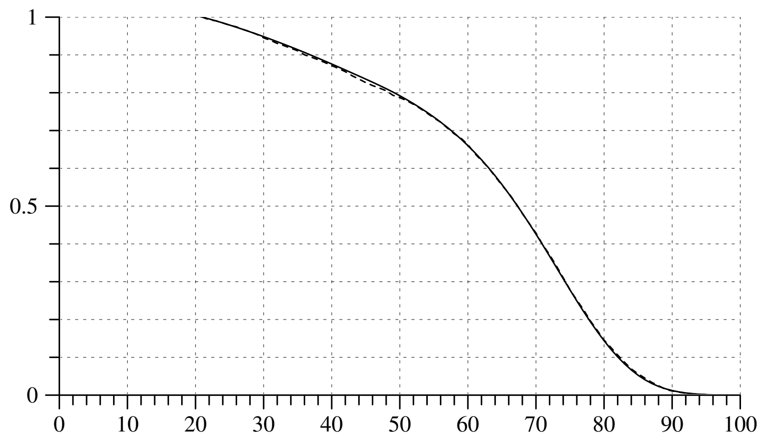



**Fig. 9.2-6** Conditional survivor function estimated from the rates $\tilde{r}^*_c(\tau)$, compared with a conditional survivor function from the 1891-1900 period life table.

However, by assumption these rates are all (approximately) identical with the death rate $\tilde{r}^*(\tau)$. Consequently, we do not need weights when we only use information from children born until $\tau$. Instead, we can directly refer to the sets of children born of women in $\Omega^*_{\tau,k}$ which can be defined by

$$\Omega^c_{\tau,k} := m^{-1}(\Omega^*_{\tau,k})$$

The death rates $\tilde{r}^*_k(\tau)$ may then be written as

$$\tilde{r}^*_k(\tau) \approx_e \frac{|\{\omega \in \Omega^c_{\tau,k} \,|\, T^*_c(\omega) = \tau\}|}{|\Omega^c_{\tau,k}|}$$

and, since these rates are approximately identical across the subsets, we might finally write

$$\tilde{r}^*(\tau) \approx_e \tilde{r}^*_c(\tau) := \frac{|\{\omega \in \Omega^c \,|\, T^*_c(\omega) = \tau, S^*_c(\omega) \leq \tau\}|}{|\{\omega \in \Omega^c \,|\, T^*_c(\omega) \geq \tau, S^*_c(\omega) \leq \tau\}|} \qquad (9.2.1)$$

where now $S^*_c(\omega)$ is the age of $\omega$'s mother at the birth of $\omega$. Notice that this approach does not require any weights and also requires no information about the mothers age at her first child-bearing.

**5.** To illustrate the argument we use again data from the simulation model. We take into account all children in $\Omega^c$ but, for the calculation of the rates $\tilde{r}^*_c(\tau)$ only use information from children born not later than $\tau$. Of course, this simply means to use all information from $\Omega^c$ and, for each $\omega \in \Omega^c$, treat the observation about $\omega$'s mother as left truncated at $S^*_c(\omega)$.[9] Figure 9.2-6

[9] One can use, therefore, any standard Kaplan-Meier procedure that allows for left truncated data. We have used TDA's `dple` procedure.

shows the conditional survivor function calculated from the rates $\tilde{r}^* c(\tau)$. This function obviously agrees quite well with the 1891-1900 period life table that was used to generate the data. Of course, the result would be basically the same if we had used as simple random sample from $\Omega^c$.

### 9.2.4 Retrospective Surveys

**1.** In the previous section we assumed that we have data from a simple random sample from the complete set of children, $\Omega^c$. However, our data actually result from a retrospective survey performed in some specific year, say $t$, and we therefore have to take into account that not all members of $\Omega^c$ survive until $t$. Fortunately, the approach to estimate $\delta^f_\tau$ via the rates $\tilde{r}^*_c(\tau)$ that was discussed in the previous section can also be applied to a retrospective sample if we make the additional assumption that children's life lengths are independent of their mother's life length.[10] To explain the argument, let $T^c$ denote the life length of children in the reference set $\Omega^c$. On a historical time axis, if mothers are born in the year $t_0$, each child $\omega \in \Omega^c$ survives until $t_0 + S^*_c(\omega) + T^c(\omega)$ (as already introduced, $S^*_c(\omega)$ is the age of the mother when $\omega$ was born). The set of children who survive at least until the year $t$ is therefore given by

$$\Omega^c[t] := \{\omega \in \Omega^c \,|\, t_0 + S^*_c(\omega) + T^c(\omega) \geq t\}$$

In the simulation model introduced in Section 9.2.1 we assumed $t_0 = 1900$. Based on this assumption, Figure 9.2-1 shows the survival of children in historical time.

**2.** Now assume a retrospective survey performed in the year $t$. The sample is then drawn from the reference set $\Omega^c[t]$. Following the approach discussed in the previous section, we can calculate rates

$$\tilde{r}^*_{c,t}(\tau) := \frac{|\{\omega \in \Omega^c[t] \,|\, T^*_c(\omega) = \tau, S^*_c(\omega) \leq \tau\}|}{|\{\omega \in \Omega^c[t] \,|\, T^*_c(\omega) \geq \tau, S^*_c(\omega) \leq \tau\}|}$$

which are defined analogously to the rates $\tilde{r}^*_c(\tau)$ introduced in (9.2.1). In order to see that the rates $\tilde{r}^*_{c,t}(\tau)$ are reasonable estimates of the rates $\tilde{r}^*_c(\tau)$, their definition might be written in the following way:

$$\tilde{r}^*_{c,t}(\tau) = \frac{|\{\omega \in \Omega^c \,|\, T^*_c(\omega) = \tau, S^*_c(\omega) \leq \tau, S^*_c(\omega) + T^c(\omega) \geq t - t_0\}|}{|\{\omega \in \Omega^c \,|\, T^*_c(\omega) \geq \tau, S^*_c(\omega) \leq \tau, S^*_c(\omega) + T^c(\omega) \geq t - t_0\}|}$$

The further argument proceeds in terms of conditional frequencies. Using

[10] This assumption, already built into the simulation model in Section 9.2.1, is probably not completely true. However, for the moment we will base our argument on this assumption.



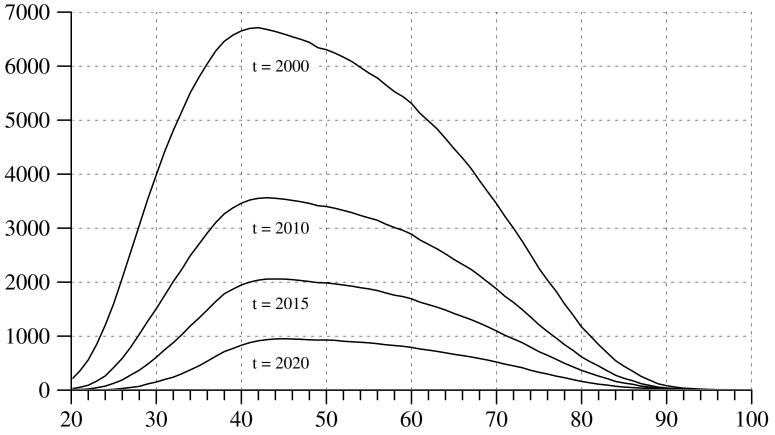



**Fig. 9.2-7** Sizes of the risk sets $\Omega^c_\tau[t]$, depending on $\tau$, calculated from four retrospective surveys of the simulated data set in the years $t = 2000, 2010, 2015$, and $2020$.

an abbreviated notation, we may write:

$$\begin{aligned}
\tilde{r}^*_{c,t}(\tau) &= \frac{\mathrm{P}(T^*_c = \tau, S^*_c \leq \tau, S^*_c + T^c \geq t - t_0)}{\mathrm{P}(T^*_c \geq \tau, S^*_c \leq \tau, S^*_c + T^c \geq t - t_0)} \\
&= \frac{\mathrm{P}(S^*_c + T^c \geq t - t_0 \,|\, T^*_c = \tau, S^*_c \leq \tau)\, \mathrm{P}(T^*_c = \tau, S^*_c \leq \tau)}{\mathrm{P}(S^*_c + T^c \geq t - t_0 \,|\, T^*_c \geq \tau, S^*_c \leq \tau)\, \mathrm{P}(T^*_c \geq \tau, S^*_c \leq \tau)} \\
&= \tilde{r}^*_c(\tau)\, \frac{\mathrm{P}(S^*_c + T^c \geq t - t_0 \,|\, T^*_c = \tau, S^*_c \leq \tau)}{\mathrm{P}(S^*_c + T^c \geq t - t_0 \,|\, T^*_c \geq \tau, S^*_c \leq \tau)}
\end{aligned}$$

Now, given the assumption mentioned at the beginning, that, conditional on $S^*_c \leq \tau$, the survival of children does not depend on the survival of their mothers, the last term on the right-hand side becomes approximately

$$\frac{\mathrm{P}(S^*_c + T^c \geq t - t_0 \,|\, S^*_c \leq \tau)}{\mathrm{P}(S^*_c + T^c \geq t - t_0 \,|\, S^*_c \leq \tau)}$$

and may be omitted.

**3.** There is, however, a further difficulty resulting from retrospective surveys. The later the year $t$ in which the survey is performed, the smaller is the number of children who might participate in the survey, and consequently also the risk set to be used for the estimation of the death rates $\tilde{r}^*_{c,t}$ becomes smaller. This is shown in Figure 9.2-7 which is based on the data from our simulation model. Shown are the functions

$$\tau \longrightarrow \Omega^c_\tau[t] := \{\omega \in \Omega^c[t] \,|\, T^*_c(\omega) \geq \tau, S^*_c(\omega) \leq \tau\}$$

as they result from four fictitious retrospective surveys performed in the

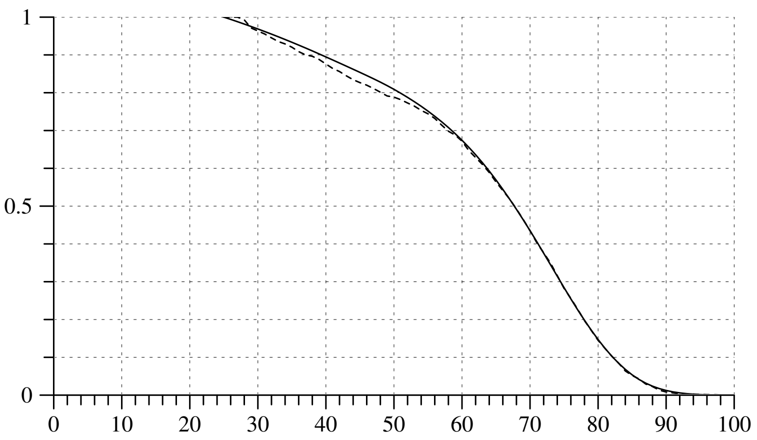



**Fig. 9.2-8** Conditional survivor functions, estimated from $\Omega^c[2010]$ (dotted line) and calculated from the 1891–1900 period life table (solid line), both beginning at $a^* = 25$.

years $t = 2000, 2010, 2015$, and 2020. The possible problem concerns estimation with left truncated data. Contrary to the standard Kaplan-Meier procedure with right censored data only, the risk set is very small at the beginning and may not allow reliable estimates of the death rates. Due to the cumulative nature of the calculation of survivor functions from these rates, any imprecisions introduced at the beginning will then propagate to values of the survivor function at later ages. To illustrate, we use the simulated data set and perform a retrospective survey in the year $t = 2010$. We assume that all children who survive this year, that is about 20 % of the 11407 children in $\Omega^c$, participate in the survey and provide information about their mothers. Nevertheless, we can only begin to estimate a conditional survivor function at $a^* = 25$ as shown in Figure 9.2-8.

## 9.3 Inferences from the GLHS and SOEP Data

We now use the methods discussed in the previous sections to draw some inferences from the GLHS and SOEP data. We begin with a brief data description, then estimate survivor functions, and finally show plots of the death rates.

### 9.3.1 Description of the Data

**1.** We briefly describe the available data. The basic figures are shown in Table 9.3-1. From the GLHS we use all studies which are currently available in the *Zentralarchiv für empirische Sozialforschung* (see Section 14.1).



a) The first study was LV I. The 2171 respondents were born in the periods 1929 – 31, 1939 – 41, and 1949 – 51; the interviews were conducted in the years 1981 – 83. In 2120 cases respondents were able to provide a valid birth year of their mother. Of these mothers, 732 died before the interview date, 1386 were still alive, and for two mothers we have no information. Complete information is therefore available for 2118 mothers. In 8 cases this information is inconsistent or implausible, for example, the birth year of the respondent is greater than the death year of the mother.[11] If we exclude these cases there finally remain 2110 cases in which we know: the birth year of the mother, whether she died before the interview date, and, if she died, also the death year. Similarly, we get valid information for 2044 fathers.

b) The second study was LV II and involved respondents born in the years 1919 – 21. This study was conducted in two parts: LV IIA with interviews during 1985 – 86, and LV IIT with interviews during 1987 – 88. In the same way as explained for LV I we get valid information about the lifetimes of $387 + 956 = 1343$ mothers and $382 + 943 = 1325$ fathers.

c) The third study was LV III and involved respondents born in the periods 1954 – 56 and 1959 – 61. From this study we get valid information about 1954 mothers and 1911 fathers.

**2.** Comparable information is available from the third wave of the SOEP conducted in 1986. All members of subsample A of the SOEP were asked to provide information about birth years of their parents, whether parents died before the interview date and, if they died, about death years. In order to get data comparable with the GLHS, we selected only persons with a German citizenship. As shown in Table 9.3-1, there are 8021 persons from which we get valid information about 7746 mothers and 7614 fathers. Taking the GLHS and SOEP data together, we finally have valid information about 13153 mothers and 12894 fathers.

**3.** We prepared two data files for further analysis, one for mothers and the other one for fathers. Both files contain values of four variables:

$B^f$ := birth year of the mother

$P^f$ := birth year of the child (respondent)

$E^f$ := 1 if mother died before the interview date, 0 otherwise

$D^f$ := mother's death year, or the year of the interview, depending on the value of $E^f$

Variables in the data file for fathers are defined accordingly and will be denoted by $B^m, P^m, E^m$, and $D^m$.

[11] In addition to inconsistent cases we also exclude cases with a life length which is greater than 105 years. For women we also require that the age at which the women gave birth to her child (the respondent) is not greater than 51 years.

**Table 9.3-1** Information about lifetimes of mothers and fathers which is available in the GLHS and SOEP data sets.

| | LV I | LV IIA | LV IIT | LV III | SOEP |
|---|---|---|---|---|---|
| Interview dates | 1981-83 | 1985-86 | 1987-88 | 1989 | 1986 |
| Respondents | 2171 | 407 | 1005 | 2008 | 8021 |
| Mothers | | | | | |
| - valid birth year | 2120 | 390 | 962 | 1954 | 7819 |
| - still alive | 1386 | 24 | 43 | 1766 | 4872 |
| - known death year | 732 | 366 | 919 | 188 | 2911 |
| - no information | 2 | 0 | 0 | 0 | 36 |
| - complete information | 2118 | 390 | 962 | 1954 | 7783 |
| - dismissed | 8 | 3 | 6 | 0 | 37 |
| - remaining cases | 2110 | 387 | 956 | 1954 | 7746 |
| - still alive | 1385 | 24 | 43 | 1766 | 4854 |
| - died | 725 | 363 | 913 | 188 | 2892 |
| Fathers | | | | | |
| - valid birth year | 2062 | 386 | 955 | 1916 | 7699 |
| - still alive | 909 | 1 | 8 | 1460 | 3586 |
| - known death year | 1150 | 384 | 945 | 451 | 4053 |
| - no information | 3 | 1 | 2 | 5 | 60 |
| - complete information | 2059 | 385 | 953 | 1911 | 7639 |
| - dismissed | 15 | 3 | 10 | 0 | 25 |
| - remaining cases | 2044 | 382 | 943 | 1911 | 7614 |
| - still alive | 909 | 1 | 8 | 1460 | 3577 |
| - died | 1135 | 381 | 935 | 451 | 4037 |

## 9.3.2 Survivor Functions of Parents

**1.** We now apply the method discussed in the previous section to the data introduced in Section 9.3.1. Since we already know that mortality conditions have substantially changed during the last 100 years, we consider birth cohorts as defined in Table 9.3-2.[12] To develop the argument we consider variables $\hat{T}_c^f$ and $\hat{T}_c^m$ representing the life length of women and men who belong to a birth cohort indexed by $c$. Derivable from the variables introduced at the end of Section 9.3.1, available data are given by variables

$$C_c^f := P_c^f - B_c^f \quad \text{and} \quad C_c^m := P_c^m - B_c^m$$

which record the ages at which persons belonging to birth cohort $c$ became mothers or fathers, and variables

$$T_c^f := D_c^f - B_c^f \quad \text{and} \quad T_c^m := D_c^m - B_c^m$$

which record the knowledge about the life length. If $E_c^f(\omega) = 1$, $T_c^f(\omega) = \hat{T}_c^f(\omega)$ is the known life length of $\omega$; otherwise, the information is censored

[12]Compared with the figures in Table 9.3-1 the total number of cases is slightly smaller because persons born before 1870 or after 1939 have been omitted.



**Table 9.3-2** Definition of birth cohorts used in the estimation of survivor functions.

| | | Mothers | | | Fathers | | |
|---|---|---|---|---|---|---|---|
| Cohort | Birth years | died | alive | total | died | alive | total |
| C1 | 1870 – 1879 | 271 | 0 | 271 | 528 | 0 | 528 |
| C2 | 1880 – 1889 | 1064 | 10 | 1074 | 1393 | 12 | 1405 |
| C3 | 1890 – 1899 | 1698 | 170 | 1868 | 1591 | 101 | 1692 |
| C4 | 1900 – 1909 | 1123 | 954 | 2077 | 1685 | 600 | 2285 |
| C5 | 1910 – 1919 | 438 | 1456 | 1894 | 907 | 1011 | 1918 |
| C6 | 1920 – 1929 | 272 | 2467 | 2739 | 464 | 2035 | 2499 |
| C7 | 1930 – 1939 | 123 | 2219 | 2342 | 196 | 1773 | 1969 |

and we only know that $\hat{T}_c^f(\omega) \geq T_c^f(\omega)$. For variables pertaining to men the interpretation is analogous.

**2.** To illustrate the calculations we refer to women belonging to birth cohort C4. The data are shown in Table 9.3-3. The column labeled (a) shows the risk sets. As discussed in the previous section, the risk set at age $\tau$ contains all women who did not die before $\tau$ and became a mother not later than $\tau$.[13] In this example, the youngest age for which we know of a child is 15; risk sets can therefore be calculated only for ages $\tau \geq \tau^* = 15$. The next column, labeled (b), shows the number of death events. Then follows column (d) providing the number of censored cases which are required to update the risk sets. As shown by the definition

$$\mathcal{R}^*(\tau) := \{\omega \,|\, T_c^f(\omega) \geq \tau, C_c^f(\omega) \leq \tau\}$$

women belong to a risk set only until the maximal value of $T_c^f$, that is, until a death event occurs or until the interview date (of their children).

**3.** The information in Table 9.3-3 suffices to calculate death rates. For example, $r^*(20) = 1/100$ and $r^*(80) = 33/507$. These rates can then be used to estimate the survivor function

$$G^*(\tau) = g_{\tau^*} \prod_{j=\tau^*}^{\tau-1} (1 - r^*(j))$$

Of course, we do not know $g_{\tau^*}$, that is, the proportion of women who survived age 14. So we can only estimate a conditional survivor function

$$G[\hat{T}_c^f | \hat{T}_c^f \geq \tau^*] \approx_e \prod_{j=\tau^*}^{\tau-1} (1 - r^*(j))$$

[13]Of course, from our data we do not know when women actually gave birth to a first child. Whether this has implications for the quality of the estimates will be discussed in a later section.

**Table 9.3-3** Mortality data for mothers belonging to birth cohort C4 in the merged GLHS and SOEP data set.

(a) Size of risk set at age $\tau$.
(b) Number of deaths at age $\tau$.
(c) Number of censored cases at age $\tau$.
(d) Values of the conditional survivor function at age $\tau$.

| $\tau$ | (a) | (b) | (c) | (d) | $\tau$ | (a) | (b) | (c) | (d) |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 1 | 0 | 0 | 1.000 | 52 | 1901 | 4 | 0 | 0.878 |
| 16 | 2 | 0 | 0 | 1.000 | 53 | 1897 | 15 | 0 | 0.876 |
| 17 | 3 | 0 | 0 | 1.000 | 54 | 1882 | 13 | 0 | 0.869 |
| 18 | 18 | 0 | 0 | 1.000 | 55 | 1869 | 21 | 0 | 0.863 |
| 19 | 45 | 0 | 0 | 1.000 | 56 | 1848 | 7 | 0 | 0.854 |
| 20 | 100 | 1 | 0 | 1.000 | 57 | 1841 | 17 | 0 | 0.850 |
| 21 | 183 | 1 | 0 | 0.990 | 58 | 1824 | 14 | 0 | 0.843 |
| 22 | 266 | 2 | 0 | 0.985 | 59 | 1810 | 18 | 0 | 0.836 |
| 23 | 347 | 0 | 0 | 0.977 | 60 | 1792 | 20 | 0 | 0.828 |
| 24 | 435 | 2 | 0 | 0.977 | 61 | 1772 | 19 | 0 | 0.818 |
| 25 | 556 | 1 | 0 | 0.973 | 62 | 1753 | 15 | 0 | 0.810 |
| 26 | 690 | 3 | 0 | 0.971 | 63 | 1738 | 23 | 0 | 0.803 |
| 27 | 781 | 4 | 0 | 0.967 | 64 | 1715 | 18 | 0 | 0.792 |
| 28 | 928 | 1 | 0 | 0.962 | 65 | 1697 | 33 | 0 | 0.784 |
| 29 | 1075 | 3 | 0 | 0.961 | 66 | 1664 | 23 | 0 | 0.769 |
| 30 | 1217 | 3 | 0 | 0.958 | 67 | 1641 | 36 | 0 | 0.758 |
| 31 | 1327 | 2 | 0 | 0.956 | 68 | 1605 | 32 | 0 | 0.741 |
| 32 | 1427 | 5 | 0 | 0.954 | 69 | 1573 | 24 | 0 | 0.727 |
| 33 | 1512 | 9 | 0 | 0.951 | 70 | 1549 | 41 | 0 | 0.715 |
| 34 | 1597 | 5 | 0 | 0.945 | 71 | 1508 | 39 | 0 | 0.697 |
| 35 | 1677 | 6 | 0 | 0.942 | 72 | 1469 | 54 | 32 | 0.679 |
| 36 | 1740 | 5 | 0 | 0.939 | 73 | 1383 | 53 | 53 | 0.654 |
| 37 | 1800 | 5 | 0 | 0.936 | 74 | 1277 | 56 | 57 | 0.629 |
| 38 | 1864 | 9 | 0 | 0.934 | 75 | 1164 | 53 | 43 | 0.601 |
| 39 | 1903 | 13 | 0 | 0.929 | 76 | 1068 | 55 | 30 | 0.574 |
| 40 | 1929 | 12 | 0 | 0.923 | 77 | 983 | 42 | 125 | 0.544 |
| 41 | 1945 | 2 | 0 | 0.917 | 78 | 816 | 57 | 96 | 0.521 |
| 42 | 1952 | 5 | 0 | 0.916 | 79 | 663 | 42 | 114 | 0.484 |
| 43 | 1957 | 7 | 0 | 0.914 | 80 | 507 | 33 | 85 | 0.454 |
| 44 | 1957 | 5 | 0 | 0.910 | 81 | 389 | 20 | 78 | 0.424 |
| 45 | 1956 | 13 | 0 | 0.908 | 82 | 291 | 12 | 70 | 0.402 |
| 46 | 1947 | 7 | 0 | 0.902 | 83 | 209 | 21 | 48 | 0.386 |
| 47 | 1943 | 10 | 0 | 0.899 | 84 | 140 | 12 | 34 | 0.347 |
| 48 | 1935 | 8 | 0 | 0.894 | 85 | 94 | 5 | 41 | 0.317 |
| 49 | 1928 | 8 | 0 | 0.891 | 86 | 48 | 0 | 39 | 0.300 |
| 50 | 1920 | 7 | 0 | 0.887 | 87 | 9 | 0 | 5 | 0.300 |
| 51 | 1913 | 12 | 0 | 0.884 | 88 | 4 | 0 | 4 | 0.300 |



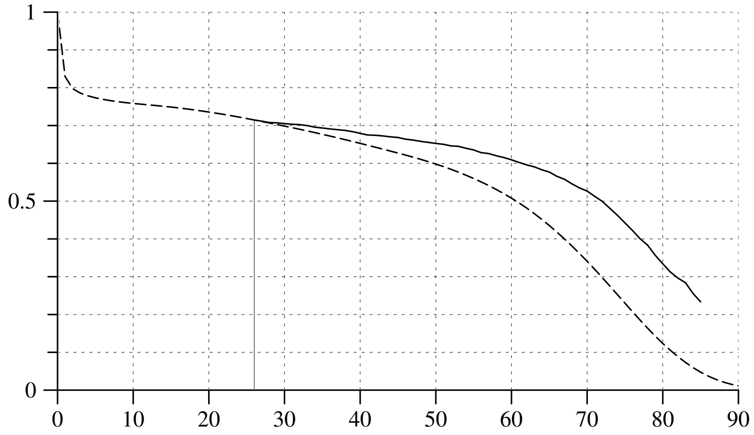



**Fig. 9.3-1** Female survivor function of the German period life table 1901/10 (dotted line) and conditional survivor function from Table 9.3-3.

which is shown in the last column of Table 9.3-3 labeled (d).

**4.** Since this approach to estimate a conditional survivor function depends on a previous estimation of rates, one should also consider the question whether these rates can be reliably estimated. Formally, one can begin at age $\tau^*$ which is 15 in our example. However, due to the small number of cases in the risk sets at ages under 20, one might question the reliability of these estimates. In fact, formally following the estimation procedure implies estimated death rates having a value of zero during ages from 15 to 19. But given our knowledge about mortality and life tables from other sources, these estimates will clearly be wrong. Moreover, the reliability of estimates of death rates not only depends on the size of the risk sets but also on the number of death events that can be observed. Therefore, regarding the data in Table 9.3-3, it might be sensible to begin an interpretation of estimated death rates only at some later age, for example, at age 26 or even later.

**5.** Conditional survivor functions can be represented graphically in two possible ways: The function can be plotted beginning at some age $\tau$ with arbitrary value $g_\tau$; or one can try to find some estimate of $g_\tau$ and then plot the conditional survivor function as part of a complete survivor function. In any case one needs to decide where to start the plotting. For our example we begin at age 26 and estimate $g_{26}$ from the female survivor function of the German period life table for the period 1901–10 (see Table 7.4-3 in Section 7.4.2). Beginning at age 26, we therefore multiply all values of column (d) in Table 9.3-3 with the factor

$$g_{26} = \frac{0.71463}{0.971} = 0.736$$

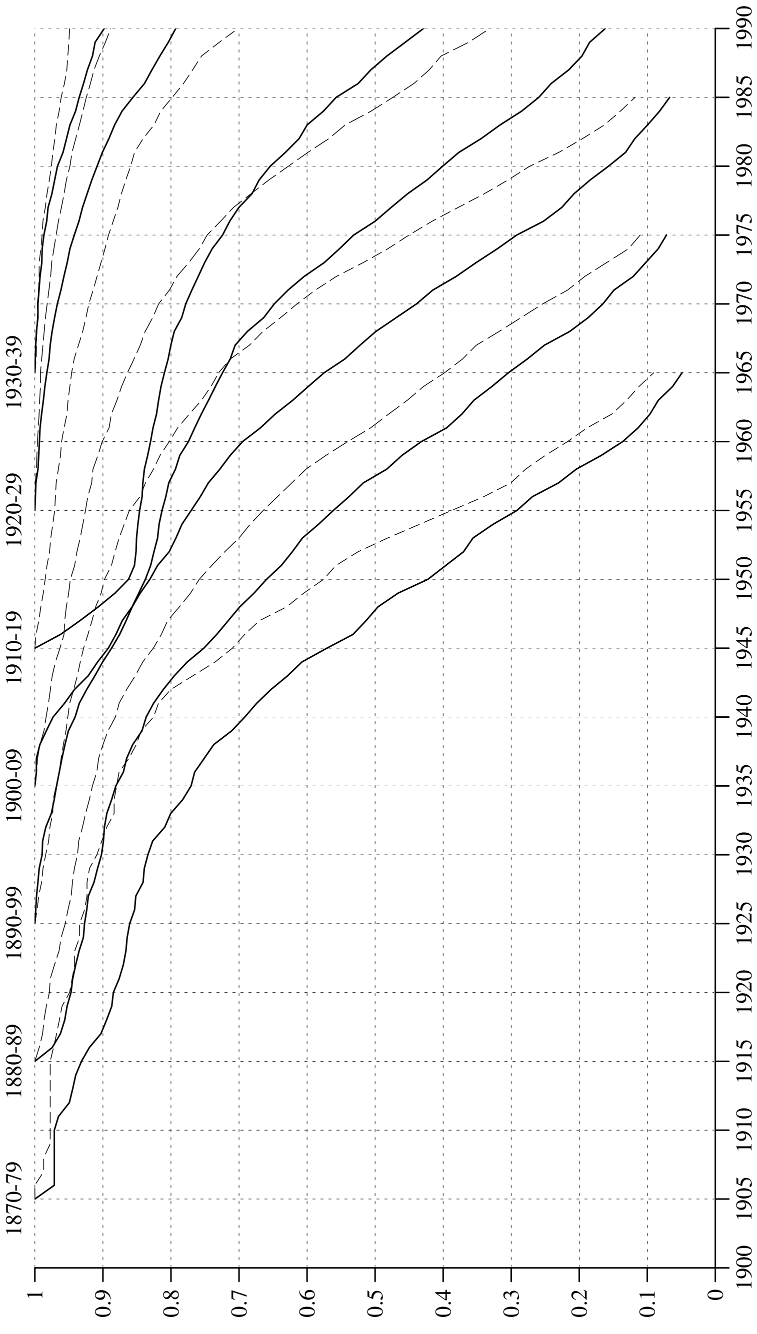



**Fig. 9.3-2** Conditional survivor functions, beginning at age 30, for men (solid lines) and women (dotted lines) belonging to specified birth cohorts.



The result is shown in Figure 9.3-1. The dotted line represents the female survivor function from the 1901–10 period life table; the solid line shows the adjusted conditional survivor function from Table 9.3-3. By definition, values are identical at age 26. The different development of both curves reflects the reduction of death rates that occurred during the period from about 1930 until the end of the century. So we might use the latest 1986–88 period life table for a further comparison. As can be estimated from Table 9.3-3, the death rate at age 80 is about 0.065. A corresponding estimate from the 1986–88 period life table is 0.066.[14] One should note, however, that values of rates calculated from sample data for single years often show high fluctuations and it might be better, therefore, to use mean values for larger age classes.

**6.** In the same way as has been discussed for women belonging to birth cohort C4 (1900–1909) one can estimate conditional survivor functions for all birth cohorts distinguished in Table 9.3-2. Results are shown in Figure 9.3-2. To allow for a comparison, all survivor functions are drawn conditional on $\tau^* = 30$. The placement onto a historical time axis was done by using the centers of the birth cohort intervals. For example, the value of the conditional survivor function for birth cohort C1 at age 30 is shown in the year $1875 + 30 = 1905$. The changing shapes of the survivor functions not only reflect a general tendency of decreasing death rates, both for men and women. Also clearly seen are period effects, especially the substantial increases of male death rates during the years of World War II. This seems not to be the case with regard to female death rates. An interpretation should consider, however, that the occurrence of death events might not be independent for mothers and their children, in particular during war time. The death events of mothers might therefore be substantially underrepresented in our data set.

### 9.3.3 Visualization of Death Rates

**1.** In order to investigate period effects it is often preferable to directly plot the rates from which (conditional) survivor functions are derived. The only drawback is that rates calculated from small samples are often highly fluctuating. As an example we refer to death rates of men belonging to birth cohort C5 (1910–1919). The solid line in Figure 9.3-3 shows the death rates as directly calculated from the data, that is, for each year of age, the number of deaths divided by the number of persons in the risk set. There obviously are big fluctuations. One should therefore apply some kind of smoothing procedure to provide a better view of the general shape of the rate function.

**2.** Many such smoothing procedures have been proposed in the literature.

---

[14]Calculated from the data in Table 7.4-4 in Section 7.4.2.

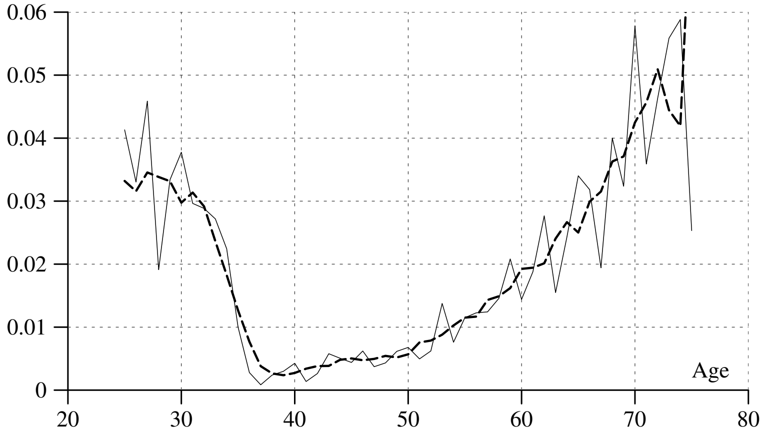



**Fig. 9.3-3** Raw values (solid line) and smoothed values (dotted line) of death rates of men belonging to birth cohort C5 (1910 – 1919).

In the present context, smoothing will only serve to visualize rate functions. It might therefore suffice to simply use moving averages. Given a series of values $r_\tau$, for $\tau = \tau_1, \ldots, \tau_n$, each value is then substituted by a mean of neighboring values. If the number of neighbors is denoted by $k$, the smoothed values are calculated as

$$r_\tau^{(k)} := \frac{1}{2k+1} \sum_{j=\tau-k}^{\tau+k} r_j$$

At both ends of the series only the actually available values are taken into account.[15] Choosing $k = 2$, this procedure was used to calculate values for the dotted line in Figure 9.3-3. It is seen how the smoothing removes the fluctuations but preserves the global shape of the rate function.

**3.** We now compare the death rates of men belonging to birth cohorts C1, ..., C6. The rate functions are shown in Figure 9.3-4 and placed onto a historical time axis. To support visibility, the rate functions are smoothed with the procedure just described (again, $k = 2$). Compared with the survivor functions shown in Figure 9.3-2, the rate functions provide a much better view of the impact of World War II.

[15]The complete formula may then be written as follows:

$$r_\tau^{(k)} := \frac{1}{\min\{\tau_n, \tau+k\} - \max\{\tau_1, \tau-k\} + 1} \sum_{j=\max\{\tau_1,\tau-k\}}^{\min\{\tau_n,\tau+k\}} r_j$$

where $\tau_1$ and $\tau_n$ refer, respectively, to the first and last element of the series.



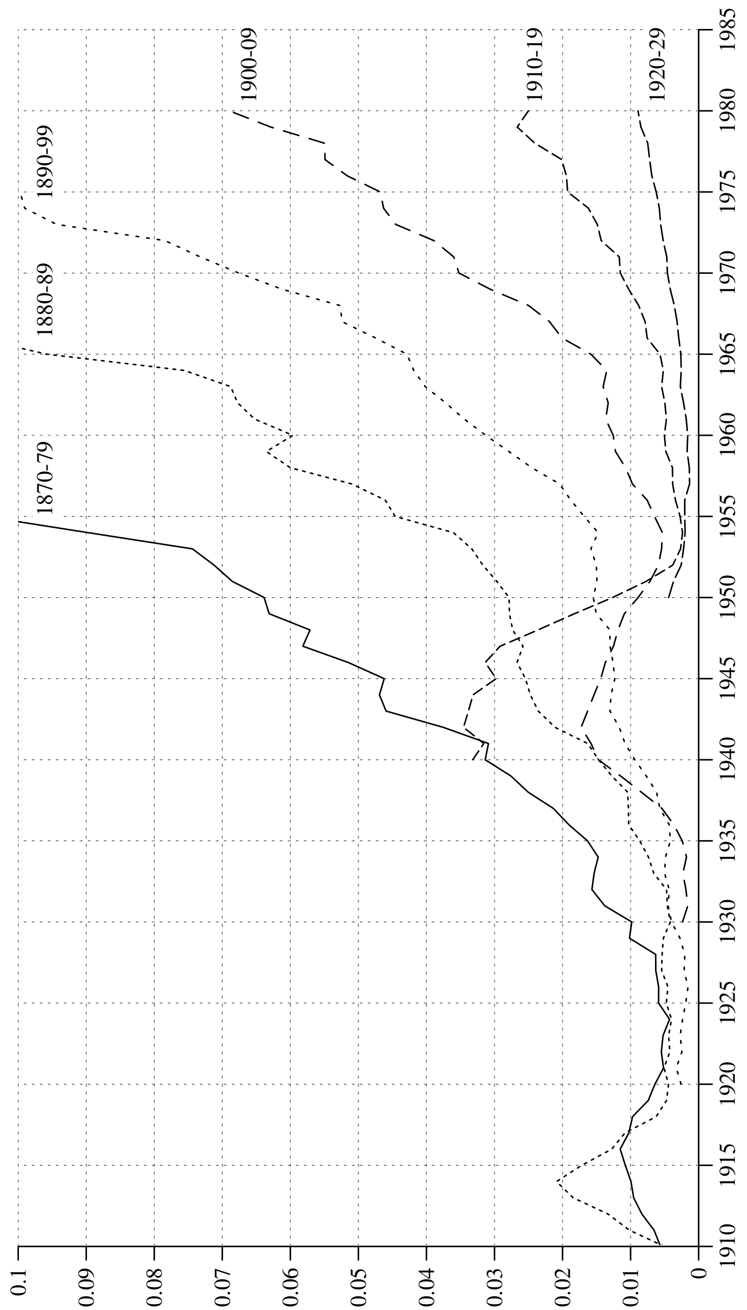



**Fig. 9.3-4** Smoothed death rates of men belonging to the indicated birth cohorts. (Moving averages with $k = 2$.)

# Chapter 10

# Parametric Mortality Curves

This chapter is not finished yet.

# Chapter 11

# Period and Cohort Birth Rates

We now leave the topic of mortality and turn to the complementary one: the birth of children. In this chapter, we begin with the standard approach that records the development of births in terms of rates. We then turn to a life course perspective which suggests to view birth events in the context of women's life courses.

## 11.1 Birth Rates

**1.** Demographers have invented a lot of measures to statistically record the fertility of a population.[1] An elementary measure parallels the crude mortality rate and is called *crude birth rate* [allgemeine Geburtenziffer].[2] It is defined as

$$\text{Crude birth rate} := \frac{b_t}{n_t} \quad \text{(multiplied by 1000)}$$

The numerator records the total number of births that occurred during the year $t$, and the denominator refers to the midyear population size in the same year. To calculate crude birth rates one can use the data from Tables 6.2-2 and 6.3-1 in Chapter 6. For example, referring to the territory of the former FRG, the crude birth rate in 1950 is $1000 \cdot 812.8/49989 = 16.26$. Figure 11.1-1 compares the development, until 1999, in the territories of the former FRG and the former GDR. The impression is that developments were quite similar until about 1973. Then, in the western part of Germany, the crude birth rate stabilized around a value of 10, while in the eastern part a temporary increase in fertility was ended by a sharp decline that began, roughly, at the time of the German unification.

**2.** Like crude mortality rates, crude birth rates neglect the age and sex composition of a population. Demographers therefore often calculate a *general birth rate* [allgemeine Geburtenrate[3]], also called a *general fertility*

[1] We mention that in the German demographic literature, and in publications of statistical offices, the literal translation of 'fertility' ['Fruchtbarkeit'] is considered obsolete; instead, one refers to birth events [Geburten] or newborn children [Geborene]. One should also notice that the terms 'fertility' and 'fecundity' are used somewhat differently in the literature. English texts most often use the term 'fertility' to refer to realized births, and the term 'fecundity' to refer to women's ability to bear children (see, e.g., Pressat 1972, p. 172, and Newell 1988, p. 35); some other authors use these words in an opposite meaning (see, e.g., Mueller 1993, p. 154).

[2] One also often finds the term 'crude fertility rate'.

[3] Also called 'allgemeine Fruchtbarkeitsziffer' in the older German literature, see, e.g., Statistisches Bundesamt 1985, p. 18.

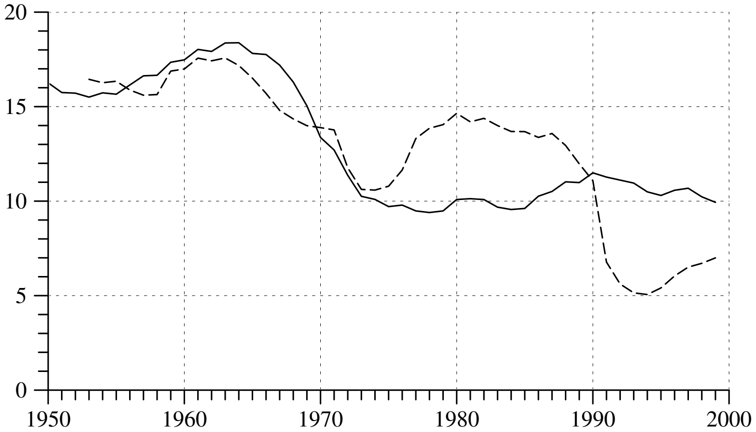



**Fig. 11.1-1** Crude birth rates in the territory of the former FRG (solid line) and the territory of the former GDR (dotted line); calculated from Tables 6.2-2 and 6.3-1 in Chapter 6.

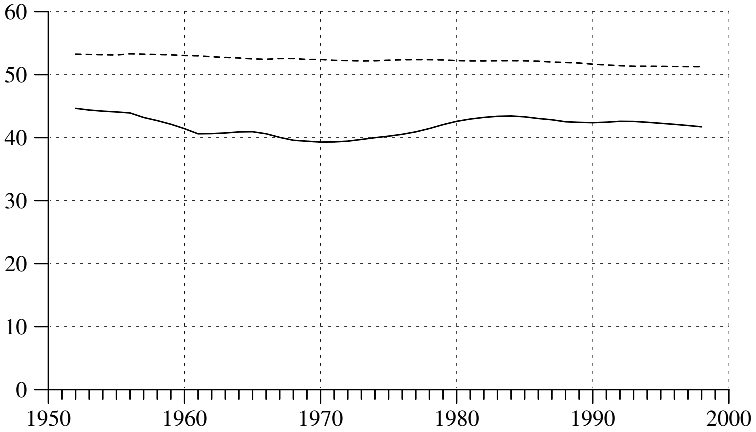



**Fig. 11.1-2** Proportion (in percent) of women in the midyear population (dotted line) and of women aged 15 to 45 in all women (solid line). Calculated from data in Segment 36 of the STATIS data base of the *Statistisches Bundesamt*.

*rate*, in which the number of births is related only to the number of women in childbearing ages. We will use the notation

$$\text{General birth rate} := \frac{b_t}{n_t^{f^*}} \quad \text{(multiplied by 1000)}$$

where the index, $f^*$, refers to women in the *reproductive period*, often assumed to be 15 to 45 years of age. However, there is no general agreement;



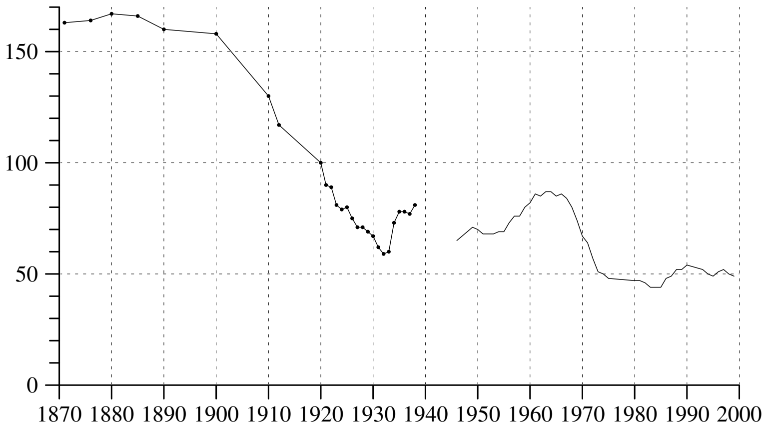



**Fig. 11.1-3** Development of the general birth rate in Germany since 1871. Data for the post-World War II period refer to the territory of the former FRG. Available data for the period before 1939 are indicated by dots. Source: Statistisches Bundesamt, Bevölkerung und Wirtschaft 1872 – 1972 (p. 109), and Fachserie 1, Reihe 1.

in publications of demographic data one also finds the periods 15 – 44, or 15 – 49, etc. We will use $\tau_a$ to denote the beginning and $\tau_b$ to denote the end of the reproductive period. As shown by the definition, the difference between a crude and a general birth rate depends on the sex ratio and the proportion of women in childbearing ages. How these proportions have changed over the years in the territory of the former FRG is shown in Figure 11.1-2. They obviously cannot explain the big changes that are visible in Figure 11.1-1.

**3.** In order to get an impression of long-term changes in childbearing both crude and general birth rates can be used. The long-term development of crude birth rates has been shown in Figure 6.3-2 in Section 6.3. A similar plot based on data on the general birth rate is shown in Figure 11.1-3. Both figures show that a long-term trend of declining fertility began in Germany roughly at the end of the nineteenth century.

**4.** A further concept is the *age-specific birth rate* [altersspezifische Geburtenziffer] which refers to women of a specific age. We will use the following definition:

$$\beta_{t,\tau} := \frac{b_{t,\tau}}{n_{t,\tau}^f}$$

The denominator refers to the midyear number of women in year $t$ aged $\tau$ (in completed years), and the numerator refers to the number of children born of these women during the year $t$. Notice that in publications from

**Table 11.1-1** Number of children born in Germany 1999 ($b_{1999,\tau}$) and number of women ($n^f_{1999,\tau}$ and $n^{f+}_{1999,\tau}$) classified according to women's age ($\tau$); also shown are age-specific birth rates ($\tilde{\beta}_{1999,\tau}$ and ($\tilde{\beta}_{1999,\tau}$). Source: Values of $b_{1999,\tau}$: Statistisches Jahrbuch 2001 (p. 71) and Segment 2070 in the STATIS data base; $n^f_{1999,\tau}$: Fachserie 1, Reihe 1, 1999 (pp. 64-65); values of $n^{f+}_{1999,\tau}$: unpublished material provided by the *Statistisches Bundesamt*.

| $\tau$ | $b_{1999,\tau}$ | $n^f_{1999,\tau}$ | $\tilde{\beta}_{1999,\tau}$ | $n^{f+}_{1999,\tau}$ | $\tilde{\beta}^+_{1999,\tau}$ |
|---|---|---|---|---|---|
| ≤14 | 80 | | | | |
| 15 | 341 | 438.4 | 0.78 | 436.782 | 0.78 |
| 16 | 1234 | 446.7 | 2.76 | 441.006 | 2.80 |
| 17 | 3085 | 453.2 | 6.81 | 452.610 | 6.82 |
| 18 | 6332 | 457.2 | 13.85 | 454.730 | 13.92 |
| 19 | 11158 | 451.3 | 24.72 | 460.706 | 24.22 |
| 20 | 15558 | 441.6 | 35.23 | 442.599 | 35.15 |
| 21 | 19693 | 441.7 | 44.58 | 440.781 | 44.68 |
| 22 | 24009 | 442.0 | 54.32 | 443.065 | 54.19 |
| 23 | 27326 | 436.7 | 62.57 | 440.361 | 62.05 |
| 24 | 30436 | 438.6 | 69.39 | 432.779 | 70.33 |
| 25 | 35493 | 449.3 | 79.00 | 444.718 | 79.81 |
| 26 | 39850 | 477.1 | 83.53 | 454.341 | 87.71 |
| 27 | 45348 | 528.2 | 85.85 | 500.610 | 90.59 |
| 28 | 52632 | 568.8 | 92.53 | 555.333 | 94.78 |
| 29 | 56566 | 604.4 | 93.59 | 582.220 | 97.16 |
| 30 | 60007 | 642.6 | 93.38 | 626.937 | 95.71 |
| 31 | 60093 | 668.1 | 89.95 | 657.849 | 91.35 |
| 32 | 56767 | 686.8 | 82.65 | 677.296 | 83.81 |
| 33 | 50623 | 697.5 | 72.58 | 696.136 | 72.72 |
| 34 | 43428 | 705.9 | 61.52 | 699.210 | 62.11 |
| 35 | 36185 | 711.7 | 50.84 | 713.016 | 50.75 |
| 36 | 28680 | 700.4 | 40.95 | 710.250 | 40.38 |
| 37 | 21055 | 687.5 | 30.63 | 690.981 | 30.47 |
| 38 | 15398 | 675.1 | 22.81 | 684.141 | 22.51 |
| 39 | 11165 | 656.2 | 17.01 | 666.236 | 16.76 |
| 40 | 7540 | 630.1 | 11.97 | 646.050 | 11.67 |
| 41 | 4627 | 608.9 | 7.60 | 614.752 | 7.53 |
| 42 | 2963 | 597.8 | 4.96 | 603.257 | 4.91 |
| 43 | 1619 | 584.7 | 2.77 | 592.163 | 2.73 |
| 44 | 789 | 575.7 | 1.37 | 577.973 | 1.37 |
| 45 | 342 | 566.9 | 0.60 | 573.468 | 0.60 |
| 46 | 163 | 561.7 | 0.29 | 560.591 | 0.29 |
| 47 | 58 | 558.4 | 0.10 | 563.369 | 0.10 |
| 48 | 48 | 556.0 | 0.09 | 553.593 | 0.09 |
| 49 | 25 | 548.5 | 0.05 | 558.612 | 0.04 |
| 50 | 12 | 517.2 | 0.02 | | |
| ≥51 | 16 | | | | |

official statistics age-specific birth rates are often multiplied by 1000, we will then use the notation $\tilde{\beta}_{t,\tau} := \beta_{t,\tau}\,1000$. Table 11.1-1 illustrates the calculation of these rates for the year 1999. We also mention that the *Statistisches Bundesamt* uses a slightly different definition and calculates women's age as the difference between the birth year of the women and the



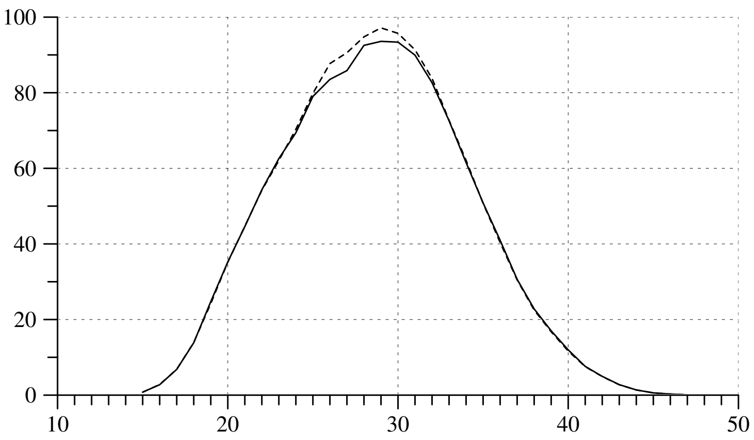



**Fig. 11.1-4** Age-specific birth rates in Germany, 1999, restricted to ages in the range from 15 to 49 years. Data are taken from columns $\tilde{\beta}_{1999,\tau}$ (solid line) and $\tilde{\beta}^+_{1999,\tau}$ (dotted line) in Table 11.1-1.

birth year of the child. This leads to slightly different birth rates as is also shown in Table 11.1-1: $n^{f+}_{1999,\tau}$ is the number of women who are born in the year $1999 - \tau$.[4] The differences are illustrated in Figure 11.1-4. However, both curves clearly show how birth rates depend on women's age.

**5.** The general birth rate can then be viewed as a weighted mean of age-specific birth rates. We mention that demographers also calculate an unweighted mean value which is called *total birth rate* [zusammengefasste Geburtenziffer].[5] The definition is

$$\text{Total birth rate} := \sum_{\tau=\tau_a}^{\tau_b} \beta_{t,\tau} \quad \text{(multiplied by 1000)}$$

where the range of summation depends on assumptions about the childbearing ages of women. For example, the calculation of total birth rates in Fachserie 1, Reihe 1 (1999, p. 50) is based on an age range from 15 to 49 years. The value for 1999, calculated for the territory of the former FRG, is 1405.8. However, while formally a mean value, this figure does not relate to any well-defined population and is therefore difficult to interprete. It is not possible, for example, to infer that the mean number of children per women (which women?) is 1.4. However, the total birth rate can also be viewed as a standardized version of the general birth rate and,

[4]We are grateful to Hans-Peter Bosse who made available these figures which are not normally published by the *Statistisches Bundesamt*.

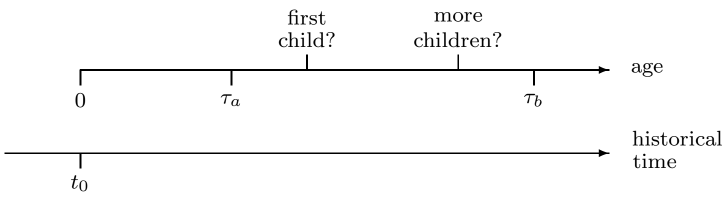
[5]One also often finds the term 'total fertility rate'.

with this understanding, used as a measure for the comparison of birth frequencies in a sequence of calendar years. This will be illustrated in a later section where we compare total birth rates with a similar measure relating to cohorts.

## 11.2 A Life Course Perspective

**1.** In order to develop a conceptual framework for recording birth events it seems sensible to refer to the life courses of women who might, or might not, give birth to children. Beginning at some age, it becomes possible for most women to bear children; but whether they will do so is basically contingent on their life courses. With the general availability of contraceptive means, women can also influence the occurrence of birth events. Therefore, whether, and when, a women will give birth to children is always a personal decision. A statistical approach cannot claim to reconstruct such individual histories in any serious sense. Nevertheless, also from a statistical point of view, one can try to relate birth events to women's life courses and their social conditions.

**2.** This is most often done by using a cohort approach. In the present context this means that we begin with a reference to birth cohorts. Using previously introduced notation, we will denote by $\mathcal{C}^f_{t_0}$ a set of women all born in the same year, $t_0$. Life courses of the members of $\mathcal{C}^f_{t_0}$ are then parallel on a calendar time axis as shown in the following graphic.




All members of $\mathcal{C}^f_{t_0}$ begin their life course in the same year, $t_0$, at age $\tau = 0$, and they can be compared with respect to their childbearing histories.

**3.** How can one record birth events of the members of $\mathcal{C}^f_{t_0}$ in terms of statistical variables? We can begin by defining a variable

$$\bar{B}_{t_0} : \mathcal{C}^f_{t_0} \longrightarrow \{0, 1, 2, 3, \ldots\}$$

which simply counts the number of children, possibly zero, born of members of $\mathcal{C}^f_{t_0}$. For each women $\omega \in \mathcal{C}^f_{t_0}$, $\bar{B}_{t_0}(\omega)$ is the number of children born of $\omega$. Of course, this number can only be known at the end of the reproductive period of the women in $\mathcal{C}^f_{t_0}$, that is, when they have reached an age $\tau > \tau_b$. In a temporal view, this means that only at the end of the year $t_0 + \tau_b$ the variable $\bar{B}_{t_0}$ gets an empirically definite meaning. Nevertheless, with this reservation, one can use $\bar{B}_{t_0}$ to define



$$\sum_{\omega \in \mathcal{C}^f_{t_0}} \bar{B}_{t_0}(\omega) \Big/ |\,\mathcal{C}^f_{t_0}\,|$$

This might be called a *cohort birth rate*: the denominator records the number of women in $\mathcal{C}^f_{t_0}$, and the numerator refers to the total number of children born of these women.

**4.** The cohort birth rate obviously does not provide any information about ages of childbearing but globally refers to the total number of children born during the reproductive period. To incorporate age information one might define variables

$$B_{t_0,\tau} : \mathcal{C}^f_{t_0} \longrightarrow \{0, 1, 2, 3, \ldots\}$$

recording the number of children born of members of $\mathcal{C}^f_{t_0}$ at the age of $\tau$. Values of these variables can be cumulated:

$$\bar{B}_{t_0,\tau}(\omega) := \sum_{j=\tau_a}^{\tau} B_{t_0,j}(\omega)$$

In particular, one finds the simple relationship $\bar{B}_{t_0}(\omega) = \bar{B}_{t_0,\tau_b}(\omega)$.

**5.** It is also helpful to introduce *age-specific cohort birth rates*. We will use the following definition:

$$\gamma_{t_0,\tau} := \frac{\sum_{\omega \in \mathcal{C}^f_{t_0,\tau}} B_{t_0,\tau}(\omega)}{|\,\mathcal{C}^f_{t_0,\tau}\,|}$$

The denominator refers to the number of members of $\mathcal{C}^f_{t_0}$ who survived age $\tau - 1$ and therefore might give birth to children at age $\tau$. The numerator refers to the number of children born of members of $\mathcal{C}^f_{t_0,\tau}$ at age $\tau$.

**6.** The rates $\gamma_{t_0,\tau}$ can be used to investigate the distribution of births during the reproductive period of women belonging to the same birth cohort. As will be illustrated later, this can be done by plotting values of $\gamma_{t_0,\tau}$ against $\tau$. Alternatively, one can plot *cumulated cohort birth rates*

$$\bar{\gamma}_{t_0,\tau} := \sum_{j=\tau_a}^{\tau} \gamma_{t_0,j}$$

However, $\bar{\gamma}_{t_0,\tau_b}$ should not be confused with the mean number of children born of members of $\mathcal{C}^f_{t_0}$ until the end of the reproductive period. In order to relate age-specific cohort birth rates to the total number of children born of members of a birth cohort, one has to take into account the women who died before the end of the reproductive period. The total number of children born of members of $\mathcal{C}^f_{t_0}$ is

$$\sum_{\tau=\tau_a}^{\tau_b} \gamma_{t_0,\tau} \, |\mathcal{C}^f_{t_0,\tau}|$$

Dividing this quantity by the number of women belonging to $\mathcal{C}^f_{t_0}$ would provide the mean number of children per women. To see explicitly the dependence on women's age-specific death rates, one can use the relationship

$$\frac{|\mathcal{C}^f_{t_0,\tau}|}{|\mathcal{C}^f_{t_0}|} = \prod_{j=0}^{\tau-1}(1-\eta^f_{t_0,j})$$

derived in Section (8.1). Using this relationship, one finds

$$\text{Mean number of children per women} = \sum_{\tau=\tau_a}^{\tau_b} \gamma_{t_0,\tau} \frac{|\mathcal{C}^f_{t_0,\tau}|}{|\mathcal{C}^f_{t_0}|} = \sum_{\tau=\tau_a}^{\tau_b} \gamma_{t_0,\tau} \prod_{j=0}^{\tau-1}(1-\eta^f_{t_0,j})$$

This discussion will be continued in Section 18.1 where we deal with reproduction rates. Here we only mention that, although cumulated cohort birth rates do not allow inferences about the mean number of children born of women belonging to the same birth cohort, they can be used as some measure of "cohort fertility". In particular, one can use $\bar{\gamma}_{t_0,\tau_b}$, commonly called a *completed cohort birth rate*. This rate would equal the cohort birth rate if all women survived the end of the reproductive period.

## 11.3 Childbearing and Marriage

**1.** Due to an unfortunate focus on marital births, official birth statistics are inadequate when dealing with questions of parity and number and timing of births. The problem is aggravated by the fact that generally not even divorces are taken into account. Counting of children starts anew with every marriage, disregarding all previous births.[6]

**2.** The confounding of childbearing and marriage behavior has a long tradition in demography. Many demographers assume that a statistical analysis of marriages and divorces should be considered an essential part of demography. The following quotation from a textbook can serve as an

[6]The latter defect has been avoided in a 10 % subsample of the 1970 census where women were asked to report the birth dates of all their marital children, regardless of their current marital status. These data will be discussed in Chapter 12.2.



example:[7]

"Marriage and divorce have been of long concern in population studies because of their recognized relationship to population composition, on the one hand, and to fertility, on the other. Next to age and sex, no characteristic is more basic to a population than its composition by marital status: its absolute and relative numbers of single, married, widowed, and divorced persons of each sex and at each age. Although children may be born outside of marriage, in every society childbearing is intimately associated with marriage and generally is viewed both as the object and as a more or less immediate consequence of marriage and conjugal relations." (Matras 1973, p. 258)

However, for several reasons we shall not adopt this view. The most important one is that a substantial proportion of women who give birth to children is not married. Table 11.3-1 provides figures that show the number of non-marital births (per 1000) in Germany since 1872. Figure 11.3-1 provides a graphical illustration. It is seen that until about 1940 the proportion of non-marital birth was already about 10 percent.[8] Then, after an initial decline after World War II, beginning in the mid-sixties, the proportion is continually rising. This trend is particularly strong in the territory of the former GDR where the proportion of non-marital births has reached almost 50 percent.[9] On the other hand, there are also married women who, for whatever reasons, remain childless. In short, being married is neither a necessary nor a sufficient condition for childbearing, nor is there any kind of causal relationship.

**3.** This is not to deny that living arrangements may play an important role in women's decisions to give birth to children. But living conditions and marriage are different concepts. This is often obscured by an unclear usage of terms. To cite Matras again:

"Basically, a *family* consists of an adult male and female living in a common residence, maintaining a socially approved sexual relationship, and sharing the residence with their offspring and sometimes with other persons united with them in some biologically based relationship. *Marriage* is the establishment of this residence and socially approved sexual relationship between the adult male and female." (Matras 1973, p. 260)

Not only does this definition of 'family' ignore the widely different forms of household types which have emerged in human history. More important

[7]As an example from the German literature see Bolte, Kappe and Schmid (1980, pp. 13 -14).

[8]Actually, at least in some parts of Germany, percentages of non-marital births were even higher in earlier periods. For example, Lindner (1900, p. 217) reports about 20 % non-marital births during the period 1825 –1868 for the *Königreich Bayern*. For an interpretation of some of the changes that occurred during the 19th century see Kottmann (1987).

[9]For a discussion see Huinink (1998).

**Table 11.3-1** Proportion of non-marital births (per 1000 births) in Germany and the territories of the former FRG and GDR. Sources: Statistisches Bundesamt, Bevölkerung und Wirtschaft 1872–1972 (pp. 107-108) for the period 1872–1938; Fachserie 1, Reihe 1, 1999 (pp. 50-51) for the period 1946–1999.

| Germany (Reichsgebiet) | | | |
|---|---|---|---|
| Year | | Year | |
| 1872 | 87.8 | 1908 | 87.7 |
| 1873 | 91.3 | 1909 | 89.2 |
| 1874 | 85.7 | 1910 | 89.6 |
| 1875 | 85.6 | 1911 | 90.8 |
| 1876 | | 1912 | 94.4 |
| 1877 | 85.8 | 1913 | 96.0 |
| 1878 | 85.7 | 1914 | 96.9 |
| 1879 | 87.6 | 1915 | 110.7 |
| 1880 | | 1916 | 109.5 |
| 1881 | 89.7 | 1917 | 114.1 |
| 1882 | 91.9 | 1918 | 129.6 |
| 1883 | 91.3 | 1919 | 110.3 |
| 1884 | 94.2 | 1920 | 112.2 |
| 1885 | 93.6 | 1921 | 105.6 |
| 1886 | 93.8 | 1922 | 106.3 |
| 1887 | 93.4 | 1923 | 103.1 |
| | | 1924 | 104.1 |
| 1893 | 90.5 | 1925 | 118.2 |
| 1894 | 92.7 | 1926 | 123.7 |
| 1895 | 89.8 | 1927 | 122.8 |
| 1896 | 92.7 | 1928 | 122.1 |
| 1897 | 91.3 | 1929 | 120.7 |
| 1898 | 90.3 | 1930 | 120.0 |
| 1899 | 88.8 | 1931 | 117.5 |
| 1900 | 86.3 | 1932 | 116.3 |
| 1901 | 84.8 | 1933 | 106.7 |
| 1902 | 83.9 | 1934 | 85.3 |
| 1903 | 82.4 | 1935 | 77.7 |
| 1904 | 83.1 | 1936 | 77.0 |
| 1905 | 84.3 | 1937 | 76.6 |
| 1906 | 84.1 | 1938 | 76.0 |
| 1907 | 86.0 | | |

| Territory of the former | | | | | |
|---|---|---|---|---|---|
| | FRG | GDR | | FRG | GDR |
| Year | | | Year | | |
| 1946 | 163.8 | 192.5 | 1973 | 62.7 | 156.4 |
| 1947 | 118.5 | 151.1 | 1974 | 62.7 | 162.9 |
| 1948 | 102.3 | 126.9 | 1975 | 61.2 | 161.4 |
| 1949 | 93.1 | 118.9 | 1976 | 63.5 | 162.1 |
| 1950 | 97.3 | 127.9 | 1977 | 64.7 | 157.7 |
| 1951 | 96.4 | 131.5 | 1978 | 69.6 | 173.4 |
| 1952 | 90.3 | 130.0 | 1979 | 71.3 | 195.9 |
| 1953 | 86.7 | 130.3 | 1980 | 75.6 | 228.4 |
| 1954 | 84.2 | 132.5 | 1981 | 79.0 | 255.8 |
| 1955 | 78.6 | 130.0 | 1982 | 84.9 | 292.9 |
| 1956 | 74.7 | 131.9 | 1983 | 88.3 | 320.4 |
| 1957 | 71.9 | 131.8 | 1984 | 90.7 | 335.5 |
| 1958 | 68.5 | 123.7 | 1985 | 94.0 | 338.1 |
| 1959 | 66.9 | 120.1 | 1986 | 95.5 | 344.3 |
| 1960 | 63.3 | 116.0 | 1987 | 97.1 | 328.0 |
| 1961 | 59.5 | 111.3 | 1988 | 100.3 | 334.4 |
| 1962 | 55.6 | 100.8 | 1989 | 102.2 | 336.4 |
| 1963 | 52.3 | 93.4 | 1990 | 104.9 | 349.9 |
| 1964 | 49.9 | 94.2 | 1991 | 111.1 | 417.2 |
| 1965 | 46.9 | 98.1 | 1992 | 115.9 | 418.2 |
| 1966 | 45.6 | 99.9 | 1993 | 118.7 | 410.9 |
| 1967 | 46.1 | 107.0 | 1994 | 124.3 | 414.4 |
| 1968 | 47.6 | 114.9 | 1995 | 128.9 | 417.7 |
| 1969 | 50.4 | 124.1 | 1996 | 136.8 | 423.9 |
| 1970 | 54.6 | 133.0 | 1997 | 142.7 | 441.0 |
| 1971 | 58.1 | 151.2 | 1998 | 159.2 | 471.5 |
| 1972 | 60.5 | 162.0 | 1999 | 176.7 | 499.4 |

for our present argument is Matras's unspecific use of the term 'marriage'. Given his definition, a marriage takes place when two people, of opposite sex, decide to start a common household (residence). However, this obscures the fact that, in modern societies, 'marriage' does not refer to some kind of household formation, but is a juridical term that gets its meaning from laws and a corresponding juridical practice. In fact, two people cannot simply decide to become married but need the approval of an official institution. In particular, only then they will be counted as being married in official statistics. In contrast to terms that refer to living conditions,



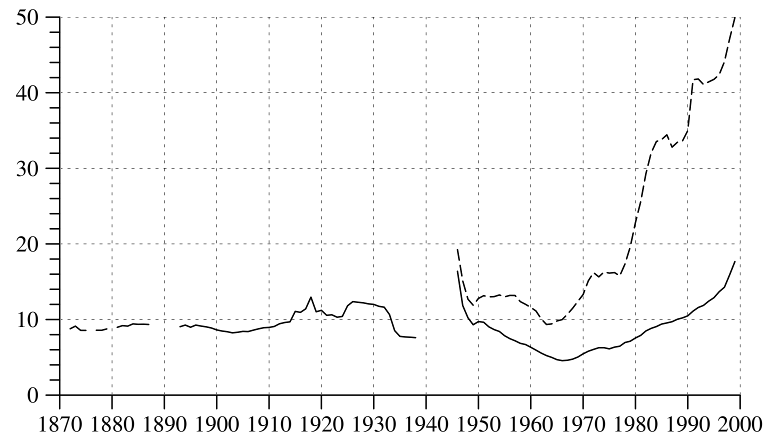



**Fig. 11.3-1** Proportion of non-marital births (in percent) in Germany and in the territory of the former FRG (solid line) and in the territory of the former GDR (dotted line); calculated from the data in Table 11.3-1.

the term 'marriage' should therefore be considered, not as a sociological category, but as belonging to the realm of administrative regulations. Of course, this does not preclude a sociological analysis of practices of marriage and divorce.

**4.** A further argument can illustrate the difference. While it seems plausible that women's decisions to give birth to children depend on their actual and expected living arrangements, this can most often not be said of marriages. Women do not bear children because they are married; but they might want to become married because they want a legally secured framework for their children.

## 11.4 Birth Rates in a Cohort View

**1.** In order to record age-specific birth rates of a cohort $\mathcal{C}^f_{t_0}$ it would be necessary to follow the cohort members from birth until the end of the reproductive period. Difficulties are the same as in the construction of cohort life tables (see Chapter 8). Mainly three surrogate methods seem possible:

a) One can approximate age-specific cohort birth rates with age-specific period data;

b) one can use data from retrospective surveys, which implies that one has to ignore cohort mortality; and

c) one can use data from panel studies which allow to follow the members of a birth cohort for a sequence of years during their life courses.

**Table 11.4-1** Age-specific birth rates of women belonging to birth cohorts 1930, ..., 1970. Source: Fachserie 1, Reihe 1, 1999 (pp. 198-200).

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

In the present section we discuss the first approach.

**2.** The basic idea is quite simple. If there were no in- and out-migration, one could identify age-specific cohort birth rates and period birth rates in the following way:

$$\gamma_{t_0,\tau} = \beta_{t_0+\tau,\tau}$$

It would be possible, then, to reconstruct the age-specific birth rates of a cohort $\mathcal{C}^f_{t_0}$ from the sequence of period birth rates

$$\beta_{t_0+\tau_a,\tau_a}, \ldots, \beta_{t_0+\tau_b,\tau_b}$$



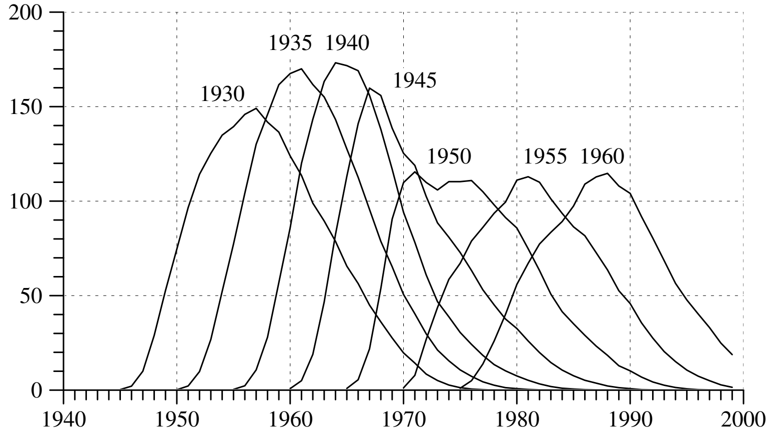



**Fig. 11.4-1** Age-specific birth rates of quasi-cohorts 1930, ..., 1960. Data are taken from Table 11.4-1.

Of course, in reality migration takes place, and the approach therefore essentially consists in the construction of age-specific birth rates for quasi-cohorts.

**3.** Age-specific birth rates for the territory of the former FRG have been published by the *Statistisches Bundesamt* in Fachserie 1, Reihe 1, 1999 (pp. 198-200) for cohorts beginning in the birth year 1930 and for ages 15 to 49. For a selection of birth cohorts the data are shown in Table 11.4-1. Figure 11.4-1 shows a plot of these age-specific quasi-cohort birth rates on a calendar time axis. It can be seen how the births of women of successive quasi-cohorts changed through historical time, both in shape and size.

**4.** Next, we consider the cumulated birth rates

$$\bar{\gamma}^*_{t_0,\tau} := \sum_{j=15}^{\tau} \beta_{t_0+j,j}$$

which are helpful to compare distributions of birth rates during the reproductive period as shown in Figure 11.4-2. The plot exhibits substantial changes in the timing of births. Using the 1930 birth cohort as an arbitrary reference, the plot suggests that the mean age of childbearing declined until birth cohorts born between 1945 and 1950, and then began to increase. This is also seen in Figure 11.4-3 showing a level plot of the cumulated rates in an age-period diagram. The mapping is

$$(t,\tau) \longrightarrow \bar{\gamma}^*_{t-\tau,\tau}$$

where $t$ refers to calendar years. Accordingly, each diagonal line refers to a 1-year birth cohort, with birth years ranging from 1930 to 1978. The

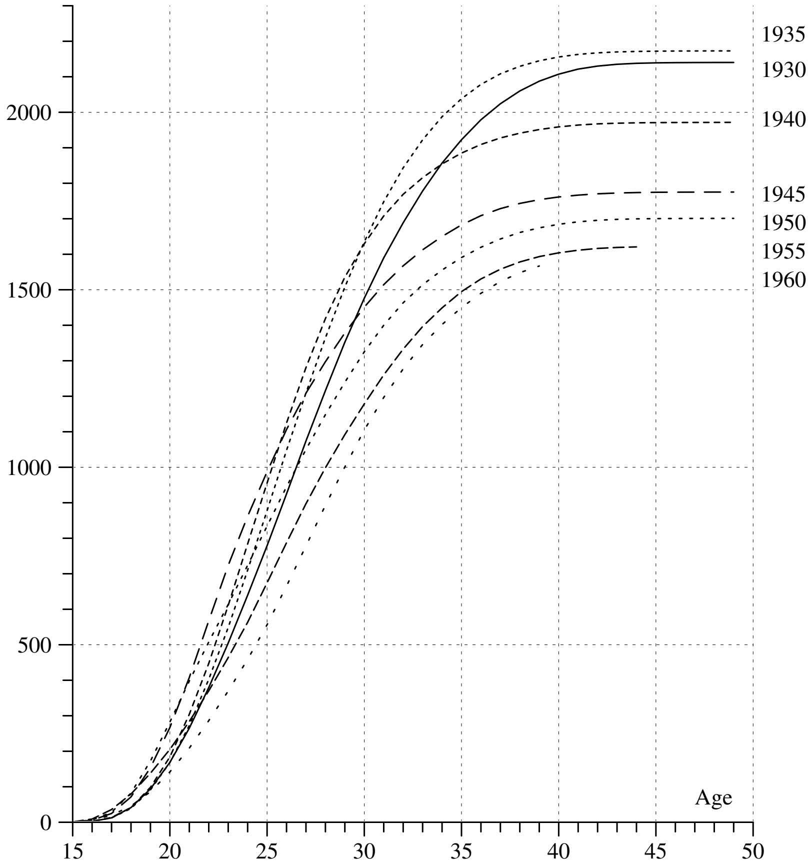



**Fig. 11.4-2** Plot of cumulated age-specific birth rates for birth cohorts 1930, ..., 1960, based on the data in Table 11.4-1.

contour lines connect cumulated birth rates (per 1000 of women) having approximately the same value. The maximal value of 2240 is reached by women belonging to birth cohort 1934 at age 49.

**5.** Figure 11.4-2 also suggests that completed cohort birth rates declined, beginning with birth years around 1935. This is also seen from the dotted line in Figure 11.4-4 which shows $\bar{\gamma}^*_{t_0,40}$ for $t_0 = 1930, \ldots, 1959$.[10] The figure also shows the development of total birth rates that, for comparability,

[10] Age 40 was chosen to allow the calculation of cumulated birth rates until birth cohort 1959, given that data are only available until calendar year 1999.



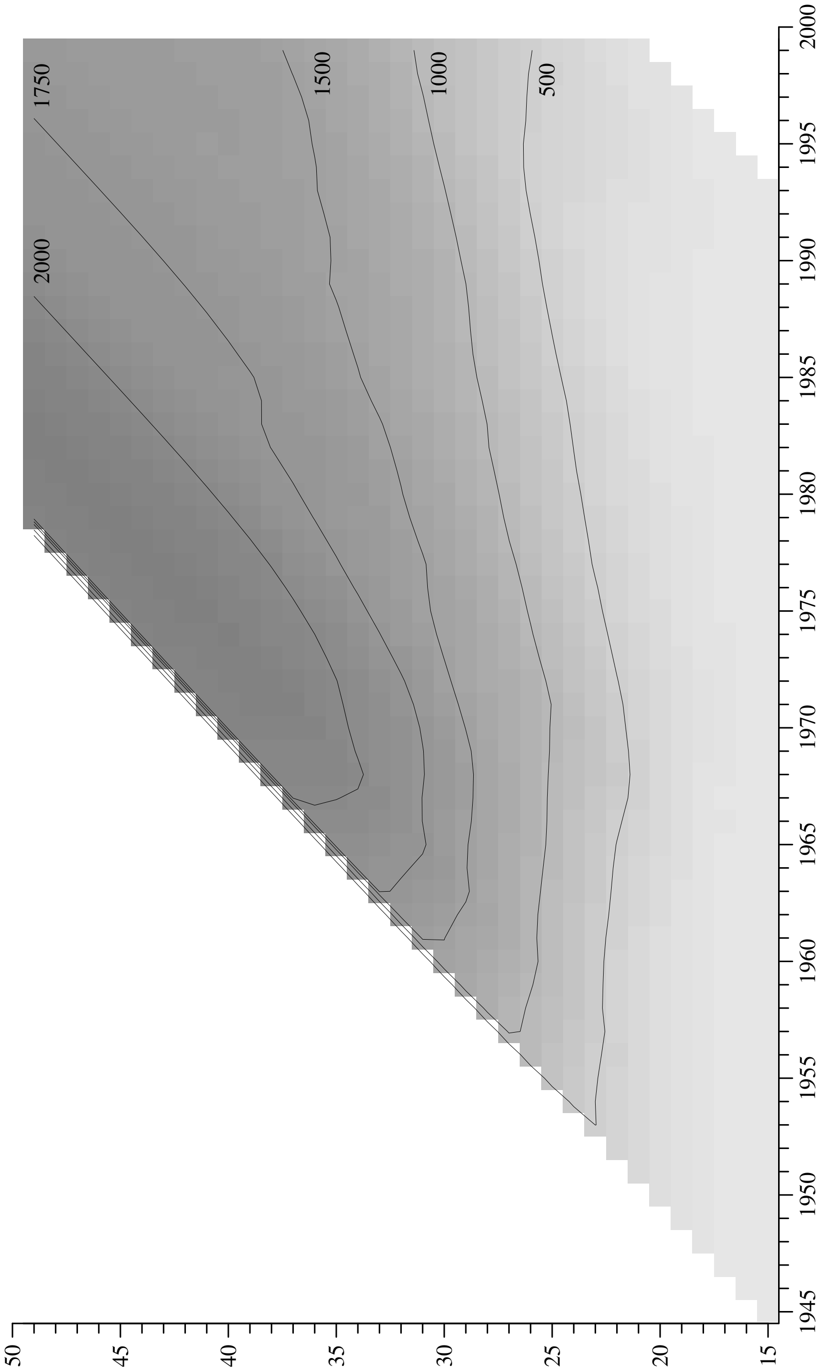



**Fig. 11.4-3** Level plot of cumulated age-specific quasi-cohort birth rates in the period 1945 – 99, based on data from Fachserie 1, Reihe 1, 1999 (pp. 198-200).

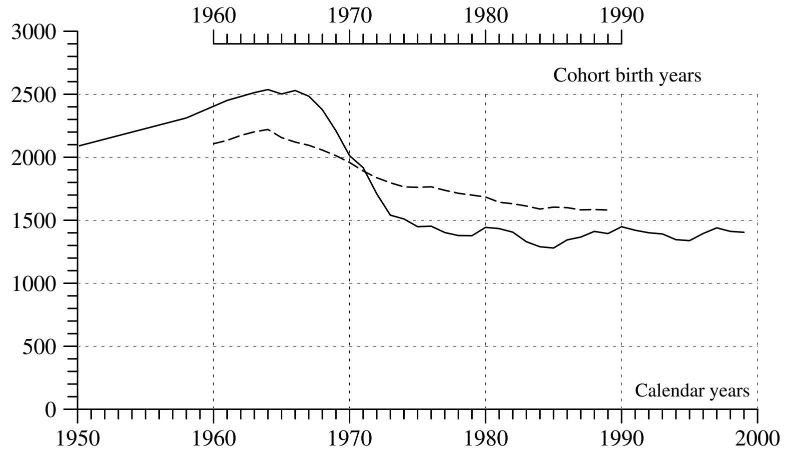



**Fig. 11.4-4** Comparison of total birth rates (solid line) and cumulated quasi-cohort birth rates (dotted line), both calculated for ages 15 – 40.

have been calculated from age-specific birth rates as

$$\sum_{\tau=15}^{40} \beta_{t,\tau}$$

for $t = 1950, \ldots, 1999$. Since completed cohort birth rates do not refer to single calendar years, both time series cannot be compared directly. Nevertheless, the figure clearly suggests that variability in completed cohort birth rates is much smaller than in total birth rates. This is explainable by the fact that total birth rates also depend on the timing of births. Women might give birth to more children in one year and to less children in another year without necessarily affecting the completed cohort birth rates. This idea will be taken up in Chapter 13.3 where we show that a substantial part of the "baby boom" that occurred in the period 1955 – 65 can be attributed to "timing effects".

# Chapter 12

# Retrospective Surveys

Even though cohort birth rates are quite informative, they do not allow to recover (a) the distribution of ages at first childbearing, (b) the proportion of women who remain childless, and (c) the distribution of the number of births. As was mentioned in the previous chapter, due to an unfortunate focus on marital births, official statistics in Germany provides only limited information on these quantities. Most investigations are therefore based on retrospective surveys in which women are asked to report about the birth dates of their children. In the present chapter we briefly discuss the conceptual framework and then use data from the 1970 census. Additional data from non-official surveys will be considered in Chapter 14.

## 12.1 Introduction and Notations

**1.** To focus the discussion, we consider the question whether, and at which age, women give birth to a first child. To allow for an investigation of changes among successive cohorts, our conceptual framework refers to birth cohorts. Using $\mathcal{C}^f_{t_0}$ to denote the birth cohort of women born in the year $t_0$ we might begin with a duration variable, $\hat{T}_{t_0}$, that records the age when members of $\mathcal{C}^f_{t_0}$ get their first child (the corresponding property space, $\tilde{\mathcal{T}} := \{0, 1, 2, \ldots\}$, being understood as a representation of ages in completed years). Obviously, there are two complications. First, not all women will give birth to a child, and in such cases there is also no duration until the birth of a first child. Secondly, some women will die before the end of the reproductive period.[1] It is therefore necessary to introduce a second variable that records which of these possibilities actually takes place. This second variable will be denoted by $\hat{D}_{t_0}$ and defined as follows:

$$\hat{D}_{t_0}(\omega) := \begin{cases} 1 & \text{if } \omega \in \mathcal{C}^f_{t_0} \text{ has given birth to at least one child} \\ 0 & \text{otherwise} \end{cases}$$

We are concerned, then, with a two-dimensional variable

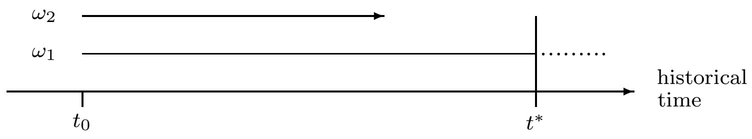
$$(\hat{T}_{t_0}, \hat{D}_{t_0}) : \mathcal{C}^f_{t_0} \longrightarrow \tilde{\mathcal{T}} \times \tilde{\mathcal{D}}$$

where the meaning of the duration variable, $\hat{T}_{t_0}$, depends on the value of $\hat{D}_{t_0}$. If $\hat{D}_{t_0}(\omega) = 1$ then $\hat{T}_{t_0}(\omega) \le \tau_b$ records the age at which $\omega$ has

[1] We use $\tau_a$ and $\tau_b$ to denote, respectively, the beginning and end of the reproductive period of women.

given birth to a first child.[2] If, on the other hand, $\hat{D}_{t_0}(\omega) = 0$, $\hat{T}_{t_0}(\omega)$ will represent the age at which $\omega$ dies, or $\tau_b$, whichever occurs first.

**2.** In order to recover the distribution of $(\hat{T}_{t_0}, \hat{D}_{t_0})$, it is necessary to follow the members of $\mathcal{C}^f_{t_0}$ at least until the end of the reproductive period. However, if only for reasons of practicability, the standard approach is to perform a *retrospective survey*. The following graphic provides an illustration:




At some date in historical time, $t^*$, which will be called the *interview date*, people are asked about their previous life courses. Of course, this can only be done with persons born before the interview date. There are, however, two further implications.

a) One can only interview people still alive at the interview date. In the picture, one might ask $\omega_1$, but not $\omega_2$ who died before $t^*$. So it is normally not possible, with a retrospective survey, to get complete information about all members of a birth cohort.[3] Whether this is a serious problem depends on the purpose of the survey. It might be a serious problem if one intends to interview people at very old ages. On the other hand, assuming a historical situation in which only few women die during the reproductive period, it might well be possible to ignore the problem of mortality when performing a survey to record information about childbearing histories.

b) A second implication concerns the fact that information about life histories is always right censored at the date of the interview. In the above picture this is shown by the person called $\omega_1$. This person is still alive at the interview date and therefore can report about his or her life course until $t^*$,[4] but cannot report about what might happen in the future. If, for example, the end of the reproductive period of the members of $\mathcal{C}^f_{t_0}$ has been reached before the interview date, it is possible to get complete records of the childbearing histories; but otherwise the recorded histories will be more or less incomplete.

[2] Of course, also twins, or triplets, might be born. For the moment we ignore this possibility and simply speak of a first child.

[3] To indicate this fact one might speak of *retrospective cohorts*. Contrary to proper birth cohorts they are defined by conditioning on survival until the interview date and living in the region where the survey is conducted.

[4] This, of course, also depends on memory. It is well possible that details of a life course have been forgotten or become confused after some while.



**3.** An immediate implication of the censoring problem is that the amount of information that can be gathered with a retrospective survey depends on the birth year of the interviewed persons. If one selects, for the survey, only persons belonging to the same birth cohort, say $\mathcal{C}^f_{t_0}$, then also the life-span until the interview date is approximately the same for all interviewed persons. But often an interest concerns differences in the life courses of persons who belong to different birth cohorts. For example, we might want to compare childbearing histories of women born in the years 1950, 1960, and 1970, and the interviews are performed in the year 2000. The childbearing histories of the women born in 1950 will then be complete, but for the younger birth cohorts they will be censored at an age of 30 or 40, respectively.

**4.** We finally need to relate the information which can be gained by a retrospective survey to the conceptual framework introduced at the beginning. We therefore think of a survey in which members of $\mathcal{C}^f_{t_0}$, who survived the interview date $t^*$, are asked whether they already gave birth to a first child and, given this was the case, at which age the birth event occurred. The data can be represented by a two-dimensional variable denoted by $(T_{t_0}, D_{t_0})$. The property spaces are again $\tilde{\mathcal{T}}$ and $\tilde{\mathcal{D}}$, respectively, but the meaning of the variables is different from $(\hat{T}_{t_0}, \hat{D}_{t_0})$. $D_{t_0}$ now records whether a women has given birth to a first child *until the interview date*. The relation is therefore as follows:

a) If $D_{t_0}(\omega) = 1$, $\omega$ has born a child before the interview date, and in this case $T_{t_0}(\omega)$ records the age of the women in the year of her first birth. So one can conclude that $\hat{D}_{t_0}(\omega) = 1$ and $\hat{T}_{t_0}(\omega) = T_{t_0}(\omega)$.

b) If, on the other hand, there was no first birth until the interview date, then $D_{t_0}(\omega) = 0$ and $T_{t_0}(\omega)$ records the age of the women at the interview date. Given the definition of $\hat{T}_{t_0}$, one can conclude that $\hat{T}_{t_0}(\omega) \geq T_{t_0}(\omega)$ but the conclusion about $\hat{D}_{t_0}$ depends on the women's age. If $T_{t_0}(\omega) > \tau_b$, one can conclude that $\hat{D}_{t_0}(\omega) = 0$; but otherwise no definite conclusion about the value of $\hat{D}_{t_0}(\omega)$ can be drawn.

Consequently, data from a retrospective survey in which not all interviewed women have already reached the end of the reproductive period, are necessarily to some extent incomplete; and so the question arises how to use the data for an assessment of the distribution of $(\hat{T}_{t_0}, \hat{D}_{t_0})$.

**5.** In any case, the available data only allow inferences for those members of $\mathcal{C}^f_{t_0}$ who survived the interview date $t^*$, or, equivalently, who survived age $\tau^* := t^* - t_0$. Using notation introduced in Section 3.4 (see also Section 8.1), this is the subset $\mathcal{C}^f_{t_0,\tau^*}$. One therefore can only consider a variable

$$(\hat{T}_{t_0,\tau^*}, \hat{D}_{t_0,\tau^*}) : \mathcal{C}^f_{t_0,\tau^*} \longrightarrow \tilde{\mathcal{T}} \times \tilde{\mathcal{D}}$$

that is restricted to the members of $\mathcal{C}^f_{t_0,\tau^*}$. For these members, the values are identical:

$$\hat{T}_{t_0,\tau^*}(\omega) = \hat{T}_{t_0}(\omega) \quad \text{and} \quad \hat{D}_{t_0,\tau^*}(\omega) = \hat{D}_{t_0}(\omega)$$

Moreover, it is also evident that the available data do not allow inferences for periods beyond $t^*$. Therefore, one can only consider the distribution of $(\hat{T}_{t_0,\tau^*}, \hat{D}_{t_0,\tau^*})$ conditional on $\hat{T}_{t_0,\tau^*} \leq \tau^*$. However, this implies the formal identity

$$\mathrm{P}[\hat{T}_{t_0,\tau^*}, \hat{D}_{t_0,\tau^*} \,|\, \hat{T}_{t_0,\tau^*} \leq \tau^*] \;=\; \mathrm{P}[T_{t_0}, D_{t_0} \,|\, T_{t_0} \leq \tau^*]$$

for all $\omega \in \mathcal{C}^f_{t_0,\tau^*}$. One therefore does not need any specific estimation procedure but can directly use the observed values of $T_{t_0}$ and $D_{t_0}$.

**6.** This result is due to the fact that censoring occurs at the same time for all members of $\mathcal{C}^f_{t_0,\tau^*}$. Slightly more complicated is the situation when interview dates extend over a longer period of time and/or cohorts are defined by comprising several birth years. As a consequence, also the censoring times extend over several years and there is no longer a definite period for reliable conclusions. If one is not willing to restrict inferences until the minimal age of censored observations, that is, $\min\{T(\omega) \,|\, D(\omega) = 0\}$, one needs some method of estimation. One possibility is to use the Kaplan-Meier procedure introduced in Section 8.3.4. Examples will be discussed in Chapter 14.

## 12.2 Data from the 1970 Census

As was mentioned in Section 11.3, information available from official birth statistics in Germany is severely limited by the fact that the parity of births [Ordnungsnummer der Geburten] is only recorded for marital births in current marriages. Somewhat better information is available from the 1970 census in which 10 % of the women were asked to report the dates of all marital births, regardless of their current marital status. In the following sections we discuss a subsample of this data set available for scientific research.

### 12.2.1 Sources and Limitations

**1.** The census of 1970 was conducted on May 27 of that year in the territory of the former FRG.[5] As part of this census a subsample of 10 % of the population was asked to provide additional information, in particular, all women with a German citizenship who participated in the 10 % subsample

[5]For a detailed description, including a presentation of the questionnaire, see Schubnell and Herberger (1970).



were asked for dates of marriage and birth dates of all their marital children, regardless of their current marital status. Some results from these additional questions were published, albeit in highly aggregated form, by the *Statistisches Bundesamt* in Fachserie A.[6] Fortunately, some years ago, official statistics in Germany agreed to make available, for scientific research, anonymised subsamples of many main surveys, including the 1970 census.[7]

**2.** The data set to be used in the present chapter consists of a 10 % subsample of the 10 % part of the 1970 census.[8] So it is a 1 % subsample of all women who lived in May 1970 in the territory of the former FRG and had a German citizenship. The number of cases is 314993; if multiplied by 100, this is roughly the number of women, with a German citizenship, who lived in the former FRG in May 1970.

**3.** For each person in our subsample we have the following information: (a) the birth year, and (b) the births years of all (up to 12) marital children. So we are able to reconstruct marital childbearing histories. The limitation is, of course, that we have no information about non-marital births. As shown in Table 11.3-1 in Section 11.3, for the period until about 1970 this amounts to about 10 % of all births. Actually, however, the birth coverage of the sample is somewhat higher than 90 % because a substantial portion of non-marital births has been "legitimized" by a following marriage. To provide an example, the total number of births during the year 1969 in the territory of the former FRG was 903456. In 852783 cases the mother had a German citizenship, and of these cases 810002 were marital births.[9] On the other hand, the number of birth in 1969, reported by women in our sample, is 8215 which is 821500 when multiplied by 100. So one can estimate that about 27 % of non-marital birth have been "legitimized" by a following marriage. Nevertheless, it is clearly important to be aware of the fact that our sample does not cover all births.

**4.** Further limitations are due to the fact that our data set results from a *retrospective* survey as was discussed in Chapter 12. Only women who survived until 1970 could have been asked about previous childbearing. This is illustrated by the distribution of birth years shown in Figure 12.2-

[6]Fachserie A. Bevölkerung und Kultur. Volkszählung vom 27. Mai 1970. Heft 7, Geburten. See also Schwarz (1974).

[7]More information on these data sets are available from the *Zentrum für Umfragen, Methoden und Analysen* (ZUMA, Mannheim), Abteilung für Mikrodaten; see: `www.gesis.org/Dauerbeobachtung/Mikrodaten`.

[8]We are grateful to Bernhard Schimpl-Neimanns (ZUMA) who prepared the tables which we have used. The tables are based on the data set: *Ergebnisse der Volks- und Berufszählung 1970 mit den Ergänzungsfragen (1 % Stichprobe der Wohnbevölkerung)*; see Bach, Handl and Müller (1980), Schimpl-Neimanns and Frenzel (1995).

[9]Fachserie 1, Reihe 1, 1999 (p. 211).

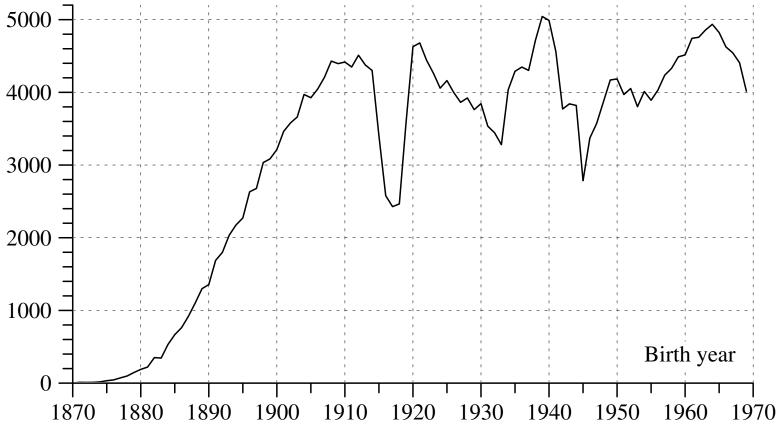



**Fig. 12.2-1** Number of women born between 1870 and 1969 in the 1 % subsample of the 1970 census.

1.[10] However, part of the problem can be circumvented by a separation into birth cohorts and reconstructing childbearing histories for each birth cohort separately. This then only requires to assume that differential mortality is not heavily correlated with childbearing. In the following sections we will make this assumption and consider all birth cohorts with birth years from 1905 to 1945. Of course, for the younger birth cohorts, beginning about 1930, childbearing histories are not completed by 1970.

**5.** Only to consider, and compare, birth cohorts is insufficient if one intends to reconstruct the historical development of births. It becomes necessary, then, to locate the birth cohorts in historical time and, in particular, take into account changes in cohort size. One aspect of this problem concerns the absolute number of children born of women who survived until 1970. This is shown by the solid line in Figure 12.2-2. For comparison, the dotted line that begins in 1946 shows the number of births as recorded in the territory of the former FRG by official statistics. The difference is mainly due to the fact that the dotted line refers to all births while our sample only reports marital births of women with a German citizenship. The important point is that both curves are nearly proportional so that it seems justified to use our sample for a reconstruction of *changes* in the development of birth rates in the post-war period. More problematic are the earlier periods. Since political boundaries have changed and no valid data are available for the years from 1939 to 1945, it is already difficult to assess the birth coverage of the sample. The first part of the dotted line in Figure 12.2-2, which ends in 1938, shows the total number of births

[10]This can also be viewed as an age distribution of the female population in 1970; see, for comparison, the age distributions which were shown in Chapter 6.5.



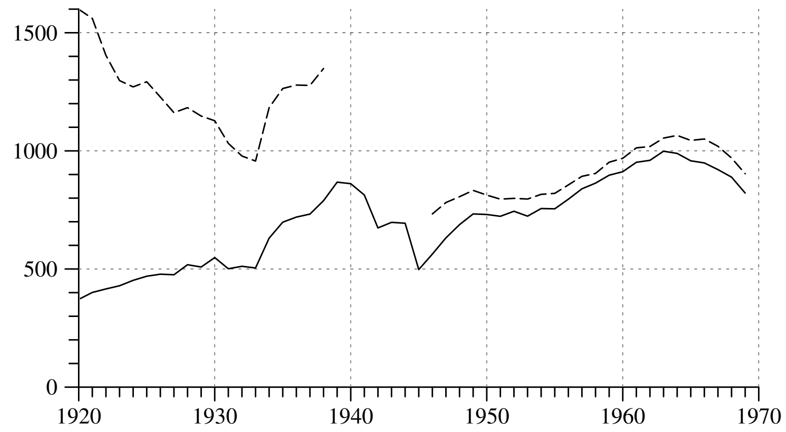



**Fig. 12.2-2** Number of children (in 1000) born during the years 1920 – 1969 in Germany. The solid line refers to the number of children reported by women in the 1 % subsample of the 1970 census. Dotted lines are based on data taken from Statistisches Bundesamt, Bevölkerung und Wirtschaft 1872 –1972 (pp. 107-9). Until 1938 these data refer to the territory of the former Deutsches Reich, beginning in 1946 they refer to the territory of the former FRG.

in the territory of the former Deutsches Reich. Obviously, at least until about 1930, there is no correspondence between the two curves, due to the fact that many women who gave birth to children before 1930 died before 1970. It might be possible, however, to use the sample data also for some conclusions about the development of births since the beginning of the 1930 s.

### 12.2.2 Age at First Childbearing

**1.** We begin with an investigation of the distribution of ages at first marital childbearing. This will be done separately for each 1-year birth cohort, C5, . . . , C45. The numbers refer to birth years, for example, C5 denotes the birth cohort of women born in 1905. For some of these birth cohorts the data are shown in Table 12.2-1. Referring to birth cohort C10 as an example, there are 5 women who reported that their first marital birth occurred in the year 1926, that is, at age 16. As can be seen from the table, all births occurred at ages between 16 and 48. Altogether, 3251 women reported a birth year for the first marital child. In addition, 1166 women had no marital children until the interview date in 1970, corresponding to an age of 60 years. These are called censored cases in the table. However, since the age at censoring is after the end of the reproductive period, one can safely assume that these women will remain without a marital

**Table 12.2-1** Number of women in the 1 % subsample of the 1970 census who reported a first marital birth at the specified age, classified according to 1-year birth cohorts. Also shown is the number of women with no marital birth until the interview date in 1970.

| $\tau$ | C5 | C10 | C15 | C20 | C25 | C30 | C35 | C40 | C45 |
|---|---|---|---|---|---|---|---|---|---|
| 15 | | | | | 1 | | | | |
| 16 | 3 | 5 | 1 | 5 | 7 | 1 | 10 | 7 | 9 |
| 17 | 8 | 15 | 18 | 29 | 12 | 10 | 30 | 57 | 39 |
| 18 | 34 | 58 | 34 | 82 | 33 | 43 | 70 | 103 | 112 |
| 19 | 75 | 82 | 80 | 119 | 73 | 136 | 152 | 218 | 176 |
| 20 | 117 | 138 | 128 | 197 | 114 | 208 | 237 | 342 | 230 |
| 21 | 166 | 167 | 169 | 245 | 177 | 259 | 292 | 401 | 238 |
| 22 | 177 | 167 | 169 | 261 | 259 | 289 | 365 | 452 | 225 |
| 23 | 204 | 227 | 228 | 297 | 311 | 280 | 390 | 439 | 229 |
| 24 | 220 | 274 | 269 | 356 | 305 | 270 | 373 | 407 | 172 |
| 25 | 211 | 302 | 272 | 201 | 292 | 270 | 320 | 409 | 58 |
| 26 | 182 | 307 | 233 | 233 | 251 | 269 | 285 | 286 | |
| 27 | 154 | 265 | 145 | 216 | 241 | 228 | 265 | 252 | |
| 28 | 166 | 255 | 170 | 203 | 207 | 181 | 183 | 201 | |
| 29 | 189 | 231 | 134 | 224 | 200 | 136 | 153 | 150 | |
| 30 | 159 | 190 | 66 | 179 | 137 | 120 | 118 | 43 | |
| 31 | 118 | 121 | 73 | 150 | 121 | 90 | 107 | | |
| 32 | 97 | 78 | 61 | 125 | 109 | 47 | 82 | | |
| 33 | 81 | 78 | 65 | 78 | 74 | 68 | 52 | | |
| 34 | 71 | 67 | 53 | 66 | 52 | 41 | 32 | | |
| 35 | 59 | 41 | 52 | 56 | 52 | 35 | 12 | | |
| 36 | 46 | 28 | 25 | 39 | 36 | 22 | | | |
| 37 | 32 | 43 | 25 | 35 | 30 | 20 | | | |
| 38 | 33 | 29 | 17 | 21 | 33 | 18 | | | |
| 39 | 27 | 26 | 9 | 17 | 19 | 12 | | | |
| 40 | 14 | 19 | 9 | 17 | 8 | 2 | | | |
| 41 | 12 | 14 | 9 | 7 | 9 | | | | |
| 42 | 8 | 6 | 4 | 4 | 7 | | | | |
| 43 | 6 | 8 | 3 | 3 | 3 | | | | |
| 44 | 2 | 3 | 4 | 3 | 3 | | | | |
| 45 | 2 | 4 | 3 | | 2 | | | | |
| 46 | | 2 | | | | | | | |
| 47 | 2 | | | | | | | | |
| 48 | | 1 | 1 | 1 | | | | | |
| 49 | | | | 1 | | | | | |
| 50 | 1 | | | | | | | | |
| Total | 2676 | 3251 | 2529 | 3470 | 3178 | 3055 | 3528 | 3767 | 1488 |
| Censored | 1250 | 1166 | 874 | 1158 | 984 | 790 | 762 | 1223 | 1297 |
| at age | 65 | 60 | 55 | 50 | 45 | 40 | 35 | 30 | 25 |
| Cohort size | 3926 | 4417 | 3403 | 4628 | 4162 | 3845 | 4290 | 4990 | 2785 |



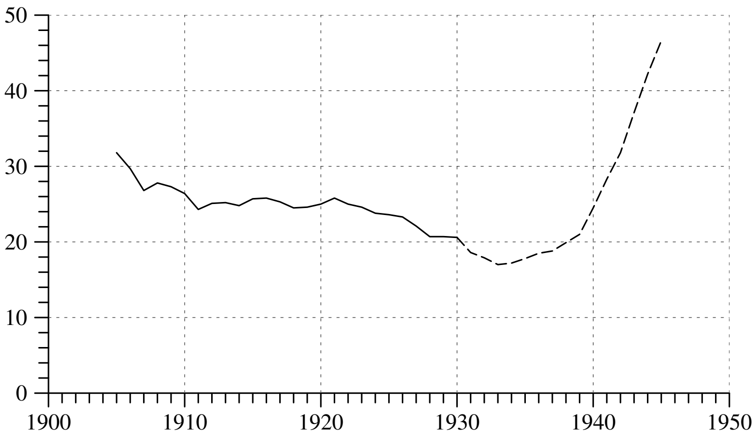



**Fig. 12.2-3** Percentage of women without marital children in the 1 % subsample of the 1970 census. The broken part of the curve can not be reliably estimated from the data.

child. This allows to calculate the proportion of women who finally remain without a marital child in the birth cohort C10 to be 26 %.

**2.** In the same way one can calculate the proportion of women without a marital child for each birth cohort. The result is shown in Figure 12.2-3. Obviously, at least until birth cohorts born around 1930 the proportion of childless women declined. For younger birth cohorts our data set does not allow any safe conclusions because these cohorts did not reach the end of the reproductive period in 1970. However, we will see in Chapter 14 that the trend of declining proportions of childless women continued until birth cohorts of women born around 1945.

**3.** Additional information can be gained by a consideration of the distribution of ages at first childbearing. This is easy because, as shown in Table 12.2-1, censored cases only occur in 1970. So one does not need the Kaplan-Meier procedure that was discussed in Section 8.3.4 but can directly calculate distribution and survivor functions. For example, referring again to C10, 298 out of 4417 women had a first child before age 20. The distribution function has therefore a value of $F(20) = 298/4417 = 0.067$, that is, about 7 % of women born in 1910 had a first marital child until age 20. For selected cohorts, corresponding survivor functions are shown in Figure 12.2-4. It is clearly seen that younger birth cohorts began childbearing at earlier ages. As discussed above, this was associated by a declining proportion of finally childless women. There is no reason, however, to believe this correlation to be stable through time. In fact, as will be shown later, a decline of the age at first childbearing can also be associated with an increase in the proportion of finally childless women.

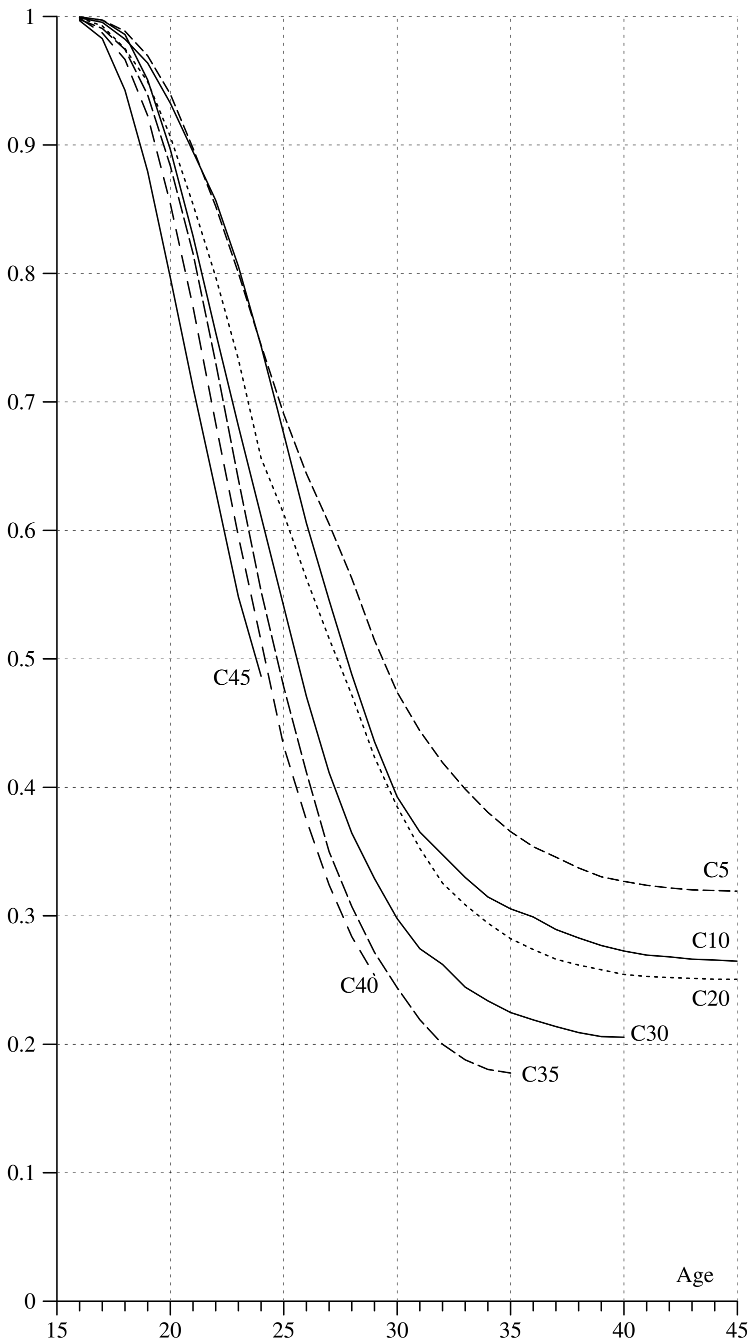



**Fig. 12.2-4** Survivor functions for the age at first marital birth, calculated from the 1 % subsample of the 1970 census.



**4.** While a plot of survivor functions, as shown in Figure 12.2-4, is well suited to compare a small number of cohorts, it becomes impractical for long time series. An alternative possibility to investigate changes in a series of distribution or survivor functions is based on the calculation of quantiles. An example is the median that was introduced in Section 7.2. Referring to a distribution function $F$, the median is a number, say $m$, such that $F(m) \approx 0.5$. By generalization, the *q-quantile* is defined as a number, say $m_q$, such that

$$F(m_q) \approx q$$

with the understanding that $q$ is some number strictly between 0 and 1. One possibility to calculate quantiles is by linear interpolation.[11] To illustrate, we calculate the median age at first childbearing for the C10 cohort. Using the data from Table 12.2-1, one finds $F(27) = 0.454$ and $F(28) = 0.512$. Therefore, by linear interpolation:

$$\frac{28 - 27}{0.512 - 0.454} = \frac{m - 27}{0.5 - 0.454}$$

from which can be derived the median $m = 27.8$. In the same way, one can calculate quantiles for any value of $q$ between 0 and 1. We have done this for the values

$$q = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$$

The result is shown in Table 12.2-2 for all birth cohorts born between 1905 and 1945. It is seen, for example, that the median age at first marital childbearing declined from 29.4, for birth cohort C5, to 23.8, for birth cohort C45. These quantiles can finally be presented graphically as shown in Figure 12.2-5.

**5.** Interpretations should keep in mind that our data set only records marital births. Results would be different if one would be able to include all first births, regardless of whether the mother is married or not. To provide an impression of the differences we use data from the *German Life History Study* (GLHS) for three birth cohorts, C20, C30, and C40.[12] Figure 12.2-6 compares the distributions; the solid lines refer to the 1 % subsample of the 1970 census, the dotted lines refer to the GLHS data set. Obviously, the proportion of childless women is much smaller than the proportion of women without a marital child.

---

[11] Since already the definition of quantiles relies on approximation, there exist several different methods to calculate quantiles. One should also note that statistical packages often use different formulas. For an overview see Hyndman and Fan (1996).

[12] The GLHS, and how we have done the calculations, will be discussed in Section 14.1.

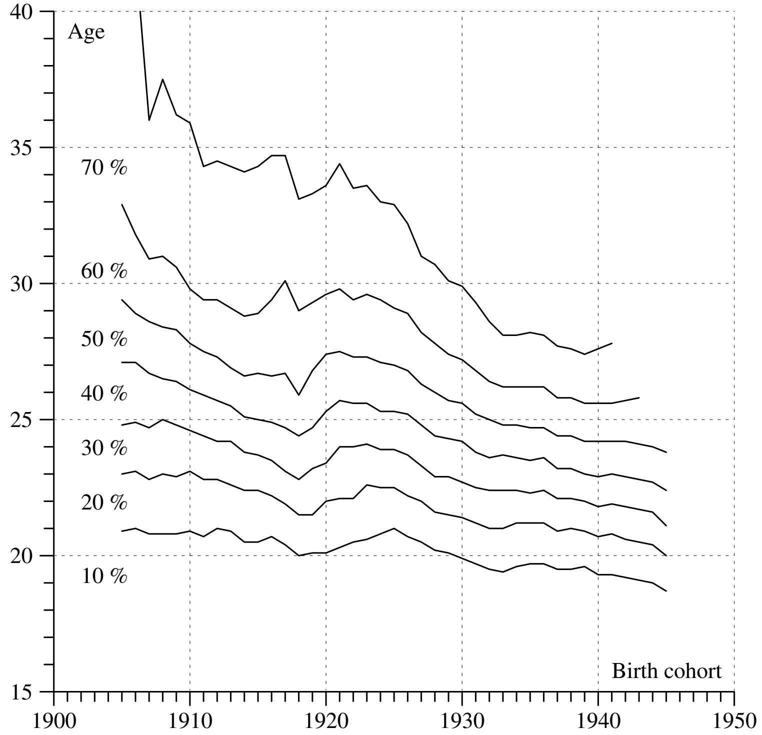



**Fig. 12.2-5** Graphical presentation of quantiles of the distribution of ages at first marital childbearing, calculated for 1-year cohorts with birth years between 1905 and 1945 from the 1 % subsample of the 1970 census.

### 12.2.3 Age-specific Birth Rates

**1.** Our next question concerns the number of children born. As was discussed in Chapter 11, some useful information can already be gained from age-specific cohort birth rates, defined as

$$\gamma_{t_0,\tau} := \frac{\text{number of children born of women at age } \tau}{\text{number of women at age } \tau}$$

with the understanding that both the numerator and the denominator refer to women born in the year $t_0$. If we assume that death rates do not depend on women's parity, such rates, restricted to martial births, can be calculated from the 1 % subsample of the 1970 census.[13] For a selection of birth cohorts, the required data are shown in Table 12.2-3. Each entry shows how many children were born of members of a specified birth cohort

[13] Actually, one also has to ignore migration. Therefore, in a strict sense, also the 1 % percent subsample of the 1970 census only allows to calculate quasi-cohort birth rates.



**Table 12.2-2** Quantiles of the distribution of ages at first marital childbearing, calculated from the 1 % subsample of the 1970 census.

| | Quantiles | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Cohort | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 |
| 1905 | 20.9 | 23.0 | 24.8 | 27.1 | 29.4 | 32.9 | | |
| 1906 | 21.0 | 23.1 | 24.9 | 27.1 | 28.9 | 31.8 | 42.1 | |
| 1907 | 20.8 | 22.8 | 24.7 | 26.7 | 28.6 | 30.9 | 36.0 | |
| 1908 | 20.8 | 23.0 | 25.0 | 26.5 | 28.4 | 31.0 | 37.5 | |
| 1909 | 20.8 | 22.9 | 24.8 | 26.4 | 28.3 | 30.6 | 36.2 | |
| 1910 | 20.9 | 23.1 | 24.6 | 26.1 | 27.8 | 29.8 | 35.9 | |
| 1911 | 20.7 | 22.8 | 24.4 | 25.9 | 27.5 | 29.4 | 34.3 | |
| 1912 | 21.0 | 22.8 | 24.2 | 25.7 | 27.3 | 29.4 | 34.5 | |
| 1913 | 20.9 | 22.6 | 24.2 | 25.5 | 26.9 | 29.1 | 34.3 | |
| 1914 | 20.5 | 22.4 | 23.8 | 25.1 | 26.6 | 28.8 | 34.1 | |
| 1915 | 20.5 | 22.4 | 23.7 | 25.0 | 26.7 | 28.9 | 34.3 | |
| 1916 | 20.7 | 22.2 | 23.5 | 24.9 | 26.6 | 29.4 | 34.7 | |
| 1917 | 20.4 | 21.9 | 23.1 | 24.7 | 26.7 | 30.1 | 34.7 | |
| 1918 | 20.0 | 21.5 | 22.8 | 24.4 | 25.9 | 29.0 | 33.1 | |
| 1919 | 20.1 | 21.5 | 23.2 | 24.7 | 26.8 | 29.3 | 33.3 | |
| 1920 | 20.1 | 22.0 | 23.4 | 25.3 | 27.4 | 29.6 | 33.6 | |
| 1921 | 20.3 | 22.1 | 24.0 | 25.7 | 27.5 | 29.8 | 34.4 | |
| 1922 | 20.5 | 22.1 | 24.0 | 25.6 | 27.3 | 29.4 | 33.5 | |
| 1923 | 20.6 | 22.6 | 24.1 | 25.6 | 27.3 | 29.6 | 33.6 | |
| 1924 | 20.8 | 22.5 | 23.9 | 25.3 | 27.1 | 29.4 | 33.0 | |
| 1925 | 21.0 | 22.5 | 23.9 | 25.3 | 27.0 | 29.1 | 32.9 | |
| 1926 | 20.7 | 22.2 | 23.7 | 25.2 | 26.8 | 28.9 | 32.2 | |
| 1927 | 20.5 | 22.0 | 23.3 | 24.8 | 26.3 | 28.2 | 31.0 | |
| 1928 | 20.2 | 21.6 | 22.9 | 24.4 | 26.0 | 27.8 | 30.7 | |
| 1929 | 20.1 | 21.5 | 22.9 | 24.3 | 25.7 | 27.4 | 30.1 | |
| 1930 | 19.9 | 21.4 | 22.7 | 24.2 | 25.6 | 27.2 | 29.9 | |
| 1931 | 19.7 | 21.2 | 22.5 | 23.8 | 25.2 | 26.8 | 29.3 | 35.2 |
| 1932 | 19.5 | 21.0 | 22.4 | 23.6 | 25.0 | 26.4 | 28.6 | 34.0 |
| 1933 | 19.4 | 21.0 | 22.4 | 23.7 | 24.8 | 26.2 | 28.1 | 33.0 |
| 1934 | 19.6 | 21.2 | 22.4 | 23.6 | 24.8 | 26.2 | 28.1 | 32.4 |
| 1935 | 19.7 | 21.2 | 22.3 | 23.5 | 24.7 | 26.2 | 28.2 | 32.0 |
| 1936 | 19.7 | 21.2 | 22.4 | 23.6 | 24.7 | 26.2 | 28.1 | 32.1 |
| 1937 | 19.5 | 20.9 | 22.1 | 23.2 | 24.4 | 25.8 | 27.7 | 31.5 |
| 1938 | 19.5 | 21.0 | 22.1 | 23.2 | 24.4 | 25.8 | 27.6 | |
| 1939 | 19.6 | 20.9 | 22.0 | 23.0 | 24.2 | 25.6 | 27.4 | |
| 1940 | 19.3 | 20.7 | 21.8 | 22.9 | 24.2 | 25.6 | 27.6 | |
| 1941 | 19.3 | 20.8 | 21.9 | 23.0 | 24.2 | 25.6 | 27.8 | |
| 1942 | 19.2 | 20.6 | 21.8 | 22.9 | 24.2 | 25.7 | | |
| 1943 | 19.1 | 20.5 | 21.7 | 22.8 | 24.1 | 25.8 | | |
| 1944 | 19.0 | 20.4 | 21.6 | 22.7 | 24.0 | | | |
| 1945 | 18.7 | 20.0 | 21.1 | 22.4 | 23.8 | | | |

at a specified age. For example, out of 3845 women belonging to birth cohort C30, 505 reported to have born a marital child at the age of 25. Thus, given the above mentioned assumption, one gets the approximation

$$\gamma_{1930,25} \approx \frac{505}{3845} = 0.1313$$

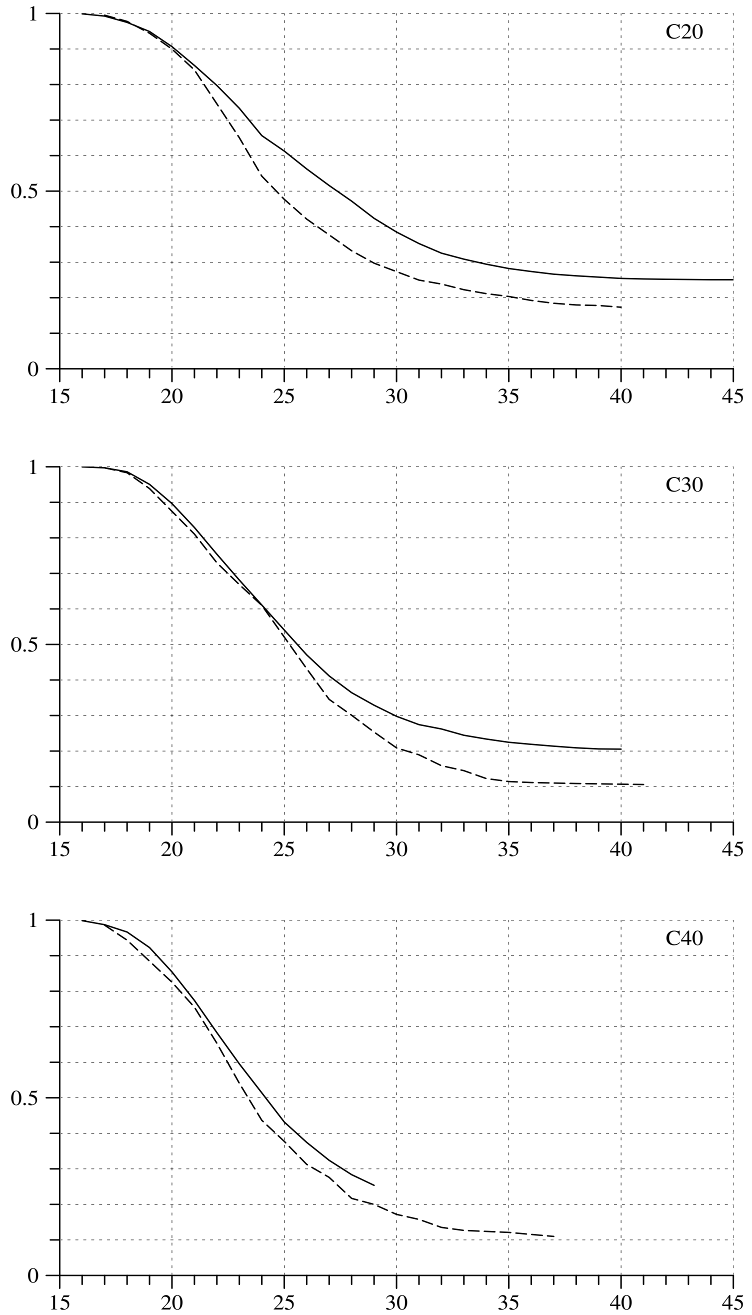



**Fig. 12.2-6** Comparison of distributions of the age at first (marital) birth, calculated from the 1 % subsample of the 1970 census (solid lines) and from the GLHS data (dotted lines).



This figure can also be compared with period data for the year 1955. As reported in Fachserie 1, Reihe 1, 1999 (p. 198), the birth rate of women at age 25 in 1955 was

$$\beta_{1955,25} = 0.1394$$

The difference of about 6 % can be attributed to the fact that the period data include all births and are not restricted to women having a German citizenship.

**2.** Again for birth cohort C30, Figure 12.2-7 compares the distribution of these rates. The solid and dotted lines show, respectively, the quantities

$$\frac{\gamma_{1930,\tau}}{\sum_{j=15}^{39} \gamma_{1930,j}} \quad \text{and} \quad \frac{\beta_{1930,\tau}}{\sum_{j=15}^{39} \beta_{1930,j}}$$

The differences between the curves that occur at younger ages might indicate that non-marital births are more frequent in these ages.

**3.** It remains the question how to graphically present the age-specific cohort birth rates for a whole sequence of birth cohorts. One possibility is to first calculate cumulated cohort birth rates, $\bar{\gamma}_{t_0,\tau}$, and then to plot values for specified ages. The required data are shown in Table 12.2-4. From these data one can calculate the cumulated cohort birth rates

$$\bar{\gamma}_{t_0,25},\ \bar{\gamma}_{t_0,30},\ \bar{\gamma}_{t_0,35},\ \bar{\gamma}_{t_0,40},\ \bar{\gamma}_{t_0,45}$$

for birth cohorts $t_0 = \text{C5}, \ldots, \text{C45}$. These rates can finally be presented graphically as shown in Figure 12.2-8.

### 12.2.4 Number of Children

**1.** Cumulated cohort birth rates refer to all children born of all members of a birth cohort until some specified age and therefore do not provide information about the distribution of the number of children among the cohort members. The latter distribution requires to calculate, separately for each birth cohort, the proportion of women with distinct numbers of children. This has been done in Table 12.2-5. For example, altogether there are 3926 women in the 1 % subsample of the 1970 census born in the year 1905. Of these, 1250 have no marital child, 828 have one marital child, 797 have two marital children, and so on. The total number of children born of these women is 6713 (equal to the number of children shown in Table 12.2-3). One should notice that these figures refer to the interview date in 1970. For birth cohorts born after about 1930, both the absolute numbers and the proportions will probably change until the end of the reproductive period.

**2.** The figures in Table 12.2-5 can be used to calculate the mean number

**Table 12.2-3** Number of marital children, born of women belonging to the specified birth cohort in the specified age; calculated from the 1 % subsample of the 1970 census.

| $\tau$ | C5 | C10 | C15 | C20 | C25 | C30 | C35 | C40 | C45 |
|---|---|---|---|---|---|---|---|---|---|
| 15 | | | | | 1 | | | | |
| 16 | 3 | 5 | 1 | 5 | 7 | 1 | 10 | 7 | 9 |
| 17 | 8 | 16 | 19 | 29 | 12 | 10 | 32 | 57 | 42 |
| 18 | 36 | 61 | 36 | 88 | 35 | 44 | 76 | 118 | 116 |
| 19 | 79 | 97 | 89 | 133 | 76 | 140 | 169 | 260 | 209 |
| 20 | 141 | 173 | 150 | 243 | 126 | 230 | 274 | 416 | 305 |
| 21 | 198 | 215 | 210 | 326 | 213 | 324 | 380 | 553 | 374 |
| 22 | 261 | 242 | 257 | 336 | 332 | 402 | 507 | 683 | 379 |
| 23 | 310 | 338 | 358 | 422 | 429 | 445 | 595 | 768 | 414 |
| 24 | 375 | 437 | 451 | 524 | 444 | 488 | 646 | 801 | 358 |
| 25 | 378 | 503 | 484 | 368 | 473 | 505 | 657 | 834 | 130 |
| 26 | 354 | 568 | 476 | 423 | 474 | 541 | 688 | 730 | |
| 27 | 338 | 568 | 331 | 452 | 488 | 555 | 700 | 743 | |
| 28 | 353 | 596 | 389 | 482 | 467 | 496 | 628 | 634 | |
| 29 | 432 | 618 | 356 | 522 | 474 | 485 | 597 | 558 | |
| 30 | 426 | 573 | 223 | 463 | 459 | 437 | 503 | 193 | |
| 31 | 393 | 510 | 247 | 428 | 383 | 424 | 474 | | |
| 32 | 386 | 383 | 237 | 408 | 389 | 339 | 408 | | |
| 33 | 374 | 368 | 262 | 372 | 367 | 338 | 339 | | |
| 34 | 382 | 341 | 256 | 315 | 319 | 269 | 269 | | |
| 35 | 321 | 193 | 221 | 277 | 289 | 225 | 86 | | |
| 36 | 273 | 183 | 187 | 232 | 233 | 218 | | | |
| 37 | 207 | 193 | 155 | 217 | 187 | 156 | | | |
| 38 | 200 | 167 | 112 | 186 | 156 | 139 | | | |
| 39 | 160 | 159 | 104 | 142 | 142 | 111 | | | |
| 40 | 103 | 118 | 90 | 99 | 91 | 28 | | | |
| 41 | 66 | 95 | 71 | 79 | 64 | | | | |
| 42 | 54 | 61 | 60 | 48 | 51 | | | | |
| 43 | 50 | 40 | 30 | 22 | 23 | | | | |
| 44 | 28 | 29 | 25 | 27 | 17 | | | | |
| 45 | 12 | 15 | 13 | 9 | 7 | | | | |
| 46 | 5 | 8 | 5 | 4 | | | | | |
| 47 | 3 | 2 | 2 | 2 | | | | | |
| 48 | 2 | 3 | 2 | 3 | | | | | |
| 49 | 1 | 4 | 1 | 2 | | | | | |
| 50 | 1 | 1 | | | | | | | |
| 51 | | 1 | | | | | | | |
| Total | 6713 | 7884 | 5910 | 7688 | 7228 | 7350 | 8038 | 7355 | 2336 |
| Cohort size | 3926 | 4417 | 3403 | 4628 | 4162 | 3845 | 4290 | 4990 | 2785 |



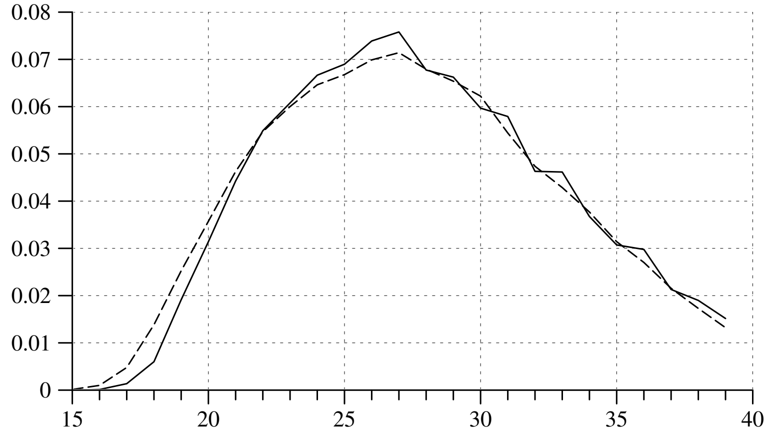



**Fig. 12.2-7** Age-specific quasi-cohort birth rates for birth cohort C30, calculated from the 1 % subsample of the 1970 census (solid line) and from Table 11.4-1 (dotted line). The ordinate shows proportions as explained in the text.

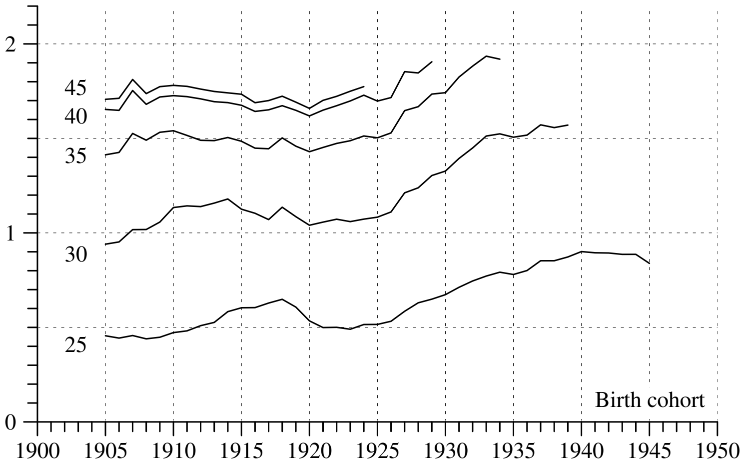



**Fig. 12.2-8** Cumulated cohort birth rates until specified ages, based on the data in Table 12.2-4.

of marital children per women. If one takes into account only women with at least one marital child, this mean value varies between 2.2 and 2.5, but does not show any substantial trend. Changes become visible, however, if one investigates the distribution of the number of children. The proportions can easily be calculated from the data in Table 12.2-5 and their development is graphically presented in Figure 12.2-9.

**Table 12.2-4** Number of women belonging to specified birth cohorts, and number of children born of these women until specified age, calculated from the 1 % subsample of the 1970 census.

| | | Children until age | | | | |
|---|---|---|---|---|---|---|
| Birth year | Cohort size | 25 | 30 | 35 | 40 | 45 |
| 1905 | 3926 | 1789 | 3692 | 5548 | 6491 | 6701 |
| 1906 | 4045 | 1794 | 3852 | 5768 | 6668 | 6926 |
| 1907 | 4208 | 1923 | 4281 | 6422 | 7378 | 7624 |
| 1908 | 4430 | 1947 | 4510 | 6603 | 7444 | 7695 |
| 1909 | 4395 | 1969 | 4647 | 6735 | 7556 | 7796 |
| 1910 | 4417 | 2087 | 5010 | 6805 | 7625 | 7865 |
| 1911 | 4349 | 2095 | 4971 | 6592 | 7485 | 7721 |
| 1912 | 4511 | 2297 | 5139 | 6721 | 7710 | 7944 |
| 1913 | 4377 | 2304 | 5067 | 6514 | 7416 | 7652 |
| 1914 | 4301 | 2509 | 5073 | 6472 | 7266 | 7488 |
| 1915 | 3403 | 2055 | 3830 | 5053 | 5701 | 5900 |
| 1916 | 2578 | 1559 | 2846 | 3734 | 4235 | 4354 |
| 1917 | 2428 | 1527 | 2599 | 3509 | 4009 | 4128 |
| 1918 | 2465 | 1599 | 2800 | 3704 | 4124 | 4248 |
| 1919 | 3595 | 2184 | 3904 | 5245 | 5927 | 6082 |
| 1920 | 4628 | 2474 | 4816 | 6616 | 7492 | 7677 |
| 1921 | 4680 | 2336 | 4947 | 6796 | 7720 | 7962 |
| 1922 | 4442 | 2223 | 4764 | 6545 | 7431 | 7652 |
| 1923 | 4265 | 2091 | 4520 | 6345 | 7242 | 7462 |
| 1924 | 4058 | 2091 | 4354 | 6139 | 7015 | 7199 |
| 1925 | 4162 | 2147 | 4509 | 6256 | 7065 | |
| 1926 | 3998 | 2129 | 4441 | 6113 | 6861 | |
| 1927 | 3863 | 2261 | 4682 | 6362 | 7159 | |
| 1928 | 3923 | 2472 | 4859 | 6543 | 7243 | |
| 1929 | 3762 | 2443 | 4904 | 6525 | 7168 | |
| 1930 | 3845 | 2589 | 5103 | 6698 | | |
| 1931 | 3535 | 2518 | 4928 | 6450 | | |
| 1932 | 3444 | 2568 | 4992 | 6483 | | |
| 1933 | 3280 | 2531 | 4961 | 6347 | | |
| 1934 | 4036 | 3197 | 6151 | 7746 | | |
| 1935 | 4290 | 3346 | 6462 | | | |
| 1936 | 4348 | 3482 | 6598 | | | |
| 1937 | 4302 | 3668 | 6762 | | | |
| 1938 | 4715 | 4020 | 7342 | | | |
| 1939 | 5044 | 4402 | 7921 | | | |
| 1940 | 4990 | 4497 | | | | |
| 1941 | 4562 | 4082 | | | | |
| 1942 | 3771 | 3370 | | | | |
| 1943 | 3842 | 3406 | | | | |
| 1944 | 3820 | 3387 | | | | |
| 1945 | 2785 | 2336 | | | | |



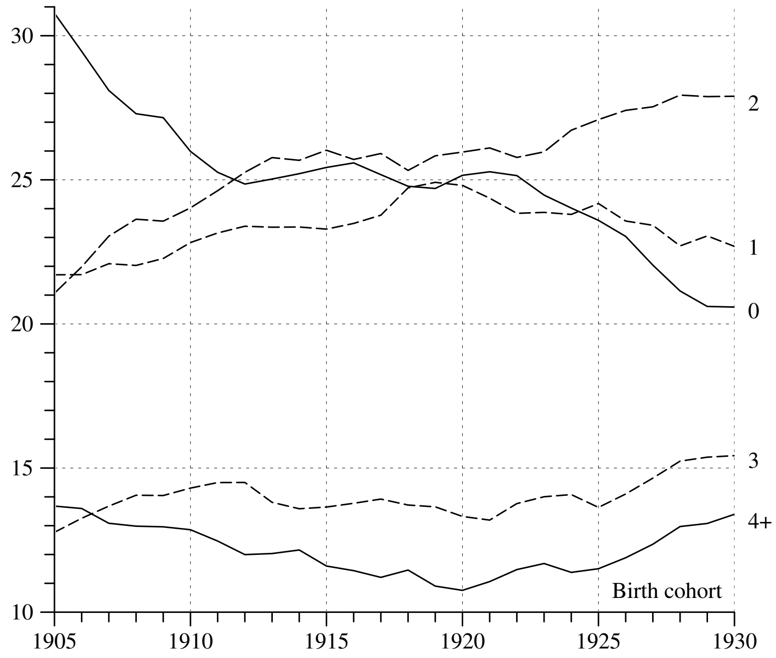



**Fig. 12.2-9** Proportion (in %) of women born in specified years with 0, 1, 2, 3, and 4 or more children; calculated from the data in Table 12.2-5.

## 12.2.5 Timing of Births

**1.** A final question concerns the timing of childbearing in women's life courses. One aspect of this question, the age at first childbearing, has been discussed in Section 12.2.2. We now discuss two further aspects. One of them concerns the temporal distance between the births of several children, often called the *spacing of childbearing*. Another one concerns the idea that there might be a relationship between age at first childbearing and the total number of children born until the end of the reproductive period.

**2.** As in the preceding sections, calculations are based on the 1 % subsample of the 1970 census. For all women with at least two children (excluding twins) we can calculate the temporal distance between the two births. Similarly, for all women with at least three children one can calculate the temporal distances between the first and the third and between the second and the third birth. Additional temporal intervals can be calculated for women with at least four children. Results of these calculations are shown in Table 12.2-6. As can be seen, at least for the birth cohorts C5, . . . , C30, there are virtually no changes in the spacing of childbearing. No conclu-

**Table 12.2-5** Number of women in the 1 % subsample of the 1970 census, and number of children born of these women, classified according to women's birth cohort.

| | | Number of children | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Birth year | Cohort size | 0 | 1 | 2 | 3 | 4 | 5+ | Total |
| 1905 | 3926 | 1250 | 828 | 797 | 505 | 266 | 280 | 6713 |
| 1906 | 4045 | 1202 | 903 | 883 | 513 | 268 | 276 | 6945 |
| 1907 | 4208 | 1128 | 914 | 1002 | 599 | 258 | 307 | 7640 |
| 1908 | 4430 | 1230 | 985 | 1041 | 626 | 272 | 276 | 7711 |
| 1909 | 4395 | 1200 | 973 | 1037 | 607 | 299 | 279 | 7812 |
| 1910 | 4417 | 1166 | 991 | 1043 | 627 | 317 | 273 | 7884 |
| 1911 | 4349 | 1055 | 1039 | 1082 | 648 | 263 | 262 | 7739 |
| 1912 | 4511 | 1133 | 1044 | 1145 | 649 | 282 | 258 | 7955 |
| 1913 | 4377 | 1102 | 1013 | 1117 | 622 | 264 | 259 | 7664 |
| 1914 | 4301 | 1066 | 1023 | 1136 | 552 | 294 | 230 | 7506 |
| 1915 | 3403 | 874 | 788 | 854 | 467 | 224 | 196 | 5910 |
| 1916 | 2578 | 666 | 591 | 685 | 371 | 154 | 111 | 4358 |
| 1917 | 2428 | 613 | 592 | 618 | 321 | 160 | 124 | 4134 |
| 1918 | 2465 | 603 | 592 | 634 | 349 | 174 | 113 | 4256 |
| 1919 | 3595 | 885 | 926 | 892 | 495 | 222 | 175 | 6093 |
| 1920 | 4628 | 1158 | 1155 | 1248 | 603 | 270 | 194 | 7688 |
| 1921 | 4680 | 1209 | 1109 | 1222 | 616 | 279 | 245 | 7978 |
| 1922 | 4442 | 1110 | 1085 | 1121 | 595 | 305 | 226 | 7665 |
| 1923 | 4265 | 1050 | 997 | 1108 | 629 | 249 | 232 | 7467 |
| 1924 | 4058 | 966 | 966 | 1083 | 563 | 265 | 215 | 7203 |
| 1925 | 4162 | 984 | 1008 | 1144 | 567 | 252 | 207 | 7228 |
| 1926 | 3998 | 933 | 980 | 1083 | 536 | 265 | 201 | 6981 |
| 1927 | 3863 | 855 | 849 | 1068 | 590 | 250 | 251 | 7283 |
| 1928 | 3923 | 810 | 933 | 1093 | 599 | 267 | 221 | 7313 |
| 1929 | 3762 | 777 | 841 | 1065 | 571 | 253 | 255 | 7199 |
| 1930 | 3845 | 789 | 885 | 1057 | 603 | 278 | 233 | 7350 |
| 1931 | 3535 | 656 | 785 | 1067 | 544 | 264 | 219 | 6923 |
| 1932 | 3444 | 617 | 727 | 1050 | 579 | 250 | 221 | 6852 |
| 1933 | 3280 | 557 | 714 | 990 | 553 | 278 | 188 | 6552 |
| 1934 | 4036 | 694 | 878 | 1290 | 695 | 267 | 212 | 7824 |
| 1935 | 4290 | 762 | 989 | 1345 | 735 | 276 | 183 | 8038 |
| 1936 | 4348 | 802 | 1020 | 1428 | 677 | 247 | 174 | 7886 |
| 1937 | 4302 | 807 | 1036 | 1376 | 700 | 239 | 144 | 7653 |
| 1938 | 4715 | 939 | 1172 | 1570 | 678 | 253 | 103 | 7927 |
| 1939 | 5044 | 1060 | 1323 | 1682 | 676 | 205 | 98 | 8063 |
| 1940 | 4990 | 1223 | 1377 | 1535 | 624 | 163 | 68 | 7355 |
| 1941 | 4562 | 1291 | 1273 | 1383 | 453 | 113 | 49 | 6119 |
| 1942 | 3771 | 1199 | 1147 | 1010 | 314 | 78 | 23 | 4543 |
| 1943 | 3842 | 1427 | 1169 | 899 | 283 | 43 | 21 | 4102 |
| 1944 | 3820 | 1613 | 1179 | 764 | 225 | 32 | 7 | 3549 |
| 1945 | 2785 | 1297 | 841 | 489 | 123 | 27 | 8 | 2336 |



**Table 12.2-6** Temporal distance in years between the $i$-th and the $j$-th birth, for birth cohorts $C5, \ldots, C30$; calculated from the 1 % subsample of the 1970 census.

| Birth year | 1 – 2 | 2 – 3 | 3 – 4 | 1 – 3 | 1 – 4 |
|---|---|---|---|---|---|
| 1905 | 5.1 | 5.3 | 4.8 | 8.7 | 10.9 |
| 1906 | 5.2 | 5.1 | 4.7 | 8.3 | 11.2 |
| 1907 | 5.1 | 5.0 | 4.7 | 8.4 | 10.7 |
| 1908 | 5.0 | 5.0 | 4.4 | 8.5 | 10.7 |
| 1909 | 5.0 | 4.8 | 4.2 | 8.2 | 10.4 |
| 1910 | 4.9 | 4.9 | 4.6 | 8.1 | 10.5 |
| 1911 | 5.0 | 4.8 | 4.3 | 8.1 | 10.4 |
| 1912 | 4.9 | 4.9 | 4.8 | 8.1 | 10.7 |
| 1913 | 5.0 | 5.1 | 4.9 | 8.4 | 11.1 |
| 1914 | 4.7 | 4.9 | 4.9 | 8.0 | 10.6 |
| 1915 | 4.9 | 5.1 | 4.9 | 8.2 | 10.6 |
| 1916 | 4.9 | 5.0 | 4.5 | 8.3 | 10.4 |
| 1917 | 4.9 | 4.8 | 4.8 | 8.2 | 10.7 |
| 1918 | 4.9 | 4.8 | 4.7 | 8.3 | 10.9 |
| 1919 | 4.8 | 5.1 | 4.9 | 8.6 | 11.0 |
| 1920 | 4.9 | 4.9 | 4.8 | 8.5 | 11.1 |
| 1921 | 5.0 | 5.2 | 4.7 | 8.5 | 10.6 |
| 1922 | 4.8 | 5.0 | 5.1 | 8.3 | 11.6 |
| 1923 | 4.8 | 5.1 | 4.7 | 8.4 | 10.8 |
| 1924 | 5.0 | 5.1 | 4.6 | 8.4 | 10.6 |
| 1925 | 4.9 | 5.1 | 4.8 | 8.2 | 10.7 |
| 1926 | 4.8 | 4.9 | 5.0 | 8.1 | 10.9 |
| 1927 | 5.0 | 5.1 | 4.5 | 8.3 | 10.3 |
| 1928 | 4.8 | 4.9 | 4.5 | 8.1 | 10.6 |
| 1929 | 4.9 | 4.8 | 4.3 | 8.0 | 10.2 |
| 1930 | 5.0 | 4.8 | 4.1 | 8.0 | 9.9 |

sions can be derived, however, for younger birth cohorts.[14]

**3.** Also based on data from the 10 % subsample of the 1970 census, similar calculations have been performed by Rückert (1975). Instead of birth cohorts, Rückert considers marriage cohorts of women in their first marriage. He therefore finds slightly shorter distances between successive births. For example, he finds a mean duration of 4 years between the birth of the first and second child, for women who married between 1940 and 1949. However, also Rückert's figures show that there have been virtually no changes in the mean durations between successive births at least for marriage cohorts of women who married between 1920 and 1949.

**4.** Rückert also investigated possible relationships between the spacing of childbearing and the final number of children born of women in their first

[14] Performing the same calculations for younger birth cohorts would result in a substantial selection bias. For example, the temporal distance between the first and second child for birth cohort C45 is just 2.5 years. But this is most probably due to the fact that members of this cohort are only observed until an age of 25 years.

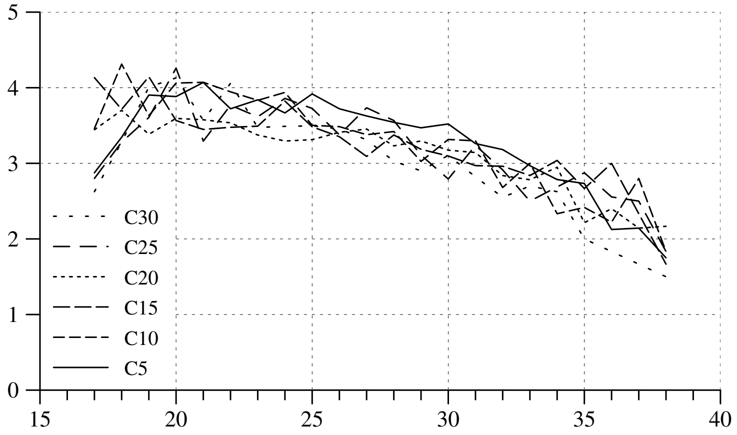



**Fig. 12.2-10** Mean number of marital children (ordinate) by age at first marital childbearing (abscissa), for birth cohorts C5, C10, C15, C20, C25, and C30, calculated from the 1 % subsample of the 1970 census.

marriage. He found that the mean interval lengths between successive births decreases with increasing completed marital fertility.[15] It is, however, questionable how to interpret this result. There is obviously no reason to assume a causal relationship. How many children a women will eventually bear does not depend on the temporal distance between her first and second child. On the other hand, given a limited period of childbearing and conditioning on the number of eventually born children, it seems not surprising to find relatively shorter distances between successive births for women who finally give birth to more children.

**5.** A similar problem occurs when one tries to find relationships between the age at first childbearing and the final number of children born. Figure 12.2-10 provides an illustration. The relationship is quite similar for all birth cohorts C5, ..., C30. Women who began childbearing at younger ages finally gave birth to relatively more children. But again, except for cases where reaching the end of the reproductive period creates definite limits to childbearing, there is no obvious causal relationship. Moreover, the relationship is actually not so stable as suggested by Figure 12.2-10. While this can not be demonstrated with the data from the 1 % subsample of the 1970 census, some additional information can be gained from period data of official statistics. Given age-specific birth rates, $\beta_{t,\tau}$, they can be used to calculate the mean age at childbearing for quasi-cohorts in the

[15] „Es gilt offensichtlich allgemein, daß der durchschnittliche Geburtenabstand um so kürzer ist, je größer die Kinderzahlen in den Ehen nach abgeschlossener Familienbildung sind.“ (Rückert 1975, p. 87)



**Table 12.2-7** Mean age at childbearing ($\bar{\tau}_{t_0}$) and until age 40 cumulated cohort birth rates ($\bar{\beta}_{t_0}$, per 1000) for birth cohorts $t_0$. Calculated from age-specific birth rates in Fachserie 1, Reihe 1, 1999 (pp. 198 -200).

| $t_0$ | $\bar{\tau}_{t_0}$ | $\bar{\beta}_{t_0}$ | $t_0$ | $\bar{\tau}_{t_0}$ | $\bar{\beta}_{t_0}$ | $t_0$ | $\bar{\tau}_{t_0}$ | $\bar{\beta}_{t_0}$ |
|---|---|---|---|---|---|---|---|---|
| 1930 | 27.7 | 2107.3 | 1940 | 26.1 | 1958.8 | 1950 | 26.1 | 1684.5 |
| 1931 | 27.7 | 2133.7 | 1941 | 26.0 | 1891.2 | 1951 | 26.2 | 1642.6 |
| 1932 | 27.6 | 2173.4 | 1942 | 25.9 | 1837.5 | 1952 | 26.4 | 1630.8 |
| 1933 | 27.4 | 2201.4 | 1943 | 25.7 | 1797.2 | 1953 | 26.6 | 1612.9 |
| 1934 | 27.1 | 2220.7 | 1944 | 25.6 | 1765.1 | 1954 | 26.8 | 1589.0 |
| 1935 | 27.1 | 2155.7 | 1945 | 25.5 | 1761.4 | 1955 | 26.9 | 1604.0 |
| 1936 | 26.9 | 2120.3 | 1946 | 25.5 | 1765.3 | 1956 | 27.1 | 1599.7 |
| 1937 | 26.7 | 2095.1 | 1947 | 25.6 | 1738.0 | 1957 | 27.3 | 1582.5 |
| 1938 | 26.5 | 2056.6 | 1948 | 25.7 | 1714.2 | 1958 | 27.5 | 1585.2 |
| 1939 | 26.3 | 2012.1 | 1949 | 25.9 | 1700.0 | 1959 | 27.6 | 1581.4 |

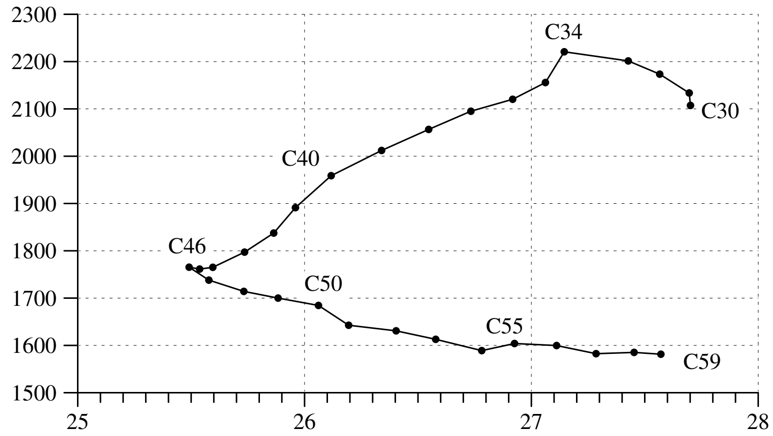



**Fig. 12.2-11** Plot of the data in Table 12.2-7. The abscissa refers to mean age at childbearing, the ordinate refers to until age 40 cumulated age-specific birth rates (per 1000).

following way:

$$\bar{\tau}_{t_0} := \frac{\sum_{\tau=\tau_a}^{\tau_b} \tau\, \beta_{t_0+\tau,\tau}}{\sum_{\tau=\tau_a}^{\tau_b} \beta_{t_0+\tau,\tau}}$$

For each cohort $t_0$, one can also calculate the cumulated cohort birth rate

$$\bar{\beta}_{t_0} := \sum_{\tau=\tau_a}^{\tau_b} \beta_{t_0+\tau,\tau}$$

so that it becomes possible to investigate changes in the relationship between the two quantities across cohorts. The age-specific birth rates published by the *Statistisches Bundesamt* allow to calculate these quantities for $t_0 = 1930, \ldots, 1959$, assuming $\tau_a = 15$ and $\tau_b = 40$.[16] Results of the calculation are shown in Table 12.2-7. The graphical view of these data in Figure 12.2-11 clearly shows that there is no simple relationship between mean age at first childbearing and cumulated birth rates.

[16]Data are taken from Fachserie 1, Reihe 1, 1999 (pp. 198-200).

# Chapter 13

# Births in the Period 1950–1970

In the previous chapter, the presentation of data from a 1 % percent subsample of the 1970 census focused on birth cohorts. This is useful for an understanding of historical changes but has limitations. The subjects of historical change are individuals, not birth cohorts. Birth cohorts are just analytical tools for the presentation of data related to life courses of individuals. These life courses are not, however, determined by a specific birth year but depend on the changing historical contexts in which they develop. The cohort approach therefore has to incorporate historical periods. There is, however, no direct connection both for a technical and a substantial reason. The technical reason refers to the fact that cohorts contribute to the whole range of historical periods during which their members live. The more substantial reason refers to the fact that the starting point for an understanding of individual behavior is a historical period from which, possibly, substantial differences between successive birth cohorts result.[1] In order to understand the relationship between cohorts and periods, we consider, in the present chapter, the development of births in the period 1950–70. As in the previous chapter, the data source is the 1 % subsample of the 1970 census.

## 13.1 Age-specific Birth Rates

**1.** In the territory of the former FRG, a substantial increase in the number of births began in the mid-fifties. This increase, sometimes called "baby boom", lasted until about the mid-sixties and was followed by a long-term decline in the number of births. In the present chapter we try to reconstruct this development, for the period 1950–70, with the data of the 1 % subsample of the 1970 census. As was shown in Figure 12.2.1-2 in Section 12.2.1, most of the births that occurred in this period were contributed by women represented in our data set.

**2.** We begin with an investigation of age-specific birth rates. Using our standard notation, the age-specific birth rate for age $\tau$ in year $t$ is defined as $\beta_{t,\tau} = b_{t,\tau}/n^f_{t,\tau}$. The denominator refers to the number of women who are of age $\tau$ in the year $t$, and the numerator refers to the number of children

[1]We therefore do not follow Ryder's (1964) idea of a "process of demographic translation". From a technical point of view this simply means to derive period measures from cohort measures of people's reproductive behavior. But this way to set up the problem confuses the order in which the facts from which cohort measures are statistically derived are brought about by people's behavior which always takes place in specific and changing historical periods.

**Table 13.1-1** Number of marital children born of mothers of specified age in specified year (1 % subsample of the 1970 census).

| $\tau$ | 1947 | 1948 | 1949 | 1950 | 1951 | 1952 | 1953 | 1954 | 1955 | 1956 | 1957 | 1958 | 1959 | 1960 | 1961 | 1962 | 1963 | 1964 | 1965 | 1966 | 1967 | 1968 | 1969 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 0 | 3 | 6 | 6 | 10 | 5 | 6 | 7 | 3 | 7 | 5 | 4 | 10 | 10 | 9 | 16 | 9 | 15 | 15 | 12 | 7 | 9 | 11 |
| 17 | 10 | 18 | 18 | 31 | 32 | 32 | 29 | 25 | 31 | 33 | 57 | 36 | 38 | 43 | 53 | 42 | 61 | 60 | 60 | 67 | 75 | 63 | 81 |
| 18 | 51 | 44 | 68 | 76 | 86 | 91 | 76 | 88 | 101 | 103 | 107 | 118 | 118 | 110 | 115 | 115 | 116 | 128 | 176 | 174 | 173 | 197 | 171 |
| 19 | 121 | 121 | 140 | 149 | 168 | 144 | 181 | 169 | 180 | 186 | 236 | 235 | 260 | 238 | 205 | 221 | 254 | 209 | 254 | 266 | 295 | 297 | 297 |
| 20 | 171 | 216 | 219 | 230 | 243 | 270 | 246 | 251 | 274 | 284 | 338 | 344 | 355 | 416 | 370 | 329 | 357 | 342 | 305 | 398 | 409 | 420 | 440 |
| 21 | 254 | 278 | 308 | 314 | 324 | 306 | 290 | 288 | 349 | 380 | 406 | 443 | 495 | 541 | 553 | 485 | 413 | 437 | 425 | 374 | 466 | 462 | 451 |
| 22 | 332 | 367 | 344 | 406 | 374 | 402 | 369 | 345 | 366 | 462 | 507 | 486 | 544 | 588 | 642 | 683 | 613 | 508 | 508 | 520 | 379 | 474 | 459 |
| 23 | 390 | 429 | 358 | 444 | 431 | 404 | 445 | 426 | 447 | 409 | 585 | 595 | 616 | 645 | 705 | 805 | 768 | 712 | 575 | 598 | 578 | 414 | 470 |
| 24 | 434 | 451 | 444 | 418 | 411 | 464 | 441 | 488 | 460 | 461 | 439 | 592 | 646 | 656 | 677 | 738 | 859 | 801 | 736 | 584 | 564 | 564 | 358 |
| 25 | 468 | 499 | 482 | 473 | 451 | 480 | 477 | 509 | 505 | 479 | 490 | 516 | 648 | 657 | 732 | 703 | 773 | 822 | 834 | 769 | 604 | 553 | 526 |
| 26 | 523 | 494 | 505 | 446 | 474 | 469 | 498 | 503 | 516 | 541 | 502 | 554 | 530 | 623 | 688 | 670 | 708 | 778 | 802 | 730 | 684 | 492 | 503 |
| 27 | 452 | 518 | 539 | 503 | 478 | 488 | 483 | 512 | 482 | 531 | 555 | 508 | 509 | 522 | 619 | 700 | 732 | 687 | 784 | 791 | 743 | 637 | 481 |
| 28 | 353 | 482 | 530 | 540 | 500 | 471 | 467 | 476 | 457 | 487 | 504 | 496 | 499 | 480 | 515 | 600 | 628 | 664 | 625 | 672 | 721 | 634 | 541 |
| 29 | 223 | 329 | 522 | 530 | 480 | 455 | 433 | 474 | 457 | 494 | 500 | 490 | 485 | 441 | 456 | 463 | 587 | 597 | 544 | 568 | 559 | 630 | 558 |
| 30 | 209 | 228 | 395 | 463 | 510 | 488 | 466 | 435 | 459 | 427 | 460 | 415 | 420 | 437 | 460 | 425 | 400 | 525 | 503 | 506 | 506 | 529 | 575 |
| 31 | 188 | 210 | 211 | 333 | 428 | 472 | 416 | 425 | 396 | 383 | 406 | 440 | 408 | 406 | 424 | 410 | 399 | 363 | 425 | 474 | 467 | 415 | 435 |
| 32 | 237 | 201 | 216 | 216 | 282 | 408 | 375 | 383 | 383 | 410 | 389 | 347 | 395 | 362 | 366 | 339 | 335 | 331 | 316 | 363 | 408 | 417 | 335 |
| 33 | 294 | 262 | 173 | 174 | 183 | 285 | 372 | 379 | 336 | 376 | 367 | 367 | 360 | 315 | 320 | 306 | 338 | 337 | 295 | 260 | 316 | 339 | 288 |
| 34 | 286 | 286 | 256 | 178 | 167 | 164 | 228 | 315 | 318 | 328 | 319 | 319 | 319 | 293 | 264 | 304 | 294 | 269 | 233 | 251 | 221 | 281 | 269 |
| 35 | 245 | 261 | 274 | 221 | 148 | 143 | 130 | 213 | 277 | 305 | 318 | 322 | 293 | 289 | 266 | 266 | 290 | 249 | 225 | 207 | 215 | 226 | 210 |
| 36 | 200 | 242 | 280 | 231 | 187 | 156 | 131 | 125 | 175 | 232 | 237 | 229 | 235 | 243 | 233 | 206 | 230 | 205 | 193 | 218 | 174 | 185 | 144 |
| 37 | 193 | 212 | 274 | 208 | 190 | 155 | 101 | 129 | 76 | 178 | 217 | 253 | 221 | 229 | 204 | 187 | 180 | 207 | 174 | 158 | 156 | 148 | 137 |
| 38 | 167 | 167 | 209 | 207 | 158 | 140 | 112 | 103 | 79 | 94 | 128 | 186 | 175 | 168 | 188 | 166 | 156 | 162 | 157 | 120 | 120 | 139 | 116 |
| 39 | 135 | 146 | 159 | 140 | 131 | 150 | 119 | 104 | 82 | 83 | 58 | 108 | 142 | 157 | 166 | 137 | 153 | 142 | 123 | 117 | 105 | 97 | 111 |
| 40 | 138 | 129 | 130 | 118 | 132 | 135 | 106 | 114 | 90 | 59 | 78 | 67 | 93 | 99 | 102 | 102 | 108 | 110 | 91 | 77 | 86 | 96 | 75 |
| 41 | 94 | 99 | 98 | 93 | 95 | 96 | 81 | 93 | 94 | 71 | 45 | 52 | 47 | 54 | 79 | 81 | 97 | 91 | 74 | 64 | 59 | 64 | 59 |
| 42 | 54 | 76 | 66 | 70 | 65 | 61 | 55 | 72 | 60 | 50 | 60 | 31 | 25 | 32 | 44 | 48 | 75 | 50 | 47 | 59 | 51 | 35 | 49 |
| 43 | 50 | 50 | 49 | 35 | 39 | 41 | 40 | 54 | 40 | 46 | 36 | 30 | 20 | 20 | 27 | 28 | 22 | 45 | 42 | 44 | 18 | 23 | 18 |
| 44 | 21 | 26 | 28 | 26 | 26 | 29 | 29 | 29 | 19 | 31 | 26 | 24 | 25 | 17 | 17 | 11 | 16 | 27 | 22 | 19 | 21 | 18 | 17 |
| 45 | 14 | 20 | 15 | 12 | 13 | 20 | 15 | 12 | 15 | 12 | 10 | 11 | 18 | 13 | 6 | 5 | 7 | 13 | 9 | 19 | 13 | 17 | 15 |



born of these women during year $t$. We assume that these rates, when restricted to marital births, can sensibly be approximated by the data in the 1 % subsample of the 1970 census for the years $t = 1950, \ldots, 1969$.[2] Values for the numerators are shown in Table 13.1-1.[3] For example, 230 children were born in the year 1950 of women in our data set who were at age 20 in that year. In order to estimate age-specific birth rates we also need to know the number of women. These numbers can be taken from Table 12.2.3-2 in Section 12.2.3. For example, there are 3845 women in our data set who were born in 1930 and therefore of age 20 in 1950. So we get the estimate

$$\beta_{1950,20} \approx \frac{230}{3845} = 0.0598$$

In the same way one can estimate age-specific birth rates for all ages $\tau = 16, \ldots, 45$ and all years in the period 1947–69. Given this period and range of ages, the birth years range from 1902 to 1953.[4]

**3.** In order to visualize age-specific birth rates one can use level plots as introduced in Section 11.4. Such a plot is shown in Figure 13.1-1. The darkness of the grey-scale corresponds to the values of the age-specific birth rates. The maximum value of 170.3 marital children per 1000 women is reached in 1963 at an age of 24 years. We have selected five levels (40, 70, 100, 130, and 150) for contour lines. It is seen that high values of age-specific birth rates concentrate in the period 1958–66 at ages between 23 and 27 years. In the same way one can visualize cumulated age-specific birth rates as shown in Figure 13.1-2. The highest value of 2406 children

[2] One cannot reliably approximate these rates for the year 1970 because the census was conducted already in May of that year.

[3] Since only very few children were born of women at ages under 16 or above 45, we restrict all calculations to ages $\tau = 16, \ldots, 45$.

[4] In order to approximate age-specific birth rates for the whole period we therefore need, in addition to the numbers from Table 12.2.3-2, also the following cohort sizes:

| Birth year | Cohort size |
|---|---|
| 1902 | 3579 |
| 1903 | 3661 |
| 1904 | 3971 |
| 1946 | 3371 |
| 1947 | 3573 |
| 1948 | 3872 |
| 1949 | 4170 |
| 1950 | 4185 |
| 1951 | 3970 |
| 1952 | 4053 |
| 1953 | 3803 |

Like Table 12.2.3-2, these figures refer to the number of women in the 1 % subsample of the 1970 census.

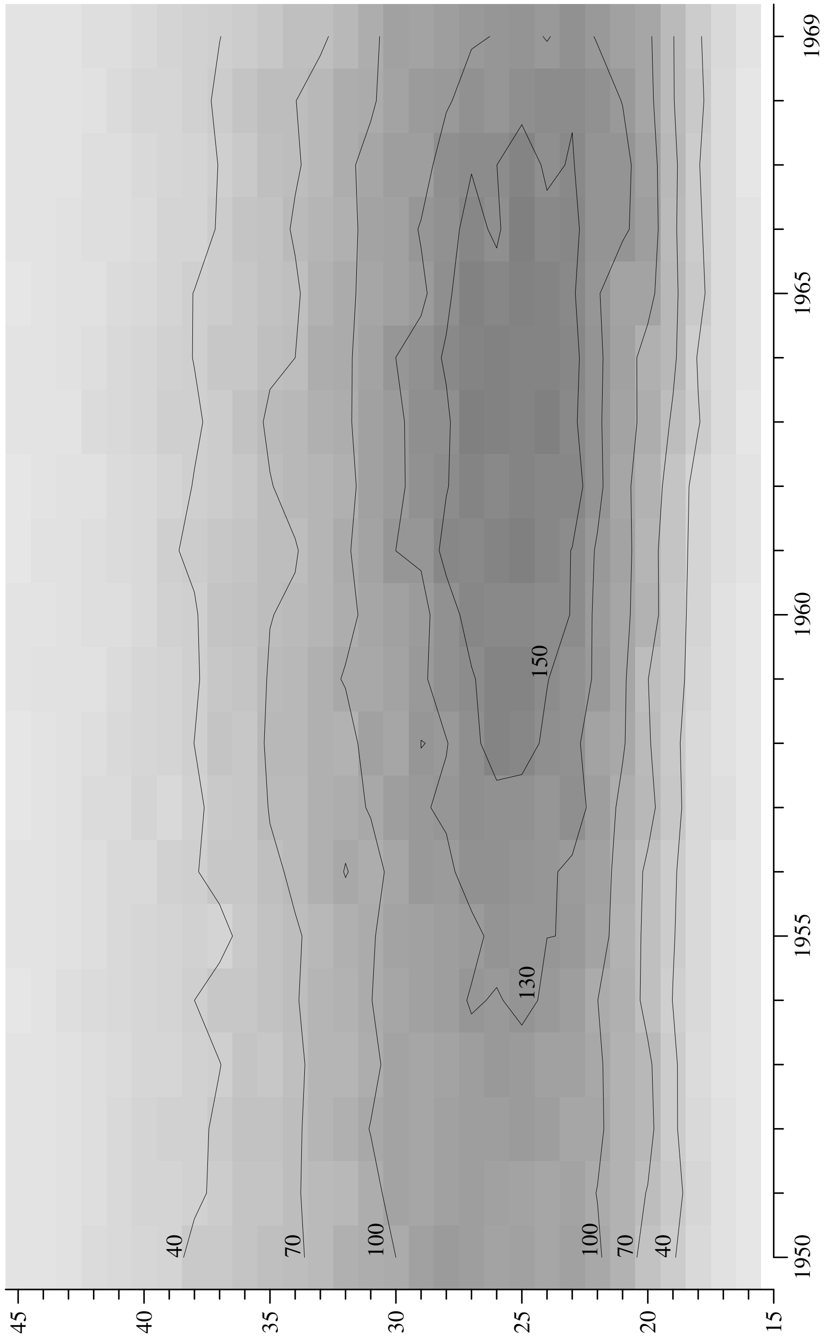



**Fig. 13.1-1** Level plot of age-specific birth rates in the period 1950 – 69, calculated from the 1 % subsample of the 1970 census.



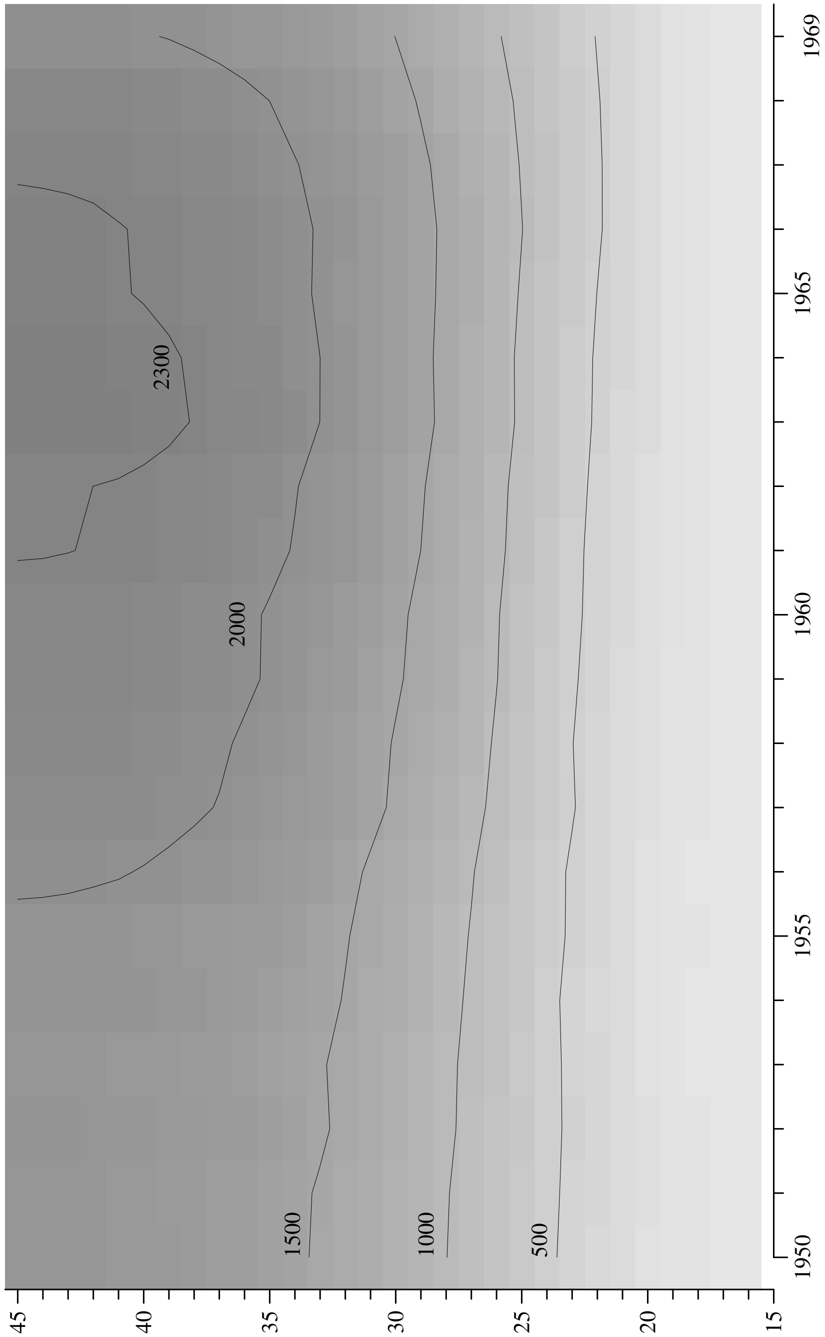



**Fig. 13.1-2** Level plot of cumulated age-specific birth rates in the period 1950 – 69, calculated from the 1 % subsample of the 1970 census.

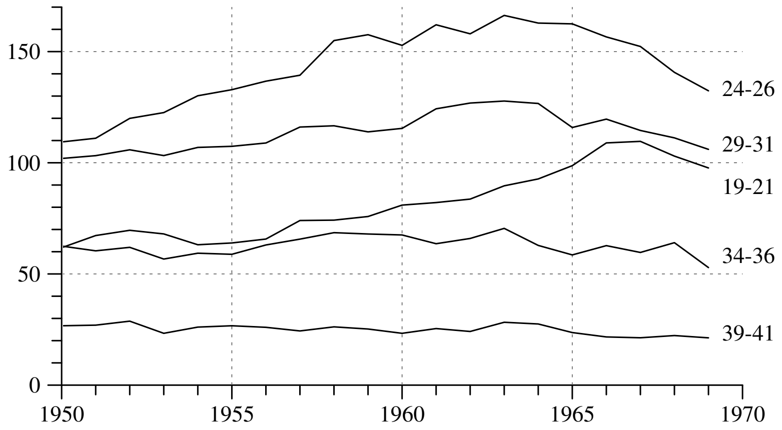



**Fig. 13.1-3** Age-specific birth rates in the period 1950 – 69. Mean values for the specified age groups calculated from our 1 % subsample of the 1970 census.

per 1000 women occurs again in the year 1963. This figure also illustrates that, until the second half of the 1960 s, women tended to get children in younger ages.

**4.** Additional information can be gained by focusing on birth rates for specific age groups. We selected five age groups and, for each group, calculated birth rates as unweighted mean values of the age-specific birth rates of the contributing ages. The result is presented in Figure 13.1-3. It is seen that mainly young women, up to an age of about 30, contributed to the rising number of births until the mid-sixties. Furthermore, with the exception of very young women, birth rates began to decline already since about 1965.

**5.** The data presented so far suggest that the increase in the number of births until the mid-sixties is mainly due to increasing birth rates. However, the number of children actually born also depends on the number and age distribution of potential mothers. One might therefore ask what part of the rising number of children can be attributed to changes in cohort sizes and age distribution, and what part can be attributed to changes in age-specific birth rates. One possibility to approach this question is by performing a hypothetical calculation based on the assumption that the number and age distribution of women remained the same as it was in 1950 for the whole period from 1950 to 1969. The actual number of births, $b_t$, can then be compared with a hypothetical number of births, calculated as

$$b_t^* := \sum_{\tau=16}^{45} \beta_{t,\tau}\, n_{1950,\tau}^f$$



**Table 13.1-2** Actual and hypothetical number of marital children born in the period 1950 – 69 of women in the 1 % subsample of the 1970 census.

| | Actual development | | | Hypothetical development | | |
|---|---|---|---|---|---|---|
| $t$ | $b_t$ | $b_t - 7291$ | cumulated | $b_t^*$ | $b_t^* - 7291$ | cumulated |
| 1950 | 7291 | 0 | 0 | 7291 | 0 | 0 |
| 1951 | 7216 | -75 | -75 | 7251 | -40 | -40 |
| 1952 | 7424 | 133 | 58 | 7479 | 188 | 148 |
| 1953 | 7217 | -74 | -16 | 7290 | -1 | 147 |
| 1954 | 7546 | 255 | 239 | 7655 | 364 | 511 |
| 1955 | 7527 | 236 | 475 | 7586 | 295 | 806 |
| 1956 | 7942 | 651 | 1126 | 7972 | 681 | 1488 |
| 1957 | 8385 | 1094 | 2220 | 8333 | 1042 | 2529 |
| 1958 | 8618 | 1327 | 3547 | 8502 | 1211 | 3741 |
| 1959 | 8949 | 1658 | 5205 | 8685 | 1394 | 5134 |
| 1960 | 9104 | 1813 | 7018 | 8741 | 1450 | 6585 |
| 1961 | 9505 | 2214 | 9232 | 9082 | 1791 | 8376 |
| 1962 | 9591 | 2300 | 11532 | 9075 | 1784 | 10160 |
| 1963 | 9978 | 2687 | 14219 | 9414 | 2123 | 12283 |
| 1964 | 9886 | 2595 | 16814 | 9377 | 2086 | 14369 |
| 1965 | 9572 | 2281 | 19095 | 9134 | 1843 | 16212 |
| 1966 | 9479 | 2188 | 21283 | 9140 | 1849 | 18061 |
| 1967 | 9193 | 1902 | 23185 | 8902 | 1611 | 19672 |
| 1968 | 8875 | 1584 | 24769 | 8643 | 1352 | 21023 |
| 1969 | 8200 | 909 | 25678 | 8023 | 732 | 21755 |

Table 13.1-2 shows the result of this calculation; Figure 13.1-4 compares the development of $b_t$ and $b_t^*$. It is seen that only a small part of the increasing number of marital children born in the period 1950 – 69 can be attributed to changes in the number and age distribution of women. So we conclude that the main part of the "baby boom" resulted from increasing birth rates, that is, women, in particular younger women, gave birth to more children.

**6.** One might try to quantify the contribution to the rising number of children which can be attributed to changes in the number and age distribution of women. A simple measure can be derived from Table 13.1-2. Column $b_t - 7291$ shows the surplus of marital children compared with 1950; the next column shows the cumulated values. The same calculations are then applied to the hypothetical development. For example, until the year 1963, the cumulated surplus of marital children amounted to 14219. Under the assumption that the number and age distribution of women had not changed since 1950, this figure would be 12283. One can therefore attribute about 14 % of the cumulated surplus until 1963 to changes in the number and age distribution of women.

**7.** An additional consideration concerns "timing effects": women can give birth to children anywhere during the reproductive period. In fact, as already mentioned several times, until about the mid-sixties, women began

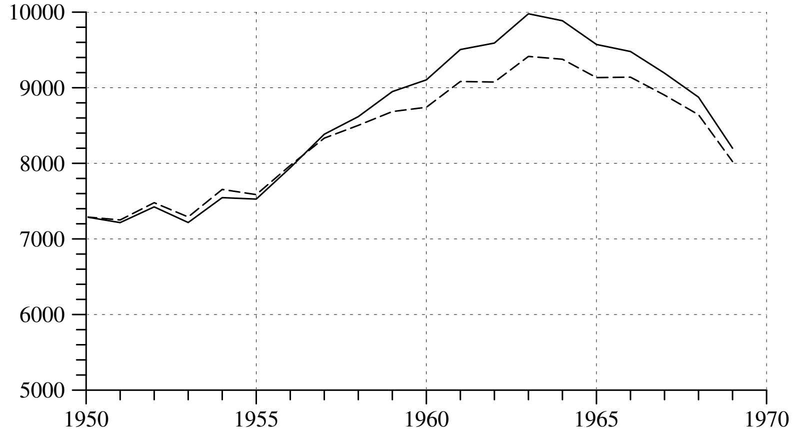



**Fig. 13.1-4** Actual (solid line) and hypothetical (dotted line) number of marital children born in the period 1950 – 69. Data are taken from columns $b_t$ and $b_t^*$ in Table 13.1-2.

childbearing in younger ages. Several authors have therefore suggested that at least some part of the baby boom is due to such “timing effects” (see, e.g., Dinkel, 1983). However, investigating this question requires more complicated considerations and will be postponed until Chapter 13.3.

## 13.2 Parity-specific Birth Rates

**1.** One can get additional information by distinguishing births with respect to parity, that is, for each women, the first child, the second child, and so on. We will use the following notation:

$$b_{t,\tau}^{(p)} := \text{number of children of parity } p \text{ born in year } t \text{ of women at age } \tau$$

Of course, women might give birth to several children at the same time (twins, triplets, ...), and in these cases parities are arbitrarily assigned. In a first step we ignore the dependence on age and simply consider

$$b_t^{(p)} := \sum_{\tau=\tau_a}^{\tau_b} b_{t,\tau}^{(p)}$$

called *parity-specific number of children*. Values can be calculated from the 1 % subsample of the 1970 census as shown in Table 13.2-1. For consistency with earlier calculations, we have taken into account, for each year $t = 1947, \dots, 1969$, only women who were of age 16 – 45 in the respective year. Once again, the figures in Table 13.2-1 only refer to marital children.



**Table 13.2-1** Number of marital children, classified with respect to parity, born by women of age 16 – 45 who are members of the 1 % subsample of the 1970 census.

| $t$ | $b_t$ | $b_t^{(1)}$ | $b_t^{(2)}$ | $b_t^{(3)}$ | $b_t^{(4)}$ | $b_t^{(5+)}$ |
|---|---|---|---|---|---|---|
| 1947 | 6307 | 2963 | 1768 | 808 | 385 | 383 |
| 1948 | 6864 | 3209 | 1949 | 940 | 381 | 385 |
| 1949 | 7316 | 3395 | 2182 | 970 | 415 | 354 |
| 1950 | 7291 | 3530 | 2086 | 932 | 360 | 383 |
| 1951 | 7216 | 3415 | 2124 | 897 | 405 | 375 |
| 1952 | 7424 | 3399 | 2245 | 1020 | 423 | 337 |
| 1953 | 7217 | 3260 | 2233 | 992 | 405 | 327 |
| 1954 | 7546 | 3358 | 2345 | 1050 | 448 | 345 |
| 1955 | 7527 | 3364 | 2266 | 1045 | 494 | 358 |
| 1956 | 7942 | 3503 | 2332 | 1183 | 518 | 406 |
| 1957 | 8385 | 3675 | 2487 | 1279 | 518 | 426 |
| 1958 | 8618 | 3744 | 2591 | 1237 | 565 | 481 |
| 1959 | 8949 | 3781 | 2768 | 1310 | 582 | 508 |
| 1960 | 9104 | 3924 | 2716 | 1316 | 633 | 515 |
| 1961 | 9505 | 4061 | 2832 | 1379 | 655 | 578 |
| 1962 | 9591 | 3989 | 2946 | 1431 | 663 | 562 |
| 1963 | 9978 | 4042 | 3141 | 1509 | 656 | 630 |
| 1964 | 9886 | 3872 | 3103 | 1525 | 725 | 661 |
| 1965 | 9572 | 3905 | 3007 | 1457 | 617 | 586 |
| 1966 | 9479 | 3662 | 3063 | 1543 | 624 | 587 |
| 1967 | 9193 | 3617 | 2965 | 1471 | 608 | 532 |
| 1968 | 8875 | 3457 | 2864 | 1435 | 587 | 532 |
| 1969 | 8200 | 3345 | 2617 | 1247 | 532 | 459 |

**2.** Figure 13.2-1 provides a graphical display of the data in Table 13.2-1. It is seen that children of all parities contributed to the general increase until about 1963. It is also seen that, depending on parity, the decline in the number of births began in different years.

**3.** In a next step we can consider parity-specific birth rates. The standard definition is

$$\beta_{t,\tau}^{(p)} := \frac{b_{t,\tau}^{(p)}}{n_{t,\tau}^{f}}$$

The denominator refers to the number of women aged $\tau$ in the year $t$, and the numerator refers to the number of children of parity $p$ who are born of these women during the year $t$. The definition has, however, a drawback in not taking into account that, except in cases of multiple births, children of parity $p$ can only be born of women who have already given birth to $p - 1$ children. It is therefore preferable to calculate *parity progression rates*.[5]

[5]Also called *parity progression ratios*, see, e.g., Newell (1988, p. 58).

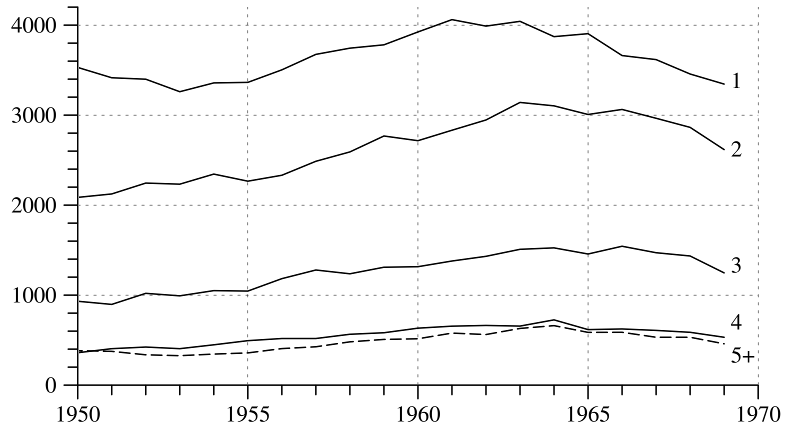



**Fig. 13.2-1** Parity-specific number of children in the period 1950–69, corresponding to the values in Table 13.2-1.

We use the following definition:

$$\beta_{t,\tau}^{(p)} := \frac{b_{t,\tau}^{(p)}}{n_{t,\tau}^{f,(p-1)}}$$

In this definition, the denominator only refers to women at age $\tau$ in year $t$ who have already given birth to $p-1$ children. For example, $\beta_{t,\tau}^{(1)}$ is the proportion of women at age $\tau$ in year $t$ who gave birth to their first child during that year. Similarly, $\beta_{t,\tau}^{(2)}$ is the proportion of women with already a first child who gave birth to a second child during the year $t$.[6]

**4.** We will try to calculate parity progression rates from the data in the 1 % subsample of the 1970 census. We begin with a simplified approach and ignore age, that is, we relate the number of children of parity $p$ born during a year $t$ to all women who might give birth to a child of this parity during the year $t$. The formal definition is

$$\beta_{t}^{(p)} := \frac{b_{t}^{(p)}}{n_{t}^{f,(p-1)}}$$

Table 13.2-2 shows values for the denominator. For example, there are 117800 women born during the years 1902–31 and therefore in an age between 16 and 45 in the year $t = 1947$. Of these women, 64498 had no child until the end of 1946 and might get a first child during the year 1947;

[6]Birg, Filip and Flöthmann (1990, p. 11) call these rates “bedingte Geburtenwahrscheinlichkeiten”. We avoid this wording because partity progression rates are simply proportions, not probabilities.



**Table 13.2-2** Number of women born in the years specified in the first column with $p$ children before year $t$, calculated from the 1 % subsample of the 1970 census.

| Birth years | $t$ | $n_t^{f,(0)}$ | $n_t^{f,(1)}$ | $n_t^{f,(2)}$ | $n_t^{f,(3)}$ | $n_t^{f,(4+)}$ | $n_t^f$ |
|---|---|---|---|---|---|---|---|
| 1902 – 1931 | 1947 | 64498 | 22644 | 16692 | 7801 | 6165 | 117800 |
| 1903 – 1932 | 1948 | 63745 | 23087 | 16948 | 7815 | 6070 | 117665 |
| 1904 – 1933 | 1949 | 62616 | 23566 | 17228 | 7922 | 5952 | 117284 |
| 1905 – 1934 | 1950 | 62028 | 23939 | 17577 | 8006 | 5799 | 117349 |
| 1906 – 1935 | 1951 | 61535 | 24556 | 17936 | 8072 | 5614 | 117713 |
| 1907 – 1936 | 1952 | 61266 | 24941 | 18282 | 8046 | 5481 | 118016 |
| 1908 – 1937 | 1953 | 61038 | 25180 | 18507 | 8044 | 5341 | 118110 |
| 1909 – 1938 | 1954 | 61260 | 25222 | 18705 | 8010 | 5198 | 118395 |
| 1910 – 1939 | 1955 | 61744 | 25261 | 18964 | 8004 | 5071 | 119044 |
| 1911 – 1940 | 1956 | 62201 | 25368 | 19141 | 7930 | 4977 | 119617 |
| 1912 – 1941 | 1957 | 62202 | 25500 | 19207 | 7947 | 4974 | 119830 |
| 1913 – 1942 | 1958 | 61162 | 25644 | 19272 | 8058 | 4954 | 119090 |
| 1914 – 1943 | 1959 | 60158 | 25783 | 19505 | 8111 | 4998 | 118555 |
| 1915 – 1944 | 1960 | 59127 | 25775 | 19826 | 8286 | 5060 | 118074 |
| 1916 – 1945 | 1961 | 57113 | 26194 | 20370 | 8505 | 5274 | 117456 |
| 1917 – 1946 | 1962 | 55757 | 26831 | 21139 | 8856 | 5666 | 118249 |
| 1918 – 1947 | 1963 | 54726 | 27284 | 22036 | 9303 | 6045 | 119394 |
| 1919 – 1948 | 1964 | 53949 | 27595 | 23035 | 9806 | 6416 | 120801 |
| 1920 – 1949 | 1965 | 53359 | 27438 | 23722 | 10113 | 6744 | 121376 |
| 1921 – 1950 | 1966 | 52479 | 27181 | 24024 | 10348 | 6901 | 120933 |
| 1922 – 1951 | 1967 | 51575 | 26672 | 24319 | 10654 | 7003 | 120223 |
| 1923 – 1952 | 1968 | 50899 | 26241 | 24690 | 10922 | 7082 | 119834 |
| 1924 – 1953 | 1969 | 50195 | 25837 | 25009 | 11143 | 7188 | 119372 |

22644 women had a first child until the end of 1946 and might get a second child during the year 1947, and so on. These numbers are used to calculate

$$\beta_t^{(1)},\ \beta_t^{(2)},\ \beta_t^{(3)},\ \text{and}\ \beta_t^{(4)}$$

The required numerators can be found in Table 13.2-1. For example,

$$\beta_{1947}^{(1)} = \frac{2963}{64498} = 0.0459$$

that is, about 4.6 % of women who might get a first marital child during 1947 actually realized this possibility.

**5.** In the same way parity progression rates, for $p = 1, \ldots, 4$, can be calculated for all years. The result is shown in Figure 13.2-2. It suggests that the decline of parity progression rates began somewhat earlier for parities 3 and 4, compared with parities 1 and 2. Since there is only a very short time-lag it seems not warranted, however, to make this a substantial point.

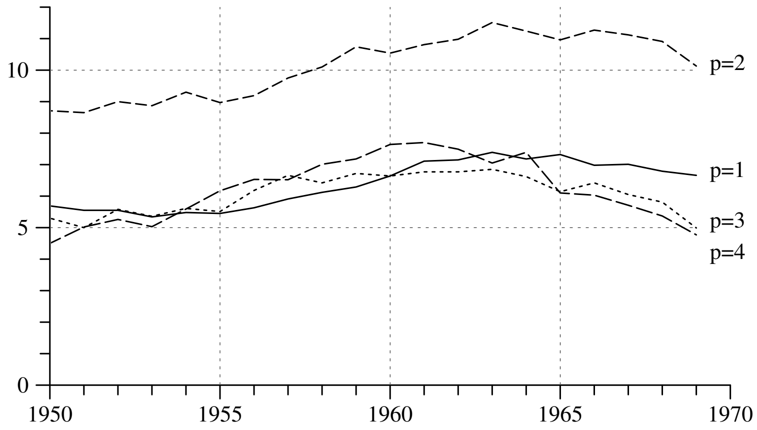



**Fig. 13.2-2** Parity progression rates of marital children in the period 1950 – 69, calculated from the data in Tables 13.2-1 and 13.2-2.



## 13.3 Understanding the Baby Boom

An instructive example of the fact that population growth not only depends on the number of newborn children but also on the timing of births and, especially, on women's age at childbearing, is the baby boom in West Germany during the period 1955 – 1965. It has been argued (e.g., by Dinkel, 1983) that this baby boom was mainly a consequence of the fact that women began childbearing at younger ages. The argument implies a comparison between the actual population growth and a hypothetical one that might have occurred if women behaved differently. One therefore needs some kind of analytical model to make the argument fully explicit.

### 13.3.1 Number and Timing of Births

**1.** In order to develop a conceptual framework we refer to birth cohorts of women denoted by $\mathcal{C}^f_{t_0}$, $t_0$ being the birth year. This allows to define age-specific cohort birth rates[7]

$$\gamma_{t_0,\tau} := \frac{b_{t_0+\tau,\tau}}{|\mathcal{C}^f_{t_0,\tau}}$$

where the denominator refers to the number of women belonging to the birth cohort $\mathcal{C}^f_{t_0}$ at age $\tau$, and the numerator records the number of children born by members of $\mathcal{C}^f_{t_0}$ at age $\tau$ (in the year $t = t_0 + \tau$). Denoting the beginning and end of the reproductive period by $\tau_a$ and $\tau_b$, respectively, one can also define cumulated cohort birth rates

$$\bar{\gamma}_{t_0,\tau} := \sum_{j=\tau_a}^{\tau} \gamma_{t_0,j}$$

**2.** These concepts can be used to compare childbearing among birth cohorts of women and, in these comparisons, distinguish between the number of children born and the timing of childbearing. The first aspect is captured by the completed cohort birth rate, $\bar{\gamma}_{t_0,\tau_b}$; the second aspect is captured by the shape of the function

$$\tau \longrightarrow \bar{\gamma}_{t_0,\tau}$$

As an example, we use data from the 1 % subsample of the 1970 census discussed in Chapter 12.2. Figure 13.3.1-1 compares birth rates of the cohorts $t_0 = 1910$ and $t'_0 = 1920$. The topmost plot (a) compares the cumulated cohort birth rates $\bar{\gamma}_{1910,\tau}$ (solid line) and $\bar{\gamma}_{1920,\tau}$ (dotted line). Assuming $\tau_b = 45$, it is seen that cohort C20 has a somewhat lower completed cohort

[7] These notions have been introduced in Section 11.2.

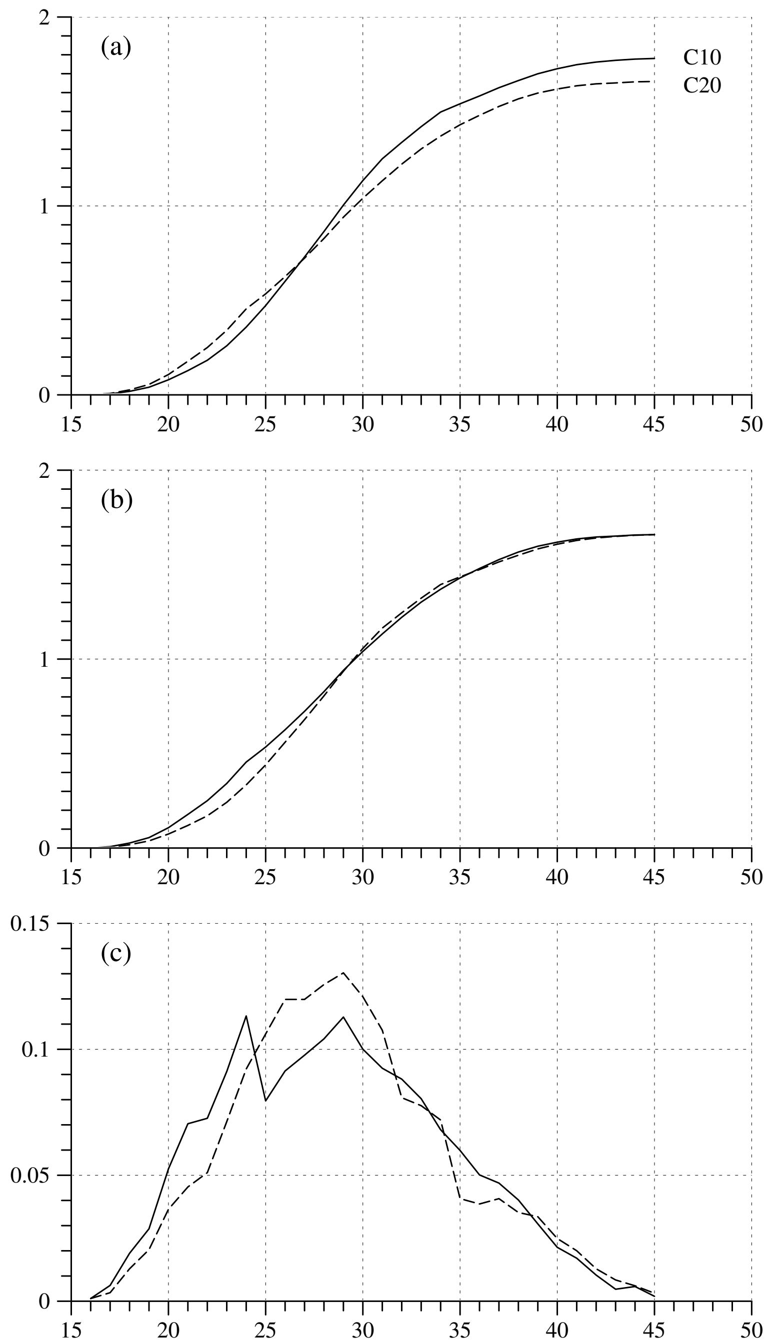



**Fig. 13.3.1-1** Comparison of age-specific cohort birth rates; see the text for explanation.



birth rate than cohort C10. Values can be calculated from Table 12.2.3-1 in Section 12.2.3: $\bar{\gamma}_{1910,45} = 1.7806$ and $\bar{\gamma}_{1920,45} = 1.6588$. Furthermore, there is also a somewhat different timing of births. Compared with C10, relatively more women belonging to C20 gave birth to children at ages under 25.

**3.** Of course, it would be strange to say that these women hastened to realize births that they anticipated to have anyway.[8] Nevertheless, in order to conceptually distinguish between the number and timing of births, one cannot avoid to apply a retrospective view and assume completed cohort birth rates as given. In our example, this allows to construct, for birth cohort C20, hypothetical cumulated cohort birth rates which have the same shape as the cumulated cohort birth rates of C10 but keep the original completed cohort birth rate of C20. The following definition shows the construction:

$$\bar{\gamma}^*_{1920,\tau} := \bar{\gamma}_{1910,\tau} \frac{\bar{\gamma}_{1920,45}}{\bar{\gamma}_{1910,45}}$$

Part (b) of Figure 13.3.1-1 compares $\bar{\gamma}_{1920,\tau}$ (solid line) and $\bar{\gamma}^*_{1920,\tau}$ (dotted line). Without changing the completed cohort birth rate, part of the births are "shifted" into higher ages. This is also seen in part (c) of the figure where the solid line refers to the age-specific birth rates $\gamma_{1920,\tau}$ and the dotted line refers to the corresponding hypothetical birth rates

$$\gamma^*_{1920,\tau} := \gamma_{1910,\tau} \frac{\bar{\gamma}_{1920,45}}{\bar{\gamma}_{1910,45}} \tag{13.3.1}$$

**4.** Finally, one can compare the actual with a hypothetical development of births. The actual development is given by the equation

$$b_t = \sum_{\tau=\tau_a}^{\tau_b} b_{t,\tau} = \sum_{\tau=\tau_a}^{\tau_b} \gamma_{t-\tau,\tau} \, |\mathcal{C}^f_{t-\tau,\tau}|$$

which shows how the number of births in year $t$ derives from the surviving cohort members, $|\mathcal{C}^f_{t-\tau,\tau}|$, and the cohort birth rates, $\gamma_{t-\tau,\tau}$, of all births cohorts $t-\tau$ ($\tau = \tau_a, \ldots, \tau_b$). The idea now is to compare this actual development of births with a hypothetical development defined by

$$b^*_t := \sum_{\tau=\tau_a}^{\tau_b} \gamma^*_{t-\tau,\tau} \, |\mathcal{C}^f_{t-\tau,\tau}| \tag{13.3.2}$$

$b^*_t$ would be the number of children born in year $t$ if the childbearing of women who might contribute to these births would follow the modified

[8] Actually, the whole argument is in statistical terms and does not relate to the behavior of individual women; and, as was discussed in Section 3.4, one also cannot sensibly speak of the behavior of a cohort.

birth rates $\gamma^*_{t-\tau,\tau}$, instead of the actually realized birth rates $\gamma_{t-\tau,\tau}$. Of course, in order to define the modified birth rates one needs to refer to one birth cohort whose timing of births provides a reference. In the example above the cohort of women born 1910 was used to defined the reference. It is quite possible, however, that the results of a comparison between $b_t$ and $b^*_t$ also depend on the choice of the reference cohort.

### 13.3.2 Performing the Calculations

**1.** We now try to compare $b_t$ and (different versions of) $b^*_t$ in the period 1950–1970 in the territory of the former FRG. Values of $b_t$ are available from official statistics (see Table 6.3-1 in Section 6.3). In order to find values of $b^*_t$ one needs to refer to all birth cohorts of women who contributed to the births in the period 1950–1970. Assuming a reproductive period from age 16 to age 45, cohort birth years range from 1905 to 1954. For each birth cohort we need values for the cohort size in 1950 and the completed cohort birth rates. Since appropriate data are not directly available from official period statistics, we try to find approximately valid quantities from the 1 % subsample of the 1970 census discussed in Chapter 12.2.

**2.** We assume that cohort sizes in 1950 are approximately proportional to the number of women, born in years from 1905 to 1954, who were still alive at the census date in 1970. These numbers, taken from the 1 % subsample of the census, are shown in the second column of Table 13.3.2-1. Completed cohort birth rates are more difficult to approximate. Cumulated cohort birth rates that can be calculated from the subsample of the 1970 census only refer to marital births. Moreover, assuming $\tau_b = 45$, completed marital cohort birth rates can only be calculated for cohorts 1905–1924. They are shown in column (b) of Table 13.3.2-1. In order to extend the period one can use the fact that cumulated cohort birth rates at age 45 are only slightly larger than at age 40. This is seen in Table 13.3.2-1 by comparing column (b) with column (a) which shows the cumulated cohort birth rates up to an age of 40. The entries for birth cohorts 1925–30, shown in column (c) of the table, have been calculated by simply multiplying the entries in column (a) by 1.026. Beginning with birth cohort 1930, official period statistics allow to calculate completed quasi-cohort birth rates. Still assuming that $\tau_b = 45$, they are shown in column (d) of Table 13.3.2-1.[9] Finally, since these values refer to all births, one needs an adjustment of the completed cohort birth rates calculated from the 1 % subsample of the 1970 census which only refer to marital births. Assuming a proportion of about 10 % non-marital births we have simply multiplied the entries in columns (b) and (c) by the factor 1.1 in order to get the entries in column (e). The values in columns (d) and (e)

[9] Data are taken from Fachserie 1, Reihe 1, 1999 (pp. 198-200). See also the discussion of these data in Section 11.4.



**Table 13.3.2-1** Calculation of completed cohort birth rates for birth cohorts 1905 – 1954. See the text for explanations.

| Birth year | Cohort size | (a) | (b) | (c) | (d) | (e) |
|---|---|---|---|---|---|---|
| 1905 | 3926 | 1.6533 | 1.7068 | | | 1.8775 |
| 1906 | 4045 | 1.6485 | 1.7122 | | | 1.8834 |
| 1907 | 4208 | 1.7533 | 1.8118 | | | 1.9930 |
| 1908 | 4430 | 1.6804 | 1.7370 | | | 1.9107 |
| 1909 | 4395 | 1.7192 | 1.7738 | | | 1.9512 |
| 1910 | 4417 | 1.7263 | 1.7806 | | | 1.9587 |
| 1911 | 4349 | 1.7211 | 1.7754 | | | 1.9529 |
| 1912 | 4511 | 1.7092 | 1.7610 | | | 1.9371 |
| 1913 | 4377 | 1.6943 | 1.7482 | | | 1.9230 |
| 1914 | 4301 | 1.6894 | 1.7410 | | | 1.9151 |
| 1915 | 3403 | 1.6753 | 1.7338 | | | 1.9072 |
| 1916 | 2578 | 1.6427 | 1.6889 | | | 1.8578 |
| 1917 | 2428 | 1.6512 | 1.7002 | | | 1.8702 |
| 1918 | 2465 | 1.6730 | 1.7233 | | | 1.8956 |
| 1919 | 3595 | 1.6487 | 1.6918 | | | 1.8610 |
| 1920 | 4628 | 1.6188 | 1.6588 | | | 1.8247 |
| 1921 | 4680 | 1.6496 | 1.7013 | | | 1.8714 |
| 1922 | 4442 | 1.6729 | 1.7226 | | | 1.8949 |
| 1923 | 4265 | 1.6980 | 1.7496 | | | 1.9246 |
| 1924 | 4058 | 1.7287 | 1.7740 | | | 1.9514 |
| 1925 | 4162 | 1.6975 | | 1.7416 | | 1.9158 |
| 1926 | 3998 | 1.7161 | | 1.7607 | | 1.9368 |
| 1927 | 3863 | 1.8532 | | 1.9014 | | 2.0915 |
| 1928 | 3923 | 1.8463 | | 1.8943 | | 2.0837 |
| 1929 | 3762 | 1.9054 | | 1.9549 | | 2.1504 |
| 1930 | 3845 | 1.9116 | | 1.9613 | 2.1395 | |
| 1931 | 3535 | | | | 2.1623 | |
| 1932 | 3444 | | | | 2.1993 | |
| 1933 | 3280 | | | | 2.2244 | |
| 1934 | 4036 | | | | 2.2395 | |
| 1935 | 4290 | | | | 2.1721 | |
| 1936 | 4348 | | | | 2.1347 | |
| 1937 | 4302 | | | | 2.1079 | |
| 1938 | 4715 | | | | 2.0695 | |
| 1939 | 5044 | | | | 2.0243 | |
| 1940 | 4990 | | | | 1.9708 | |
| 1941 | 4562 | | | | 1.9025 | |
| 1942 | 3771 | | | | 1.8490 | |
| 1943 | 3842 | | | | 1.8089 | |
| 1944 | 3820 | | | | 1.7771 | |
| 1945 | 2785 | | | | 1.7746 | |
| 1946 | 3371 | | | | 1.7791 | |
| 1947 | 3573 | | | | 1.7513 | |
| 1948 | 3872 | | | | 1.7286 | |
| 1949 | 4170 | | | | 1.7145 | |
| 1950 | 4185 | | | | 1.7003 | |
| 1951 | 3970 | | | | 1.6578 | |
| 1952 | 4053 | | | | 1.6464 | |
| 1953 | 3803 | | | | 1.6287 | |
| 1954 | 4012 | | | | 1.6057 | |

**Table 13.3.2-2** Age-specific cohort birth rates for birth cohorts 1910, 1920, and 1930.

| $\tau$ | 1910 | 1920 | 1930 | $\tau$ | 1910 | 1920 | 1930 |
|---|---|---|---|---|---|---|---|
| 16 | 0.00113 | 0.00108 | 0.0021 | 31 | 0.11546 | 0.09248 | 0.1136 |
| 17 | 0.00362 | 0.00627 | 0.0100 | 32 | 0.08671 | 0.08816 | 0.0989 |
| 18 | 0.01381 | 0.01901 | 0.0289 | 33 | 0.08331 | 0.08038 | 0.0895 |
| 19 | 0.02196 | 0.02874 | 0.0527 | 34 | 0.07720 | 0.06806 | 0.0787 |
| 20 | 0.03917 | 0.05251 | 0.0748 | 35 | 0.04369 | 0.05985 | 0.0656 |
| 21 | 0.04868 | 0.07044 | 0.0968 | 36 | 0.04143 | 0.05013 | 0.0564 |
| 22 | 0.05479 | 0.07260 | 0.1142 | 37 | 0.04369 | 0.04689 | 0.0450 |
| 23 | 0.07652 | 0.09118 | 0.1253 | 38 | 0.03781 | 0.04019 | 0.0361 |
| 24 | 0.09894 | 0.11322 | 0.1349 | 39 | 0.03600 | 0.03068 | 0.0276 |
| 25 | 0.11388 | 0.07952 | 0.1394 | 40 | 0.02671 | 0.02139 | 0.0197 |
| 26 | 0.12859 | 0.09140 | 0.1459 | 41 | 0.02151 | 0.01707 | 0.0143 |
| 27 | 0.12859 | 0.09767 | 0.1491 | 42 | 0.01381 | 0.01037 | 0.0085 |
| 28 | 0.13493 | 0.10415 | 0.1418 | 43 | 0.00906 | 0.00475 | 0.0051 |
| 29 | 0.13991 | 0.11279 | 0.1365 | 44 | 0.00657 | 0.00583 | 0.0027 |
| 30 | 0.12973 | 0.10004 | 0.1239 | 45 | 0.00340 | 0.00194 | 0.0013 |
| | | | | Total | 1.7806 | 1.6588 | 2.1395 |

will then be used in the following simulations.

**3.** A further question concerns the birth cohort to be used as a reference for the assessment of timing effects. Since simulation results might well depend on the choice of a reference cohort, we perform the calculations separately for three reference cohorts with birth years 1910, 1920, and 1930, respectively. The age-specific birth rates for these cohorts that we have used for the simulations are shown in Table 13.3.2-2. For birth cohorts 1910 and 1920, the rates refer to marital birth and are calculated from the 1 % percent subsample of the 1970 census. For birth cohort 1930 they are taken from official period statistics (Fachserie 1, Reihe 1, 1999, p. 198) and refer to all births. This difference may be neglected, however, because in the simulation the age-specific rates are only used to provide a standard shape for the timing of childbearing. In order to calculate hypothetical birth rates with formula (13.3.1), one only needs to use the appropriate completed cohort birth rates as shown in the last row of Table 13.3.2-2.

**4.** So we finally have at least some approximations for all values required to calculate hypothetical developments of births with formula (13.3.2). The result is shown in Figure 13.3.2-1. The solid line shows the actual number of births in the period 1950–1970.[10] The dotted lines show corresponding hypothetical developments. Since the calculation is based on a 1 % percent subsample of the 1970 census, the simulated figures have been multiplied by 100 in order to make the hypothetical developments roughly comparable with the actual development. However, regardless of the exact level, it is clearly seen that the development of birth would have been quite different without the changes in the timing of births which actually occurred. In fact, the plot suggests that the baby boom that occurred in the period 1955–65 can mainly be attributed to changes in the timing of childbearing. It is also remarkable that the hypothetical developments, on the whole, do not depend on the birth cohort that is used to provide a shape for the timing of births. This is consistent with the fact, discussed in Section 11.4, that substantial shifts of births towards younger ages occurred in cohorts with birth years roughly between 1930 and 1945.

[10] Data are taken from Table 6.3-1 in Section 6.3.



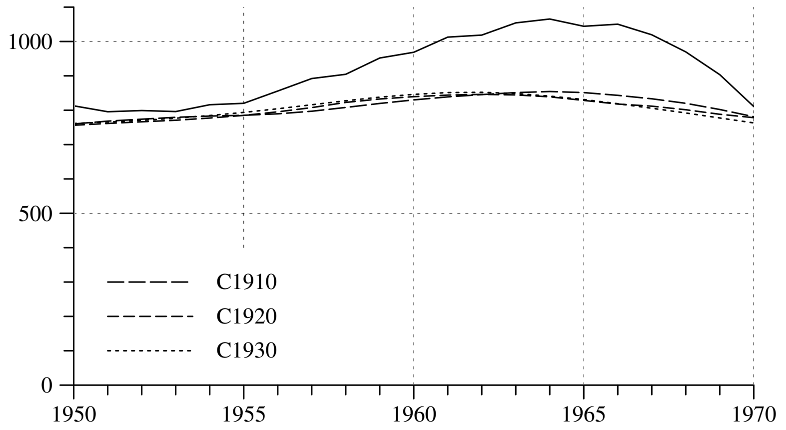



**Fig. 13.3.2-1** Number of birth (in 1000) in the territory of the former FRG (solid line) and hypothetical developments (dotted lines) which assume a timing of births according to birth cohorts 1910, 1920, and 1930, respectively.

## 13.3.3 Extending the Simulation Period

This section is not finished yet.

# Chapter 14

# Data from Non-official Surveys

As was mentioned in Section 11.3, if one wants to investigate the timing and distribution of birth events, data from official statistics are of only limited use. A closer investigation requires data which allow to relate birth events to women's life courses. Such data can be gathered with retrospective surveys in which women are asked about the birth dates of their children. One example, a subsample of the 1970 census, has been discussed in the two preceeding chapters. In addition, several non-official surveys are available that provide data on childbearing histories.[1] In the present chapter we consider data from the following non-official surveys that, in particular, provide information about number and birth dates of children:

- the German Life History Study (GLHS),
- the Socio-economic Panel (SOEP),
- the Fertility and Family Survey (FFS), and
- the DJI Family Survey (DJIFS).

The main questions to be discussed in the present chapter concern age at first childbearing, the proportion of childless women, and the distribution of the number of children. We also calculate cumulated cohort birth rates to allow comparisons with data from official statistics.

## 14.1 German Life History Study

**1.** The *German Life History Study* (GLHS) is a long-term project conducted by the Max Planck Institute for Human Development (Berlin). The main data source of this project is a series of retrospective surveys in which members of selected birth cohorts were asked to provide detailed information about their life courses. Part of these data are available for

[1] We speak of *non-official surveys* in order to signify that these surveys are conducted, not by official statistics, but by a variety of institutions of social research. Additional differences depend on circumstances. Most often the sample size of non-official surveys is much smaller than the sample size of official surveys. Furthermore, while some official surveys (e.g., the *Mikrozensus*) are based on an obligation to give information, participation in non-official surveys is always a matter of free decision. Consequently, there is often a substantial proportion of non-respondents in non-official surveys; see, e.g., Porst (1996).



the general scientific public:[2]

a) Data from the first survey (LV I) were sampled during the years 1981–83 and included 2171 members of the birth cohorts 1929–31, 1939–41, and 1949–51.

b) Data from a second survey (LV II) were sampled in two parts, both relating to persons born in the years 1919–21; a first part was conducted in 1985–86 and included 407 persons (LV IIA), a second part was conducted in 1987–88 and included 1005 persons (LV IIT).

c) Data from a third survey (LV III) were sampled in 1989 and included 2008 members of the birth cohorts 1954–56 and 1959–61.

All surveys were conducted in the territory of the former FRG. For our present study we take into account all female respondents from the surveys LV I, LV IIT, and LV III (only cohort 1959–61). The case numbers and how they distribute over the five cohorts is shown in the following table:[3]

| Birth cohort | Birth years | Male | Female | Interview date |
|---|---|---|---|---|
| C20 | 1919 – 21 | 373 | 632 | 1987 – 88 |
| C30 | 1929 – 31 | 349 | 359 | 1981 – 83 |
| C40 | 1939 – 41 | 375 | 355 | 1981 – 83 |
| C50 | 1949 – 51 | 365 | 368 | 1981 – 83 |
| C60 | 1959 – 61 | 512 | 489 | 1989 |

We also mention that all members of our subsample have a German citizenship. — In the remainder of this section we use this data set to investigate changes in the distribution of ages at first childbearing and the number of children across the five birth cohorts.[4]

### Age at First Childbearing

**2.** Denoting our subsample of the GLHS by $\Omega$, we can define a three-dimensional variable

$$(C, T, D) : \Omega \longrightarrow \tilde{\mathcal{C}} \times \tilde{\mathcal{T}} \times \tilde{\mathcal{D}}$$

[2] For an overview, see Wagner (1996). The data are available from the *Zentralarchiv für empirische Sozialforschung* (Köln). We thank Karl Ulrich Mayer, the director of the GLHS, for the permission to use the data sets.

[3] Of the 632 women of birth cohort C20 three did not give valid birth years for their children and will be excluded in further calculations.

[4] We mention that the GLHS data have already been used in quite a large number of earlier studies. Concerning the questions of the present section, see, in particular, Huinink (1987, 1988, 1989), Blossfeld and Huinink (1989), Tuma and Huinink (1990).

**Table 14.1-1** Age at first childbearing in our GLHS subsample.

| | C20 | | C30 | | C40 | | C50 | | C60 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ |
| 15 | | | | | | | | | 1 | |
| 16 | 1 | | | | | | 3 | | 3 | |
| 17 | 2 | | 1 | | 5 | | 4 | | 6 | |
| 18 | 11 | | 5 | | 15 | | 10 | | 10 | |
| 19 | 21 | | 16 | | 21 | | 30 | | 16 | |
| 20 | 28 | | 23 | | 21 | | 34 | | 14 | |
| 21 | 37 | | 23 | | 25 | | 36 | | 22 | |
| 22 | 60 | | 29 | | 36 | | 26 | | 24 | |
| 23 | 60 | | 22 | | 40 | | 22 | | 17 | |
| 24 | 68 | | 21 | | 37 | | 16 | | 21 | |
| 25 | 41 | | 32 | | 21 | | 13 | | 27 | |
| 26 | 35 | | 32 | | 23 | | 20 | | 26 | |
| 27 | 28 | | 31 | | 13 | | 22 | | 35 | 12 |
| 28 | 28 | | 16 | | 21 | | 14 | | 27 | 85 |
| 29 | 22 | | 17 | | 6 | | 12 | | 8 | 80 |
| 30 | 15 | | 16 | | 10 | | 13 | 21 | 1 | 54 |
| 31 | 15 | | 7 | | 5 | | 3 | 22 | | |
| 32 | 7 | | 11 | | 8 | | 3 | 35 | | |
| 33 | 10 | | 5 | | 3 | | 1 | 8 | | |
| 34 | 7 | | 8 | | | | | | | |
| 35 | 5 | | 3 | | 2 | | | | | |
| 36 | 7 | | 1 | | 2 | | | | | |
| 37 | 5 | | | | 2 | | | | | |
| 38 | 3 | | 1 | | | | | | | |
| 39 | 1 | | | | | | | | | |
| 40 | 3 | | | | | 7 | | | | |
| 41 | | | 1 | | | 11 | | | | |
| 42 | | | | | | 13 | | | | |
| 43 | | | | | | 8 | | | | |
| 50 | | | | 9 | | | | | | |
| 51 | | | | 16 | | | | | | |
| 52 | | | | 7 | | | | | | |
| 53 | | | | 6 | | | | | | |
| 66 | | 4 | | | | | | | | |
| 67 | | 47 | | | | | | | | |
| 68 | | 34 | | | | | | | | |
| 69 | | 24 | | | | | | | | |
| Total | 520 | 109 | 321 | 38 | 316 | 39 | 282 | 86 | 258 | 231 |

$C$, with property space $\tilde{\mathcal{C}} := \{\text{C20}, \dots, \text{C60}\}$, records the birth cohort; $D$, with property space $\tilde{\mathcal{D}} := \{0, 1\}$, records whether a women has given birth to at least one child;[5] and $T$, with property space $\tilde{\mathcal{T}} := \{0, 1, 2, \dots\}$, records the age of the women which, depending on the value of $D$, is the

[5]The GLHS allows to distinguish women's own children, step children, and adoptive children. For the present investigation we only take into account women's own children.



age of first childbearing (if $T(\omega) = 1$) or the age in the interview year (if $T(\omega) = 0$). The distribution of this three-dimensional variable, in terms of absolute frequency, is shown in Table 14.1-1.[6] For example, there are 68 women in birth cohort C20 who gave birth to a first child at age 24, 41 at age 25, and so on. In total, 520 women of this birth cohort had at least one child, and 109 remained childless.

**3.** The data from Table 14.1-1 can be used to estimate distributions of the age at first childbearing. We use the formal framework introduced in Chapter 12 and refer to a duration variable

$$\hat{T}_c : \Omega_c \longrightarrow \tilde{\mathcal{T}} := \{0, 1, 2, 3, \dots\}$$

where the index $c$ specifies one of the birth cohorts in our sample. Since each birth cohort comprises three birth years, and the interviews extend over up to three years, also the censoring times extend over several ages. However, as seen from Table 14.1-1, for birth cohorts C20, C30, and C40, censoring only occurs after the last observed event (first childbearing). For these birth cohorts, the data can therefore directly be used to calculate a frequency distribution of $\hat{T}_c$:

$$\mathrm{P}[\hat{T}_c](\tau) = \frac{|\{\omega \in \Omega_c \,|\, T(\omega) = \tau\}|}{|\Omega_c|}$$

For example, referring to birth cohort C20, one immediately finds

$$\mathrm{P}[\hat{T}_{\text{C20}}](25) = \frac{41}{629} = 0.065$$

that is, 6.5 % of the members of C20 gave birth to a first child at age 25. These values can then be used for the calculation of distribution functions, survivor functions, and rate functions.

**4.** The situation is slightly different for birth cohorts C50 and C60 where event times and censoring times overlap in some years. To illustrate, we refer to birth cohort C50. Obviously, for ages under 30, one can calculate frequencies directly. For example, for $\tau = 25$, one gets

$$\mathrm{P}[\hat{T}_{\text{C50}}](25) = \frac{13}{368} = 0.035$$

However, this direct calculation is no longer possible for ages $\tau \geq 30$. We therefore use the Kaplan-Meier procedure introduced in Section 8.3.4. Table 14.1-2 illustrates the calculations for birth cohort C50. Notice that, until age 29, results are identical with those from a direct calculation of frequencies.

[6]Note that the ages are not contiguous because the table refers to the realized property spaces. Note also that birth cohort C20 only contains 629 members because we have excluded three cases with unknown birth years of children.

**Table 14.1-2** Kaplan-Meier procedure to calculate the survivor function for the age at first childbearing. Data refer to cohort C50 in Table 14.1-1.

| $\tau$ | at risk | events | censored | rate | 1 − rate | survivor function |
|---|---|---|---|---|---|---|
| 16 | 368 | 3 | 0 | 0.0082 | 0.9918 | 1.0000 |
| 17 | 365 | 4 | 0 | 0.0110 | 0.9890 | 0.9918 |
| 18 | 361 | 10 | 0 | 0.0277 | 0.9723 | 0.9809 |
| 19 | 351 | 30 | 0 | 0.0855 | 0.9145 | 0.9537 |
| 20 | 321 | 34 | 0 | 0.1059 | 0.8941 | 0.8722 |
| 21 | 287 | 36 | 0 | 0.1254 | 0.8746 | 0.7798 |
| 22 | 251 | 26 | 0 | 0.1036 | 0.8964 | 0.6820 |
| 23 | 225 | 22 | 0 | 0.0978 | 0.9022 | 0.6114 |
| 24 | 203 | 16 | 0 | 0.0788 | 0.9212 | 0.5516 |
| 25 | 187 | 13 | 0 | 0.0695 | 0.9305 | 0.5081 |
| 26 | 174 | 20 | 0 | 0.1149 | 0.8851 | 0.4728 |
| 27 | 154 | 22 | 0 | 0.1429 | 0.8571 | 0.4185 |
| 28 | 132 | 14 | 0 | 0.1061 | 0.8939 | 0.3587 |
| 29 | 118 | 12 | 0 | 0.1017 | 0.8983 | 0.3206 |
| 30 | 106 | 13 | 21 | 0.1226 | 0.8774 | 0.2880 |
| 31 | 72 | 3 | 22 | 0.0417 | 0.9583 | 0.2527 |
| 32 | 47 | 3 | 35 | 0.0638 | 0.9362 | 0.2422 |
| 33 | 9 | 1 | 8 | 0.1111 | 0.8889 | 0.2267 |
| 34 | | | | | | 0.2015 |

**5.** The survivor functions for all five birth cohorts are shown in Figure 14.1-1. Several points are remarkable.

a) Until an age of about 27, the distribution for cohort C30 is quite similar to the distribution for cohort C20. After this age, that is, beginning at the end of the nineteen-fifties, a substantially greater proportion of the women belong to cohort C30 give birth to a child. Eventually, the proportion of childless women is quite smaller in C30 than in C20.

b) Compared with C30, members of birth cohort C40 begin childbearing at younger ages, but overall, both distributions are quite similar. In particular, in both cohorts, a high proportion of women, about 90 %, have at least one child.

c) Like the members of C40, also the members of C50 begin childbearing at younger ages. However, beginning in the mid-sixties, birth rates begin to decline, and it might be supposed that the proportion of women who eventually remain childless will be substantially greater than it was in the two preceeding cohorts.

d) Finally, members of birth cohort C60 delay the birth of a first child, and although the data do not allow definite conclusions, it seems quite possible that the proportion of finally childless women will again be greater than in the preceeding cohorts.



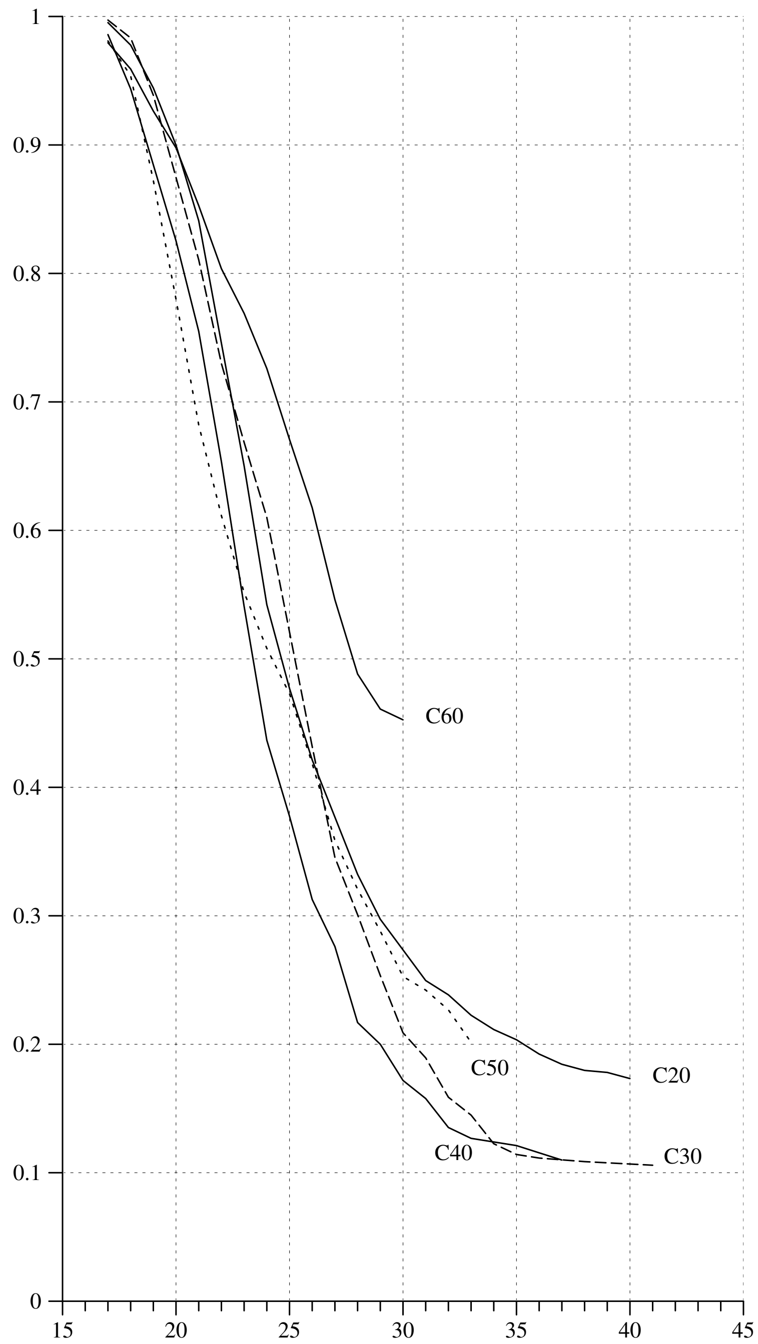



**Fig. 14.1-1** Distribution of age at first childbearing described by survivor functions, calculated from the data in Table 14.1-1.

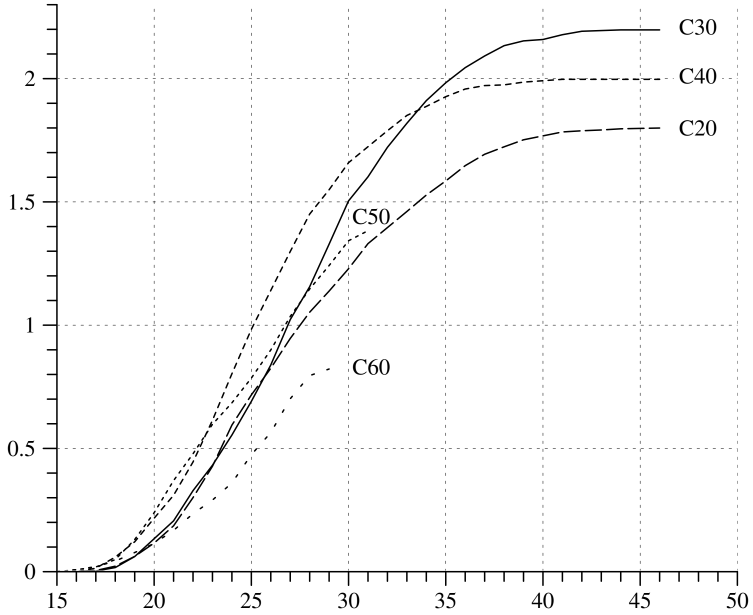



**Fig. 14.1-2** Cumulated cohort birth rates calculated from the data in Table 14.1-3.

The results can be compared with the distribution of age at first marital childbearing. This was already done in Section 12.2.2.

## Number of Children

**6.** The next step is to investigate the number of children born of women in the GLHS subsample. We begin with the calculation of cumulated cohort birth rates. Table 14.1-3 shows the data. For example, 44 women belonging to birth cohort C20 have given birth to a child at an age of 21. The data can be used to calculate cumulated cohort births rates. The following table shows these rates, denoted by CCBR($\tau$), until age $\tau$ as specified in the second column:

| Cohort | $\tau$ | CCBR($\tau$) | CCBR*($\tau$) |
|---|---|---|---|
| C20 | 45 | 1.80 | |
| C30 | 43 | 2.19 | 2.15 |
| C40 | 40 | 1.99 | 1.96 |
| C50 | 31 | 1.38 | 1.39 |
| C60 | 29 | 0.82 | 0.99 |

The final column, labeled CCBR*($\tau$), shows corresponding cumulated co-



**Table 14.1-3** Number of children in the GLHS subsample, classified with respect to mother's birth cohort and age ($\tau$).

| $\tau$ | C20 | C30 | C40 | C50 | C60 |
|---|---|---|---|---|---|
| 15 | | | | | 1 |
| 16 | 1 | | | 3 | 3 |
| 17 | 2 | 1 | 5 | 4 | 6 |
| 18 | 11 | 5 | 16 | 11 | 11 |
| 19 | 25 | 16 | 22 | 30 | 17 |
| 20 | 34 | 26 | 34 | 40 | 18 |
| 21 | 44 | 26 | 33 | 48 | 26 |
| 22 | 72 | 44 | 48 | 40 | 33 |
| 23 | 80 | 37 | 60 | 44 | 27 |
| 24 | 105 | 44 | 67 | 32 | 34 |
| 25 | 77 | 49 | 64 | 37 | 56 |
| 26 | 68 | 53 | 56 | 42 | 44 |
| 27 | 76 | 66 | 56 | 50 | 67 |
| 28 | 67 | 48 | 54 | 41 | 46 |
| 29 | 54 | 62 | 35 | 35 | 13 |
| 30 | 57 | 63 | 39 | 37 | 2 |
| 31 | 64 | 35 | 23 | 15 | |
| 32 | 41 | 43 | 23 | 9 | |
| 33 | 41 | 35 | 22 | 1 | |
| 34 | 42 | 33 | 13 | | |
| 35 | 36 | 26 | 14 | | |
| 36 | 39 | 22 | 11 | | |
| 37 | 29 | 17 | 5 | | |
| 38 | 19 | 15 | 1 | | |
| 39 | 18 | 7 | 4 | | |
| 40 | 10 | 2 | 2 | | |
| 41 | 10 | 7 | 2 | | |
| 42 | 3 | 5 | | | |
| 43 | 2 | 1 | | | |
| 44 | 3 | 1 | | | |
| 45 | 1 | | | | |
| 46 | 1 | | | | |
| Total | 1132 | 789 | 709 | 519 | 404 |

hort birth rates calculated from official statistics.[7] Except for the youngest cohort, the rates are surprisingly similar. The difference for the youngest cohort is possibly due to the fact that the official statistics also includes births of immigrants.

**7.** Figure 14.1-2 presents a graphical view of the cumulated cohort birth rates. It is remarkable that we do not find a simple relationship between age at first childbearing and completed cohort birth rates. This can be seen, for example, by comparing cohorts C30 and C40. Although members of C40 begin childbearing at younger ages, compared with members of C30

[7]These are mean values of the year-specific rates published in Fachserie 1, Reihe 1 (1999, p. 198 -200). No official data are available for C20; for C30, the mean value refers to the years 1930 and 1931.

**Table 14.1-4** Number of women with 0, 1, 2, 3, 4, and 5 or more children, calculated from the data in the GLHS subsample. Percentage values relate to all women in each of the cohorts who have at least one child. Percentage values in brackets provide the proportion of finally childless women.

| | C20 | | C30 | | C40 | | C50 | | C60 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Children | N | % | N | % | N | % | N | % | N | % |
| 0 | 109 | (17) | 38 | (11) | 39 | (11) | 86 | | 231 | |
| 1 | 185 | 35.6 | 75 | 23.4 | 78 | 24.7 | 106 | 37.6 | 145 | 56.2 |
| 2 | 168 | 32.3 | 126 | 39.3 | 139 | 44.0 | 134 | 47.5 | 86 | 33.3 |
| 3 | 104 | 20.0 | 61 | 19.0 | 64 | 20.3 | 30 | 10.6 | 21 | 8.1 |
| 4 | 40 | 7.7 | 36 | 11.2 | 23 | 7.3 | 7 | 2.5 | 6 | 2.3 |
| $\geq 5$ | 23 | 4.4 | 23 | 7.2 | 12 | 3.8 | 5 | 1.8 | | |

(see Figure 14.1-1), the completed cohort birth rate is lower for C40 than for C30. Of course, a delay of childbearing might be accompanied by a decline in the total number of births; this will probably be true for cohort C60. However, a decline of birth rates can not be explained by simply referring to changes in the distribution of ages at first childbearing.[8]

**8.** Cumulated and completed cohort birth rates provide information about the total number of children born, but not about the distribution of the number of children. So we should finally also look at the number of births *per women*. The data are shown in Table 14.1-4. Since members of birth cohorts C50 and C60 have not reached the end of the reproductive period by the time when the interviews were performed, an interpretation should be confined to the cohorts C20, C30, and C40.

a) Compared with C20, more women of C30 gave birth to at least one child. Moreover, the proportion of women with only one child declined, resulting in an increase of the mean number of children per women, from 2.2 in C20 to 2.5 in C30. The substantial increase in the completed cohort birth rate is therefore a result of both, the decline in the proportion of childless women and the increase in the mean number of children per women.

b) The proportion of childless women in C40 remains roughly the same as it was in C30. There is, however, a tendency to reduce the number of children per women. In particular, the proportion of women with four or more children declines while the proportion of women with two children increases. The result is a decline of the mean number of children per women, from 2.5 in C30 to 2.3 in C40, and consequently also a decline in the completed cohort birth rate.

It remains to be investigated how these tendencies continued in younger

[8]See also the discussion in Section 12.2.5.



birth cohorts. We already know from official statistics that the completed cohort birth rates continued to decline at least until birth cohort C60 (see Section 11.4). However, the data in our GLHS subsample do not allow to identify the changes in the distribution of children from which this tendency results.

## 14.2 Socio-economic Panel

**1.** Our second data source is the Socio-economic Panel (SOEP), already introduced in Section 8.4. In the present section we discuss data from the second wave (1985), in which participants were asked about children and their birth dates.[9] Our data set will be confined to women who belong to the subsample A of the SOEP which are mainly persons with a German citizenship.[10] In total, 4353 women with birth years from 1892 to 1968 participated in this subsample. For the data set to be used in the present section we take into account all of these women who are born not earlier than 1908 and not later than 1957. The resulting number of 3203 women is partitioned into 5-year birth cohorts as shown in the following table:

| Birth cohort | Birth years | Number of women |
|---|---|---|
| C10 | 1908 – 12 | 209 |
| C15 | 1913 – 17 | 189 |
| C20 | 1918 – 22 | 263 |
| C25 | 1923 – 27 | 322 |
| C30 | 1928 – 32 | 325 |
| C35 | 1933 – 37 | 338 |
| C40 | 1938 – 42 | 439 |
| C45 | 1943 – 47 | 339 |
| C50 | 1948 – 52 | 395 |
| C55 | 1953 – 57 | 384 |

**2.** As was done in the previous section, we begin with an investigation of the distribution of the age at first childbearing. Data are shown in Tables 14.2-1a and 14.2-1b. As in Table 14.1-1, columns labeled $d = 1$ provide numbers of women who have given birth to a first child at the corresponding age, and columns labeled $d = 0$ provide numbers of women who remained childless until the interview date. The survivor functions that can be calculated from these data are shown in Figure 14.2-1.[11]

**3.** Before any interpretations, the results should be compared with those

[9]For an earlier analysis of these data see Klein (1989).

[10]This is done in order to make our subsample comparable with the other surveys to be discussed in this chapter. This selection also allows to ignore sampling weights.

[11]Since the distributions for C25 and C30 are very similar we have omitted C25.

**Table 14.2-1a** Age at first childbearing in our SOEP subsample.

| | C10 | | C15 | | C20 | | C25 | | C30 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ |
| 15 | | | | | | | | | 1 | |
| 16 | 1 | | | | | | | | | |
| 17 | 2 | | 1 | | 3 | | 2 | | 1 | |
| 18 | 4 | | 3 | | 4 | | 4 | | 4 | |
| 19 | 2 | | 5 | | 8 | | 7 | | 7 | |
| 20 | 3 | | 8 | | 9 | | 9 | | 13 | |
| 21 | 11 | | 15 | | 17 | | 18 | | 21 | |
| 22 | 7 | | 9 | | 18 | | 17 | | 25 | |
| 23 | 10 | | 12 | | 16 | | 32 | | 23 | |
| 24 | 16 | | 19 | | 20 | | 27 | | 30 | |
| 25 | 8 | | 11 | | 15 | | 21 | | 22 | |
| 26 | 13 | | 12 | | 18 | | 23 | | 25 | |
| 27 | 14 | | 11 | | 16 | | 23 | | 15 | |
| 28 | 16 | | 15 | | 10 | | 21 | | 20 | |
| 29 | 14 | | 6 | | 12 | | 16 | | 18 | |
| 30 | 11 | | 5 | | 14 | | 13 | | 12 | |
| 31 | 7 | | 6 | | 10 | | 12 | | 5 | |
| 32 | 3 | | 2 | | 7 | | 7 | | 6 | |
| 33 | 4 | | 2 | | 4 | | 2 | | 8 | |
| 34 | 4 | | 9 | | 5 | | 3 | | 6 | |
| 35 | | | 2 | | 2 | | 2 | | 3 | |
| 36 | 3 | | | | 3 | | 3 | | 5 | |
| 37 | 1 | | 1 | | 3 | | 4 | | 2 | |
| 38 | 1 | | 1 | | 5 | | 1 | | 1 | |
| 39 | 3 | | | | 1 | | 1 | | | |
| 40 | | | | | | | 4 | | 2 | |
| 41 | 1 | | | | 1 | | 2 | | | |
| 42 | | | | | | | 1 | | | |
| 43 | 1 | | | | | | | | 1 | |
| 53 | | | | | | | | | | 4 |
| 54 | | | | | | | | | | 14 |
| 55 | | | | | | | | | | 11 |
| 56 | | | | | | | | | | 12 |
| 57 | | | | | | | | | | 8 |
| 58 | | | | | | | | 9 | | |
| 59 | | | | | | | | 11 | | |
| 60 | | | | | | | | 12 | | |
| 61 | | | | | | | | 7 | | |
| 62 | | | | | | | | 8 | | |
| 63 | | | | | | 8 | | | | |
| 64 | | | | | | 11 | | | | |
| 65 | | | | | | 12 | | | | |
| 66 | | | | | | 9 | | | | |
| 67 | | | | | | 2 | | | | |
| 68 | | | | 3 | | | | | | |
| 69 | | | | 7 | | | | | | |
| 70 | | | | 6 | | | | | | |
| 71 | | | | 10 | | | | | | |
| 72 | | | | 8 | | | | | | |
| 73 | | 10 | | | | | | | | |
| 74 | | 9 | | | | | | | | |
| 75 | | 8 | | | | | | | | |
| 76 | | 14 | | | | | | | | |
| 77 | | 8 | | | | | | | | |
| Total | 160 | 49 | 155 | 34 | 221 | 42 | 275 | 47 | 276 | 49 |



**Table 14.2-1b** Age at first childbearing in our SOEP subsample.

| | C35 | | C40 | | C45 | | C50 | | C55 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\tau$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ |
| 16 | | | 1 | | | | 1 | | 2 | |
| 17 | 2 | | 5 | | 4 | | 8 | | 7 | |
| 18 | 9 | | 9 | | 12 | | 20 | | 8 | |
| 19 | 18 | | 21 | | 19 | | 32 | | 19 | |
| 20 | 24 | | 29 | | 28 | | 35 | | 20 | |
| 21 | 23 | | 33 | | 39 | | 33 | | 22 | |
| 22 | 24 | | 34 | | 39 | | 29 | | 20 | |
| 23 | 26 | | 37 | | 32 | | 25 | | 27 | |
| 24 | 35 | | 45 | | 24 | | 22 | | 21 | |
| 25 | 24 | | 41 | | 25 | | 22 | | 27 | |
| 26 | 21 | | 32 | | 12 | | 21 | | 33 | |
| 27 | 18 | | 26 | | 23 | | 21 | | 20 | |
| 28 | 16 | | 16 | | 7 | | 21 | | 12 | 27 |
| 29 | 14 | | 11 | | 13 | | 9 | | 8 | 33 |
| 30 | 16 | | 17 | | 6 | | 14 | | 4 | 30 |
| 31 | 7 | | 8 | | 6 | | 6 | | 5 | 24 |
| 32 | 7 | | 6 | | 4 | | 7 | | | 15 |
| 33 | 3 | | 5 | | 5 | | | 20 | | |
| 34 | 2 | | 5 | | 2 | | 2 | 11 | | |
| 35 | 1 | | 4 | | 2 | | 2 | 15 | | |
| 36 | 5 | | 4 | | 4 | | | 11 | | |
| 37 | 1 | | 2 | | | | | 8 | | |
| 38 | | | 1 | | 1 | 5 | | | | |
| 39 | 1 | | 1 | | 2 | 6 | | | | |
| 40 | | | 2 | | | 5 | | | | |
| 41 | | | | | | 8 | | | | |
| 42 | | | | | | 6 | | | | |
| 43 | | | | 10 | | | | | | |
| 44 | | | | 9 | | | | | | |
| 45 | | | | 10 | | | | | | |
| 46 | | | | 7 | | | | | | |
| 47 | | | | 8 | | | | | | |
| 48 | | 8 | | | | | | | | |
| 49 | | 7 | | | | | | | | |
| 50 | | 13 | | | | | | | | |
| 51 | | 5 | | | | | | | | |
| 52 | | 8 | | | | | | | | |
| Total | 297 | 41 | 395 | 44 | 309 | 30 | 330 | 65 | 255 | 129 |

from the GLHS data discussed in the previous section. This can be done for cohorts C20, C30, C40, and C50, as shown in Figure 14.2-2. It comes without surprise that the survivor functions describing the distribution of ages at first childbearing are not identical. Since the data result from surveys and the sampled cohort sizes are small, one might have expected even greater differences. In particular for cohorts C20 and C40, both data sets provide essentially the same estimates of the proportion of finally childless women, about 16 % for C20 and 10 % for C40. An exception is

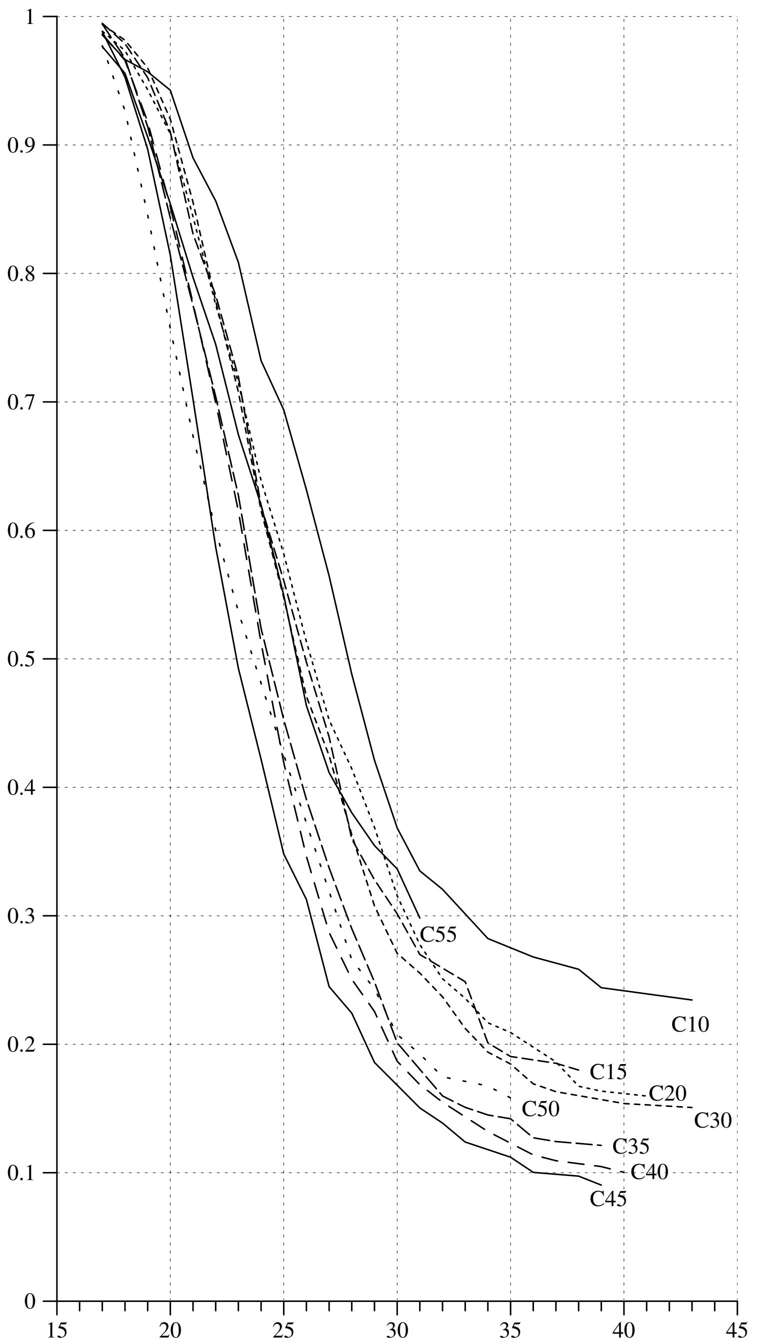



**Fig. 14.2-1** Distribution of age at first childbearing described by survivor functions, based on the data in Tables 14.2-2a and 14.2-2b.



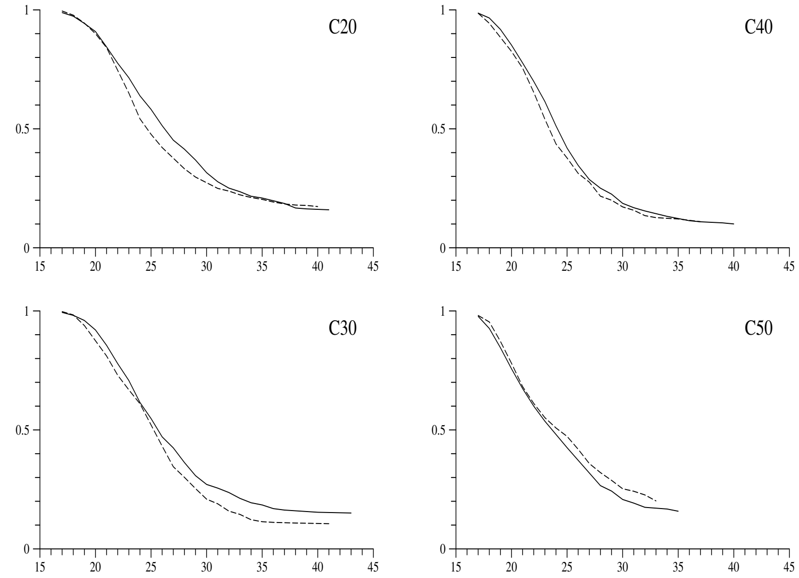



**Fig. 14.2-2** Comparison of survivor functions for the age at first childbearing estimated, respectively, with SOEP data (solid line) and GLHS data (dotted line).

the cohort C30. As will be seen below, based on a comparison of cumulated cohort birth rates, the data from the GLHS are probably more reliable than the SOEP data for this cohort.

**4.** Given that the data sets provide comparable results, Figure 14.2-1 can be used to supplement some conclusions already drawn from Figure 14.1-1. The most remarkable point is that the tendency to begin childbearing at younger ages already began with birth cohort following C10. Compared with this cohort, already women belonging to C15 had their first child at younger ages. It is also seen that this tendency holds at least until birth cohort C45, roughly corresponding to the end of the baby boom in the mid-sixties.

**5.** Further information can also be gained about the proportion of finally childless women. The following table summarizes the results from the GLHS and SOEP data:

| | C10 | C15 | C20 | C25 | C30 | C35 | C40 | C45 |
|---|---|---|---|---|---|---|---|---|
| GLHS | | | 17 | | 11 | | 11 | |
| SOEP | 23 | 18 | 16 | 15 | 15 | 12 | 10 | $\leq 9$ |

Leaving aside the SOEP result for C30, the figures indicate a long-term

**Table 14.2-2** Number of children in the SOEP subsample, classified with respect to mother's birth cohort and age ($\tau$).

| $\tau$ | C10 | C15 | C20 | C25 | C30 | C35 | C40 | C45 | C50 | C55 |
|---|---|---|---|---|---|---|---|---|---|---|
| 15 | | | | | 1 | | | | | |
| 16 | 1 | | | | | | 1 | | 1 | 2 |
| 17 | 2 | 1 | 3 | 2 | 1 | 2 | 6 | 4 | 8 | 7 |
| 18 | 4 | 3 | 6 | 4 | 4 | 9 | 11 | 13 | 23 | 8 |
| 19 | 2 | 5 | 8 | 7 | 7 | 18 | 23 | 24 | 35 | 20 |
| 20 | 5 | 9 | 11 | 11 | 16 | 28 | 42 | 38 | 44 | 22 |
| 21 | 15 | 20 | 21 | 23 | 24 | 33 | 52 | 47 | 40 | 26 |
| 22 | 7 | 18 | 24 | 22 | 34 | 40 | 54 | 53 | 45 | 31 |
| 23 | 16 | 22 | 27 | 38 | 36 | 42 | 61 | 57 | 44 | 40 |
| 24 | 24 | 29 | 32 | 40 | 47 | 63 | 75 | 52 | 41 | 38 |
| 25 | 20 | 24 | 23 | 42 | 44 | 57 | 79 | 61 | 41 | 51 |
| 26 | 26 | 21 | 35 | 43 | 48 | 50 | 75 | 48 | 52 | 65 |
| 27 | 27 | 21 | 34 | 51 | 40 | 43 | 68 | 46 | 45 | 43 |
| 28 | 30 | 24 | 33 | 43 | 39 | 49 | 60 | 19 | 55 | 41 |
| 29 | 38 | 21 | 35 | 46 | 46 | 50 | 52 | 35 | 29 | 24 |
| 30 | 22 | 17 | 27 | 49 | 39 | 64 | 53 | 19 | 40 | 14 |
| 31 | 26 | 17 | 29 | 44 | 35 | 27 | 41 | 17 | 26 | 12 |
| 32 | 17 | 12 | 21 | 48 | 20 | 26 | 20 | 20 | 29 | 1 |
| 33 | 16 | 15 | 20 | 27 | 27 | 36 | 25 | 14 | 15 | |
| 34 | 19 | 15 | 24 | 30 | 22 | 27 | 17 | 13 | 6 | |
| 35 | 15 | 14 | 9 | 23 | 26 | 18 | 15 | 9 | 8 | |
| 36 | 12 | 7 | 19 | 21 | 23 | 16 | 10 | 14 | 3 | |
| 37 | 13 | 9 | 14 | 20 | 11 | 10 | 13 | 11 | 1 | |
| 38 | 11 | 9 | 12 | 10 | 6 | 5 | 8 | 7 | | |
| 39 | 10 | 5 | 10 | 11 | 11 | 8 | 4 | 4 | | |
| 40 | 7 | 11 | 7 | 18 | 12 | 3 | 5 | 1 | | |
| 41 | 5 | 4 | 7 | 6 | 1 | | 3 | | | |
| 42 | 2 | 2 | 2 | 6 | 3 | 2 | | | | |
| 43 | 4 | | 3 | 4 | 3 | 2 | 1 | | | |
| 44 | 1 | | 1 | 1 | 1 | 1 | | | | |
| 45 | | | 1 | 2 | | | | | | |
| 46 | 1 | | 1 | | 1 | | | | | |
| 47 | 1 | | | | | | | | | |
| 48 | | | | | | | | | | |
| 49 | | | | | | | | | | |
| 50 | | | | | | | | | | |
| 51 | | | | | | | | | | |
| 52 | | 1 | | | | | | | | |
| Total | 399 | 356 | 499 | 692 | 628 | 729 | 874 | 626 | 631 | 445 |

decrease in the proportion of finally childless women at least until birth cohort C45. Of course, in interpretating these figures one has to consider the fact that the data result from retrospective surveys and consequently only provide information about women who survived the interview dates in the 1980 s. The proportions of finally childless women would presumably quite higher if related to all women of the respective birth cohorts.



**Table 14.2-3** Cumulated cohort birth rates up to an age of $\tau$, calculated from SOEP data ($\mathrm{CCBR}_s(\tau)$), from GLHS data ($\mathrm{CCBR}_g(\tau)$), and from official statistics ($\mathrm{CCBR}^*(\tau)$). Fachserie 1, Reihe 1, 1999 (pp. 198-200).

| Birth cohort | $\tau$ | $\mathrm{CCBR}_s(\tau)$ | $\mathrm{CCBR}_g(\tau)$ | $\mathrm{CCBR}^*(\tau)$ |
|---|---|---|---|---|
| C10 | 45 | 1.90 | | |
| C15 | 45 | 1.88 | | |
| C20 | 45 | 1.89 | 1.80 | |
| C25 | 45 | 2.15 | | |
| C30 | 43 | 1.93 | 2.19 | 2.15 |
| C35 | 43 | 2.15 | | 2.18 |
| C40 | 40 | 1.98 | 1.99 | 1.96 |
| C45 | 38 | 1.83 | | 1.75 |
| C50 | 31 | 1.44 | 1.38 | 1.39 |
| C55 | 29 | 1.09 | | 1.09 |
| C60 | 29 | | 0.82 | 0.99 |

**6.** The next step is to investigate the number of children that were born of women in our SOEP subsample. We begin with the calculation of cumulated cohort birth rates. Table 14.2-2 shows the data and is organized in the same way as Table 14.1-2. With the exception of cohort C30, plotting the cumulated cohort birth rates would show mainly the same cross-cohort changes as have been visible in Figure 14.1-2. We therefore only compare the cumulated cohort birth rates up to some higher ages as shown in Table 14.2-3. Also shown are comparable rates calculated from official statistics.[12] The comparison suggests that the SOEP data for birth cohort C30 are, in fact, somewhat exceptional and that, for this cohort, the GLHS data might be more reliable. However, more interesting is the additional information that can be gained for birth cohorts born before 1930. Since official statistics only allows to calculate completed cohort birth rates beginning with birth year 1930, one might easily get the impression of a long-term decline of these rates that began with birth cohorts following C35 (see Section 11.4). Quite to the contrary, our survey data suggest that the birth rates of cohorts with birth years roughly between 1925 and 1935 were exceptional high.

**7.** Finally, we can distinguish women with regard to parity. Results from the SOEP subsample are shown in Table 14.2-4. In the same way as was done in Table 14.1-4 in the previous section, the lower panel of Table 14.2-4 shows the distribution of parities in subsets of women having at least one child. This allows to separate the parity distribution from effects that result from a changing proportion of finally childless women. As an example, we consider the proportion of women having four or more children. How this proportion developed is shown graphically in Figure

[12]These rates are calculated as mean values for 3-year periods in the same way as was explained in the previous section.

**Table 14.2-4** Upper panel: Number of women with 0, 1, 2, 3, 4, and 5 or more children, calculated from the data in the SOEP subsample. Lower panel: Percentage values relating to all women in each of the cohorts who have at least one child.

| Children | C10 | C15 | C20 | C25 | C30 | C35 | C40 | C45 | C50 | C55 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 49 | 34 | 42 | 47 | 49 | 41 | 44 | 30 | 65 | 129 |
| 1 | 46 | 41 | 79 | 74 | 80 | 59 | 101 | 89 | 99 | 106 |
| 2 | 55 | 66 | 71 | 95 | 106 | 129 | 185 | 146 | 173 | 115 |
| 3 | 30 | 23 | 41 | 52 | 47 | 62 | 62 | 57 | 48 | 27 |
| 4 | 14 | 16 | 15 | 25 | 28 | 26 | 31 | 12 | 8 | 7 |
| 5+ | 15 | 9 | 15 | 29 | 15 | 21 | 16 | 5 | 2 | |
| **Children** | **C10** | **C15** | **C20** | **C25** | **C30** | **C35** | **C40** | **C45** | **C50** | **C55** |
| 1 | 28.8 | 26.5 | 35.7 | 26.9 | 29.0 | 19.9 | 25.6 | 28.8 | | |
| 2 | 34.4 | 42.6 | 32.1 | 34.5 | 38.4 | 43.4 | 46.8 | 47.2 | | |
| 3 | 18.8 | 14.8 | 18.6 | 18.9 | 17.0 | 20.9 | 15.7 | 18.4 | | |
| 4 | 8.8 | 10.3 | 6.8 | 9.1 | 10.1 | 8.8 | 7.8 | 3.9 | | |
| 5+ | 9.4 | 5.8 | 6.8 | 10.5 | 5.4 | 7.1 | 4.1 | 1.6 | | |

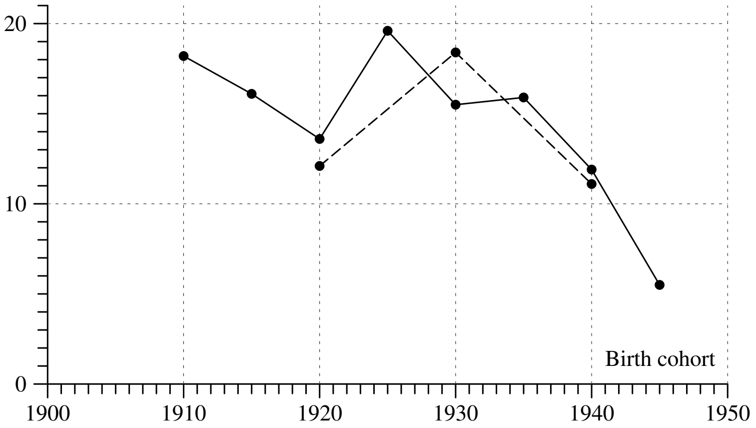



**Fig. 14.2-3** Percentages of women with four or more children belonging to birth cohorts C10, . . . , C45; calculated from SOEP data (solid line) and from GLHS data (dotted line).

14.2-3. The solid line connects figures calculated from the SOEP data, the dotted line connects comparable figres from the GLHS. Again, the value for the C30 cohort in the SOEP data should be considered as exceptional. However, the remarkable result is that we do not find a continuous long-term decline in the proportion of women with four or more children. To the contrary, an initial decline was superseded by rising proportions in birth cohorts with birth years roughly between 1920 and 1930. A repeated decline only began roughly at the time when the baby boom ended in the second half of the 1960 s.



## 14.3 Fertility and Family Survey

**1.** Even if surveys refer to the same region and historical period they are likely to provide more or less different data. So it is always a good idea to consider all possibly informative data sources and compare the information. In the present section we use data from the German part of the *Fertility and Family Survey* (FFS). The FFS project was initiated by the Population Activities Unit (PAU) of the United Nations Economic Commission for Europe (UNECE) in order to conduct comparable Fertility and Family Surveys in about 20 ECE member countries.[13] The German FFS was conducted by the *Bundesinstitut für Bevölkerungsforschung* (BiB, Wiesbaden) in 1992.[14] While several studies using these data have already been performed and published,[15] the data set is now generally available for scientific research.[16]

**2.** The sampling design intended to get data from 10000 persons, 5000 in the territory of the former FRG ("West") and 5000 in the territory of the former GDR ("East"). In both territories, 3000 women and 2000 men of age 20 to 39, having a German citizenship, should be included.[17] The field work was done during the period May to September in 1992 using a random route method to select persons for the survey. The final sample includes data from interviews with 10012 persons. The number of male and female sample members in both regions of Germany is shown in the left part of the following table:

| | All sample members | | With valid birth year | |
|---|---|---|---|---|
| Region | Male | Female | Male | Female |
| West | 2024 | 3012 | 2016 | 3005 |
| East | 1992 | 2984 | 1982 | 2971 |

Since for 38 persons neither a valid birth year nor a valid age at the time of the interview is known, the number of cases reduces as shown in the right part of the table. All remaining persons are born between 1952 and

[13]See `www.unece.org/ead/pau/ffs/`. Festy and Prioux (2002) provide an overview and evaluation.

[14]The basic data documentation is by Pohl (1995). For additional information see the homepage of the BiB: `www.bib-demographie.de`.

[15]Hullen (1998), Roloff and Dorbritz (1999).

[16]We thank Gert Hullen (BiB) who provided us with a copy of the data. The data set is also available from the *Zentralarchiv für empirische Sozialforschung* (Köln).

[17]For more details on the sampling design see Pohl (1995, pp. 7-8).

**Table 14.3-1** Age at first childbearing in the FFS subsample.

| | West | | | | | | East | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C55 | | C60 | | C65 | | C55 | | C60 | | C65 | |
| $\tau$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ | $d=1$ | $d=0$ |
| 15 | 1 | | 2 | | | | 3 | | 1 | | 1 | |
| 16 | 6 | | 5 | | 3 | | 3 | | 6 | | 5 | |
| 17 | 12 | | 6 | | 4 | | 7 | | 14 | | 9 | |
| 18 | 32 | | 13 | | 13 | | 35 | | 36 | | 23 | |
| 19 | 28 | | 22 | | 13 | | 59 | | 69 | | 57 | |
| 20 | 49 | | 21 | | 28 | | 67 | | 106 | | 88 | |
| 21 | 38 | | 32 | | 25 | | 94 | | 108 | | 95 | |
| 22 | 30 | | 35 | | 19 | | 94 | | 90 | | 91 | |
| 23 | 34 | | 31 | | 53 | | 68 | | 62 | | 76 | |
| 24 | 40 | | 39 | | 23 | | 54 | | 52 | | 49 | |
| 25 | 39 | | 50 | | 32 | 141 | 42 | | 48 | | 30 | 52 |
| 26 | 36 | | 32 | | 26 | 103 | 25 | | 33 | | 30 | 45 |
| 27 | 30 | | 47 | | 17 | 101 | 17 | | 17 | | 5 | 42 |
| 28 | 30 | | 39 | | 8 | 96 | 11 | | 15 | | 6 | 19 |
| 29 | 28 | | 33 | | 2 | 59 | 13 | | 6 | | | 26 |
| 30 | 33 | | 14 | 91 | | | 7 | | 8 | 26 | | |
| 31 | 20 | | 9 | 46 | | | 7 | | 3 | 24 | | |
| 32 | 13 | | 4 | 67 | | | 4 | | 2 | 15 | | |
| 33 | 12 | | 1 | 42 | | | 5 | | | 9 | | |
| 34 | 6 | | | 41 | | | 2 | | | 17 | | |
| 35 | 8 | 47 | | | | | 2 | 19 | | | | |
| 36 | 3 | 37 | | | | | 2 | 10 | | | | |
| 37 | | 35 | | | | | | 11 | | | | |
| 38 | 1 | 34 | | | | | | 12 | | | | |
| 39 | | 46 | | | | | | 30 | | | | |
| Total | 529 | 199 | 435 | 287 | 266 | 500 | 621 | 82 | 676 | 91 | 565 | 184 |

1972. Since our interest concerns births we only consider female sample members. In order to allow comparisons with the GLHS and SOEP we only consider women who belong to one of the birth cohorts shown in the following table:

| Birth cohort | Birth years | West | East |
|---|---|---|---|
| C55 | 1953 – 57 | 728 | 704 |
| C60 | 1958 – 62 | 723 | 767 |
| C65 | 1963 – 67 | 768 | 751 |

**3.** As was done in the previous sections, we begin with an investigation of the distribution of ages at first childbearing. Table 14.3-1, organized in the same way as Tables 14.1-1 and 14.2-1, shows the data and can be used to calculate survivor functions.[18] For the cohorts C55 (West) and

[18]Like the GLHS, also the FFS allows to distinguish women's own children from step children and adoptive children. For creating the data in Table 14.3-1 we have only



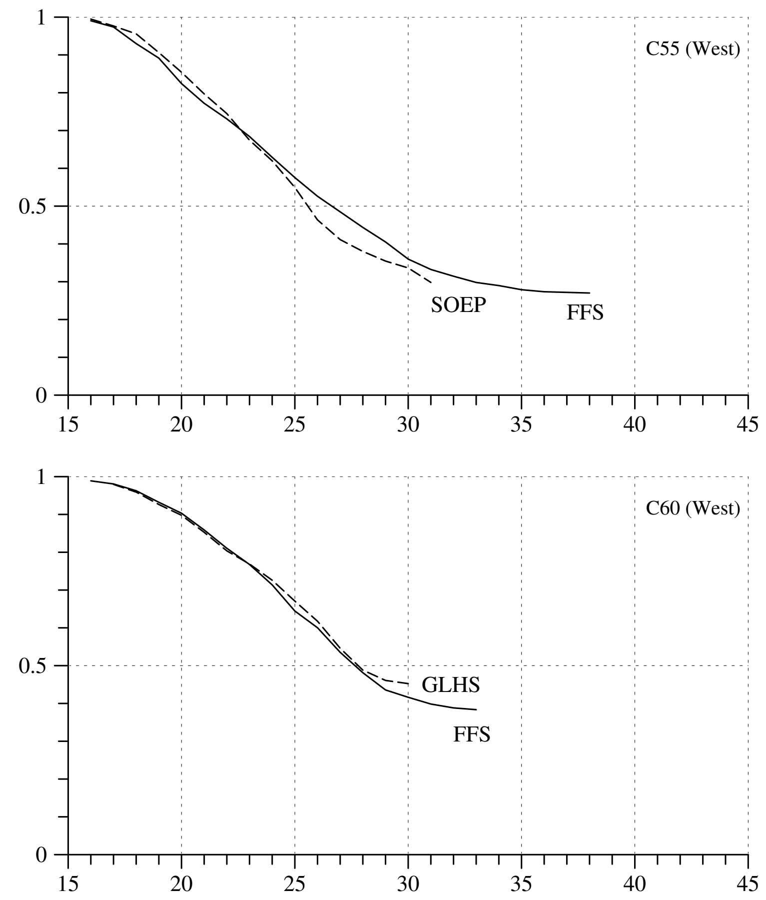



**Fig. 14.3-1** Comparison of survivor functions for the age at first childbearing. FFS survivor functions are calculated from the data in Table 14.3-1. The SOEP and GLHS survivor functions are taken from Figures 14.1-1 and 14.2-1, respectively.

C60 (West) they can be compared with corresponding survivor functions from the SOEP and GLHS respectively. As can be seen in Figure 14.3-1, the curves agree quite well. So we can turn to a comparison of all six age distributions that can be calculated with the data in Table 14.3-1. The result is shown in Figure 14.3-2. Quite remarkable is the difference between the distributions in both territories. In the former GDR, women began childbearing at substantially younger ages, and also the proportion

considered women's own children. One should note, however, that in a few cases no valid birth year for the first child is available, the number of women referred to in Table 14.3-1 is therefore slightly smaller than in the table in paragraph 2.

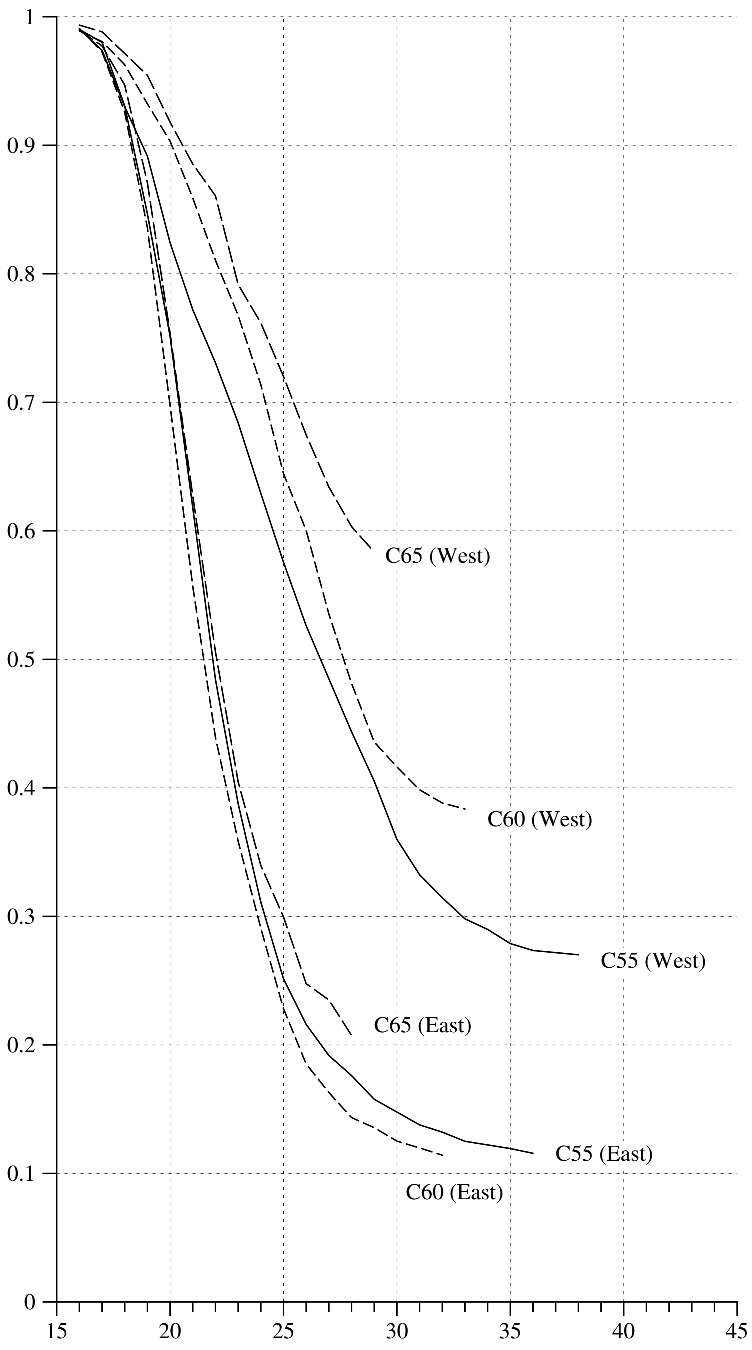



**Fig. 14.3-2** Distribution of age at first childbearing described by survivor functions, calculated from the data in Table 14.3-2.



**Table 14.3-2** Number of children in the FFS subsample, classified with respect to mother's birth cohort and age ($\tau$).

| | West | | | East | | |
|---|---|---|---|---|---|---|
| $\tau$ | C55 | C60 | C65 | C55 | C60 | C65 |
| 15 | 1 | 2 | | 3 | 1 | 1 |
| 16 | 7 | 5 | 4 | 4 | 7 | 5 |
| 17 | 12 | 6 | 6 | 8 | 14 | 10 |
| 18 | 36 | 17 | 13 | 37 | 39 | 26 |
| 19 | 31 | 23 | 13 | 62 | 75 | 61 |
| 20 | 61 | 27 | 38 | 79 | 119 | 100 |
| 21 | 50 | 42 | 29 | 113 | 143 | 114 |
| 22 | 51 | 58 | 31 | 128 | 142 | 138 |
| 23 | 49 | 55 | 69 | 116 | 127 | 118 |
| 24 | 70 | 60 | 48 | 105 | 112 | 108 |
| 25 | 71 | 75 | 60 | 113 | 111 | 69 |
| 26 | 62 | 62 | 47 | 86 | 110 | 71 |
| 27 | 61 | 85 | 32 | 74 | 81 | 30 |
| 28 | 64 | 84 | 13 | 53 | 63 | 13 |
| 29 | 62 | 70 | 9 | 41 | 36 | 1 |
| 30 | 69 | 56 | | 47 | 29 | |
| 31 | 52 | 38 | | 28 | 19 | |
| 32 | 40 | 17 | | 26 | 8 | |
| 33 | 38 | 13 | | 23 | 3 | |
| 34 | 26 | 2 | | 8 | 1 | |
| 35 | 25 | | | 8 | | |
| 36 | 11 | | | 6 | | |
| 37 | 1 | | | 5 | | |
| 38 | 4 | | | 1 | | |
| 39 | 1 | | | | | |
| Total | 955 | 797 | 412 | 1174 | 1240 | 865 |
| Total* | 981 | 826 | 429 | 1213 | 1253 | 887 |

of childless women was much smaller than in the former FRG. Furthermore, the distribution is quite similar for all three cohorts. In contrast, the tendency of delaying childbearing into older ages continues in the western part of Germany. Of course, at least for the birth cohorts C60 and C65, the data to not allow to reliably estimate the proportion of eventually childless women.

**4.** We now turn to the number of children and begin with cumulated cohort birth rates. The data are shown in Table 14.3-2. As in Table 14.3-1, we have only considered women's own children. We also note that the FFS questionnaire only asked for birth years of up to four children. However, the number of women with more than four children is quite small (seven women have five, and four women have six children). More important is the number of cases where, for one or more children, there is no valid birth year. This is documented in the last two rows of Table 14.3-2. The row labeled Total* has been calculated from women's report on the total number of their own children, so that the difference between both rows

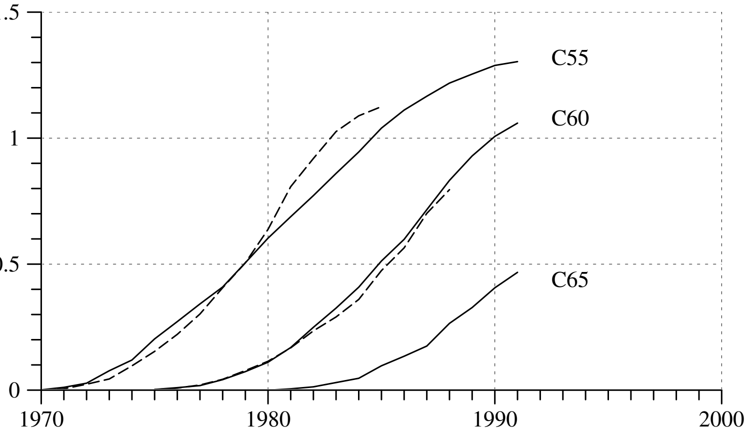



**Fig. 14.3-3** Cumulated cohort birth rates calculated from Table 14.3-2 for three cohorts in the western part of Germany (solid lines). The dotted lines show corresponding rates calculated from the SOEP (C55) and the GLHS (C60).

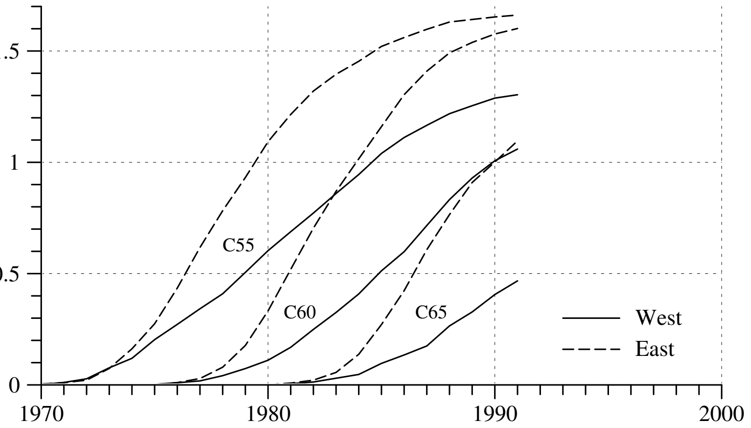



**Fig. 14.3-3** Cumulated cohort birth rates calculated from the data in Table 14.3-2 for the western part (solid lines) and the eastern part (dotted lines) of Germany.

amounts to the number of children without a valid birth year. However, the impact of these missing values on *cumulated* cohort birth rates is quite limited, and so the data can nevertheless be used for further investigation. Figure 14.3-3 shows these cumulated rates for the cohorts in the western part of Germany and, for cohorts C55 and C60, also provides a comparison with the results from the SOEP and the GLHS data, respectively. Figure 14.3-4 compares the rates between both territories.



## 14.4 DJI Family Surveys

**1.** A further source of information about childbearing histories in Germany is a series of surveys conducted by the *Deutsches Familieninstitut* (DJI, München). Data sets are available from the *Zentralarchiv für empirische Sozialforschung* (Köln). In the present section we use data from a survey conducted in the territory of the former FRG in 1988. The sample refers to persons with a German citizenship who, at the interview date in 1988, lived in private households and were between 18 and 55 years old.[19] The final sample size is 10043, 4554 men and 5489 women. The following table shows the distribution of birth years of the female participants:

| Birth year | Number | Birth year | Number | Birth year | Number |
|---|---|---|---|---|---|
| 1933 | 130 | 1946 | 96 | 1959 | 161 |
| 1934 | 121 | 1947 | 128 | 1960 | 169 |
| 1935 | 129 | 1948 | 165 | 1961 | 158 |
| 1936 | 153 | 1949 | 143 | 1962 | 158 |
| 1937 | 143 | 1950 | 158 | 1963 | 165 |
| 1938 | 139 | 1951 | 173 | 1964 | 179 |
| 1939 | 149 | 1952 | 172 | 1965 | 145 |
| 1940 | 142 | 1953 | 153 | 1966 | 147 |
| 1941 | 123 | 1954 | 166 | 1967 | 112 |
| 1942 | 129 | 1955 | 185 | 1968 | 118 |
| 1943 | 148 | 1956 | 157 | 1969 | 114 |
| 1944 | 137 | 1957 | 182 | 1970 | 66 |
| 1945 | 96 | 1958 | 180 | | |

For compatibility with the data discussed in previous sections we consider the following birth cohorts:

| Birth cohort | Birth years | Number of women |
|---|---|---|
| C35 | 1933 – 1937 | 676 |
| C40 | 1938 – 1942 | 682 |
| C45 | 1943 – 1947 | 605 |
| C50 | 1948 – 1952 | 811 |
| C55 | 1953 – 1957 | 843 |
| C60 | 1958 – 1962 | 826 |

Women born later than 1962 will not be considered because their age in 1988 does not allow any reliable conclusions about childbearing histories.

**2.** As was done in the previous sections, we begin with an investigation of ages at first childbearing. This is easy because the data set already contains

[19]For a description of the sampling design see Alt (1991).

**Table 14.4-1** Age at first childbearing in our DJI subsample.

| $\tau$ | C35 $d{=}1$ | C35 $d{=}0$ | C40 $d{=}1$ | C40 $d{=}0$ | C45 $d{=}1$ | C45 $d{=}0$ | C50 $d{=}1$ | C50 $d{=}0$ | C55 $d{=}1$ | C55 $d{=}0$ | C60 $d{=}1$ | C60 $d{=}0$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 1 | | 2 | | | | | | 3 | | | |
| 16 | 4 | | 1 | | | | 3 | | 6 | | 4 | |
| 17 | 10 | | 10 | | 5 | | 15 | | 7 | | 13 | |
| 18 | 15 | | 12 | | 22 | | 32 | | 38 | | 19 | |
| 19 | 37 | | 25 | | 48 | | 50 | | 42 | | 31 | |
| 20 | 46 | | 51 | | 41 | | 56 | | 41 | | 40 | |
| 21 | 35 | | 53 | | 55 | | 49 | | 39 | | 49 | |
| 22 | 69 | | 65 | | 56 | | 56 | | 53 | | 55 | |
| 23 | 54 | | 68 | | 51 | | 52 | | 65 | | 39 | |
| 24 | 58 | | 59 | | 40 | | 65 | | 51 | | 54 | |
| 25 | 65 | | 46 | | 33 | | 55 | | 57 | | 69 | |
| 26 | 55 | | 50 | | 28 | | 48 | | 63 | | 54 | 83 |
| 27 | 39 | | 31 | | 26 | | 46 | | 59 | | 44 | 66 |
| 28 | 20 | | 29 | | 23 | | 45 | | 39 | | 32 | 50 |
| 29 | 32 | | 28 | | 26 | | 38 | | 39 | | 11 | 56 |
| 30 | 9 | | 14 | | 17 | | 17 | | 26 | | 7 | 50 |
| 31 | 15 | | 18 | | 13 | | 23 | | 21 | 44 | | |
| 32 | 6 | | 7 | | 11 | | 18 | | 10 | 31 | | |
| 33 | 9 | | 5 | | 9 | | 14 | | 6 | 36 | | |
| 34 | 6 | | 4 | | 5 | | 9 | | 4 | 39 | | |
| 35 | 1 | | 6 | | 8 | | 7 | | 2 | 22 | | |
| 36 | 4 | | 3 | | 4 | | 7 | 21 | | | | |
| 37 | 1 | | 2 | | 1 | | 3 | 32 | | | | |
| 38 | 1 | | 1 | | | | 1 | 22 | | | | |
| 39 | | | 1 | | | | 2 | 11 | | | | |
| 40 | | | 3 | | 3 | | | 13 | | | | |
| 41 | | | | | 1 | 14 | | | | | | |
| 42 | | | 1 | | | 16 | | | | | | |
| 43 | 1 | | | | | 8 | | | | | | |
| 44 | 1 | | | | | 22 | | | | | | |
| 45 | | | | | | 19 | | | | | | |
| 46 | | | | 10 | | | | | | | | |
| 47 | | | | 15 | | | | | | | | |
| 48 | | | | 15 | | | | | | | | |
| 49 | | | | 28 | | | | | | | | |
| 50 | | | | 18 | | | | | | | | |
| 51 | | 11 | | | | | | | | | | |
| 52 | | 22 | | | | | | | | | | |
| 53 | | 14 | | | | | | | | | | |
| 54 | | 15 | | | | | | | | | | |
| 55 | | 18 | | | | | | | | | | |
| Total | 594 | 80 | 595 | 86 | 526 | 79 | 711 | 99 | 671 | 172 | 521 | 305 |



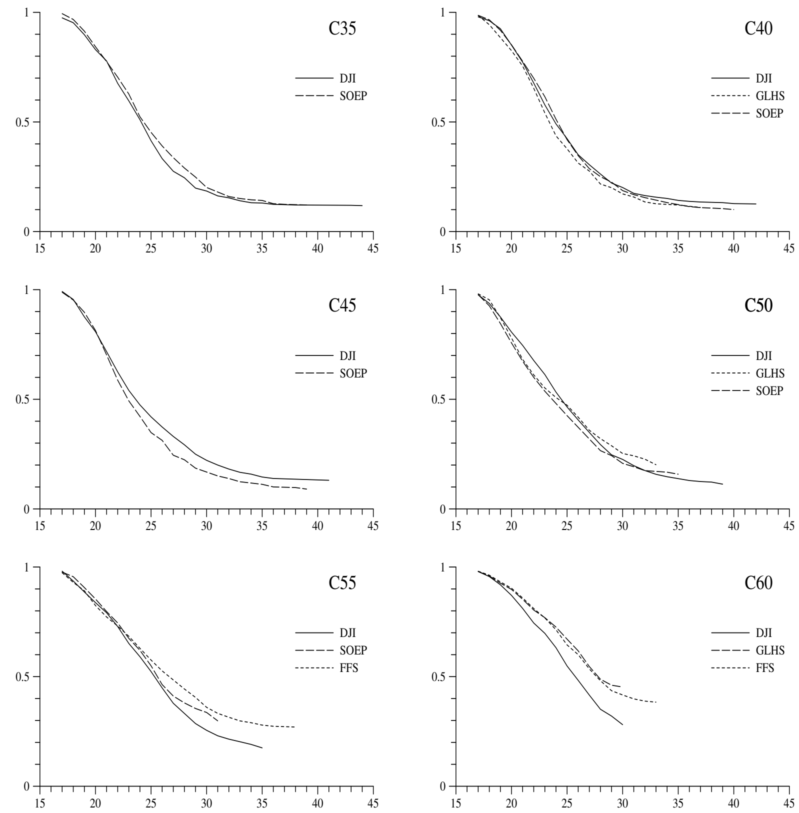



**Fig. 14.4-1** Comparison of survivor functions for the age at first childbearing for birth cohorts C35, C40, C45, C50, C55, and C60.

a variable providing the age of women at first childbearing.[20] Table 14.4-1 shows, separately for birth cohorts, how many women of specified age have given birth to a child ($d = 1$) or are censored at the interview date ($d = 0$).[21] These data can be used to estimate survivor functions as in the previous sections. Figure 14.4-1 compares the survivor functions with estimates based on the GLHS SOEP and FFS data. For birth cohorts C35, C40, C45, and C50, the results are quite similar. Substantial differences only occur for the two younger cohorts, C55 and C60.

[20] We have used the SPSS file `fall88.sav`. The variable providing age at first childbearing is `F275_ALT`.

[21] Notice that for some birth cohorts the totals are slightly smaller than the number of cases tabulated in the preceeding paragraph because we have dropped cases with a reported age at first childbearing below 15.

## Chapter 15

# Birth Rates in East Germany

This chapter is not finished yet.

## Chapter 16

# In- and Out-Migration

This chapter is not finished yet.

# Chapter 17

# An Analytical Modeling Approach

In the present chapter we begin with the discussion of an analytical model that can support modal reasoning about demographic processes. We begin with a version of the model that takes into account births and deaths but ignores migration. How to extend the model in order to include migration will be discussed in Section 18.6.

## 17.1 Conceptual Framework

**1.** To introduce a conceptual framework for the model, we refer to a demographic process, $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}}^*, \Omega_t)$, as discussed in Section 3.2. $\boldsymbol{\mathcal{S}}$ provides the spatial context, $\boldsymbol{\mathcal{T}}^*$ is the time axis, and $\Omega_t$ represents the population living in the space $\boldsymbol{\mathcal{S}}$ in the temporal location $t \in \boldsymbol{\mathcal{T}}^*$. The numbers of men and women in $\Omega_t$ aged $\tau$ will be denoted by $n^m_{t,\tau}$ and $n^f_{t,\tau}$, respectively; the total number of persons aged $\tau$ will be denoted by $n_{t,\tau} := n^m_{t,\tau} + n^f_{t,\tau}$. To simplify notations we will assume that age is measured in the same time units that are used in the definition of $\boldsymbol{\mathcal{T}}^*$. For example, if $\boldsymbol{\mathcal{T}}^*$ refers to calendar years, it will be assumed that age is measured in completed years. We also assume a maximal age which will be denoted by $\tau_m$.[1]

**2.** To formulate the model it is now helpful to use matrix notations.[2] Classified by age, the male and female population will be represented, respectively, by the vectors

$$\mathbf{n}^m_t := \begin{bmatrix} n^m_{t,1} \\ \vdots \\ n^m_{t,\tau_m} \end{bmatrix} \quad \text{and} \quad \mathbf{n}^f_t = \begin{bmatrix} n^f_{t,1} \\ \vdots \\ n^f_{t,\tau_m} \end{bmatrix} \tag{17.1.1}$$

In addition, we represent the total population by the vector $\mathbf{n}_t := \mathbf{n}^m_t + \mathbf{n}^f_t$. Notice that the count of vector elements begins with 1, not with 0, so that only persons who have reached an age of one time unit will be given an explicit representation.

**3.** The purpose of a demographic model is to provide a conceptual framework for thinking about possible developments of a population:

$$\mathbf{n}_0 \longrightarrow \mathbf{n}_1 \longrightarrow \mathbf{n}_2 \longrightarrow \cdots$$

[1]This is not a serious limitation because $\tau_m$ can be given an arbitrarily high value; also, in practical applications, $\tau_m$ can be assumed to be an open-ended age class.

[2]For a brief introduction to matrix notations and elementary rules see Rohwer and Pötter (2002a, Appendix A).



that begin in some arbitrary temporal location with an initial population $\Omega_0$, here represented by the vector $\mathbf{n}_0$. This requires the introduction of rules that can be used to derive $\mathbf{n}_1$ from $\mathbf{n}_0$, $\mathbf{n}_2$ from $\mathbf{n}_1$, and so on. Since we ignore migration (think of $\boldsymbol{\mathcal{S}}$ as a closed region), it suffices to take into account birth and death events. However, only women can give birth to children, and so it is necessary to represent the process in the following way:

$$\begin{array}{ccccccc} \mathbf{n}^m_0 & \longrightarrow & \mathbf{n}^m_1 & \longrightarrow & \mathbf{n}^m_2 & \longrightarrow & \cdots \\ & \nearrow & & \nearrow & & \nearrow & \\ \mathbf{n}^f_0 & \longrightarrow & \mathbf{n}^f_1 & \longrightarrow & \mathbf{n}^f_2 & \longrightarrow & \cdots \end{array}$$

**4.** In order to formulate rules we use age-specific birth and death rates. Death rates for men and women at age $\tau$ in temporal location $t$ will be denoted, respectively, by

$$\delta^m_{t,\tau} \quad \text{and} \quad \delta^f_{t,\tau}$$

Given these rates, the number of men and women dying in $t$ at age $\tau$ is $\delta^m_{t,\tau}\, n^m_{t,\tau}$ and $\delta^f_{t,\tau} n^f_{t,\tau}$, respectively. Notice that the assumption of a maximal age $\tau_m$ implies that $\delta^m_{t,\tau_m} = \delta^f_{t,\tau_m} = 1$.

**5.** Age-specific birth rates will be denoted by $\beta^*_{t,\tau}$.[3] In order to simplify the formulation of the model these rates will be interpreted as follows: $\beta^*_{t,\tau} n^f_{t,\tau}$ is the number of children, born of women at age $\tau$ in temporal location $t$, who survived the first time unit and are consequently members of $\Omega_{t+1}$. Of course, since only women can bear children, these birth rates need not be indexed with respect to sex. However, one has to take into account differences in the percentages of male and female births. We use $\sigma_m$ and $\sigma_f$ to denote the proportions ($\sigma_m + \sigma_f = 1$). Therefore, if $n_{t+1,1}$ is the total number of children born in $t$, the number of male children is $n^m_{t+1,1} = \sigma_m n_{t+1,1}$ and the number of female children is $n^f_{t+1,1} = \sigma_f n_{t+1,1}$. To ease notations, we assume that the sex ratio at birth is independent of mother's age and constant over time.

**6.** Since we only consider children who survived the first time unit we also do not explicitly model death rates of children during the temporal location in which they are born. There is, however, a simple relationship between $\beta^*_{t,\tau}$ and the birth rates $\beta_{t,\tau}$, introduced in Section 11.1:

$$\beta^*_{t,\tau} = \beta_{t,\tau}\,(1 - \delta_{t,0})$$

In this formulation, $\delta_{t,0} = \sigma_m \delta^m_{t,0} + \sigma_f \delta^f_{t,0}$ is a weighted mean of the death rates of male and female children during their first year of life.

[3]We assume that these birth rates are defined for all ages and have a value of zero at ages outside the reproductive period of women.

**7.** Assuming that birth and death rates are given, one can derive some elementary rules for the development of the population. First, the total number of children born in temporal location $t$ and still alive in $t+1$ can be derived from $\mathbf{n}_t^f$ and the age-specific birth rates as follows:

$$n_{t+1,1} \;=\; \sum_{\tau=1}^{\tau_m} \beta^*_{t,\tau} n^f_{t,\tau}$$

Secondly, the relation between the number of men and women at ages $\tau \geq 1$ in two successive temporal locations can be derived from death rates:

$$n^m_{t+1,\tau+1} \;=\; (1-\delta^m_{t,\tau})\, n^m_{t,\tau} \quad \text{and} \quad n^f_{t+1,\tau+1} \;=\; (1-\delta^f_{t,\tau})\, n^f_{t,\tau}$$

Together, the three equations allow to derive $\mathbf{n}_t^m$ and $\mathbf{n}_t^f$ from $\mathbf{n}_0^m$ and $\mathbf{n}_0^f$ for all $t>0$. Of course, this requires to think of the birth and death rates, and also the proportions of male and female births, as given and known parameters of the demographic process.

**8.** We now proceed with matrix notation. First, we define $(\tau_m,\tau_m)$ matrices

$$\mathbf{B}_t \;:=\; \begin{bmatrix} \beta^*_{t,1} & \beta^*_{t,2} & \cdots & \beta^*_{t,\tau_m} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

which comprise the age-specific birth rates. The number of male and female children in $t+1$ is then given, respectively, by

$$\begin{bmatrix} n^m_{t+1,1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_m\, \mathbf{B}_t\, \mathbf{n}_t^f \quad \text{and} \quad \begin{bmatrix} n^f_{t+1,1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_f\, \mathbf{B}_t\, \mathbf{n}_t^f$$

Secondly, we define $(\tau_m,\tau_m)$ matrices $\mathbf{D}_{m,t}$ and $\mathbf{D}_{f,t}$ which comprise the death rates of men and women:

$$\mathbf{D}_{m,t} \;:=\; \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1-\delta^m_{t,1} & 0 & \cdots & 0 & 0 \\ 0 & 1-\delta^m_{t,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-\delta^m_{t,\tau_m-1} & 0 \end{bmatrix}$$

$\mathbf{D}_{f,t}$ is of the same form but has $\delta^f_{t,\tau}$ instead of $\delta^m_{t,\tau}$. Using these matrices, the three equations derived in the previous paragraph can be written as

$$\begin{aligned} \mathbf{n}^m_{t+1} \;&=\; \mathbf{D}_{m,t}\, \mathbf{n}_t^m + \sigma_m\, \mathbf{B}_t\, \mathbf{n}_t^f \\ \mathbf{n}^f_{t+1} \;&=\; \mathbf{D}_{f,t}\, \mathbf{n}_t^f + \sigma_f\, \mathbf{B}_t\, \mathbf{n}_t^f \end{aligned} \qquad (17.1.2)$$



## 17.2 The Stable Population

**1.** The model framework introduced in the previous section can be used to speculate about possible population developments. This, of course, requires additional assumptions about birth and death rates and how they change over time. The simplest assumption is that the rates are constant over time. This assumptions leads to the idea of a *stable population*, an idea first developed by Alfred J. Lotka (1907, 1922). In the present section we illustrate the idea by an example; some mathematical details will be discussed in the next section.

**2.** To begin with, we distinguish between the size of a population and its age distribution. Changes of size can be described by growth rates. Denoting the size of the male and female population by $n_t^m := \Sigma_\tau n^m_{t,\tau}$ and $n_t^f := \Sigma_\tau n^f_{t,\tau}$, respectively, the growth rates are

$$\rho_{m,t} \;:=\; \frac{n^m_{t+1} - n^m_t}{n^m_t} \quad \text{and} \quad \rho_{f,t} \;:=\; \frac{n^f_{t+1} - n^f_t}{n^f_t}$$

The growth rate of the whole population, $n_t := n_t^m + n_t^f$, is then a weighted mean, namely

$$\rho_t \;=\; \frac{\rho_{m,t}\, n_t^m + \rho_{f,t}\, n_t^f}{n_t^m + n_t^f}$$

**3.** We now assume that birth and death rates are constant over time. This implies that also the matrices $\mathbf{B}_t$, $\mathbf{D}_{m,t}$, and $\mathbf{D}_{f,t}$ are independent of time and may simply be denoted by $\mathbf{B}$, $\mathbf{D}_m$, and $\mathbf{D}_f$. Using the definition $\mathbf{F} := \mathbf{D}_f + \sigma_f \mathbf{B}$, we get

$$\mathbf{n}^f_{t+1} \;=\; \mathbf{F}\, \mathbf{n}^f_t \qquad (17.2.1)$$

Using this equation and starting with an initial female population $\mathbf{n}_0^f$, we may write:

$$\mathbf{n}_1^f \;=\; \mathbf{F}\mathbf{n}_0^f,\;\; \mathbf{n}_2^f \;=\; \mathbf{F}\mathbf{n}_1^f \;=\; \mathbf{F}^2\mathbf{n}_0^f,\;\; \mathbf{n}_3^f \;=\; \mathbf{F}\mathbf{n}_2^f \;=\; \mathbf{F}^3\mathbf{n}_0^f$$

and so on. This leads to the general equation

$$\mathbf{n}_t^f \;=\; \mathbf{F}^t\mathbf{n}_0^f \qquad (17.2.2)$$

which allows to calculate $\mathbf{n}_t^f$ from a knowledge of the initial population $\mathbf{n}_0^f$ and the matrix $\mathbf{F}$.

**4.** To investigate the development of $\mathbf{n}_t^f$ if $\mathbf{n}_0^f$ and $\mathbf{F}$ are given, we begin

**Tab. 17.2-1** Development of $\mathbf{n}_t^f$ in our example.

| $t$ | $n_{t1}^f$ | $n_{t2}^f$ | $n_{t3}^f$ | $n_{t4}^f$ | $n_{t1}^{f,p}$ | $n_{t2}^{f,p}$ | $n_{t3}^{f,p}$ | $n_{t4}^{f,p}$ | $n_t^f$ | $\rho_{f,t}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 | 1.00 | 0.25 | 0.25 | 0.25 | 0.25 | 4.00 | -0.0750 |
| 1 | 1.60 | 0.80 | 0.70 | 0.60 | 0.43 | 0.22 | 0.19 | 0.16 | 3.70 | -0.0595 |
| 2 | 1.22 | 1.28 | 0.56 | 0.42 | 0.35 | 0.37 | 0.16 | 0.12 | 3.48 | 0.0989 |
| 3 | 1.62 | 0.98 | 0.90 | 0.34 | 0.42 | 0.25 | 0.23 | 0.09 | 3.82 | 0.0531 |
| 4 | 1.51 | 1.29 | 0.68 | 0.54 | 0.38 | 0.32 | 0.17 | 0.13 | 4.03 | 0.0500 |
| 5 | 1.70 | 1.21 | 0.91 | 0.41 | 0.40 | 0.29 | 0.21 | 0.10 | 4.23 | 0.0658 |
| 6 | 1.75 | 1.36 | 0.85 | 0.54 | 0.39 | 0.30 | 0.19 | 0.12 | 4.51 | 0.0509 |
| 7 | 1.87 | 1.40 | 0.95 | 0.51 | 0.39 | 0.30 | 0.20 | 0.11 | 4.74 | 0.0613 |
| 8 | 1.98 | 1.50 | 0.98 | 0.57 | 0.39 | 0.30 | 0.20 | 0.11 | 5.03 | 0.0551 |
| 9 | 2.09 | 1.58 | 1.05 | 0.59 | 0.39 | 0.30 | 0.20 | 0.11 | 5.30 | 0.0583 |
| 10 | 2.21 | 1.67 | 1.11 | 0.63 | 0.39 | 0.30 | 0.20 | 0.11 | 5.61 | 0.0568 |
| 11 | 2.33 | 1.77 | 1.17 | 0.66 | 0.39 | 0.30 | 0.20 | 0.11 | 5.93 | 0.0574 |
| 12 | 2.47 | 1.87 | 1.24 | 0.70 | 0.39 | 0.30 | 0.20 | 0.11 | 6.27 | 0.0573 |
| 13 | 2.61 | 1.97 | 1.31 | 0.74 | 0.39 | 0.30 | 0.20 | 0.11 | 6.63 | 0.0572 |
| 14 | 2.76 | 2.09 | 1.38 | 0.78 | 0.39 | 0.30 | 0.20 | 0.11 | 7.01 | 0.0573 |
| 15 | 2.92 | 2.21 | 1.46 | 0.83 | 0.39 | 0.30 | 0.20 | 0.11 | 7.41 | 0.0572 |
| 16 | 3.08 | 2.33 | 1.54 | 0.88 | 0.39 | 0.30 | 0.20 | 0.11 | 7.84 | 0.0573 |
| 17 | 3.26 | 2.47 | 1.63 | 0.93 | 0.39 | 0.30 | 0.20 | 0.11 | 8.28 | 0.0573 |
| 18 | 3.45 | 2.61 | 1.73 | 0.98 | 0.39 | 0.30 | 0.20 | 0.11 | 8.76 | 0.0573 |
| 19 | 3.64 | 2.76 | 1.83 | 1.04 | 0.39 | 0.30 | 0.20 | 0.11 | 9.26 | 0.0573 |
| 20 | 3.85 | 2.91 | 1.93 | 1.10 | 0.39 | 0.30 | 0.20 | 0.11 | 9.79 | |

with a small example. We assume that there are only four age groups ($\tau_m = 4$), birth rates are given by

$$\beta_1^* = 0,\ \beta_2^* = 2,\ \beta_3^* = 1.2,\ \beta_4^* = 0$$

and female death rates are given by

$$\delta_1^f = 0.2,\ \delta_2^f = 0.3,\ \delta_3^f = 0.4,\ \delta_4^f = 1$$

Furthermore, it will be assumed that the proportion of female births is $\sigma_f = 0.5$. From these assumptions one can calculate the matrix

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0.6 & 0 \\ 0.8 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \end{bmatrix}$$

Now, assuming arbitrarily some initial female population $\mathbf{n}_0^f = (1,1,1,1)'$, one can use equation (17.2.2) to calculate $\mathbf{n}_t^f$ for all subsequent temporal locations $t > 0$. Table 17.2-1 shows the result of the calculation for $t = 1,\ldots,20$. The total size of the female population is seen in the column labeled $n_t^f$ and its growth rate in the last column. Obviously, the growth rates converge to a fixed value, $\rho_f^* \approx 5.73\,\%$, in this example. This is the first remarkable result.



**5.** A second result concerns the age distribution. This is seen if we explicitly distinguish between the population size and the age distribution. Since the size of the population is given by $n_t^f$, the age distribution can be represented by the vector

$$\mathbf{n}_t^{f,p} := \frac{1}{n_t^f}\,\mathbf{n}_t^f$$

whose components show the relative frequencies of persons in the age groups. As seen in Table 17.2-1, also these frequencies converge to some fixed values, in our example:

$$\mathbf{n}_t^{f,p} \longrightarrow \mathbf{n}^{f,p} \approx (0.39, 0.30, 0.20, 0.11)'$$

**6.** To summarize the findings from this example, the demographic process eventually reaches some kind of equilibrium which is fully described by a time-independent growth rate, $\rho_f^*$, and a time-independent age distribution, $\mathbf{n}^{f,p}$. If this equilibrium is approximately reached in some temporal location $t$, then

$$\mathbf{n}_{t+k}^f \approx (1+\rho_f^*)^k\, n_t^f\, \mathbf{n}^{f,p} \qquad (\text{for } k = 1,2,3,\ldots)$$

$\rho_f^*$ is then called the *intrinsic growth rate* of the demographic process, and $\mathbf{n}^{f,p}$ is called its *stable (female) age distribution.*

## 17.3 Mathematical Supplements

We now discuss under which conditions intrinsic growth rates and stable age distributions do exist, and whether they depend on the initial population vector $\mathbf{n}_0^f$ or only on the matrix $\mathbf{F}$.

### Existence of a Stable Population

**1.** We begin with the first question, whether one can construct an intrinsic growth rate and a stable age distribution for some given matrix $\mathbf{F}$. This depends on the coefficients of $\mathbf{F}$. As introduced in the previous section, $\mathbf{F}$ has the following structure:[4]

$$\mathbf{F} = \begin{bmatrix} \sigma_f\beta_1^* & \sigma_f\beta_2^* & \cdots & \sigma_f\beta_{\tau_m-1}^* & \sigma_f\beta_{\tau_m}^* \\ 1-\delta_1^f & 0 & \cdots & 0 & 0 \\ 0 & 1-\delta_2^f & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-\delta_{\tau_m-1}^f & 0 \end{bmatrix}$$

[4]Matrices having this structure are often called *Leslie matrices* to remind of P. H. Leslie who has first provided an extensive discussion with demographic applications, see Leslie (1945).

One can be sure that $\mathbf{F} \geq 0$, meaning that all coefficients of $\mathbf{F}$ are non-negative. One can also safely assume that $0 < \delta^f_\tau < 1$, for $\tau = 1, \ldots, \tau_m - 1$, and consequently all entries in the subdiagonal of $\mathbf{F}$ are greater than zero. But a question concerns the birth rates $\beta^*_\tau$. Since the reproductive period of women is limited and, in general, $\tau_b < \tau_m$, we can assume that $\beta^*_{\tau_b} > 0$ but need to observe that $\beta^*_\tau = 0$ for $\tau > \tau_b$, implying that $\mathbf{F}$ has less than full rank.

**2.** We can proceed, however, in two steps. In a first step we consider only the first $\tau_b$ rows and and columns of $\mathbf{F}$, that is, the matrix

$$\tilde{\mathbf{F}} := \begin{bmatrix} \sigma_f \beta^*_1 & \sigma_f \beta^*_2 & \cdots & \sigma_f \beta^*_{\tau_b - 1} & \sigma_f \beta^*_{\tau_b} \\ 1 - \delta^f_1 & 0 & \cdots & 0 & 0 \\ 0 & 1 - \delta^f_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 - \delta^f_{\tau_b - 1} & 0 \end{bmatrix}$$

This is now a non-negative matrix which has full rank.[5] Furthermore, $\tilde{\mathbf{F}}$ is an irreducible matrix.[6] This allows to apply a famous mathematical theorem by G. Frobenius.[7] The theorem guarantees that $\tilde{\mathbf{F}}$ has at least one real positive eigenvalue, say $\lambda^*$, also called a dominant eigenvector of $\tilde{\mathbf{F}}$, with a corresponding eigenvector, say $\mathbf{v}^* = (v^*_1, \ldots, v^*_{\tau_b})'$, whose coefficients are all real and positive. So we can write the equation

$$\tilde{\mathbf{F}}\, \mathbf{v}^* = \lambda^*\, \mathbf{v}^* \tag{17.3.1}$$

A further implication of the theorem that will be used below in the discussion of our second question is that all eigenvalues of $\tilde{\mathbf{F}}$ have an absolute value (modulus) which is less than, or equal to, $\lambda^*$.

**3.** We can now derive a stable age distribution and an intrinsic growth rate. The intrinsic growth rate can be simply defined by $\rho^*_f := \lambda^* - 1$. The derivation of the stable age distribution is in two steps. In a first step we define components of a vector $\mathbf{n}^{f,*}$ by

$$n^{f,*}_\tau := \begin{cases} v^*_\tau & \text{for } \tau = 1, \ldots, \tau_b \\ \frac{1 - \delta^f_{\tau - 1}}{\lambda^*}\, v^*_{\tau - 1} & \text{for } \tau = \tau_b + 1, \ldots, \tau_m \end{cases}$$

[5]This is seen by the determinant of $\tilde{\mathbf{F}}$ which is

$$\det(\tilde{\mathbf{F}}) = \pm \sigma_f\, \beta^*_{\tau_b} \prod_{\tau=1}^{\tau_b - 1} (1 - \delta^f_\tau) \neq 0$$

The sign depends on whether $\tau_b$ is even or odd.

[6]By this is meant that, for any two indices $i$ and $j$ $(1 \leq i < j \leq \tau_b)$, one can find further indices, say $k_1, \ldots, k_m$, such that $a_{ik_1}\, a_{k_1 k_2} \cdots a_{k_m j} > 0$.

[7]We refer to Gantmacher (1971, ch. xxiii).



From equation (17.3.1) and the structure of $\mathbf{F}$ it then follows that

$$\mathbf{F}\, \mathbf{n}^{f,*} = \lambda^*\, \mathbf{n}^{f,*} = (1 + \rho^*_f)\, \mathbf{n}^{f,*} \tag{17.3.2}$$

showing that the age distribution which is represented by $\mathbf{n}^{f,*}$ will not change when multiplied by $\mathbf{F}$; all components of $\mathbf{n}^{f,*}$ will grow, or shrink, with the same rate, $\rho^*_f$. Therefore, to get the stable age distribution one only has to transform $\mathbf{n}^{f,*}$ into proper proportions:

$$n^{f,p}_\tau := n^{f,*}_\tau \Big/ \sum_{j=1}^{\tau_m} n^{f,*}_j$$

**4.** To illustrate the argument we use the example of the previous section. In this example the matrix $\tilde{\mathbf{F}}$ is given by

$$\tilde{\mathbf{F}} = \begin{bmatrix} 0 & 1 & 0.6 \\ 0.8 & 0 & 0 \\ 0 & 0.7 & 0 \end{bmatrix}$$

Calculating eigenvalues and eigenvectors can be done with the following TDA script:[8]

```
mdef(F,3,3) = 0.0,1.0,0.6,
              0.8,0.0,0.0,
              0.0,0.7,0.0;
mev(F,ER,EI,EVR,EVI);
mpr(ER);
mpr(EI);
mpr(EVR);
mpr(EVI);
```

One finds that the dominant eigenvalue is $\lambda^* = 1.0573$ and the corresponding eigenvector is

$$\mathbf{v}^* = (0.7405, 0.5603, 0.3710)'$$

The eigenvalue provides the intrinsic growth rate, $\rho^*_f = 0.0573$, which is identical with the value found in the previous section. The eigenvector can be used to calculate the components of $\mathbf{n}^{f,p}$:

$$n^{f,*}_1 = 0.7405,\; n^{f,*}_2 = 0.5603,\; n^{f,*}_3 = 0.3710, \text{ and}$$

$$n^{f,*}_4 = \frac{0.6}{1.0573}\, 0.3710 = 0.2105$$

[8]More detailed explanations of the practical calculations will be given in Section 17.5.1.

Of course, equation (17.3.2) does not change if $\mathbf{n}^{f,*}$ is multiplied by an arbitrary scalar value. So we can rescale $\mathbf{n}^{f,*}$ to get a frequency distribution with components adding to unity. The result is

$$\mathbf{n}^{f,p} = (0.39, 0.30, 0.20, 0.11)'$$

and equals the age distribution found in the previous section.

**5.** It would suffice to calculate the dominant eigenvalue of $\tilde{\mathbf{F}}$ because the corresponding eigenvector, and consequently the stable age distribution, can be derived from the death rates. Let the dominant eigenvalue, $\lambda^*$, be given. Since the corresponding eigenvector, $\mathbf{v}^*$, is determined only up to an arbitrary multiplicative factor, we can set $v_1^* = 1$. All further elements of $\mathbf{v}^*$ can be calculated recursively with the formula

$$v_\tau^* = \frac{1-\delta_{\tau-1}^f}{\lambda^*}\, v_{\tau-1}^* \qquad (\text{for } \tau = 2, \ldots, \tau_m)$$

The argument also shows that, if $\lambda^* = 1$, the stable age distribution depends only on the death rates, not on the birth rates. But, of course, $\lambda^*$ also depends on birth rates.

### Convergence to a Stable Age Distribution

**6.** We now turn to the second question, whether, beginning with an arbitrary initial female population $\mathbf{n}_0^f$, the sequence $\mathbf{n}_t^f = \mathbf{F}^t \mathbf{n}_0^f$ finally converges to an equilibrium defined by the intrinsic growth rate, $\rho_f^*$, and the stable age distribution, $\mathbf{n}^{f,p}$.[9] As will be shown, the answer is positive under quite general conditions. To develop the argument, we first consider the sub-matrix $\tilde{\mathbf{F}}$ which consists of the first $\tau_b$ rows and columns of $\mathbf{F}$. Correspondingly, we refer to the first $\tau_b$ elements of $\mathbf{n}_t^f$ by the vector $\mathbf{n}_t^{f,a}$. Since $\tilde{\mathbf{F}}$ is an upper block-diagonal matrix, it follows that

$$\mathbf{n}_t^{f,a} = \tilde{\mathbf{F}}^t \mathbf{n}_0^{f,a} \tag{17.3.3}$$

We now show that, given an additional assumption to be explained below, $\mathbf{n}_t^{f,a}$ converges to a vector which is proportional to $\mathbf{v}^*$, that is, the eigenvector corresponding to the dominant eigenvalue of $\tilde{\mathbf{F}}$.

**7.** This requires to refer to all eigenvalues of $\tilde{\mathbf{F}}$ which will be denoted by $\lambda_j$, with corresponding eigenvectors $\mathbf{v}_j$, for $j = 1, \ldots, \tau_b$. One of these eigenvalues, say $\lambda_{j^*} = \lambda^*$, is the dominant one and has the corresponding eigenvector $\mathbf{v}_{j^*} = \mathbf{v}^*$. So we can write the equations

$$\tilde{\mathbf{F}}\mathbf{v}_j = \lambda_j \mathbf{v}_j \qquad (\text{for } j = 1, \ldots, \tau_b)$$

[9]It will be assumed that there is at least one woman of an age under, or equal to, $\tau_b$.



which, by defining $\mathbf{\Lambda} := \operatorname{diag}(\lambda_1, \ldots, \lambda_{\tau_b})$ and $\mathbf{V} := (\mathbf{v}_1, \ldots, \mathbf{v}_{\tau_b})$, may also be written as a single matrix equation

$$\tilde{\mathbf{F}}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}$$

As mentioned above, $\tilde{\mathbf{F}}$ has full rank and its eigenvectors are therefore linear independent. This implies that $\mathbf{V}$ is an invertible matrix and we may write $\tilde{\mathbf{F}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, from which it follows that

$$\tilde{\mathbf{F}}^t = \mathbf{V}\mathbf{\Lambda}^t\,\mathbf{V}^{-1}$$

This then allows to write

$$\mathbf{n}_t^{f,a} = \tilde{\mathbf{F}}^t\,\mathbf{n}_0^{f,a} = \mathbf{V}\mathbf{\Lambda}^t\mathbf{V}^{-1}\mathbf{n}_0^{f,a} = \mathbf{V}\mathbf{\Lambda}^t\mathbf{u}$$

where, for the last equation, we have used the abbreviation $\mathbf{u} := \mathbf{V}^{-1}\mathbf{n}_0^{f,a}$. In a next step this equation can be written in the following way:

$$\mathbf{n}_t^{f,a} = (\mathbf{v}_1, \ldots, \mathbf{v}_{\tau_b}) \begin{bmatrix} \lambda_1^t u_1 \\ \vdots \\ \lambda_{\tau_b}^t u_{\tau_b} \end{bmatrix} = \sum_{j=1}^{\tau_b} (\lambda_j^t u_j)\,\mathbf{v}_j$$

which shows that $\mathbf{n}_t^{f,a}$ is a weighted mean of the eigenvectors of $\tilde{\mathbf{F}}$. Finally, dividing by $\lambda_{j^*}^t$, we get

$$\frac{1}{\lambda_{j^*}^t}\,\mathbf{n}_t^{f,a} = \sum_{j=1}^{\tau_b} \left(\frac{\lambda_j^t}{\lambda_{j^*}^t}\,u_j\right)\mathbf{v}_j = u_{j^*}\mathbf{v}_{j^*} + \sum_{j\neq j^*} \left(\frac{\lambda_j}{\lambda_{j^*}}\right)^t u_j\,\mathbf{v}_j \tag{17.3.4}$$

**8.** This equation can be used to think about the convergence problem. From the theorem of Frobenius we already know that $\lambda_{j^*} \geq |\lambda_j|$ for all $j = 1, \ldots, \tau_b$. We now introduce a further assumption, to be discussed below, that $\lambda_{j^*} > |\lambda_j|$ for all $j \neq j^*$. Given this assumption, it follows that the second term on the right-hand side of equation (17.3.4) will converge to zero and this, in turn, implies the convergence

$$\frac{1}{\lambda_{j^*}^t}\,\mathbf{n}_t^{f,a} \longrightarrow u_{j^*}\,\mathbf{v}_{j^*}$$

This shows that, for sufficiently large $t$,

$$\mathbf{n}_{t+1}^{f,a} \approx \lambda_{j^*}\mathbf{n}_t^{f,a}$$

and $\mathbf{n}_t^{f,a}$ will be approximately proportional to the eigenvector $\mathbf{v}^*$. Moreover, also the remaining components of $\mathbf{n}_t^f$ will converge to a stable age distribution. This is seen from the fact that these remaining components only depend on the growth of the female population at age $\tau_b$ and the death rates at ages greater than, or equal to, $\tau_b$. Therefore, if eventually the number of women at age $\tau_b$ grows, or shrinks, with a constant (intrinsic) rate, this will propagate to all higher ages. The stable age distribution for all ages may then be calculated as shown in the first part of this section.

**9.** It remains to discuss the assumption that the dominant eigenvalue of $\tilde{\mathbf{F}}$ is greater, in magnitude, than all other eigenvalues. This is not necessarily the case. For example, the matrix

$$\tilde{\mathbf{F}} := \begin{bmatrix} 0 & 1 \\ 0.8 & 0 \end{bmatrix}$$

has two real eigenvalues, 0.8944 and -0.8944, having the same magnitude. In this example, as shown by equation (17.3.4), $\mathbf{n}_t^{f,a}$ will not converge to a unique stable age distribution but oscillate between two different distributions. Such cases are, however, exceptional. A sufficient condition for the existence of a dominant eigenvalue which is greater, in magnitude, than all other eigenvalues is that there are at least two successive ages with a positive birth rate.[10] Therefore, cyclical solutions will only occur if one uses a highly aggregated Leslie matrix; for instance, a matrix that only distinguishes three age groups, below $\tau_a$, between $\tau_a$ and $\tau_b$, and above $\tau_b$. If one distinguishes at least two age groups in the reproductive period one can safely assume the existence of a stable age distribution.

## 17.4 Female and Male Populations

**1.** So far we have only considered the development of a female population. Assuming time-constant birth and death rates, it was shown that the development of a female population eventually reaches an equilibrium which is characterized by a constant growth rate, $\rho^*$, and a stable age distribution, $\mathbf{n}^{f,p}$. So the question remains how a corresponding male population will develop. To find an answer one can begin with equations (17.1.2) which have been derived at the end of Section 17.1. Assuming time-constant birth and death rates, they can be written as follows:

$$\mathbf{n}_{t+1}^m = \mathbf{D}_m\,\mathbf{n}_t^m + \sigma_m\,\mathbf{B}\,\mathbf{n}_t^f \quad \text{and} \quad \mathbf{n}_{t+1}^f = \mathbf{D}_f\,\mathbf{n}_t^f + \sigma_f\,\mathbf{B}\,\mathbf{n}_t^f$$

If $\sigma_m = \sigma_f$ and the age-specific death rates were identical for men and women, both the male and female population would eventually reach the same stable age distribution. However, as we have seen in Part II, both assumptions are not valid. Instead, most often $\sigma_m > \sigma_f$, and in most of the age groups death rates are higher for men than for women.

[10] This is mentioned by Anton and Rorres (1991, p. 654) where one can also find a good introduction to much of the mathematics behind the model. For a statement, and proof, of sufficient and necessary conditions see Demetrius (1971).



**2.** Since only women can bear children it is easy, however, to derive the development of the male population from the development of the female population. The argument goes in two steps. The first step concerns newborn male children. As shown in Section 17.1, their number is given by

$$n_{t+1,1}^m = \sigma_m \sum_{\tau=\tau_a}^{\tau_b} \beta_\tau^*\, n_{t,\tau}^f$$

and therefore only depends on the number and age distribution of women in the reproductive period. Consequently, if the female population eventually has a stable age distribution and grows, or shrinks, with a constant rate $\rho^*$, also the number of newborn male children will grow, or shrink, with the same rate, that is, we can write

$$n_{t+1,1}^m = (1+\rho^*)\, n_{t,1}^m$$

But this will then propagate to all further ages, and the age distribution of the male population will only depend on male death rates. For example, $n_{t+2,2}^m = n_{t+1,1}^m(1-\delta_1^m)$, and

$$n_{t+3,3}^m = n_{t+2,2}^m(1-\delta_2^m) = n_{t+1,1}^m(1-\delta_1^m)(1-\delta_2^m)$$

So we may write for all ages $\tau > 1$ the equation

$$n_{t+\tau,\tau}^m = n_{t+1,1}^m \prod_{j=1}^{\tau-1}(1-\delta_j^m)$$

Therefore, if $n_{t+1,1}^m$ grows, or shrinks, with a constant rate $\rho^*$, the same will be true for the number of men at all ages. Consequently, also the male population will eventually reach a stable age distribution which can be derived from male death rates in the same way as was shown in the previous section for females. To repeat the method of calculation, one begins with an arbitrary value for the number of males in age 1, say $v_1^* = 1$. Then one can calculate recursively

$$v_\tau^* = \frac{1-\delta_{\tau-1}^m}{1+\rho^*}\, v_{\tau-1}^* \tag{17.4.1}$$

for $\tau = 2,\ldots,\tau_m$. The frequencies in the stable age distribution are then simply $n_\tau^{m,p} = v_\tau^*/\Sigma_j v_j^*$.

## 17.5 Practical Calculations

In the present section we discuss how one can practically calculate intrinsic growth rates and stable age distributions with real data.

### 17.5.1 Two Calculation Methods

**1.** For the calculations we use matrix commands available in the computer program TDA.[11] There are two possible approaches. The first one relies on a direct calculation of the eigenvalues (and eigenvectors) of the matrix $\tilde{\mathbf{F}}$ introduced in Section 17.3. For an illustration we use the example from Section 17.2. The TDA script is shown in Box 17.5-1. The `mdef` command is used to define the matrix

$$\tilde{\mathbf{F}} = \begin{bmatrix} 0 & 1 & 0.6 \\ 0.8 & 0 & 0 \\ 0 & 0.7 & 0 \end{bmatrix}$$

called `F` in the script. Then the `mev` command is used to calculate eigenvalues and eigenvectors of this matrix. The command gets `F` as input and creates two vectors (`ER` and `EI`) and two matrices (`EVR` and `EVI`) as output. `ER` and `EI` contain, respectively, the real and imaginary parts of the eigenvalues, and `EVR` and `EVI` contain, respectively, the real and imaginary parts of the eigenvectors. Their contents are shown in the lower part of Box 17.5-1. Most important is the dominant eigenvalue which is 1.0573 in this example. As shown in Section 17.3, one can immediately derive the intrinsic growth rate and the stable age distribution.

**2.** An alternative calculation method relies on the fact that, beginning with an arbitrary female population vector $\mathbf{n}_0^f$, one can iteratively calculate new population vectors

$$\mathbf{n}_t^f = \mathbf{F}\mathbf{n}_0^f$$

which finally converge to a stable population vector. Compared with the first method, there are two advantages. One does not need to use the reduced matrix $\tilde{\mathbf{F}}$ but can directly work with the complete Leslie matrix $\mathbf{F}$. And one gets, in addition, information about the number of iterations required to approximately reach the stable distribution.

**3.** To ease the application of this method TDA provides the `mpit` command. As input, the command requires information about the matrix $\mathbf{F}$, the initial population vector $\mathbf{n}_0^f$, and the number of iterations to be performed, say $t_n$. The command has the following syntax:

```
mpit(A,N,T,R)
```

[11] This program is freely available via `www.stat.ruhr-uni-bochum.de/tda.html`.



**Box 17.5-1** TDA script to create the matrix $\tilde{\mathbf{F}}$ and calculate its eigenvalues and eigenvectors.

```
silent = -1;                  # echo commands
mfmt = 7.4;                   # set the print format
mdef(F,3,3) = 0.0,1.0,0.6,    # define the matrix F
              0.8,0.0,0.0,
              0.0,0.7,0.0;
mev(F,ER,EI,EVR,EVI);         # calculate eigenvalues and eigenvectors
                              # of F
mpr(ER);                      # print real part of eigenvalues
mpr(EI);                      # print imaginary part of eigenvalues
mpr(EVR);                     # print real part of eigenvectors
mpr(EVI);                     # print imaginary part of eigenvectors


     ER       EI                 EVR                       EVI
-------  -------  -----------------------  -----------------------
-0.5286   0.1958   0.4305  0.4305  0.7405  -0.3697  0.3697  0.0000
-0.5286  -0.1958  -0.7552 -0.7552  0.5603   0.2798 -0.2798  0.0000
 1.0573   0.0000   1.0000  1.0000  0.3710  -0.0000  0.0000  0.0000
```

`T` is a scalar that provides the number of iterations, $t_n$. `N` is a column vector with $\tau_m$ components (equal to the number of rows of the Leslie matrix $\mathbf{F}$) and contains the initial population vector. `A` is a matrix with $\tau_m$ rows and two columns; the first column contains the age-specific birth rates and the second column contains the age-specific survivor rates. Using notations introduced in Section 17.1, the matrix `A` and the vector `N` are assumed to be defined as follows:

$$\mathtt{A} = \begin{bmatrix} \sigma_f \beta_1^* & 1 - \delta_1^f \\ \vdots & \vdots \\ \sigma_f \beta_{\tau_m}^* & 1 - \delta_{\tau_m}^f \end{bmatrix} \quad \text{and} \quad \mathtt{N} = \begin{bmatrix} n_{0,1}^f \\ \vdots \\ n_{0,\tau_m}^f \end{bmatrix}$$

As output, the command creates the matrix `R` with $t_n + 1$ rows and $\tau_m$ columns. The $t$-th row (for $t = 0, \ldots, t_n$) contains the elements of the vector $\mathbf{n}_t^f$.

**4.** The TDA script in Box 17.5-2 illustrates the `mpit` command with the same example used above. In order to replicate the values shown in Table 17.2-1 in Section 17.2, the initial population vector `N` has all components set to 1. The `mpit` command then performs 20 iterations and saves the result in the matrix `R`. By adding the rows of `R` one gets the vector `NT` containing the population sizes which can be used, then, to calculate age distributions (in `D`) and the growth rates (in `RT`). Of course, the value in the last component of `RT` is not a valid growth rate.

**Box 17.5-2** TDA script to illustrate the `mpit` command.

```
silent = -1;              # echo commands
mfmt = 5.2;               # print format
mdef(A,4,2) = 0, 0.8,     # define matrix A containing birth
              1, 0.7,     # rates in the first column and
            0.6, 0.6,     # survivor rates in the second column
              0, 0.0;

mdefc(4,1,1,N);           # define unit vector N (initial population)
mpit(A,N,20,R);           # perform 20 iterations, save result in R
mpr(R);                   # print the resulting matrix R
mmul(R,N,NT);             # sum rows of R, save result in vector NT
mpr(NT);                  # print NT
mexpr(R/NT,D);            # calculate age distributions in D
mpr(D);                   # print age distributions

mexpr((lag(NT,1) - NT) / NT,RT); # calculate growth rates in RT
mfmt = 7.4;                      # new print format
mpr(RT);                         # print growth rates


          R                NT            D                 RT
----------------------    ----   ----------------------   -------
1.00  1.00  1.00  1.00    4.00   0.25  0.25  0.25  0.25   -0.0750
1.60  0.80  0.70  0.60    3.70   0.43  0.22  0.19  0.16   -0.0595
1.22  1.28  0.56  0.42    3.48   0.35  0.37  0.16  0.12    0.0989
1.62  0.98  0.90  0.34    3.82   0.42  0.26  0.23  0.09    0.0531
1.51  1.29  0.68  0.54    4.03   0.38  0.32  0.17  0.13    0.0500
1.70  1.21  0.90  0.41    4.23   0.40  0.29  0.21  0.10    0.0658
1.75  1.36  0.85  0.54    4.51   0.39  0.30  0.19  0.12    0.0509
1.87  1.40  0.95  0.51    4.74   0.40  0.30  0.20  0.11    0.0613
1.98  1.50  0.98  0.57    5.03   0.39  0.30  0.20  0.11    0.0551
2.09  1.58  1.05  0.59    5.30   0.39  0.30  0.20  0.11    0.0583
2.21  1.67  1.11  0.63    5.61   0.39  0.30  0.20  0.11    0.0568
2.33  1.77  1.17  0.66    5.93   0.39  0.30  0.20  0.11    0.0574
2.47  1.87  1.24  0.70    6.27   0.39  0.30  0.20  0.11    0.0573
2.61  1.97  1.31  0.74    6.63   0.39  0.30  0.20  0.11    0.0572
2.76  2.09  1.38  0.78    7.01   0.39  0.30  0.20  0.11    0.0573
2.92  2.21  1.46  0.83    7.41   0.39  0.30  0.20  0.11    0.0572
3.08  2.33  1.54  0.88    7.84   0.39  0.30  0.20  0.11    0.0573
3.26  2.47  1.63  0.93    8.28   0.39  0.30  0.20  0.11    0.0573
3.45  2.61  1.73  0.98    8.76   0.39  0.30  0.20  0.11    0.0573
3.64  2.76  1.83  1.04    9.26   0.39  0.30  0.20  0.11    0.0573
3.85  2.91  1.93  1.10    9.79   0.39  0.30  0.20  0.11   -1.0000
```



### 17.5.2 Calculations for Germany 1999

**1.** The intrinsic growth rate and stable female and male age distributions pertaining to Germany in the year 1999 can be calculated from Tables 7.1-1 and 11.1-1. In order to prepare the required data, shown in Table 17.5-1, we assumed that the reproductive age of women begins at $\tau_a = 14$ and ends at $\tau_b = 51$.[12] Since in 1999 the number of male and female births was 396292 and 374448,[13] one gets $\sigma_m = 0.514$ and $\sigma_f = 0.486$. The survivor rate during the first year of life can then be calculated as

$$1 - (0.514 \cdot 0.004952 + 0.486 \cdot 0.004010) = 0.9955$$

and this can be used to calculate the birth rates $\beta^*_\tau$, which are used for our model, from the birth rates $\beta_\tau$ which are shown in Table 17.5-1:

$$\beta^*_\tau = 0.9955\,\beta_\tau$$

**2.** Beginning with the first of the two calculation methods discussed in the previous section, the next step is to create the matrix $\tilde{\mathbf{F}}$ which, in the current application, has 51 rows and columns, and calculate its dominant eigenvalue. We have done this with the TDA script shown in Box 17.5-3. Input is a data file, `spm1.dat`, that contains the data shown in Table 17.5-1.[14] The dominant eigenvalue is approximately $\lambda^* = 0.985$ corresponding to a negative intrinsic growth rate of $\rho^* = -1.5\,\%$. The interpretation is: If the birth and death rates of 1999 would remain constant in the future, and if migration would not take place, the population would eventually decline with a rate of -1.5 % per year.

**3.** In order to calculate the stable age distribution we use the method described at the end of Section 17.4 for the male population but, of course, can also be applied to find the stable female age distribution. We begin with $v^*_1 := 1$ and then recursively apply formula (17.4.1). For example,

$$v^*_2 = \frac{1 - 0.000421}{0.985}\, v^*_1 = 1.0148, \quad v^*_3 = \frac{1 - 0.000304}{0.985}\, v^*_2 = 1.0299$$

and so on, will result in a vector that is proportional to the stable age distribution of men. Using the female death rates instead will produce a vector

---

[12] The number of 80 births at age 14 or below has been related to the midyear number of women at age 14, which was 437300 in 1999, and the number of 16 births at age 51 or above has been related to the midyear number of women at age 51, which was 483000 in 1999.

[13] Fachserie 1, Reihe 1, 1999 (p. 42).

[14] In addition, the data file contains two more columns containing, respectively, age-specific numbers of men and women in 1999 in Germany, taken from Table 7.1-1 in Section 7.1. The female population vector will be used below to illustrate the second calculation method.

**Table 17.5-1** Birth and death rates in Germany in 1999, calculated from Tables 7.1-1 and 11.1-1.

| $\tau$ | $\delta_\tau^m$ | $\delta_\tau^f$ | $\beta_\tau$ | $\tau$ | $\delta_\tau^m$ | $\delta_\tau^f$ | $\beta_\tau$ |
|---|---|---|---|---|---|---|---|
| 0 | 0.004952 | 0.004010 | 0 | 46 | 0.003603 | 0.001953 | 0.000290 |
| 1 | 0.000421 | 0.000352 | 0 | 47 | 0.003927 | 0.002034 | 0.000104 |
| 2 | 0.000304 | 0.000212 | 0 | 48 | 0.004295 | 0.002198 | 0.000086 |
| 3 | 0.000223 | 0.000165 | 0 | 49 | 0.004574 | 0.002407 | 0.000046 |
| 4 | 0.000198 | 0.000143 | 0 | 50 | 0.005180 | 0.002618 | 0.000023 |
| 5 | 0.000127 | 0.000106 | 0 | 51 | 0.005445 | 0.002928 | 0.000033 |
| 6 | 0.000166 | 0.000107 | 0 | 52 | 0.006376 | 0.003213 | 0 |
| 7 | 0.000152 | 0.000126 | 0 | 53 | 0.006121 | 0.003275 | 0 |
| 8 | 0.000163 | 0.000100 | 0 | 54 | 0.007317 | 0.003737 | 0 |
| 9 | 0.000115 | 0.000117 | 0 | 55 | 0.007989 | 0.004061 | 0 |
| 10 | 0.000137 | 0.000090 | 0 | 56 | 0.008473 | 0.004049 | 0 |
| 11 | 0.000144 | 0.000090 | 0 | 57 | 0.009509 | 0.004582 | 0 |
| 12 | 0.000157 | 0.000118 | 0 | 58 | 0.009642 | 0.004596 | 0 |
| 13 | 0.000168 | 0.000112 | 0 | 59 | 0.011211 | 0.005307 | 0 |
| 14 | 0.000247 | 0.000149 | 0.000183 | 60 | 0.012309 | 0.005793 | 0 |
| 15 | 0.000313 | 0.000192 | 0.000778 | 61 | 0.013351 | 0.006136 | 0 |
| 16 | 0.000410 | 0.000242 | 0.002762 | 62 | 0.014959 | 0.006751 | 0 |
| 17 | 0.000654 | 0.000327 | 0.006807 | 63 | 0.016750 | 0.007453 | 0 |
| 18 | 0.001012 | 0.000348 | 0.013850 | 64 | 0.018706 | 0.008709 | 0 |
| 19 | 0.000959 | 0.000372 | 0.024724 | 65 | 0.020027 | 0.009375 | 0 |
| 20 | 0.000941 | 0.000310 | 0.035231 | 66 | 0.022121 | 0.010193 | 0 |
| 21 | 0.001015 | 0.000353 | 0.044585 | 67 | 0.025004 | 0.011762 | 0 |
| 22 | 0.000895 | 0.000283 | 0.054319 | 68 | 0.028132 | 0.013154 | 0 |
| 23 | 0.000879 | 0.000270 | 0.062574 | 69 | 0.030690 | 0.014562 | 0 |
| 24 | 0.000961 | 0.000308 | 0.069394 | 70 | 0.033592 | 0.016121 | 0 |
| 25 | 0.000803 | 0.000345 | 0.078996 | 71 | 0.035825 | 0.017972 | 0 |
| 26 | 0.000885 | 0.000300 | 0.083525 | 72 | 0.038527 | 0.020385 | 0 |
| 27 | 0.000849 | 0.000320 | 0.085854 | 73 | 0.042586 | 0.022606 | 0 |
| 28 | 0.000883 | 0.000352 | 0.092532 | 74 | 0.047512 | 0.024907 | 0 |
| 29 | 0.000820 | 0.000354 | 0.093590 | 75 | 0.051429 | 0.028595 | 0 |
| 30 | 0.000895 | 0.000366 | 0.093382 | 76 | 0.056174 | 0.032308 | 0 |
| 31 | 0.000880 | 0.000394 | 0.089946 | 77 | 0.063623 | 0.037003 | 0 |
| 32 | 0.000909 | 0.000448 | 0.082654 | 78 | 0.070017 | 0.041655 | 0 |
| 33 | 0.000989 | 0.000495 | 0.072578 | 79 | 0.086292 | 0.051891 | 0 |
| 34 | 0.001067 | 0.000517 | 0.061521 | 80 | 0.077474 | 0.048273 | 0 |
| 35 | 0.001145 | 0.000606 | 0.050843 | 81 | 0.093884 | 0.061502 | 0 |
| 36 | 0.001428 | 0.000618 | 0.040948 | 82 | 0.103886 | 0.071132 | 0 |
| 37 | 0.001479 | 0.000784 | 0.030625 | 83 | 0.111364 | 0.075476 | 0 |
| 38 | 0.001583 | 0.000828 | 0.022808 | 84 | 0.134642 | 0.094869 | 0 |
| 39 | 0.001882 | 0.000959 | 0.017015 | 85 | 0.140858 | 0.100747 | 0 |
| 40 | 0.002010 | 0.001059 | 0.011966 | 86 | 0.155596 | 0.113830 | 0 |
| 41 | 0.002212 | 0.001178 | 0.007599 | 87 | 0.171477 | 0.129769 | 0 |
| 42 | 0.002506 | 0.001315 | 0.004957 | 88 | 0.184898 | 0.146592 | 0 |
| 43 | 0.002818 | 0.001467 | 0.002769 | 89 | 0.208687 | 0.164650 | 0 |
| 44 | 0.003008 | 0.001549 | 0.001371 | 90 | 1.000000 | 1.000000 | 0 |
| 45 | 0.003426 | 0.001729 | 0.000603 | | | | |



**Box 17.5-3** TDA script to calculate the intrinsic growth rate corresponding to data for Germany 1999.

```
silent = -1;                 # echo commands
mfmt = 7.4;                  # set print format
nvar(                        # read the data file spm1.dat
    dfile = spm1.dat,
    AGE  [2.0] = c1,
    DF<8>[8.6] = 1 - c3,
    B1<8>[8.6] = c4,
    BF<8>[8.6] = 0.486 * 0.9955 * B1,
);

tsel = AGE[1,,51];           # select ages
mdef(BRF) = BF;              # birth rates of female children

tsel = AGE[1,,50];           # select ages
mdef(DRF) = DF;              # survivor rates of women

mdiag(DRF,A);                # create diagonal matrix
mdefc(50,1,0,N);             # create a null vector
mcath(A,N,A);                # concatenate with A
mtransp(BRF,BRFT);           # make BRF a row vector
mcatv(BRFT,A,F);             # concatenate with A to get F

mev(F,ER,EI,EVR,EVI);        # calculate eigenvalues and eigenvectors
mpr(ER);                     # print real part of eigenvalues
mpr(EI);                     # print imaginary part of eigenvalues
```

that is proportional to the stable age distribution of women. Finally, one only needs to normalize these vectors in order to get distributions, i.e., proportions adding to unity. The resulting stable age distributions are shown in Figures 17.5-1 and 17.5-2 and compared with the actual age distributions of men and women in Germany 1999.[15] It is seen that a prolongation of the current birth and death rates would result in a substantial increase in the proportion of older people.

**4.** We now use TDA's `mpit` command to perform the calculations. The script is shown in Box 17.5-4. The input data are again taken from the data file `spm1.dat`. Survivor rates and adjusted birth rates are created as explained above. In addition, we use column 6 of the data file to get the female population in 1999, classified by age. The script then creates the matrix `A` and the vector `N` to be used as input for the `mpit` command. The vector `U` is used to get the row sums of `R`. The result, the vector `NT`, contains the female population size at the 200 iterations. This vector is finally used to calculate the growth rates. Investigating the output, one

[15] The data are taken from Table 7.1-1 in Section 7.1.

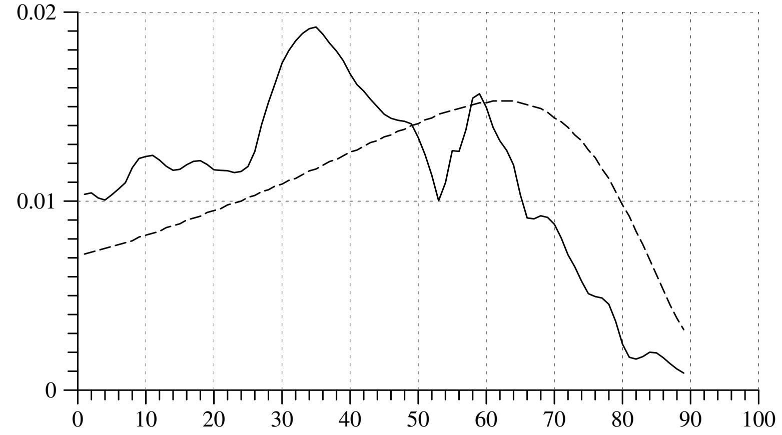



**Fig. 17.5-1** Frequency curves (restricted to ages less than 90) representing the age distribution of men in Germany 1999 (solid line) and the corresponding stable age distribution (dotted line).

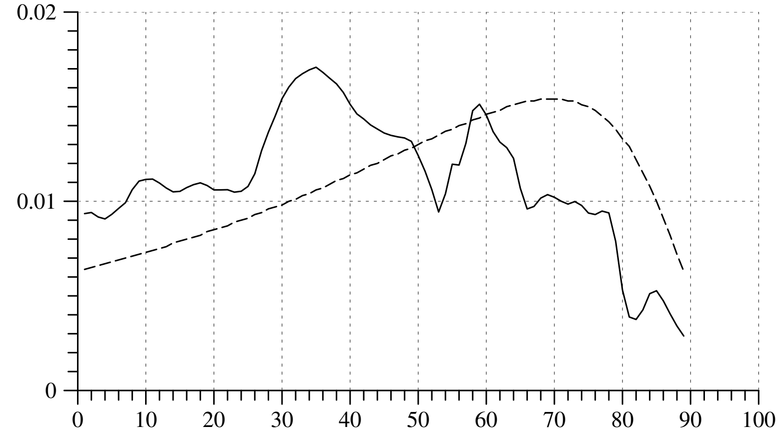



**Fig. 17.5-2** Frequency curves (restricted to ages less than 90) representing the age distribution of women in Germany 1999 (solid line) and the corresponding stable age distribution (dotted line).

finds that a stable growth rate of about -1.5 % is reached in about 100 iterations.

**5.** How long it takes to approximately reach an equilibrium depends on the extent to which the initial (current) and the final (stable) age distribution differ. As shown by Figure 17.5-2, the differences are quite substantial and it therefore requires many iterations to reach, at least approximately, the stable distribution. In our application one would need about 50–100



**Box 17.5-4** TDA script to calculate the intrinsic growth rate and stable female age distribution corresponding to data for Germany 1999.

```
silent = -1;              # echo commands
mfmt =  8.4;              # set print format
nvar(                     # read the data file spm1.dat
    dfile = spm1.dat,
    AGE  [2.0] = c1,
    DF<8>[8.6] = 1 - c3,
    B1<8>[8.6] = c4,
    BF<8>[8.6] = 0.486 * 0.9955 * B1,
    NF   [5.1] = c6,                     # female population in 1999
);
tsel = AGE[1,,90];        # select ages
mdef(A) = BF,DF;          # create the A matrix
mdef(N) = NF;             # create population vector
mpit(A,N,200,R);          # perform iterations
mpr(R);                   # show result

mdefc(90,1,1,U);          # create a unit vector
mmul(R,U,NT);             # calculate population size
mpr(NT);                  # print NT

mexpr((lag(NT,1) - NT) / NT,RT);    # calculate growth rates
mpr(RT);                            # print growth rates
```

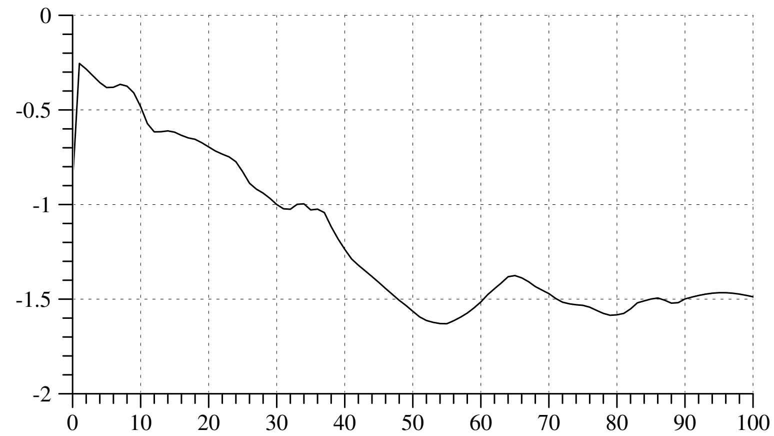



**Fig. 17.5-3** Year-to-year growth rates of the female population in Germany resulting from 100 iterations of the current female age distribution, based on birth and death rates in 1999.

iterations (years). This is illustrated in Figure 17.5-3 which shows the first 100 elements of the vector RT calculated by the script in Box 17.5-4. Correspondingly, one might calculate how the population size would decrease.

Of course, the results of these calculations should not be mistaken for a population projection. They simply serve to investigate the implications of the current birth and death rates under the fictitious assumption that they will not change and that neither in- nor out-migration will take place.

# Chapter 18

# Conditions of Population Growth

The previous chapter has introduced a general framework for analytical models and, as one application, has discussed the question how the population in Germany would develop if current birth and death rates would not change and migration would not take place. Of course, the question is hypothetical, and so is the answer. In the present chapter we continue with this kind of hypothetical question but try to get a somewhat closer understanding of how the intrinsic growth rate depends on birth and death rates. In Section 18.6 we also take into account migration.

## 18.1 Reproduction Rates

**1.** We begin with a discussion of reproduction rates. The total birth rate [zusammengefasste Geburtenziffer] in the year $t$, introduced in Section 11.1, is defined as[1]

$$\mathrm{TBR}_t := \sum_{\tau=\tau_a}^{\tau_b} \beta_{t,\tau} \quad \text{(multiplied by 1000)}$$

where the age-specific birth rates are denoted by $\beta_{t,\tau}$. It is simply the sum of the age-specific birth rates and shows how many children would be born of 1000 women if their childbearing would conform to the current birth rates and mortality would not take place until the end of the reproductive period. Table 18.1-1 shows values for both territories of Germany, Figure 18.1-1 provides a graphical illustration.[2] Obviously, since about 1970, the number of births is below a replacement level which would require a total birth rate of about 2000.

**2.** Reproduction rates are modifications of the total birth rate which refer to only female births and take into account the mortality of women until the end of the reproductive period. The first variant, called *gross reproduction rate* [Bruttoreproduktionsrate], is defined as

$$\mathrm{GRR}_t := \sigma_{t,f}\, \mathrm{TBR}_t$$

where $\sigma_{t,f}$ is the proportion of female births in year $t$. The idea behind this

[1] In the literature, the total birth rate is also termed 'total fertility rate' and accordingly abbreviated by TFR.

[2] Calculation of total birth rates for the territory of the former FRG is based on a reproductive period from 15 to 49 years. For the territory of the former GDR, the age range is 15 – 45 until 1988, 15 – 44 in 1989, and 15 – 40 since 1990.

**Table 18.1-1** Total birth rates in the territory of the former FRG ($\mathrm{TBR}^a$) and in the territory of the former GDR ($\mathrm{TBR}^b$). Source: Fachserie 1, Reihe 1, 1999 (pp. 50-51).

| $t$ | $\mathrm{TBR}^a$ | $\mathrm{TBR}^b$ | $t$ | $\mathrm{TBR}^a$ | $\mathrm{TBR}^b$ | $t$ | $\mathrm{TBR}^a$ | $\mathrm{TBR}^b$ |
|---|---|---|---|---|---|---|---|---|
| 1950 | 2100.2 | | 1967 | 2489.6 | 2337.9 | 1984 | 1290.6 | 1735.4 |
| 1951 | 2067.7 | | 1968 | 2382.1 | 2296.8 | 1985 | 1280.8 | 1734.2 |
| 1952 | 2078.8 | 2398.5 | 1969 | 2214.0 | 2235.7 | 1986 | 1345.3 | 1699.9 |
| 1953 | 2053.5 | 2369.8 | 1970 | 2016.3 | 2192.5 | 1987 | 1368.0 | 1739.9 |
| 1954 | 2101.8 | 2350.3 | 1971 | 1920.8 | 2131.0 | 1988 | 1412.5 | 1670.2 |
| 1955 | 2108.4 | 2346.7 | 1972 | 1712.9 | 1786.0 | 1989 | 1395.4 | 1572.3 |
| 1956 | 2204.3 | 2262.3 | 1973 | 1543.5 | 1576.8 | 1990 | 1450.1 | 1517.7 |
| 1957 | 2300.9 | 2208.2 | 1974 | 1512.5 | 1539.7 | 1991 | 1421.8 | 977.2 |
| 1958 | 2290.1 | 2205.4 | 1975 | 1451.3 | 1541.7 | 1992 | 1401.6 | 830.4 |
| 1959 | 2368.1 | 2346.9 | 1976 | 1454.8 | 1636.8 | 1993 | 1392.6 | 774.9 |
| 1960 | 2365.7 | 2328.3 | 1977 | 1404.6 | 1850.6 | 1994 | 1347.2 | 772.2 |
| 1961 | 2456.8 | 2397.0 | 1978 | 1380.7 | 1899.0 | 1995 | 1339.3 | 838.2 |
| 1962 | 2440.7 | 2415.1 | 1979 | 1379.1 | 1894.6 | 1996 | 1395.9 | 947.7 |
| 1963 | 2518.4 | 2469.5 | 1980 | 1444.9 | 1941.8 | 1997 | 1440.6 | 1039.0 |
| 1964 | 2542.5 | 2507.6 | 1981 | 1435.2 | 1853.9 | 1998 | 1413.1 | 1086.7 |
| 1965 | 2507.5 | 2483.4 | 1982 | 1407.2 | 1858.2 | 1999 | 1405.8 | 1148.4 |
| 1966 | 2534.6 | 2424.4 | 1983 | 1330.9 | 1789.8 | | | |

definition is that only female births can contribute to further population growth. However, since the proportion of female births is close to 0.5 without much variation, the development of the gross reproduction rate is most often quite similar to the development of the total birth rate.

**3.** A next step is to take into account mortality of women until the end of the reproductive period. The idea is that the age-specific birth rate $\beta_{t,\tau}$ only refers to women who are still alive at age $\tau$. To formally introduce the definition, we use $G^f_{t,\tau}$ to denote the proportion of women who reach at least age $\tau$. These proportions can be derived from period life tables or directly from female death rates in the year $t$. While the *Statistisches Bundesamt* uses data from life tables,[3] we prefer to use the female death rates, $\delta^f_{t,\tau}$.[4] The proportion of women still alive at age $\tau$ is then calculated as

$$G^f_{t,\tau} = \prod_{j=0}^{\tau-1} (1 - \delta^f_{t,j})$$

This leads to the definition of a *net reproduction rate* [Nettoreproduktionsrate]:

$$\mathrm{NRR}_t := \sigma_{t,f} \sum_{\tau=\tau_a}^{\tau_b} \beta_{t,\tau}\, G^f_{t,\tau} = \sigma_{t,f} \sum_{\tau=\tau_a}^{\tau_b} \beta_{t,\tau} \prod_{j=0}^{\tau-1} (1 - \delta^f_{t,j})$$

[3] See, e.g., Fachserie 1, Reihe 1, 1999 (p. 53).

[4] As will be shown in the next section, this allows to easily connect the calculations with the modeling framework introduced in Chapter 17.



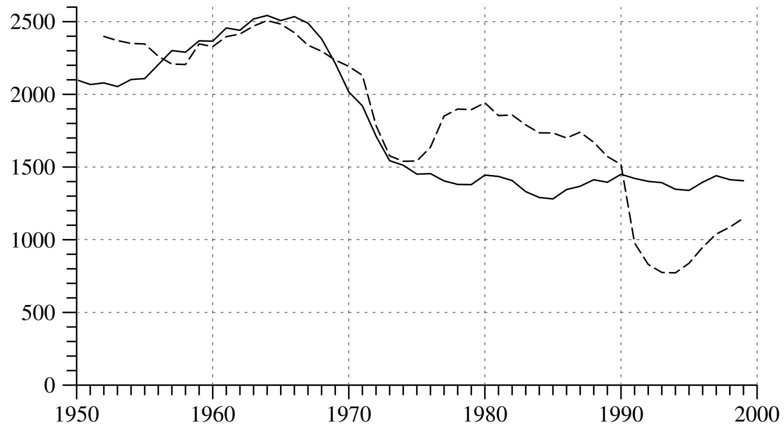



**Fig. 18.1-1** Total birth rates in the territory of the former FRG (solid line) and in the territory of the former GDR (dotted line). Data are taken from Table 18.1-1.

Its value provides the mean number of female births per women assuming that the current birth and death rates apply until the end of the reproductive period.

**4.** Based on the general life table 1986/88 and assuming a reproductive period from 15 to 50 years, the *Statistisches Bundesamt* has calculated a value of 0.651 for the net reproduction rate in Germany in the year 1999.[5] Since, in Germany, female mortality until the end of the childbearing period is very low, the net reproduction rate is only slightly lower than the gross reproduction rate:

$$\mathrm{GRR}_{1999} = \sigma_{1999,f}\, \mathrm{TBR}_{1999}/1000 = 0.486 \cdot 1360.9/1000 = 0.661$$

In fact, a plot of the net reproduction rates would be very similar to the total birth rates shown in Figure 18.1-1.

## 18.2 Relationship with Growth Rates

**1.** Reproduction rates are hypothetical constructs. Their interpretation is based on the assumption that the current birth and death rates prevail for an indefinite period of time. This is similar to the model introduced in Chapter 17 and, in fact, there is a close relationship between the net reproduction rate and the intrinsic growth rate that derives from this model. In order to discuss this relationship we refer to the matrix $\mathbf{F} = \mathbf{D}_f + \sigma_f \mathbf{B}$ that

[5] Fachserie 1, Reihe 1, 1999 (p. 53). Using the definition given above, one can derive a value of 0.645 from the data in Table 17.5-1 in Section 17.5.2.

was defined in Section 17.2. The first row of this matrix contains adjusted age-specific birth rates, $\beta^*_\tau$, and the subdiagonal contains the age-specific female survivor rates $(1-\delta^f_\tau)$.[6] The intrinsic growth rate, $\rho^*$, depends on these rates. In fact, as shown in Section 17.3, it suffices to consider the sub-matrix $\tilde{\mathbf{F}}$ which consists of the first $\tau_b$ rows and columns of $\mathbf{F}$ and has the following structure:

$$\tilde{\mathbf{F}} = \begin{bmatrix} \sigma_f\beta^*_1 & \sigma_f\beta^*_2 & \cdots & \sigma_f\beta^*_{\tau_b-1} & \sigma_f\beta^*_{\tau_b} \\ 1-\delta^f_1 & 0 & \cdots & 0 & 0 \\ 0 & 1-\delta^f_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-\delta^f_{\tau_b-1} & 0 \end{bmatrix}$$

The intrinsic growth rate is $\rho^* = \lambda^* - 1$, $\lambda^*$ being the dominant eigenvalue of $\tilde{\mathbf{F}}$. So we have to investigate how $\lambda^*$ depends on the elements of $\tilde{\mathbf{F}}$.

**2.** We first mention that the elements of $\tilde{\mathbf{F}}$ can be used to calculate the net reproduction rate. As shown in Section 17.1, the relationship between the rates $\beta^*_\tau$, which are used for the model formulation, and the age-specific birth rates $\beta_{t,\tau}$ is given by

$$\beta^*_\tau := \beta^*_{t,\tau} = \beta_{t,\tau}\,(1-\delta^f_{t,0})$$

So we get the net reproduction rate in the following way:

$$\mathrm{NRR}_t = \sigma_{t,f}\sum_{\tau=\tau_a}^{\tau_b}\beta_{t,\tau}\,G^f_{t,\tau} = \sigma_{t,f}\sum_{\tau=\tau_a}^{\tau_b}\beta^*_{t,\tau}\,G^f_{t,\tau}/(1-\delta^f_{t,0})$$

$$= \sigma_{t,f}\sum_{\tau=\tau_a}^{\tau_b}\beta^*_{t,\tau}\prod_{j=1}^{\tau-1}(1-\delta^f_{t,j})$$

Using the fact that $\beta^*_{t,\tau} = 0$ for $\tau < \tau_a$, and omitting the period index $t$, one arrives at the formulation[7]

$$\mathrm{NRR} = \sigma_f\sum_{\tau=\tau_a}^{\tau_b}\beta^*_\tau\prod_{j=1}^{\tau-1}(1-\delta^f_j)$$

This then shows how the net reproduction rate is related to the elements of the matrix $\tilde{\mathbf{F}}$.

**3.** For the next step we need a mathematical fact which will be stated without proof: For any $(n,n)$ matrix $\mathbf{A}$, its eigenvalues are the roots of the so-called characteristic equation

$$\det(\lambda\mathbf{I} - \mathbf{A}) = 0$$

[6] As in the previous chapter, in order to simplify notations we omit the period index $t$.

[7] For $\tau = 1$ the product term is assumed to be 1.



In this formulation, $\mathbf{I}$ is an identity matrix and $\det(\lambda\mathbf{I}-\mathbf{A})$ is the determinant of $(\lambda\mathbf{I}-\mathbf{A})$ considered as a polynomial in $\lambda$, also called the characteristic polynomial of $\mathbf{A}$. We further state without proof that

$$\det(\lambda\mathbf{I} - \tilde{\mathbf{F}}) = \lambda^{\tau_b} - \sigma_f\sum_{\tau=\tau_a}^{\tau_b}\beta^*_\tau\,\lambda^{\tau_b-\tau}\prod_{j=1}^{\tau-1}(1-\delta^f_j)$$

We can find, therefore, the eigenvalues of $\tilde{\mathbf{F}}$ as the solutions of the equation

$$\lambda^{\tau_b} - \sigma_f\sum_{\tau=\tau_a}^{\tau_b}\beta^*_\tau\,\lambda^{\tau_b-\tau}\prod_{j=1}^{\tau-1}(1-\delta^f_j) = 0$$

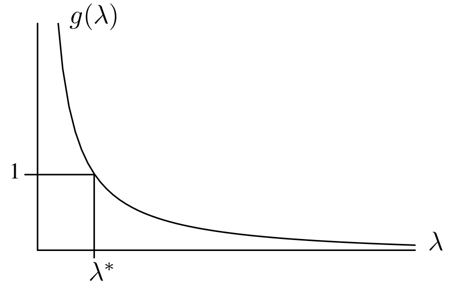
Writing this equation in the form

$$\sigma_f\sum_{\tau=\tau_a}^{\tau_b}\beta^*_\tau\,\lambda^{\tau_b-\tau}\prod_{j=1}^{\tau-1}(1-\delta^f_j) = \lambda^{\tau_b}$$

and dividing both sides by $\lambda^{\tau_b}$, we get, for $\lambda \neq 0$,

$$g(\lambda) := \sigma_f\sum_{\tau=\tau_a}^{\tau_b}\beta^*_\tau\,\lambda^{-\tau}\prod_{j=1}^{\tau-1}(1-\delta^f_j) = 1 \qquad (18.2.1)$$

**4.** In general, the equation $g(\lambda) = 1$ has $\tau_b$, possibly complex, roots. However, we are only interested in the dominant eigenvalue of $\tilde{\mathbf{F}}$ which is real and positive. Its existence is guaranteed by the theorem of Frobenius that was invoked in Section 17.3 but can also be shown directly.[8] Because $\beta^*_\tau \geq 0$ and also $(1-\delta^f_\tau) \geq 0$, $g(\lambda)$ is a monotonically decreasing, continuous function for all $\lambda > 0$. A possible graph of $g$ is shown below:




Furthermore, $g(\lambda) \to \infty$ if $\lambda \to 0$ and $g(\lambda) \to 0$ if $\lambda \to \infty$. It follows that there is a unique real and positive value, $\lambda^*$, where $g(\lambda^*) = 1$ and, since no larger positive root exists, this is the dominant eigenvalue of $\tilde{\mathbf{F}}$.

[8] See also Anton and Rorres (1991, p. 653).

**5.** Equation (18.2.1) also shows how the dominant eigenvalue depends on the birth and death rates: If one or more of the birth rates increase, or one or more of the death rates decrease, the dominant eigenvalue, and consequently the intrinsic growth rate, increases. A special case occurs if the net reproduction rate equals 1. This implies that the dominant eigenvalue also has the value 1, and the intrinsic growth rate will be zero. Of course, this argument concerns the intrinsic growth rate. The value of the actual growth rate also depends on the current age distribution and so it can happen that a population might well grow for some time although the net reproduction rate is already less than 1. However, if the net reproduction rate is below 1 and there is no immigration, the population eventually declines.

## 18.3 The Distance of Generations

**1.** In general, there is no simple and direct relationship between the net reproduction rate and the intrinsic growth rate. An exception is the case when the NRR has a value of 1. The intrinsic growth rate is then zero and also independent of the distribution of the age-specific birth rates. Except for this special case, the growth rate also depends on the timing of births. In particular, in the case of a positive net reproduction rate: if the mean age at childbearing increases the growth rate will decline and, conversely, if the mean age at childbearing decreases the growth rate will increase.

**2.** To illustrate this argument we consider the two matrices where, for simplicity, we assume zero death rates:

$$\tilde{\mathbf{F}}_a := \begin{bmatrix} 0 & 0.5 & 0.7 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{F}}_b := \begin{bmatrix} 0 & 0.7 & 0.5 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

In both cases the net reproduction rate is 1.2, the difference is in the timing of births. In case (a) more children are born at an older age, in case (b) more children are born at a younger age of their mothers. Calculating the dominant eigenvalues, we find $\lambda_a^* = 1.0734$ and $\lambda_b^* = 1.0787$ which shows that the intrinsic growth rate is higher in the second case.

**3.** One should notice, however, that this depends on whether the net reproduction rate is above or below 1. If less than 1, the relationship becomes reversed as shown by the following example:

$$\tilde{\mathbf{F}}_c := \begin{bmatrix} 0 & 0.5 & 0.3 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{F}}_d := \begin{bmatrix} 0 & 0.3 & 0.5 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$



In both cases the net reproduction rate has now a value of 0.8, implying a negative intrinsic growth rate. Calculating the dominant eigenvalues, we find $\lambda_c^* = 0.9107$ and $\lambda_d^* = 0.9188$. So case (d) has actually a relatively higher growth rate $\rho_d^* = -8.12\,\%$, compared with $\rho_c^* = -8.93\,\%$ in case (c). This can be explained by referring to the stable age distribution. If the population growth is positive, as in cases (a) and (b), there will be relatively more women in younger age classes and a shift of birth rates to these younger age classes will increase the growth rate. On the other hand, if the population growth is negative, there will be relatively more women in older age classes and a shift of birth rates to these older age classes will increase the growth rate.

**4.** The argument can also be formulated in terms of a *mean generational distance*, which is formally identical to the mean childbearing age of women, restricted to women who give birth to at least one child, and can be defined as

$$\frac{\sum_{\tau=\tau_a}^{\tau_b} \tau\, \beta_\tau\, G_\tau^f}{\sum_{\tau=\tau_a}^{\tau_b} \beta_\tau\, G_\tau^f}$$

It is often argued that, if this mean generational distance increases, the population growth rate will decrease. But this is actually only true if the net reproduction rate is greater than 1. Otherwise, if the population growth is negative, an increase in the mean generational distance will result in a less negative growth rate.

## 18.4 Growth Rates and Age Distributions

**1.** The argument in the previous section has shown that age distributions play a significant role in the analysis of population growth. On the other hand, the age distribution also depends on population growth. This is most easily shown by referring to the stable female age distribution. As has been discussed in Section 17.3, this age distribution is proportional to the eigenvector, $\mathbf{v}^*$, that corresponds to the dominant eigenvalue $\lambda^*$ and, if $\lambda^*$ is known, can easily be computed from the age-specific death rates: one begins with an arbitrary positive value for $v_1^*$ and then recursively applies the formula

$$v_\tau^* = \frac{1 - \delta_{\tau-1}^f}{\lambda^*}\, v_{\tau-1}^* \qquad (\text{for } \tau = 2, \ldots, \tau_m)$$

If the net reproduction rate is 1 ($\lambda^* = 1$), the formula shows that the age distribution only depends on the age-specific death rates.[9] But if

[9] In this case the age distribution would equal the life table age distribution discussed in Section 7.4.4.

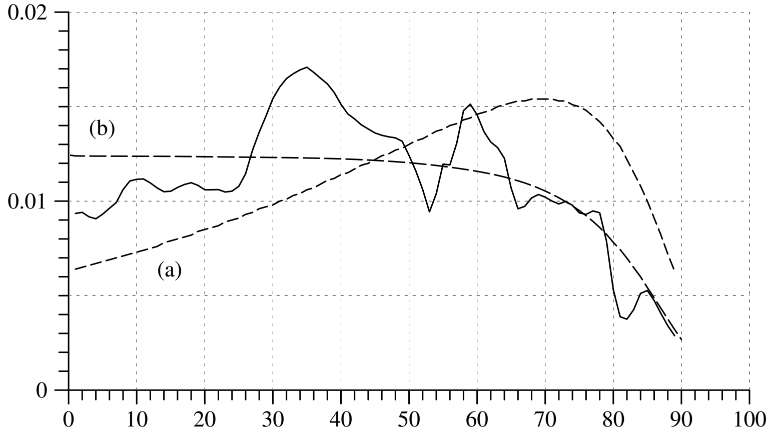



**Fig. 18.4-1** Solid line: Female age distribution in Germany 1999. Dotted lines: (a) Stable age distribution calculated from current birth and death rates. (b) Stable age distribution if the net reproduction rate would be 1.

population growth is positive, or negative, this is no longer the case. To show this we rewrite the formula in the following way (assuming that $v_1^* = 1 - \delta_0^f$):

$$v_\tau^* = v_1^* \left(\frac{1}{\lambda^*}\right)^{\tau-1} \prod_{j=1}^{\tau-1} (1 - \delta_j^f) = \left(\frac{1}{\lambda^*}\right)^{\tau-1} G_\tau^f$$

This shows that, if $\lambda^* > 1$, the frequencies of the higher age classes are multiplied by a factor that decreases with age and consequently become relatively smaller. Conversely, if $\lambda^* < 1$, the multiplicative factor increases with age, and this then implies that frequencies of the higher age classes become relatively larger.

**2.** As an illustration we consider again the stable female age distribution that was calculated in Section 17.5.2 for Germany in 1999. Two of the three frequency curves shown in Figure 18.4-1 are identical with the curves shown in Figure 17.5-2. The solid line depicts the actual female age distribution in 1999, the dotted curve (a) is the stable age distribution calculated from the birth and death rates in 1999. Since these rates imply a net reproduction rate which is far below 1, there is a huge shift towards the older age classes. The dotted curve (b) is calculated from the assumption that the female death rates have their actual values but the birth rates have values to ensure a net reproduction rate of 1. The age distribution is then solely determined by the current death rates.



## 18.5 Declining Importance of Death Rates

**1.** In general, the intrinsic growth rate depends both on birth and death rates. However, death rates are only important until the end of the reproductive period. Furthermore, in modern societies these death rates are already very low. For example, referring to the period life table for the year 1999 (see Table 7.3-1 in Section 7.3.2), out of 1000 women only 23 died until an age of 45. One can expect, therefore, that further progress in diminishing death rates will not have any substantial consequences for the intrinsic growth rate.

**2.** To illustrate the argument we refer again to the year 1999. As was shown in Section 13.3.2, the birth and death rates of that year imply an intrinsic growth rate of -1.51 %. We now assume that death rates were zero until the end of the reproductive period. The corresponding intrinsic growth rate would then be -1.48 %.

## 18.6 Population Growth with Immigration

**1.** A further question concerns the effects of immigration on population growth. To provide a brief discussion we extend the female population model introduced in Chapter 17 to include female net immigration. Remember the original model formulation: $\mathbf{n}_{t+1}^f = \mathbf{F}\mathbf{n}_t^f$, where $\mathbf{n}_t^f$ is a female population vector for the year $t$, and $\mathbf{F}$ is the Leslie matrix assumed to be time-independent. We now consider an additional vector

$$\mathbf{m}_t^f := \begin{bmatrix} m_{t,1}^f \\ \vdots \\ m_{t,\tau_m}^f \end{bmatrix}$$

where $m_{t,\tau}^f$ is the net immigration of women aged $\tau$ in the year $t$. Of course, components might be negative if out-migration exceeds in-migration. Using this vector, an extended model can be written as follows:

$$\mathbf{n}_{t+1}^f = \mathbf{F}\mathbf{n}_t^f + \mathbf{m}_t^f \tag{18.6.1}$$

The formulation assumes that there is a single Leslie matrix $\mathbf{F}$ that provides the birth and death rates both for native and immigrant women.[10]

**2.** A simple solution is possible if we assume a time-constant immigration vector $\mathbf{m}^f \geq 0$. Beginning with a base year $t = 0$, we find:

$$\mathbf{n}_1^f = \mathbf{F}\mathbf{n}_0^f + \mathbf{m}^f$$
$$\mathbf{n}_2^f = \mathbf{F}\mathbf{n}_1^f + \mathbf{m}^f = \mathbf{F}^2\mathbf{n}_0^f + \mathbf{F}\mathbf{m}^f + \mathbf{m}^f$$

[10] For a similar approach to include migration into a Leslie model see Lilienbecker (1991), further possibilities have been discussed by Sivamurthy (1982).

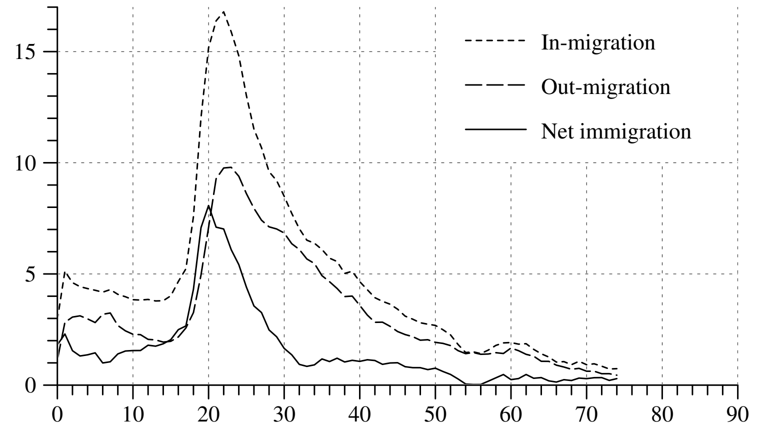



**Fig. 18.6-1** Age distribution of female immigrants and emigrants in Germany 1999. Data are taken from Table 18.6-1.

and so on, in general:[11]

$$\mathbf{n}_t^f = \mathbf{F}^t \mathbf{n}_0^f + \sum_{j=0}^{t-1} \mathbf{F}^j \mathbf{m}^f$$

This equation can be used to think about equilibrium conditions. A sufficient condition is that the intrinsic growth rate implied by $\mathbf{F}$ is negative. Then, if $t$ becomes larger, $\mathbf{F}^t \mathbf{n}_0^f$ converges to zero, and the population vector $\mathbf{n}_t^f$ converges to

$$\bar{\mathbf{n}}^f := (\mathbf{I} - \mathbf{F})^{-1}\, \mathbf{m}^f$$

This also implies that, in the long run, the growth rate becomes zero and the time-constant population $\bar{\mathbf{n}}^f$ only depends on the net immigration and the parameters of the Leslie matrix $\mathbf{F}$.

**3.** For an illustration we continue with the data used in Section 17.5.2 providing the Leslie matrix $\mathbf{F}$ and the initial female population vector $\mathbf{n}_0^f$ for the year 1999. In addition, we use the data shown in Table 18.6-1 about female immigration and emigration in Germany in the same year. The age class 75* is open-ended and covers also all higher ages. Altogether, 369049 women immigrated and 248108 women emigrated during the year 1999 resulting in a net immigration of 120941 women. As shown in Figure 18.6-1, both in- and out-migration mainly take place in younger ages. For the net immigration vector $\mathbf{m}^f$ we therefore only use the figures from Table 18.6-1 until the age 74 ($m_\tau^f = 0$ for $\tau \geq 75$), in total about 121000 persons. According to the model (18.6.1), we assume that the same female in-migration takes place also in all years following 1999. Then, after 51 iterations of the model, one finds the projected female population vector for the year 2050. The total female population would then be about 41.3 million, instead of 34.4 million as projected by a model without immigration. Since mainly young women immigrate, also the age distribution would be quite different. This is illustrated in Figure 18.6-2. The solid line shows the female age distribution in 1999; the two other lines show, respectively,

[11] By convention, $\mathbf{F}^0$ equals the identity matrix $\mathbf{I}$.



**Table 18.6-1** Female in-migration ($m_{t,\tau}^{f,i}$), out-migration ($m_{t,\tau}^{f,o}$), and net immigration ($m_{t,\tau}^{f}$), classified according to age $\tau$, in the year $t$ = 1999 in Germany. Source: Fachserie 1, Reihe 1, 1999 (pp. 116-117).

| $\tau$ | $m_{t,\tau}^{f,i}$ | $m_{t,\tau}^{f,o}$ | $m_{t,\tau}^{f}$ | $\tau$ | $m_{t,\tau}^{f,i}$ | $m_{t,\tau}^{f,o}$ | $m_{t,\tau}^{f}$ |
|---|---|---|---|---|---|---|---|
| 0 | 2942 | 1131 | 1811 | 38 | 5024 | 3983 | 1041 |
| 1 | 5123 | 2823 | 2300 | 39 | 5121 | 4006 | 1115 |
| 2 | 4609 | 3061 | 1548 | 40 | 4652 | 3583 | 1069 |
| 3 | 4428 | 3117 | 1311 | 41 | 4290 | 3150 | 1140 |
| 4 | 4342 | 2978 | 1364 | 42 | 3932 | 2825 | 1107 |
| 5 | 4266 | 2814 | 1452 | 43 | 3768 | 2832 | 936 |
| 6 | 4181 | 3184 | 997 | 44 | 3641 | 2646 | 995 |
| 7 | 4296 | 3245 | 1051 | 45 | 3416 | 2410 | 1006 |
| 8 | 4083 | 2680 | 1403 | 46 | 3103 | 2274 | 829 |
| 9 | 3975 | 2441 | 1534 | 47 | 2966 | 2180 | 786 |
| 10 | 3841 | 2287 | 1554 | 48 | 2805 | 2021 | 784 |
| 11 | 3826 | 2269 | 1557 | 49 | 2739 | 2041 | 698 |
| 12 | 3853 | 2058 | 1795 | 50 | 2686 | 1924 | 762 |
| 13 | 3785 | 2033 | 1752 | 51 | 2490 | 1877 | 613 |
| 14 | 3796 | 1937 | 1859 | 52 | 2260 | 1779 | 481 |
| 15 | 4015 | 1971 | 2044 | 53 | 1828 | 1551 | 277 |
| 16 | 4661 | 2166 | 2495 | 54 | 1466 | 1415 | 51 |
| 17 | 5231 | 2570 | 2661 | 55 | 1497 | 1466 | 31 |
| 18 | 7577 | 3263 | 4314 | 56 | 1419 | 1387 | 32 |
| 19 | 12043 | 4957 | 7086 | 57 | 1576 | 1399 | 177 |
| 20 | 15172 | 7095 | 8077 | 58 | 1787 | 1458 | 329 |
| 21 | 16383 | 9280 | 7103 | 59 | 1899 | 1422 | 477 |
| 22 | 16788 | 9766 | 7022 | 60 | 1915 | 1668 | 247 |
| 23 | 15885 | 9796 | 6089 | 61 | 1843 | 1548 | 295 |
| 24 | 14811 | 9393 | 5418 | 62 | 1863 | 1385 | 478 |
| 25 | 13025 | 8615 | 4410 | 63 | 1610 | 1301 | 309 |
| 26 | 11514 | 7953 | 3561 | 64 | 1410 | 1072 | 338 |
| 27 | 10679 | 7416 | 3263 | 65 | 1262 | 1070 | 192 |
| 28 | 9600 | 7119 | 2481 | 66 | 1036 | 898 | 138 |
| 29 | 9199 | 7023 | 2176 | 67 | 1057 | 814 | 243 |
| 30 | 8495 | 6842 | 1653 | 68 | 920 | 712 | 208 |
| 31 | 7724 | 6360 | 1364 | 69 | 1071 | 754 | 317 |
| 32 | 7035 | 6102 | 933 | 70 | 918 | 626 | 292 |
| 33 | 6518 | 5674 | 844 | 71 | 960 | 626 | 334 |
| 34 | 6381 | 5470 | 911 | 72 | 851 | 513 | 338 |
| 35 | 6101 | 4923 | 1178 | 73 | 726 | 514 | 212 |
| 36 | 5710 | 4650 | 1060 | 74 | 740 | 447 | 293 |
| 37 | 5559 | 4348 | 1211 | 75* | 5050 | 3721 | 1329 |

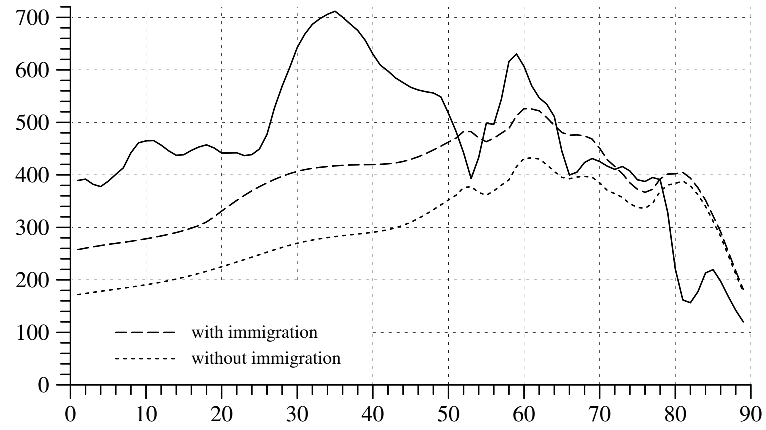



**Fig. 18.6-2** Age distribution of the female population in 1999 (solid line) and of the projections for the year 2050 with and without immigration (dotted lines). The ordinate refers to absolute frequencies.

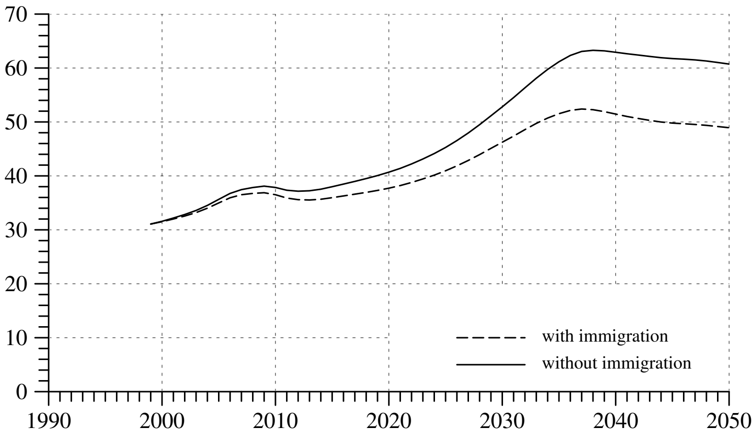



**Fig. 18.6-3** Projected development of female old age dependency ratios until the year 2050, with and without immigration.

the age distribution of the projected female population in the year 2050 with and without immigration. A simple summary measure is the female *old age dependency ratio* [Altenquotient] defined as[12]

$$\frac{\text{number of women aged 65 and over}}{\text{number of women aged 20 to 64}}$$

In our example, we can calculate this measure with a numerator that refers

[12] We mention that there is no general convention where to beginn the "old ages".



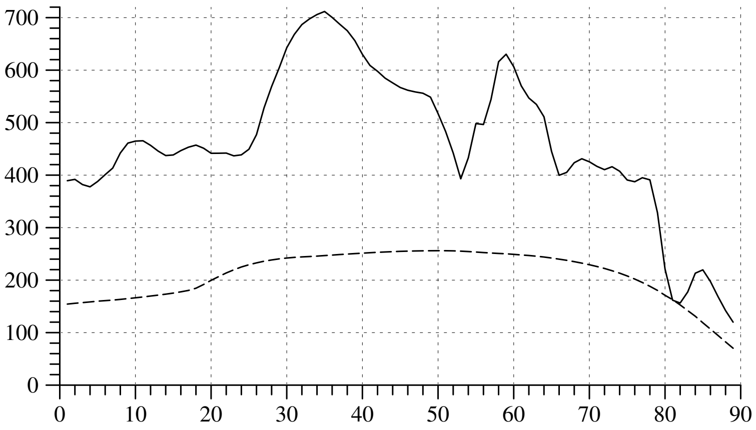



**Fig. 18.6-4** Age distribution, in absolute frequencies, of the female population in Germany 1999 (solid line) and stable age distribution derived from the 1999 Leslie matrix and a constant net immigration according to Table 18.6-1 (dotted line).

to women aged 65 to 89. Figure 18.6-3 compares the development of this ratio in models with and without immigration.

**4.** Since the Leslie matrix for Germany in 1999 implies a negative intrinsic growth rate of about -1.5 %, without immigration the population would vanish in the long run. On the other hand, a constant net immigration would not only slow down the population shrinkage but eventually stabilize the population at a constant level. In our model, this long-term level is given by the population vector $\bar{\mathbf{n}}^f$ and can easily be calculated from the Leslie matrix $\mathbf{F}$ and the net immigration vector $\mathbf{m}^f$. Using the data for Germany in 1999, the total number of female persons aged 1 to 89 would eventually stabilize at about 18.7 million. One should also note that the equation

$$\bar{\mathbf{n}} = (\mathbf{I} - \mathbf{F})^{-1}\mathbf{m}$$

is linear; a proportional increase, or decrease, of the immigration vector would result in the same proportional increase, or decrease, of the final population size.

**5.** The long-run equilibrium also implies a stable age distribution. Figure 18.6.4 compares this stable age distribution with the actual age distribution of the female population in 1999. The female old age dependency ratio would be 41 % compared with 31 % in 1999. However, to put these figures into perspective one should compare the models with and without immigration. This is done in Figure 18.6-5 that compares the stable age distributions from models with and without immigration. The stable age

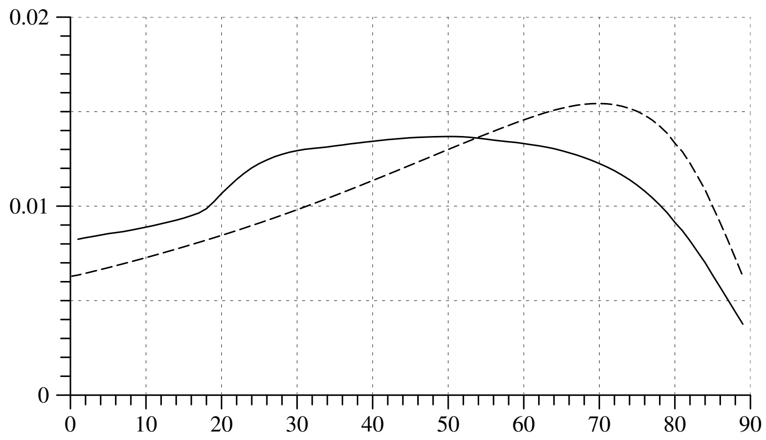



**Fig. 18.6-5** Stable age distributions (relative frequencies) implied by the Leslie models with immigration (solid line) and without immigration (dotted line).

distribution of the model without immigration would correspond to a female old age dependency ratio of 62 %. Remarkably, only a small part of the much lesser ratio of 41 % in the model with immigration is due to the fact that mainly young women immigrate. If, instead of the figures in Table 18.6-1, we assume ages of immigrants equally distributed between 1 and 50 years, the final old age dependency ratio would only slightly increase to 42.4 %.

# Appendix A

# Appendix

This appendix has two sections. Section A.1 provides some hints about how to find data and additional information from official statistics in Germany. Section A.2 briefly summarizes some notation from set theory that is used in the main text.

## A.1 Data from Official Statistics

This section is not finished yet.

## A.2 Sets and Functions

Throughout the text we stressed the fact that statistical variables are to be regarded as functions and that statistical distributions are functions on sets of sets. Sets and functions thus play a fundamental role in all statistical constructs. The following two sections summarize the basic notations for sets and functions.

### Notations from Set Theory

**1.** The basic idea is that people are able to comprehend arbitrary objects into a *set* [*Menge*]. Georg Cantor (1845–1918), the originator of set theory, gave the following explanation:

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

In accordance with this explanation, the construction of a set is a mental operation without any specific implication for the ontological status of the resulting set. Furthermore, there is no restriction in the kinds of objects that can be considered to be elements of a set.

**2.** We generally use capital letters to denote sets. The elements of the set, i.e. the entities belonging to the set, are written in small letters.[1] Thus, $A := \{a_1, a_2, a_3\}$ defines the set $A$ to be the collection made up by the elements $a_1$, $a_2$ and $a_3$.

**3.** Most of the sets that appear in this text have a finite number of elements. For a set $A$ with a finite number of elements we use the abbreviation $|A|$ for 'the number of elements of $A$'. If $A := \{a_1, a_2, a_3\}$, then $|A| = 3$.

**4.** We us the symbol $\in$ as an abbreviation for "belongs to". Thus we write $a \in A$. Similarly, we use the symbol $\notin$ as an abbreviation for "does not belong to". Two sets are equal if both sets have the same elements. In other words, for two sets $A$ and $B$, $A = B$ if each element of $A$ is also an element of $B$ and each element of $B$ is also an element of $A$. Sets are therefore completely determined when its elements are given, while the order in which elements are given is irrelevant: $\{a_1, a_2, a_3\} = \{a_2, a_3, a_1\}$.

**5.** When the order of elements in a collection is of importance we write

$$(a_1, a_2, a_3)$$

In this case, the order of the three elements makes a difference, i.e.

$$(a_1, a_2, a_3) \neq (a_2, a_1, a_3)$$

[1] We try to follow this convention throughout the text. But occasionally we will have to refer to sets whose members are themselves sets.



We call such an ordered collection a pair if it has two elements. If the collection contains three elements, we call it a triple. Generally, we call an ordered collection of $n$ elements $(a_1, \ldots, a_n)$ an *n-tuple*.

**6.** When a set, say $B$, has been defined one can build a new set by the construction

$$C := \{b \in B \,|\, b \text{ has the property } \ldots\}$$

Here, $C$ is the name of the set consisting of all those elements of $B$ which have the property given after the vertical line. The new set is a *subset* of the set $B$. We write $C \subseteq B$ if $C$ is a subset of $B$, i.e. if each element of $C$ is also an element of $B$. Consequently, $B \subseteq B$ is always true. We write $C \subset B$ if there are elements of $B$ that do not belong to $C$.

**7.** Given two sets $A$ and $B$, one can define new sets by the operations of union and intersection: The *union* $A \cup B$ is the set of elements belonging to at least one of the sets $A$ and $B$. The *intersection* $A \cap B$ is the set of elements belonging both to $A$ and $B$. It might happen that the intersection contains no elements at all. If this happens we call the two sets *mutually exclusive* or *disjoint*. We call a set with no elements empty. But according to our definition of the equality of sets there is only one empty set. We call it the *empty set* and denote it by $\emptyset$.

**8.** As a direct consequence of the definitions, union and intersection are commutative

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

and distributive

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

**9.** If $B$ is a subset of $A$ then $B^c := \{a \in A \,|\, a \notin B\}$ is the *complement* of $B$ in $A$. In general, if $A$ and $B$ are sets, $A \setminus B := \{a \in A \,|\, a \notin B\}$ is the complement of $A \cap B$ in $A$.

**10.** Given a set $A$, a *partition* of $A$ is a set of subsets of $A$ with elements $A_1, \ldots, A_m$, such that the union of all these sets is equal to $A$ ($A_1 \cup \ldots \cup A_m = A$) and such that all distinct pairs $A_i, A_j$ are mutually exclusive ($A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, m\}$ provided that $i \neq j$). For example, if $A := \{a_1, a_2, a_3\}$, then

$$\{\{a_1\}, \{a_2, a_3\}\}$$

is a partition of $A$.

**11.** The *power set* of a set $A$ is the set of all its subsets. We use the symbol $\mathcal{P}(A)$. Both the empty set $\emptyset$ and the set $A$ itself are elements of the power set. Using once again $A := \{a_1, a_2, a_3\}$ we have

$$\mathcal{P}(A) = \{\emptyset, \{a_1\}, \{a_2\}, \{a_3\}, \{a_1, a_2\}, \{a_1, a_3\}, \{a_2, a_3\}, \{a_1, a_2, a_3\}\}$$

The number of elements belonging to a power set is $|\mathcal{P}(A)| = 2^{|A|}$.

**12.** Another elementary notion is that of a *Cartesian product* of two or more sets. Given two sets $A$ and $B$, the Cartesian product $A \times B$ is the set of all ordered pairs that can be constructed from elements of $A$ and $B$. As an example, if

$$A := \{1, 2\} \quad \text{and} \quad B := \{3, 4, 5\}$$

then

$$A \times B = \{(1,3), (1,4), (1,5), (2,3), (2,4), (2,5)\}$$

The Cartesian product of three and more sets is constructed similarly. For example, if $C := \{6\}$ then

$$A \times B \times C = \{(1,3,6), (1,4,6), (1,5,6), (2,3,6), (2,4,6), (2,5,6)\}$$

One might also construct the Cartesian product of a set with itself:

$$\begin{aligned} A \times A \times A = \{&(1,1,1), (1,1,2), (1,2,1), (1,2,2), \\ &(2,1,1), (2,1,2), (2,2,1), (2,2,2)\} \end{aligned}$$

In this case we use the following abbreviation

$$A^n := \underbrace{A \times \cdots \times A}_{n \text{ times}}$$

**13.** The Cartesian product operates distributively on unions and intersections:

$$\begin{aligned} A \times (B \cup C) &= (A \times B) \cup (A \times C) \\ A \times (B \cap C) &= (A \times B) \cap (A \times C) \end{aligned}$$

In particular,

$$A \times \emptyset = \emptyset \times A = \emptyset$$

But the Cartesian product is not, in general, commutative:

$$B \times A = \{(3,1), (4,1), (5,1), (3,2), (4,2), (5,2)\} \neq A \times B$$



## The Notion of Function

**1.** The notion of function is fundamental to statistics. We use the word in the same sense as it is now used in mathematics. Given two sets $A$ and $B$, a *function* relates to each element of $A$ a unique element of $B$. We write

$$f : A \longrightarrow B$$

where $f$ is the name of the function, $A$ is called the *domain* of the function, and $B$ is the *counterdomain* of the function. If $a \in A$ is an element of the domain of the function $f$, we write $f(a)$ for the unique element of $B$ which is related to $a$ through the function $f$. We call $f(a)$ the *value* of the function evaluated at the *argument* $a$.

**2.** Given the two sets $A := \{1, 2\}$ and $B := \{3, 4, 5\}$ we might define a function $f : A \longrightarrow B$ by defining $f(1) = 3, f(2) = 4$, that is by giving its values for all the arguments. Next we must say when two functions are to be regarded as equal. We will say that two functions $f : A \longrightarrow B$ and $g : C \longrightarrow D$ are equal if $A = C$, $B = D$, and $f(a) = g(a)$ for all $a \in A$. For example, if

$$g : \{1, 2\} \longrightarrow \{3, 4\}$$

with $g(1) = 3$ and $g(2) = 4$, then $f \neq g$.

**3.** With a function $f : A \longrightarrow B$ we can associate a further function, called a *set function*, that takes subsets of the domain as its argument. In a slight abuse of notation we will denote that function by the same symbol $f$. Thus we write

$$f : \mathcal{P}(A) \longrightarrow \mathcal{P}(B)$$

where the set function relates a subset $C \subseteq A$ to the unique subset

$$f(C) := \{b \in B \,|\, \text{there is an } a \in C \text{ such that } f(a) = b\}$$

of $B$. In shorter notation,

$$f(C) = \{f(a) \,|\, a \in C\}$$

We call $f(C)$ the *image of $C$ under $f$*. Especially, the image of $A$ is called the *range* of the function. Obviously, $f(A) \subseteq B$; but as in the example above we may have $f(A) \neq B$.

**4.** The set function associated with a function $f : A \longrightarrow B$ always has an *inverse set function* defined by

$$f^{-1} : \mathcal{P}(B) \longrightarrow \mathcal{P}(A)$$

which to each subset of the counterdomain of $f$ relates a unique subset according to

$$f^{-1}(C) := \{a \in A \mid f(a) \in C\}$$

where $C$ is an arbitrary element of $\mathcal{P}(B)$. We call $f^{-1}(C)$ the *preimage* of $C$ with respect to $f$. For example, if $f : \{1,2\} \longrightarrow \{3,4,5\}$ is defined by $f(1) = 3$ and $f(2) = 4$, then

$$f^{-1}(\{3\}) = \{1\},\ f^{-1}(\{4\}) = \{2\},\ f^{-1}(\{5\}) = \emptyset,$$
$$f^{-1}(\{3,4\}) = \{1,2\},\ f^{-1}(\{3,5\}) = \{1\},\ f^{-1}(\{4,5\}) = \{2\},$$
$$f^{-1}(\{3,4,5\}) = \{1,2\},\ f^{-1}(\emptyset) = \emptyset$$

The union and intersection operations are preserved under inverse set functions:

$$f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$$
$$f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$$

where $C$ and $D$ are arbitrary subsets of $B$.

**5.** It should be clear that the mathematical notion of a function is fundamentally different from the use of the word in connection with purposes and aims. Even if this is fairly obvious from the definitions, we should stress, first, that functions are created by the human mind. It is the scientist who conceptualizes sets, and the scientist who constructs relations and functions between sets. Neither functions nor sets are empirical facts. Secondly, however, there is a difference between the uses of the concepts in mathematics and in statistics. In mathematics, one might create sets and functions without regard to empirical facts. In contrast, statistical methods are constructed in order to support reflections on empirical facts. Thus, in statistics, the usefulness of sets and functions will not only depend on their formal properties as such but much more on the intended meaning of the sets and functions.

# References

Note: Sources from Official Statistics cited in the main text are not contained in this list of references.


Alt, C. 1991. Stichprobe und Repräsentativität. In: H. Bertram (Hg.), Die Famlie in Westdeutschland, 497-531. Opladen: Leske + Budrich.

Anderson, R. N. 1999. Method for Constructing Complete Annual U.S. Life Tables. National Center for Health Statistics. Vital Health Statistics Series 2, No. 129.

Anton, H., Rorres, C. 1991. Elementary Linear Algebra. Applications Version. New York: Wiley.

Bach, W, Handl, J., Müller, W. 1980. Volks- und Berufszählung 1970. Codebuch und Grundauszählung. Mannheim: VASMA-Projekt.

Balzer, W. 1997. Die Wissenschaft und ihre Methoden. Grundsätze der Wissenschaftstheorie. München: Alber.

Baumol, W. J. 1966. Economic Models and Mathematics. In: S. R. Krupp (ed.), The Structure of Economic Science, 88 – 101. Englewood Cliffs: Prentice-Hall. [A German translation appeared in: H. Albert (ed.), Theorie und Realität, 153 – 168. Tübingen: Mohr 1972.]

Birg, H., Filip, D., Flöthmann, E.-J. 1990. Paritätsspezifische Kohortenanalyse des generativen Verhaltens in der Bundesrepublik Deutschland nach dem 2. Weltkrieg. IBS-Materialien Nr. 30. Institut für Bevölkerungsforschung und Sozialpolitik (IBS) an der Universität Bielefeld.

Blossfeld, H.-P., Huinink, J. 1989. Die Verbesserung der Bildungs- und Berufschancen von Frauen und ihr Einfluß auf den Prozeß der Familienbildung. Zeitschrift für Bevölkerungswissenschaft 15, 383 – 404.

Bolte, K. M., Kappe, D., Schmid, J. 1980. Bevölkerung. Statistik, Theorie, Geschichte und Politik des Bevölkerungsprozesses. Opladen: Leske + Budrich.

Borst, A. 1990. Computus. Zeit und Zahl in der Geschichte Europas. Berlin: Wagenbach. [English translation: The Ordering of Time. From the Ancient Computus to the Modern Computer. Cambridge: Polity Press 1993.]

Bortkiewicz, L. v. 1911. Sterblichkeit und Sterblichkeitstafeln. In: J. Conrad, L. Elster, W. Lexis, E. Loening (Hg.), Handwörterbuch der Staatswissenschaften, Bd. 7, 930 – 944. Jena: Gustav Fischer.

Bortkiewicz, L. v. 1919. Bevölkerungswesen. Leipzig: Teubner.

Brand, M. 1982. Physical Objects and Events. In: W. Leinfellner, E. Kraemer, J. Schank (eds.), Language and Ontology, 106 – 116. Wien: Hölder-Pichler-Temsky.

Bürgin, G., Schnorr-Bäcker, S. 1986. ISI-“Declaration on Professional Ethics” – Internationaler Berufskodex für Statistiker aus der Sicht der Bundesstatistik. Wirtschaft und Statistik 34, 573 – 581.

Cantor, G. 1962. Gesammelte Abhandlungen mathematischen und philosophischen Inhalts. Hrsg. von E. Zermelo. Hildesheim: Georg Olms.

Coleman, J. S. 1968. The Mathematical Study of Change. In: H. M. Blalock, A. B. Blalock (eds.), Methodology in Social Research, 428 – 478. New York:

McGraw Hill.

Danto, A. C. 1985. Narration and Knowledge (including: Analytical Philosophy of History). New York: Columbia University Press. [There is a German translation of the previously written Analytical Philosophy of History: Analytische Philosophie der Geschichte. Frankfurt: Suhrkamp 1980.]

Demetrius, L. 1971. Primitivity Conditions for Growth Matrices. Mathematical Biosciences 12, 53 – 58.

Dinkel, R. 1983. Analyse und Prognose der Fruchtbarkeit am Beispiel der Bundesrepublik Deutschland. Zeitschrift für Bevölkerungswissenschaft 9, 47 – 72.

Dinkel, R. 1984. Sterblichkeit in Perioden- und Kohortenbetrachtung. Zeitschrift für Bevölkerungswissenschaft 10, 477 – 500.

Dinkel, R. H. 1989. Demographie. Bd. 1: Bevölkerungsdynamik. München: Verlag Franz Vahlen.

Dinkel, R. H. 1992. Kohortensterbetafeln für die Geburtsjahrgänge ab 1900 bis 1962 in den beiden Teilen Deutschlands. Zeitschrift für Bevölkerungswissenschaft 18, 96 – 116.

Dinkel, R. H., Meinl, E. 1991. Die Komponenten der Bevölkerungsentwicklung in der Bundesrepublik Deutschland und der DDR zwischen 1950 und 1987. Zeitschrift für Bevölkerungswissenschaft 17, 115–134.

Dinkel, R. H., Höhn, C., Scholz, R. (eds.) 1996. Sterblichkeitsentwicklung – unter besonderer Berücksichtigung des Kohortenansatzes. München: Harald Boldt Verlag.

Esenwein-Rothe, I. 1982. Einführung in die Demographie. Bevölkerungsstruktur und Bevölkerungsprozeß aus der Sicht der Statistik. Wiesbaden: Franz Steiner Verlag.

Esenwein-Rothe, I. 1992. Wilhelm Lexis. Demograph und Nationalökonom. Frankfurt: Haag + Herchen.

Feichtinger, G. 1973. Bevölkerungsstatistik. Berlin: de Gruyter.

Feichtinger, G. 1979. Demographische Analyse und populationsdynamische Modelle. Wien: Springer-Verlag.

Festy, P., Prioux, F. 2002. An Evaluation of the Fertility and Family Surveys Project. New York and Geneva: United Nations.

Fisher, R. A. 1922. On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society of London. Series A, Vol. 222, 309 – 368.

Flaskämper, P. 1962. Bevölkerungsstatistik. Hamburg: Verlag Richard Meiner.

Fliegel, H. F., Flandern, T. C. van 1968. A Machine Algorithm for Processing Calendar Dates. Communications of the ACM 11, 657.

Frege, G. 1990. Schriften zur Logik und Sprachphilosophie. 3. Aufl., hrsg. von G. Gabriel. Hamburg: Felix Meiner.

Frey, G. 1961. Symbolische und Ikonische Modelle. In: H. Freudenthal (ed.), The Concept and the Role of the Model in Mathematics and Natural and Social Sciences, 89 – 97. Dordrecht: Reidel.

Fürst, G. 1972. Wandlungen im Programm und in den Aufgaben der amtlichen Statistik in den letzten 100 Jahren. In: Statistisches Bundesamt, Bevölkerung und Wirtschaft 1872 – 1972, pp. 11 – 83. Wiesbaden: Kohlhammer.

Galton, F. 1889. Natural Inheritance. London: Macmillan.

Gantmacher, F. R. 1971. Matrizenrechnung (Teil II). Berlin: Deutscher Verlag der Wissenschaften.

Glenn, N. D. 1977. Cohort Analysis. Beverly Hills: Sage.

Hacker, P. M. S. 1982. Events and Objects in Space and Time. Mind 91, 1 – 19.

Hauser, P. M., Duncan, O. D. 1959. Overview and Conclusions. In: P. M. Hauser, O. D. Duncan (eds.), The Study of Population, 1 – 26. Chicago: University of Chicago Press.

Hendry, D. F., Richard, J.-F. 1982. On the Formulation of Empirical Models in Dynamic Econometrics. Journal of Econometrics 20, 3 – 33.

Höhn, C. 1984. Generationensterbetafeln versus Periodensterbetafeln. In: Neuere Aspekte der Sterblichkeitsentwicklung. Dokumentation der Jahrestagung 1983 der Deutschen Gesellschaft für Bevölkerungswissenschaft e.V., 117 – 143. Wiesbaden: Selbstverlag der Deutschen Gesellschaft für Bevölkerungswissenschaft e.V.

Huinink, J. 1987. Soziale Herkunft, Bildung und das Alter bei der Geburt des ersten Kindes. Zeitschrift für Soziologie 16, 367 – 384.

Huinink, J. 1988. Die demographische Analyse der Geburtenentwicklung mit Lebensverlaufsdaten. Allgemeines Statistisches Archiv 72, 359 – 377.

Huinink, J. 1989. Das zweite Kind. Sind wir auf dem Weg zur Ein-Kind-Familie? Zeitschrift für Soziologie 18, 192 – 207.

Huinink, J. 1998. Ledige Elternschaft junger Frauen und Männer in Ost und West. In: R. Metze, K. Mühler, K.-D. Opp (eds.), Der Transformationsprozess. Analysen und Befunde aus dem Leipziger Institut für Soziologie, 301 – 320. Leipzig: Leipziger Universitätsverlag.

Hullen, G. 1998. Lebensverläufe in West- und Ostdeutschland. Längsschnittanalysen des deutschen FFS. Opladen: Leske + Budrich.

Hyndman, R. J., Fan, Y. 1996. Sample Quantiles in Statistical Packages. The American Statistician 50, 361 – 365.

Imhof, A. E., Gehrmann, R., Kloke, I. E., Roycroft, M., Wintrich, H. 1990. Lebenserwartungen in Deutschland vom 17. bis 19. Jahrhundert (Life Expectancies in Germany from the 17th to the 19th Century). Weinheim: VCH – Acta humaniora.

International Statistical Institute 1986. Declaration on Professional Ethics. International Statistical Review 54, 227 – 242.

Kaplan, E. L., Meier, P. 1958. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53, 457 – 481.

Kendall, M., Stuart, A. 1977. The Advanced Theory of Statistics, vol. 1 (4th ed.). London: Charles Griffin & Comp.

Kertzer, D. I. 1983. Generation as a Sociological Problem. Annual Review of Sociology 9, 125 – 149.

Keyfitz, N. 1977. Applied Mathematical Demography. New York: Wiley.

Klein, T. 1988. Mortalitätsveränderungen und Sterbetafelverzerrungen. Zeitschrift für Bevölkerungswissenschaft 14, 49 – 67.

Klein, T. 1989. Bildungsexpansion und Geburtenrückgang. Kölner Zeitschrift für Soziologie und Sozialpsychologie 41, 483 – 503.

Klein, T. 1993. Soziale Determinanten der Lebenserwartung. Kölner Zeitschrift für Soziologie und Sozialpsychologie 45, 712 – 730.

Knodel, J. E. 1974. The Decline of Fertility in Germany, 1871 – 1939. Princeton: Princeton University Press.

Knodel, J. 1975. Ortssippenbücher als Quelle für die Historische Demographie. Geschichte und Gesellschaft 1, 288–324.

Kottmann, P. 1987. Verrechtlichung und Bevölkerungsweisen im industriellen Deutschland. Historical Social Research, No. 41, 28–39.

Leibniz, G. W. 1985. Kleine Schriften zur Metaphysik. Philosophische Schriften, Band 1, hrsg. von H. H. Holz. Darmstadt: Wissenschaftliche Buchgesellschaft.

Leslie, P. H. 1945. On the Use of Matrices in Certain Population Mathematics. Biometrika 33, 183 – 212.

Lexis, W. 1875. Einleitung in die Theorie der Bevölkerungsstatistik. Strassburg: Trübner.

Lilienbecker, T. 1991. Konstante Migrationsströme im Modell der stabilen Bevölkerung. In: G. Buttler, H.-J. Hoffmann-Nowotny, G. Schmitt-Rink (eds.), Acta Demographica 1991, 63–80. Heidelberg: Physica-Verlag.

Lindner, F. 1900. Die unehelichen Geburten als Sozialphänomen. Ein Beitrag zur Statistik der Bevölkerungsbewegung im Königreich Bayern. Leipzig: Deichert'sche Verlagsbuchhandlung.

Lombard, L. B. 1986. Events. A Metaphysical Study. London: Routledge.

Lorimer, F. 1959. The Development of Demography. In: P. M. Hauser, O. D. Duncan (eds.), The Study of Population, 124 – 179. Chicago: University of Chicago Press.

Lotka, A. J. 1907. Relation Between Birth Rates and Death Rates. Science N.S. 26, 21 – 22.

Lotka, A. J. 1922. The Stability of the Normal Age Distribution. Proceedings of the National Academy of Science of the USA 8, 339 – 345.

Mannheim, K. 1952. The Problem of Generations. In: Essays on the Sociology of Knowledge, 276 – 320. London: Routledge & Kegan Paul. [This essay first appeared in German: Das Problem der Generationen. Kölner Vierteljahreshefte für Soziologie 7 (1928), 157 – 185, 309 – 330.]

Marschalck, P. 1984. Bevölkerungsgeschichte Deutschlands im 19. und 20. Jahrhundert. Frankfurt: Suhrkamp.

Matras, J. 1973. Population and Societies. Englewood Cliffs: Prentice Hall.

Mayer, K. U., Huinink, J. 1990. Age, Period, and Cohort in the Study of the Life Course: A Comparison of Classical A-P-C-Analysis with Event History Analysis. In: D. Magnusson, L. R. Bergman (eds.), Data Quality in Longitudinal Research, 211 – 232. Cambridge: Cambridge University Press. [A German version of this paper appeared in: K. U. Mayer (Hg.), Lebensverläufe und sozialer Wandel, 442 – 459. Opladen: Westdeutscher Verlag 1990.]

Merrell, M. 1947. Time-specific Life Tables Contrasted with Observed Survivorship. Biometrics Bulletin 3, 129–136. Reprinted in: P. M. Hauser, O. D. Duncan (eds.), The Study of Population, 108 – 114. Chicago: University of Chicago Press 1956.





Meyer, K, Rückert, G. R. 1974. Allgemeine Sterbetafel 1970/72. Wirtschaft und Statistik, Heft 7, 465 – 475, 392* – 395*.

Meyer, K., Paul, C. 1991. Allgemeine Sterbetafel 1986/88. Wirtschaft und Statistik, Heft 6, 371 – 381, 234* – 241*.

Mueller, U. 1993. Bevölkerungsstatistik und Bevölkerungsdynamik. Berlin: de Gruyter.

Mueller, U. 2000. Die Maßzahlen der Bevölkerungsstatistik. In: U. Mueller, B. Nauck, A. Diekmann (Hg.), Handbuch der Demographie, Bd. 1, 1 – 91. Berlin: Springer-Verlag.

Mueller, U., Nauck, B., Diekmann, A. (Hg.) 2000. Handbuch der Demographie. Berlin: Springer-Verlag.

Namboodiri, K., Suchindran, C. M. 1987. Life Table Techniques and their Applications. New York: Academic Press.

Newell, C. 1988. Methods and Models in Demography. New York: Guilford Press.

Olkin, I., Gleser, L. J., Derman, C. 1980. Probability Models and Applications. New York: Macmillan Publ.

Pfeil, E. 1967. Der Kohortenansatz in der Soziologie. Ein Zugang zum Generationsproblem? Kölner Zeitschrift für Soziologie und Sozialpsychologie 19, 645 – 657.

Pohl, K. 1995. Design und Struktur des deutschen FFS. Materialien zur Bevölkerungswissenschaft, Heft 82a. Wiesbaden: Bundesinstitut für Bevölkerungsforschung.

Porst, R. 1996. Ausschöpfungen bei sozialwissenschaftlichen Umfragen. Die Sicht der Institute. ZUMA-Arbeitsbericht 96/07. Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Pressat, R. 1972. Demographic Analysis. Methods, Results, Applications. Transl. from French by J. Matras. Foreword by N. Keyfitz. Chicago: Aldine & Atherton.

Proebsting, H. 1984. Entwicklung der Sterblichkeit. Wirtschaft und Statistik, Heft 1, 13 – 24, 438* – 440*.

Richards, E. G. 1998. Mapping Time. The Calendar and its History. Oxford: Oxford University Press.

Riley, M. W. 1986. Overview and Highlights of a Sociological Perspective. In: A. B. Sørensen, F. E. Weinert, L. R. Sherrod (eds.), Human Development and the Life Course: Multidisciplinary Perspectives, 153 – 175. Hillsdale: Lawrence Erlbaum Ass.

Rinne, H. 1996. Wirtschafts- und Bevölkerungsstatistik. 2. Aufl. München: Oldenbourg.

Rives, N. W., Serow, W. J. 1984. Introduction to Applied Demography. London: Sage.

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.

Roloff, J., Dorbritz, J. (Hg.) 1999. Familienbildung in Deutschland Anfang der 90er Jahre. Ergebnisse des deutschen Family and Fertility Survey. Opladen: Leske + Budrich.

Rosow, I. 1978. What Is a Cohort and Why? Human Development 21, 65 – 75.

Rückert, G.-R. 1975. Zur Bedeutung der Veränderungen der Geburtenabstände in der Bundesrepublik Deutschland. Zeitschrift für Bevölkerungswissenschaft 1, 85 – 93.

Russell, B. 1996. The Principles of Mathematics (first edition 1903). London: Norton & Comp.

Ryder, N. B. 1964. The Process of Demographic Translation. Demography 1, 74 – 82.

Ryder, N. B. 1965. The Cohort as a Concept in the Study of Social Change. American Sociological Review 30, 843 – 861.

Ryder, N. B. 1968. Cohort Analysis. International Encyclopedia of the Social Sciences. Vol. 2, 546 – 550.

Samuelson, P. A. 1952. Economic Theory and Mathematics – An Appraisal (with Discussion). American Economic Review, Papers and Proceedings, 56 – 73.

Schepers, J., Wagner, G. 1989. Soziale Differenzen der Lebenserwartung in der Bundesrepublik Deutschland – Neue empirische Analysen. Zeitschrift für Sozialreform 35, 670 – 682.

Schimpl-Neimanns, B., Frenzel, H. 1995. 1-Prozent-Stichprobe der Volks- und Berufszählung 1970 – Datei mit Haushalts- und Familiennummern und revidierter Teilstichprobe für West-Berlin. Dokumentation der Datenaufbereitung. Mannheim: ZUMA-Technischer Bericht T95/06.

Schmid, C. 1993. Der Zugang zu den Daten der Demographie. ZUMA-Arbeitsbericht 93/07. Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Schmid, C. 2000. Zugang zu den Daten der Demographie. In: U. Mueller, B. Nauck, A. Diekmann (Hg.), Handbuch der Demographie, Band 1, 476 – 523. Berlin: Springer-Verlag.

Schubnell, H. 1973. Der Geburtenrückgang in der Bundesrepublik Deutschland. Schriftenreihe des Bundesministers für Jugend, Familie und Gesundheit, Band 6. Bonn – Bad Godesberg: Bundesminister für Jugend, Familie und Gesundheit.

Schubnell, H. , Herberger, L. 1970. Die Volkszaehlung am 27. Mai 1970. Wirtschaft und Statistik 22, Heft 4, 179 – 185.

Schütz, W. 1977. 100 Jahre Standesämter in Deutschland. Kleine Geschichte der bürgerlichen Eheschließung und der Buchführung des Personenstandes. Frankfurt: Verlag für Standesamtswesen.

Schwarz, K. 1964. Allgemeine Sterbetafel für die Bundesrepublik Deutschland 1960/62. Wirtschaft und Statistik, Heft 7.

Schwarz, K. 1973. Veränderung der Geburtenabstände und Auswirkungen auf die Geburtenentwicklung. Wirtschaft und Statistik, Heft 11, 638 – 641.

Schwarz, K. 1974. Die Frauen nach der Kinderzahl. Ergebnis der Volkszählung am 27. Mai 1970. Wirtschaft und Statistik 26, Heft 6, 404 – 410.



Shryock, H. S., Siegel, J. S. 1976. The Methods and Materials of Demography. Condensed Edition by E. G. Stockwell. New York: Academic Press.

Oaks: Sage.

Sivamurthy, M. 1982. Growth and Structure of Human Population in the Presence of Migration. New York: Academic Press.

Statistisches Bundesamt 1972. Bevölkerung und Wirtschaft 1872 – 1972. Herausgegeben anläßlich des 100 jährigen Bestehens der zentralen amtlichen Statistik. Wiesbaden: Kohlhammer.

Statistisches Bundesamt 1985. Bevölkerung gestern, heute und morgen. Bearbeitet von Helmut Proebsting. Wiesbaden: Kohlhammer.

Tuma, N. B., Huinink, J. 1990. Postwar Fertility Patterns in the Federal Republic of Germany. In: K. U. Mayer, N. B. Tuma (eds.), Event History Analysis in Life Course Research, 146 – 169. Madison: University of Wisconsin Press.

United Nations 1958. Multilingual Demographic Dictionary. Prepared by the Demographic Dictionary Committee of the International Union for the Scientific Study of Population. English Section. New York: United Nations, Department of Economic and Social Affairs.

Wagner, M. 1996. Lebensverläufe und gesellschaftlicher Wandel: Die westdeutschen Teilstudien. ZA-Informationen 38, 20 – 27.

Wagner, M. 2001. Kohortenstudien in Deutschland. In: Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hg.), Wege zu einer besseren informationellen Infrastruktur. Baden-Baden: Nomos.

White, A. R. 1975. Modal Thinking. Oxford: Basil Blackwell.

Winkler, W. 1960. Mehrsprachiges demographisches Wörterbuch. Deutschsprachige Fassung bearbeitet auf der Grundlage der von einer Wörterbuchkommission der Union Internationale pour L'Etude Scientifique de la Population erstellten und von den Vereinten Nationen veröffentlichten französischen, englischen und spanischen Ausgaben. Universität Hamburg: Deutsche Akademie für Bevölkerungswissenschaft.

Würzberger, P., Störtzbach, B., Stürmer, B. 1986. Volkszählung 1987. Rechtliche Grundlagen nach dem Urteil des Bundesverfassungsgerichts vom 15. Dezember 1983. Wirtschaft und Statistik, Heft 12, 927 – 957.

Wunsch, G. J., Termote, M. G. 1978. Introduction to Demographic Analysis. Principles and Methods. New York: Plenum Press.

Young, C. M. 1978. Cohort Analysis of Mortality – An Historical Survey of the Literature. Working Papers in Demographie No. 10. Department of Demography, Research School of Social Sciences, Australian National University, Canberra.

# Name Index

# Subject Index