**Research and Development** – Methodology reports from Statistics Sweden 2009:2

# Demographic Methods for the Statistical Office

*Michael Hartmann*

Statistiska centralbyrån Statistics Sweden

The series entitled "**Research and Development – Methodology Reports from Statistics Sweden**" presents results from research activities within Statistics Sweden. The focus of the series is on development of methods and techniques for statistics production. Contributions from all departments of Statistics Sweden are published and papers can deal with a wide variety of methodological issues.

Previous publication:

2006:1 Quantifying the quality of macroeconomic variables

2006:2 Stochastic population projections for Sweden

2007:1 Jämförelse av röjanderiskmått för tabeller

2007:2 Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator.

2007:3 Kartläggning av felkällor för bättre aktualitet

2008:1 Optimalt antal kontaktförsök i en telefonundersökning

2009:1 Design for estimation: Identifying auxiliary vectors to reduce nonresponse bias

.

**Research and Development** – Methodology reports from Statistics Sweden 2009:2

# Demographic Methods for the Statistical Office

*Michael Hartmann*

Statistiska centralbyrån
2009

**Research and Development** – Methodology reports from Statistics Sweden 2009:2

# Demographic Methods for the Statistical Office

Statistics Sweden
2009

Producer Statistics Sweden, Research and Development Department
SE-701 89 ÖREBRO
+ 46 19 17 60 00

Inquiries Anders Ljungberg, +46 8 506 946 15
anders.ljungberg@scb.se



Cover Ateljén, SCB

URN:NBN:SE:SCB-2009-X103BR0902_pdf (pdf)
*This publication is only published electronically on Statistics Sweden's website www.scb.se*

## Preface

This book combines lecture notes prepared by the author that over the years have been used in Statistics Sweden for in-service training in demographic methods. They have also been used on several ICO projects as well as in the Demographic Unit in the University of Stockholm. The focus is on both practical and theoretical issues. Several numerical examples are given. The book is written in a coherent language addressing readers with a novice background in demography and with only a limited background in mathematical statistics. The book is dedicated to those who are about to begin their exploratory journey into the world of population statistics.

The author wishes to acknowledge with gratitude the comments from referees and others who have assisted in the project.

Statistics Sweden 2009

Folke Carlsson

Stina Andersson

# Content

# Summary

This publication outlines a selection of demographic methods commonly used in statistical offices. It covers both elementary and advanced methodologies. Several numerical illustrations are given. Chapter 1 gives a brief historical introduction to demography. It is noted that demography is the study of statistics on mortality, fertility and migration, and that these statistics usually derive from population censuses, the vital registration system, special surveys and registers. Chapter 2 discusses basic statistical principles such as probability, random variables, statistical means and expectations, variances and correlation. Then, in chapter 3, the important notion of the rate is discussed. Chapter 4 illustrates the use of the Lexis diagram, and how it can be used e.g., to visualize a stationary population. Measurement of mortality (the life table) is discussed in chapter 5. In support of survey taking, this chapter also discusses the sampling variance of the life expectancy. In addition, it outlines how to conduct simple simulations. Stationary and stable populations are discussed in chapter 6. Chapter 7 discusses standardization of mortality. Measures of fertility and reproduction are discussed in chapter 8. This chapter also discusses how the variance of the total fertility rate can be determined by means of simulations or by application of a large-sample approximation. Chapter 9 discusses ordinary migration statistics. Chapter 10 is devoted to population projections and illustrates the notion of structural effects. Chapter 11 gives an elementary introduction to time-series and forecasting. Chapter 12 discusses demographic models such as the Brass logit survival and the Lee-Carter methods. Indirect demographic estimation techniques are discussed and illustrated in chapter 13. Chapter 14 outlines logistic regression. Finally, chapter 15 gives a short introduction to differentiation and integration.

# 1.0 Introduction

## 1.1 John Graunt

The first scientific study of population took place in England during the era of mercantilism, a European 17th century economic doctrine primarily concerned with maximizing national wealth. National wealth was mainly perceived as bouillon (gold) reserves. Mercantilism involved, among other things, that the nation state should maintain trading monopolies, especially with respect to precious metals. Mercantilism opposed free trade and advocated high reproduction ensuring a steadily growing labor force.

John Graunt, the founder of population studies, was born on April 24, 1620. He was a wealthy and influential businessman. He is referenced in Samuel Pepys' diaries. John Aubrey (who authored several contemporary bibliographies) was familiar with Graunt and wrote his bibliography (Glass *in* Benjamin, Brass and Glass, 1963, pp. 2-37).

Europe during the 16th and 17th centuries was affected by numerous plague epidemics that killed hundreds of thousands of people. Beginning toward the end of the 16th century listings giving the number of deceased persons in London were issued to the public. These were known as the bills of mortality. The bills of mortality, along with listings of christenings for a population of about half a million London souls, were the main source of Graunt's research.

In 1662 John Graunt presented his study to the Royal Society[1] and was immediately elected a member. His study was entitled "Natural and Political Observations on the Bills of Mortality". There is considerable innovation in Graunt's research (Smith and Keyfitz, 1977, pp. 11-21). Despite the fact that his data were no more than rudimentary, he established several demographic regularities e.g., that for about every 205 live births 100 are girls and 105 boys. He was close to estimating a life table, a considerable feat at that time. He established a system for classifying causes of death. Moreover, he found a way of correcting birth estimates for underregistration of births (see e.g., Brass' comments in Benjamin, Brass and Glass, 1963,

[1] Founded in 1640, it is the oldest continuing scientific society in the world.

p. 65). John Graunt died a poor man on April 18, 1674. He was buried under the piewes (alias hoggsties), Aubrey wrote (Glass in Benjamin, Brass and Glass, 1963, p. 6).

## 1.2 William Petty

Sir William Petty (1623-1687) was a close and life-long friend of John Graunt's. His educational background was in medicine but it was not in this area that he made his main contributions. While Graunt is recognized as the founding father of population studies, Petty stands recognized as the founder of political science. Among his major works are "A Treatise on Taxes and Contributions" and several pioneering essays. He took a great deal of interest in reforming education. He worked as a surveyor and drew several maps of England and Ireland. In 1650 with the help of John Graunt he became professor of music at Gresham College.

Petty coined the expression "Political Arithmetick" for population and economic studies. He was a strong advocate of mercantilism. It was not until 1855 that the Frenchman Guillard introduced the designation *Demographie* (Duncan and Hauser, 1972, p. 158). Petty recommended that every nation should have a statistical office providing government with data for intelligent governance. Although the beginning of demography is 17$^{th}$ century England, Sweden was the first country to implement systematic collection of population data. This was accomplished by a royal decree in 1748 which led to the creation of a statistical office in Stockholm in 1749.

Durand (Bogue, 1969, p. 9) in commenting on the work of Petty writes:

*"It is remarkable how many questions Petty tackled, with which demographers and statisticians are still wrestling today, particularly in studies of the problems of under-developed countries. Among other things, he was concerned with population projections, the economics of urbanization, population structure and the labor force, unemployment and under-employment, and the measure of national income."*

Both Graunt and Petty responded to the ideas of their time. There was in the first place a strong belief in mercantilism that called for government to be knowledgeable about society's wealth-producing capacity. In the second, there was the new outlook that science should be based on observations (empiricism). The discoveries of Galileo Galilei (1564-1642) in Pisa laid the foundation of modern science. Galileo became known as the father of modern observation-al astronomy and physics. Galileo was the first to show that heavy and light objects are accelerated equally much by gravitation (save for air resistance, heavy and light objects fall to the ground equally fast). His contributions include the discovery of the four largest satellites of Jupiter, named the Galilean moons in his honor, and the observation and analysis of sunspots. The discoveries of Galileo inspired Isaac Newton (1643-1728) to establish the fundamental laws of physics. Isaac Newton and Wilhelm von Leibniz (1646-1716) created the modern mathematics of the 17$^{th}$ century (integral and differential calculus). In addition, as we have seen, the 17$^{th}$ century also initiated the pedigree of social science.

## 1.3 What is demography?

Ordinarily one would not explain the meaning of physics or chemistry. We have an adequate understanding of what the physical sciences deal with. Demography however spans several areas of intertwined academic disciplines. For this reason it is desirable to clarify its areas of study. Duncan and Hauser (1972, p. 2) write:

*"Demography is the study of the size, territorial distribution, and composition of population, changes therein, and the components of such changes, which may be identified as natality, mortality, territorial movement (migration), and social mobility (change of status). Three features of this definition merit brief explanation. First, the omission of reference to population "quality" is deliberate, to avoid bringing normative considerations into play. "Population composition" encompasses consideration of variation in the characteristics of a population, including not only age, sex, marital status, and the like, but also such "qualities" as health, mental capacity, and attained skills or qualifications. Second, interest in "social mobility" is made explicit because population composition changes through movements by individuals from one status to another, e.g., from "single" to "married," as well as through natality, mortality, and migration. Third, the term "territorial movement" is preferred to "migration" because the latter ordinarily applies to movements from or to arbitrarily defined areal units rather than to the totality of movements."*

In the past, the fundamental materials for demographic studies were censuses and vital registration. After World War II, sample surveys have been used with increasing success to provide the data for demographic studies. Today, the main bulk of useful and insight providing demographic data come from surveys.

## 1.4 Thomas Robert Malthus and Karl Marx

The Reverend Thomas Robert Malthus (1766-1834) is best known for his assertion that "the power of population is greater than the power in the earth to produce subsistence for man." In his "Essay on the Principle of Population" published in 1798 he argued that while a human population grows geometrically (1, 2, 4, 8, 16, etc), the food supply can only grow arithmetically (1, 2, 3, 4, 5, etc). It was his thesis that famine, pestilence, misery and vice prevents (check) the human population from outstripping itself of resources; a population will continue to grow until it is miserable enough to stop its growth.

Malthus advocated moral restraint (birth limitation) through deferred marriage. While, at times, his theories have faded away they have always shown a remarkable tendency to reappear. The population discussion concerning "run-away population growth" in developing countries after World War II was greatly inspired by the views of Malthus. In contrast, Karl Marx (1818-83) was intensely opposed to the Malthusian outlook. The main view in communism was that there is no natural population law; poverty comes about because of inadequate redistribution of wealth and essential resources. The French demographer Sauvy (1969) has written extensively on general theories of population. It was Sauvy who coined the expression the *Third World*. In recent years serious concern has been expressed about the possible environmental degradation that the future world population may come to live with.

## 1.5 Censuses and vital registration

Heretofore, the population census has been a principal source of information for demographic studies. As a result, the main bulk of demographic methods have been designed for use with census data[2]. While e.g., in the Scandinavian countries traditional population censuses have now been replaced by continuous population registers and demographic databases, globally the population census and the vital registration system (supplemented by surveys) remain the major sources for population studies.

[2] Census data are often referred to as cross-sectional because they give a snap shot of the population at the time of the census. Population registers, on the other hand, often permit longitudinal studies that bring to light time changes in demographic variables.

In most parts of the world, vital registration is incomplete and censuses usually suffer from underenumeration and other defects. This means that many of the standard methods of demographic estimation that are commonly used in industrialized countries cannot be applied successfully to the majority of the world population. During the 1960s the British demographer William Brass (1921-99) and his associates developed estimation methods for use with deficient and incomplete demographic data. These became known as methods of indirect demographic estimation[3].

Censuses are taken for several purposes. At the time of the first census in England and Wales in 1801 it was noted by the British parliament that the census served two objectives. The first was to ascertain the number of persons, families and houses and to obtain a broad indication of the occupations in which the people were engaged; the second was to get information that, in the absence of data from a previous [local] enumeration, would enable some view to be formed on the question whether the population was increasing or decreasing; -- interest in population growth was and still remains an important reason for taking population censuses. Sweden took its first census in 1749 and Denmark in 1769. The United States took its first census in 1790 and France its first in 1876. The Russian Empire took its first and only census in 1897.

## 1.6 World population growth and prospects

During the 20th century the world population reached magnitudes never previously attained! About 1900 the world population was less than 2 billion. Around 1950 it had approached some 2.6 billion. In the year 2010 it is estimated to be about 6.8 billion, and by the year 2050 about 9 billion. Nevertheless, historical evidence suggests that no population can increase unabated over long periods; sooner or later its growth will taper. Such changes in growth behavior have been observed on many occasions.

Associated with the increase in world population is a steadily increasing proportion of elderly people. Around 1950 the proportion aged 65 and over in the world population was about 5 percent. By 2050 it is estimated to have increased to about 15 percent. In Europe

---

[3] Originally, it was perhaps not the intention that these methods should be used for many decades. However, because vital registration in many parts of the world continues to be too incomplete for demographic estimation, indirect methods are used as much today as they were during the 1970s.

the percentage aged 65 and over around 1950 was about 8 percent. In 2050 it is estimated to have risen to about 30 percent. Alongside these changes have been increases in life expectancy, especially in industrialized nations. Since the 1950s life expectancies have increased by more than 15 years in the United States and many parts of Europe. Nevertheless, in some countries, especially on the African continent, increases in longevity have been modest (in large measure due to malnutrition and the aids/hiv epidemic). The main global demographic feature during the past decades has been falling fertility. Falling fertility is the leading reason for increasing proportions of elderly people. Moreover, in the future the global labor force is almost certain to be much older than at the moment because increasingly many people will be working after they have reached what we presently determine to be retirement age.

# 2.0 Statistical Concepts

## 2.1 Randomness

Because demography is the statistical study of mortality, fertility, marital status and migration (alluding to the aforementioned definition by Duncan and Hauser) it follows that to undertake demographic studies one should be familiar with statistical concepts and methods. A central concept is randomness. Cramér (1945, pp. 138-139) writes:

*"It does not seem possible to give a precise definition of what is meant by the word "random". The sense of the word is best conveyed by some examples. If an ordinary coin is rapidly spun several times, and if we take care to keep conditions of the experiment as uniform as possible in all respects, we shall find that we are unable to predict whether, in a particular instance, the coin will fall "heads" or "tails". If the first throw has resulted in heads and if, in the following throw, we try to give the coin exactly the same initial state of motion, it will still appear that it is not possible to secure another case of heads. Even if we try to build a machine throwing the coin with perfect regularity, it is not likely that we shall succeed in predicting the results of individual throws. On the contrary, the result of the experiment will always fluctuate in an uncontrollable way from one instance to another. At first, this may seem rather difficult to explain. If we accept a deterministic point of view, we must maintain that the result of each throw is uniquely determined by the initial state of motion of the coin (external conditions, such as air resistance and physical properties of the table, being regarded as fixed). Thus, it would seem theoretically possible to make an exact prediction, as soon as the initial state is known, and to produce any desired result by starting from an appropriate initial state. A moment's reflection will, however, show that even extremely small changes in the initial state of motion must be expected to have a dominating influence on the result. In practice, the initial state will never be exactly known, but only to a certain approximation. Similarly, when we try to establish a perfect uniformity of initial states during the course of a sequence of throws, we shall never be able to exclude small variations, the magnitude of which depends on the precision of the mechanism used for making the throws. Between the limits determined by the closeness of the approximation, there will always be room for various initial states, leading to both the possible final results of heads and tails, and thus an exact prediction will always be practically impossible."*

If “something” is random, this “something”, by definition, cannot be predicted with certainty. Hardly any demographic event can be predicted with certainty; following marriage, we expect the birth of a child, yet we cannot be certain that a married couple will have children. If they do have children, we cannot, in advance, tell how they will fare in life, and so on.

## 2.2 Probability

Imagine a coin that can be tossed a large number of times without damaging its physical constitution. The outcome of tossing the coin is either heads or tails. While it is impossible for us to predict with certainty the outcome of any particular toss, we feel confident in arguing that for a large number of tosses about half will yield heads. In fact, intuition tells us that for e.g., a million tosses, the ratio $p = f/1,000,000$ where f is the corresponding number of heads, for practical purposes, will be the same as the corresponding ratio resulting from 2,000,000 tosses (in both cases, we would expect $p \approx 0.5$).

**Table. 2.1. Frequency of boy and girl births in Poland, 1927-32**

| Year | Boys | Girls | Both sexes | Proportions | |
|---|---|---|---|---|---|
| | | | | Boys | Girls |
| 1927 | 496,544 | 462,189 | 958,733 | 0.518 | 0.482 |
| 1928 | 513,654 | 477,339 | 990,993 | 0.518 | 0.482 |
| 1929 | 514,765 | 479,336 | 994,101 | 0.518 | 0.482 |
| 1930 | 528,072 | 494,739 | 1,022,811 | 0.516 | 0.484 |
| 1931 | 496,986 | 467,587 | 964,573 | 0.515 | 0.485 |
| 1932 | 482,431 | 452,232 | 934,663 | 0.516 | 0.484 |

Source: Fisz, 1963, p. 4.

Stated otherwise, intuitively, if N is the number of tosses and f is the number of heads, the ratio $p = f/N$ will approach a constant when N increases. We call this hypothetical constant the probability of the coin showing heads in a toss. We speak of the *stability of relative frequencies*, a notion upheld by empirical experience.

A perfect coin, that is, a coin where it is as likely that it shows heads as tails when tossed is called a symmetrical coin. In reality, no such coin exists. However, in the real world there are coins that are nearly symmetrical. Now, whether a coin is symmetrical or not, we can imagine that it has attributed to it a number p, $0 \le p \le 1$, where p is the theoretical probability that the coin will yield heads in a toss. Notice that here p is an unobservable or abstract attribute of the coin; for we can never establish with total accuracy what the value of p is. We can, however, toss the coin a large number of times and use the relative frequency of heads as an approximation to p. These arguments provide us with an intuitive understanding of what probability is.

If nature did not uphold the principle of probability, probability calculus would remain an abstract mathematical exercise (likely, it would not even exist). The whole point is that nature, in fact, very much upholds the notion of probability (probabilities are <u>not</u> chaotic). As an example, consider table 2.1 showing the numbers of boy and girl births in Poland between 1927 and 1932. It will be seen that the proportions of boys and girls remain virtually constant (the sex ratio at birth is known as a demographically invariant entity). Hence, we can argue that a non-interrupted pregnancy results in a live born boy with probability 0.52, and a live born girl with probability 0.48. Sex ratios at birth for other countries and periods are virtually the same. The constancy of the sex ratio at birth is an example of *demographic regularity*. Graunt observed this and other regularities using data on christenings for London during the early $17^{th}$ century.

## 2.3 Random experiments

A random experiment, by definition, is an experiment that can be conducted a large number of times under the same conditions. Hence, tossing a die and observing the outcome is a random experiment. One may object that nothing in this world can be repeated under the exact same circumstances. After all, the second time the die is cast, its molecular constitution, as well as other physical characteristics, have changed for which reason the probabilistic mechanisms underlying its outcomes also have changed. Here the answer is that these changes are so minuscule that they evade numerical measurement and, consequently, are of no practical importance. The long and the short of it is that mathematical definitions are abstract; in the real world we deal with approximations. What we take interest in is <u>not</u> total precision (which can never be obtained) but <u>adequate</u> precision which can usually be attained from a practical point of view. Penrose (2005) discusses this issue in more detail.

A realization of a random experiment (such as tossing a coin) is called a trial. The outcome of a random experiment or trial is called an event. It is typical of a random experiment E that it may result in individual events $e_1, e_2, ..., e_n$ . These are called elementary events.

## 2.4 The concept of statistical distribution

Consider a random experiment E with elementary events $e_1, e_2, ..., e_n$. Let $p_k$ be the probability that the event $e_k$, k = 1, … , n, occurs when E is performed. We say that $p_1, p_2, ..., p_n$ is the probability distribution for E. The probabilities $p_k$ share the property that $0 \leq p_k \leq 1$ and that

$$\sum_{k=1}^{n} p_k = 1.$$

As an example, consider the random experiment of throwing a die. The die may show 1, 2, … , 6 dots. Hence, the numbers 1, 2, … , 6 are elementary events. If the die is symmetrical, that is, if any event is as likely as any other, then the probability of each event is $p_k = 1/6$. Notice that when we throw the die, it is a certain event that either we get one, two, three, four, five or six dots. Hence, probabilities across all elementary events sum to unity.

In passing, we introduce the notion of a random variable. Let X be the number of dots resulting from throwing the die[4]. Because X takes on its values randomly, we say that X is a random variable. We can now write P{X=k} = 1/6 for k = 1, … , 6. When events do not influence one another we speak of independent events. For example, in an experiment where a die and a coin are tossed the corresponding outcomes are independent. When we speak of independent observations $x_1, ..., x_n$ it is understood that no observation can influence the value of another. For example, if every observation were proportional to the preceding one, then the observations would <u>not</u> be independent.

## 2.5 The mean value of a random variable

The mean value of a discrete random variable X, which can take on values $x_1, ..., x_n$, with probability distribution $p_1, ..., p_n$ is defined as

$$E(X) = \mu = \sum_{k=1}^{n} x_k \, p_k \qquad (2.1)$$

[4] Mathematically, X is a map from the set of dots on the die {1, … , 6 dots} to the set of integers {1, … , 6}.

As an example, let X be the random variable corresponding to the experiment of throwing a die. If all events 1, … , 6 are equally likely, the mean value of X is

$$E(X) = \mu = \sum_{k=1}^{6} k \frac{1}{6} = 3.5$$

We also refer to (2.1) as the expected value of the random variable X. The mean or expected value of a random variable is so important that it is in place to discuss it in more detail.

Consider a game where a die is thrown and you get as many dollars as the number of resulting dots; if the throw results in five dots, you get five dollars. The die is tossed and we ask, how many dollars do you expect to receive? The answer is $ 3.5. How did we arrive at that result? Perhaps by saying that you can get either one, two, three, four, five or six dots when the die is thrown and that each outcome (event) is equally likely. In other words, on average, you expect to receive $\frac{1}{6}(1+2+3+4+5+6) = 3.5$ dollars. This, indeed, is the expected value calculated in agreement with (2.1).

If we ask how many out of 635 newborns are expected to die during infancy, given that the probability of infant death is 0.035, the answer is 0.035 x 635 = 22 infants. A coin is tossed 737 times. How many heads do you expect? Answer: 0.5 x 737 = 369. Stated otherwise, expectations are often found by simple multiplication. What may seem peculiar is that the expected value often is such that it does not match any particular outcome. For example, you throw a die and calculate the expectation to be 3.5. Yet, no outcome gives this number of dots. Moreover, individual outcomes may vary a great deal from their expectation. To better understand this feature, we now introduce the notion of the variance of a random variable.

## 2.6 The variance of a random variable

The variance of a discrete random variable X, which can take on values $x_1,...,x_n$, with probability distribution $p_1,...,p_n$, is defined as

$$\mathrm{Var}(X) = \sigma^2 = E(X - \mu)^2 = \sum_{k=1}^{n} (x_k - \mu)^2 p_k \qquad (2.2)$$

Notice that the variance of a random variable is the expectation of its squared deviation from its mean. The square root of the variance is called the standard deviation. Hence, the standard deviation of X is

$$sd(X) = \sigma = \sqrt{\sum_{k=1}^{n} (x_k - \mu)^2 p_k} \quad (2.3)$$

Given observations $x_1, x_2, ..., x_n$, the mean of their distribution is estimated as

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} x_k \quad (2.4)$$

that is, as a straightforward average of the observations. The variance of their distribution is estimated as

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \hat{\mu})^2 \quad (2.5)$$

Occasionally, we write

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \hat{\mu})^2$$

for the estimated variance. For large n, the two estimates of the variance are the same.

## 2.7 Covariance and correlation

Let X and Y be two random variables with means $m_x$ and $m_y$, respectively. The expected value

$$\text{Cov}(X, Y) = E[(X - m_x)(Y - m_y)] \quad (2.6)$$

is called the covariance of X and Y. From this definition, we can infer that a positive covariance means that when X is above its mean value then, likely, Y is also above its mean value. Similarly, if X is below its mean value then, likely, Y is also below its mean value. If the covariance is negative then there is a tendency for Y to be smaller than its mean value when X is higher than its mean value or, alternatively, when X is below its mean value then Y is likely to be above its mean value.

A more convenient measure is the covariance between X and Y following standardization. To standardize a random variable, we subtract its mean and divide by its standard deviation. Hence $u = (X - m_x)/\sigma_x$ where $\sigma_x$ is the standard deviation of X is the standardization of X. Similarly, $v = (Y - m_y)/\sigma_y$ is the standardization of Y. The expectation

$$\rho(X, Y) = E\left[\frac{(X - m_x)(Y - m_y)}{\sigma_x \sigma_y}\right] = E(u\, v) \qquad (2.7)$$

is called the correlation between X and Y. The correlation is always such that $-1 \le \rho \le 1$. Given paired observations $(x_1, y_1), \ldots, (x_n, y_n)$, the estimated correlation is

$$r(x, y) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{m}_x)(y_i - \hat{m}_y)}{\hat{\sigma}_x \hat{\sigma}_y} \qquad (2.8)$$

Table 2.2 shows ten hypothetical observations on two related random variables X and Y. The estimated means are $\hat{m}_x = 5.57$ and $\hat{m}_y = 6.91$. The products $(x_i - \hat{m}_x)(y_i - \hat{m}_y)$ are calculated. The sum of the products divided by the number of observations is $\frac{1}{10}\sum_{i=1}^{10}(x_i - 5.57)(y_i - 6.91) = 8.58$. The estimated standard deviations are $\hat{\sigma}_x = 2.854$ and $\hat{\sigma}_y = 3.516$. Using (2.8), we find that the estimated correlation between X and Y is r = 0.85.

In practice, we do not calculate means, variances and the like manually. These are tasks that we delegate to statistical software packages. Nevertheless, it is instructive to carry out the calculations manually because it gives us a much better understanding of the mechanisms involved than by pushing buttons on a panel.

About correlations, Bernard Shaw wrote (Hald, 1962, p. 21):

*"… it is easy to prove that the wearing of tall hats and the carrying of umbrellas enlarges the chest, prolongs life, and confers comparative immunity from disease; for the statistics shew that the classes which use these articles are bigger, healthier, and live longer than the class which never dreams of*

*possessing such things. It does not take much perspicacity to see what really makes this difference is not the tall hat and the umbrella, but the wealth and nourishment of which they are evidence, and that a gold watch or membership of a club in Pall Mall might be proved in the same way to have the like sovereign virtues."*

**Table 2.2. Covariance and correlation calculation**

| i | $x_i$ | $y_i$ | $(x_i - 5.57)(y_i - 6.91)$ |
|---|---|---|---|
| 1 | 1.10 | 1.40 | 24.630 |
| 2 | 2.30 | 2.20 | 15.402 |
| 3 | 3.10 | 4.60 | 5.706 |
| 4 | 4.60 | 5.60 | 1.271 |
| 5 | 5.50 | 7.00 | -0.006 |
| 6 | 6.20 | 8.10 | 0.750 |
| 7 | 6.90 | 11.00 | 5.440 |
| 8 | 7.00 | 7.20 | 0.415 |
| 9 | 9.00 | 11.00 | 14.029 |
| 10 | 10.00 | 11.00 | 18.119 |
| Means | 5.57 | 6.91 | 8.58 |

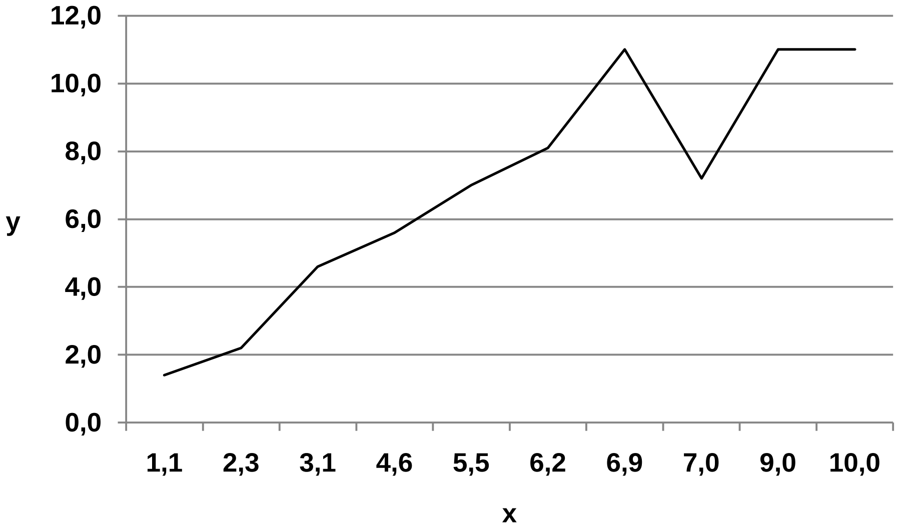
Fig. 2.1 shows Y plotted as a function of X. We see that as X increases so does Y, but only in the sense of an average or trend.

$\hat{\sigma}_x = 2.854$

$\hat{\sigma}_y = 3.516$

$$r = \frac{8.58}{2.854 \text{ x } 3.516} = 0.85.$$

**Fig. 2.1. Relationship between variables x and y**




When the correlation between two variables is r = ±1, the two variables are tied to one another in a perfect linear relationship. The correlation between X and X (that is, with itself) is r = 1. The correlation between X and –X is r = –1. This is easily verified by replacing $Y - m_y$ by $\pm(X - m_x)$ in (2.7).

## 2.8 Probability theory, random experiments and hidden variables

Textbooks on probability theory usually discuss random experiments by alluding to the toss of a coin or a die. To the social science student these examples often appear puerile; after all, social science is hardly the question of tossing a coin or throwing a die. In science the data we deal with derive from the both enigmatic and complex mechanisms that make up nature. Certainly, whether a girl baby just born eventually shall reach adulthood, marry and have three children, of which one is a girl, is a process (biological, sociological, psychological, etc) the complexities of which are far beyond that of throwing a die or a coin.

Albert Einstein (1879-1955) said that *"God does not play dice"*. To Einstein, nature was deterministic in the sense that when we fail to make perfect predictions it is because of the existence of *hidden variables* over which we have no control.

In social science, the general belief has always been, I believe, that if only we know *more* then we can make safer predictions. This however is a belief that often stands contradicted by empirical evidence. For example, consider two coin-tossing experiments where (i) the coin is dropped from a height of one meter and (ii) the coin is dropped from a height of two meters. If the two experiments are carried out with similar coins, the two resulting observation series become statistically indistinguishable. In fact, we get two series of independent binomially distributed observations with the same probability of success. From these experiments we would conclude that knowing the height of the experiment does not improve the certainty with which we predict the next outcome of tossing a coin. If rather than letting height vary between the two experiments we let air temperature vary, the same result would be obtained, that is, knowing the air temperature does not improve the certainty with which we predict the next outcome of the experiment. Continuing in this manner using many different variables that can be measured accurately, we would find that even if, in a sense, they are associated with the experiments, they are of no use for improving the certainty of the predictions. In the case of coin tossing the variables that steer the outcome do not seem possible to find; they are hidden to us. When population forecasts fail (which they do most of the time, especially in the long run!), it is because we have incomplete understanding of the processes that underlie the temporal unfolding of the population (failure in finding the hidden variables).

# 3.0 The Rate

## 3.1 The meaning of a rate

The rate is a concept closely linked to time. A simple example may illustrate this. If it is known that the rate for staying in a hotel is $ 100 per night, then having spent one night in the hotel, room charges are $ 100. Having spent two nights charges are $ 200. Stated otherwise, for every 24 hours spent in the hotel, the guest pays $ 100. If the hotel is fair to a customer who has stayed in the hotel for 58 hours, it would charge 58/24 times one hundred dollars = $ 242.

If rather than speaking of time spent in the hotel, we speak of exposure time (exposure to paying $ 100 for every 24 hours spent in the hotel), we have that room charge is exposure time multiplied by hotel rate. This is a simple enough explanation of what we mean by rate. To be more specific, let us narrow down what we mean by exposure time. To this end let us also introduce the notion of observation plan. Clearly, to make observations we must have a plan that stipulates how this is accomplished.

Consider an observation plan where we study the survival of ten newborns, born at the same time, during their first year of life. After a year of observation, we note that nine children survived and that one died exactly one week after it was born. The total exposure time lived by the nine surviving newborns during the period of observation is 9 years. The time lived by the infant that died is 1/52 of a year or 0.019 year. The total exposure time, therefore, is 9.019 years. We now reason that it required 9.019 years to bring about one infant death for which reason we expect 1/9.019 = 0.1109 deaths per year of exposure during infancy. If we formalize this reasoning, we arrive at the following definition of a rate: A rate is the number of events divided by the amount of exposure time that yielded the events (the speed with which the events took place). Therefore, if subject to an observation plan we study an event A and this occurs D times during the observation plan while, at the same time, an exposure R is consumed, then

$$\mu = \frac{D}{R} \qquad (3.1)$$

is the rate for the occurrence of the event A. Notice that this yields that

$$D = \mu R \tag{3.2}$$

so that the number of events we expect to occur given an exposure R is D. We can use (3.2) to help us calculate the room charges mentioned above. If charges are $ 100 per 24 hours, then room charges are $ 4.17 per hour. Having spent 58 hours in the hotel, D = 4.17 times 58 = $ 242 would be the total charges.

## 3.2 The relationship between rate and probability

Consider a longitudinal observation plan whereby we follow 1,000 newborns during infancy, that is, for a period of one year. At the end of the observation plan we note that 970 babies survived to age one and that 30 died during infancy. For simplicity, assume that the 30 newborns who died lived for half a week on the average. This means that the exposure time consumed by the newborns who died is $30\frac{1}{2}\frac{1}{52}$ = 0.29 years. The total exposure time therefore is 970.29 years. The infant mortality rate is estimated as

$$\hat{\mu} = \frac{30}{970.29} = 0.031 \tag{3.3}$$

On the other hand, out of 1,000 newborns 30 of them died during infancy so that the probability of dying during infancy is estimated as

$$\hat{q} = \frac{30}{1,000} = 0.030 \tag{3.4}$$

These two estimates are nearly the same, but they are not identical; for in the case of the rate (3.3) the denominator is 970.29 while in the case of the probability (3.4) the denominator is 1,000.

It is important to give some thought to the assumptions that underlie estimation of the infant mortality rate (3.3) and the probability of death (3.4). Remember that when we discussed random experiments, we argued that the probability of an event happening should remain the same during the trials. The same applies to (3.4). If we wish to estimate the probability of death during infancy for the 1,000 children under observation it is implicit that we assume that all children share the same theoretical or underlying probability of death during infancy (it is this unobservable probability we wish to estimate). The same applies to (3.3). When we estimate a rate of infant death, we assume that all the children under observation share the same theoretical or underlying rate. When a rate stays constant on an age interval, we say that it is piecewise constant.

It can be shown that when the rate of infant mortality is constant (constant across the first year of life), the corresponding probability of infant death is

$$q = 1 - e^{-\mu} \quad (3.5)$$

Notice that power series expansion implies that $q \approx \mu - \mu^2/2 \approx \mu$ for small μ. If in (3.5) we let $\mu = \frac{30}{970.29} = 0.030919$ then q = 0.030 (disregarding rounding). We shall justify the formula (3.5) later on.

## 3.3 Standard terminology and notation

In demography, a probability such as q in (3.4) is called a mortality rate. In actuarial literature[5], the rate in (3.3) is usually called a central death rate, age-specific mortality rate, mortality intensity or hazard[6]. Terminologies vary between authors. For the most part, we shall call q in (3.4) a mortality rate (although it is a conditional probability) and we shall call μ a central death rate. In addition, it should be noted that it is wise to distinguish between theoretical quantities and estimated ones. In a concrete situation of estimation, we may write $\hat{\mu} = \hat{D}/\hat{R}$ for the estimated rate and $\mu = D/R$ for the theoretical or underlying one. It is true that this may seem somewhat pedantic and, all authors do not make this distinction. It should also be mentioned that in demographic literature the expression person-years means exposure time.

## 3.4 Crude birth and death rates, and the rate of natural growth

The crude birth rate is defined as the total number of live births that have occurred during a calendar year divided by the midyear population for this calendar year. The midyear population is the population as of July 1 (hypothetically). Similarly, the crude death rate is defined as the total number of deaths that have occurred during a calendar year divided by the midyear population. Both crude rates,

[5] Actuarial literature addresses analysis of mortality in the context of life insurance and other kinds of insurance.

[6] It is also known as the force of mortality.

it will be noted, are calculated for both sexes. When we speak of the midyear population, we refer to the resident population at the middle of the year. This population times one year is an approximation to the person-years (exposure time) associated with the total number of births and deaths. Hence, the crude birth rate is

$$CBR = \frac{B}{P} \quad (3.6)$$

where B is the total number of births that took place during a calendar year and P is the corresponding midyear population. The crude death rate is

$$CDR = \frac{D}{P} \quad (3.7)$$

where D is the total number of deaths that occurred during a calendar year and P is the corresponding midyear population. Usually the crude birth and death rates are given per 1,000 population.

The difference between the crude birth rate and the crude death rate is the natural rate of population growth, which is often denoted by r, hence

$$r = CBR\text{-}CDR \quad (3.8)$$

## 3.5 Rate of attrition and longevity

If it is meaningful to attribute a "rate of attrition" to an object then it is also meaningful to attribute to it the notion of longevity. Consider a bottle that contains one liter of water. We are told that the water is tapped at a rate of one tenth of a liter per hour. We ask: When is the bottle empty? The answer is simple enough, after ten hours. Hence, it would appear that the longevity of the bottle is one divided by the rate of water use. This, in fact, is correct. To follow up on this reasoning, let us agree that when water is tapped from the bottle at a certain rate per hour, then it is reasonable to refer to the rate as a "rate of attrition".

Consider an object the rate of attrition of which is a constant m (independent of time). The expected time T required for the object to have been completely destroyed is m T = 1. Hence, T = 1/m is the life expectancy or lifetime of the object. In (3.3) we estimated an infant mortality rate at $\hat{\mu} = 0.031$. If this mortality rate were to apply throughout the life of the child, the child's life expectancy would be $e_0 = 32.3$ years.

As an aside, rate is a word that evidently only exists in English. As noted, it is a measure of change, specifically a measure of how fast something changes with time. In French, Spanish, German, Arabic and the Scandinavian languages there is no exact word for rate. In French demographic literature, a rate is usually defined as the ratio or quotient between two numbers. In demographic literature it is often explained how the rate is calculated but not what is its deeper semantic meaning (it is left for the reader to understand this). In the long run however, it is far more insight inducing to work with definitions that not only explain how something is calculated but also why it is calculated (its rationale). Alas, as is always the case, there are exceptions to be caught. As we shall see, the total fertility rate (the number of children a woman is expected to have if she survives through the reproductive ages and has specified age-specific fertility) is not a rate in the sense discussed in this chapter.

# 4.0 The Lexis Diagram

## 4.1 Its origin

The Lexis diagram is a grid of lines used to illustrate a flow of demographic events. This way of graphically portraying demographic events is often attributed to the German economist and statistician Wilhelm Lexis (1837-1914). Lexis made several important contributions to statistics. He was one of the first to work with time-series. In addition, he was the director of the first actuarial institute in Germany. Two other contemporary German statisticians or economists, Becker and Zeuner, as far as can be gathered, were the main inventors of the diagram to be discussed below. The reason why the diagram is named the Lexis diagram is that it was Wilhelm Lexis who introduced it in his correspondence with American colleagues. It was in the United States that the diagram received its name. Here we use the Lexis diagram as an illustration of infant mortality and the stationary population.

## 4.2 Infant and child mortality

Table 4.1 gives hypothetical data for estimating infant mortality. The data is illustrated by the Lexis diagram in fig. 4.1. For example, notice that of the 998 births that took place in 1970, 18 infants died during this year (1970), and 8 died during the following year (1971). When statistics show how many children died the same year they were born and how many died during the following year, we speak of double-classification (this is particularly the case in French literature). In the case of double-classification two different kinds of estimates can be obtained, namely period and cohort[7] estimates.

Table 4.1 also tells us how many children died during the calendar years. For example, in 1970 the total number of infants dying was 24. The most common definition of infant mortality is

$$\text{IMR} = \frac{\text{number of infant deaths during calendar year}}{\text{number of births during calendar year}}$$

which is known as the infant mortality rate (usually abbreviated IMR). Fig. 4.1 illustrates these figures. An advantage of the diagram

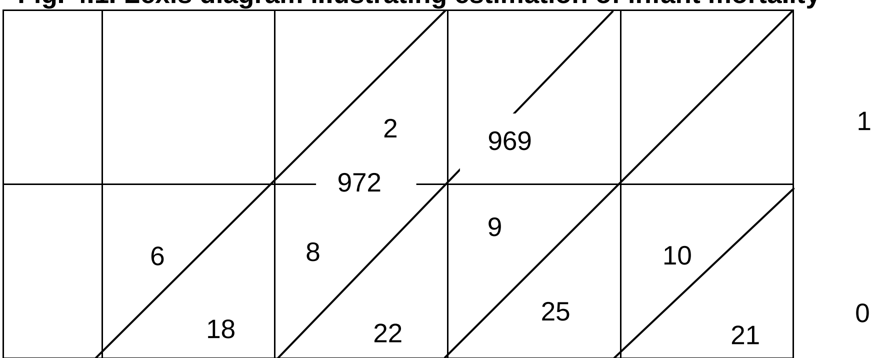
[7] A cohort is the same as a generation.

is that it clearly shows that infant deaths among a birth cohort fall partly in the cohort year, partly during the following year.

**Table 4.1. Infant mortality by double-classification.**

| Calendar year | Total births during year | Died during same year | Died during following year | Total deaths during year | Total deaths in cohort | IMR | Cohort estimate |
|---|---|---|---|---|---|---|---|
| 1970 | 998 | 18 | 8 | 24 | 26 | 0.024 | 0.026 |
| 1971 | 1,022 | 22 | 9 | 30 | 31 | 0.029 | 0.030 |
| 1972 | 1,007 | 25 | 10 | 34 | 35 | 0.034 | 0.035 |
| 1973 | 893 | 21 | na | 31 | na | 0.035 | na |

**Fig. 4.1. Lexis diagram illustrating estimation of infant mortality**




| Births | 998 | 1,022 | 1,007 | 893 |
|---|---|---|---|---|
| Year | 1970 | 1971 | 1972 | 1973 |

In statistical publications infant mortality is usually given as IMR, that is, as the number of infants dying during a calendar year divided by the total number of live births during the same calendar year. This is a simple estimation procedure, indeed often the only one that can be attempted.

When deaths are given both by year of birth of the deceased and by calendar year (double classification), it is possible to get cohort estimates.

Consider the 998 children born in 1970. We notice that 18 of those died in 1970 and 8 in 1971. Hence, for this birth cohort infant mortality was $\hat{q}_0^{1970} = \frac{26}{998} = 0.026$. This estimate is slightly different from

$\hat{IMR}_{1970} = \frac{24}{998} = 0.024$. Notice that while $\hat{q}_0^{1970}$ is a cohort estimate in that it builds on children born during 1970 and how many of them died within one year of life, this is not true of $\hat{IMR}_{1970}$ which blends the deaths from two different birth cohorts, namely the 1969 and 1970 birth cohorts. In that sense, $\hat{IMR}_{1970}$ is a composite statistic[8].

If every calendar year the same number of children is born and if infant mortality $q_0$ stays the same over time, IMR would be the same as $q_0$. In reality, both fertility and mortality change from year to year, hence we shall never expect IMR to be the same as $q_0$ (as estimated from a cohort experience).

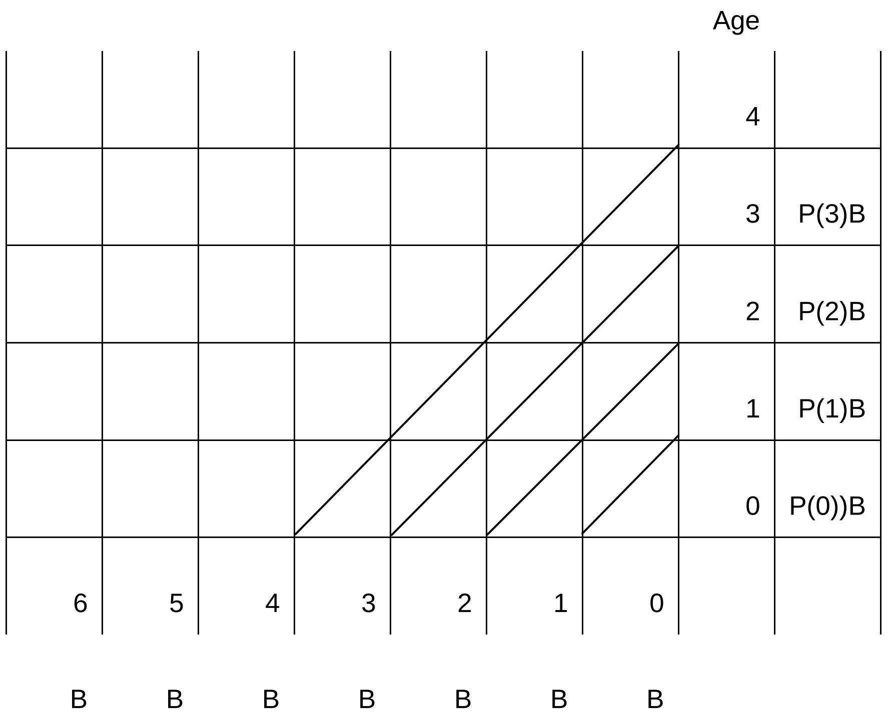
Excellent illustrations of the Lexis diagram are found in "Demographic Analysis" by the French demographer Roland Pressat. The original French version dates back to 1961. For many years, it was considered required reading for every novice student of demography. Because of its elementary and highly non-mathematical approach to explaining demographic methods, it fell in the shadow of more mathematically oriented literature. Nevertheless, even today it remains one of few practical textbooks on demographic techniques, especially for staff in statistical offices. As a novice student of demography, one should consult this textbook which is surprisingly rich in demographic insight. Another recommended textbook on demography is Spiegelman (1980).

We shall return to a discussion of the stationary population and show that it is a special case of what is known as a stable population, a concept widely used in demographic analysis. We shall also return for a discussion of the stationary population and show that its crude birth and death rates are the inverse of its life expectancy.

Fig. 4.3 illustrates a hypothetical situation where each year B children are born. A census is taken in year 0 on December 31. The population aged between 0 and 1 at the end of year 0 is not B because during year 0 some of the children died. Instead, we count P(0) B surviving children aged between 0 and 1 where P(0) is a fraction, $0 < P(0) < 1$.

[8] A statistic is a function of observations. The mean, for example, is a statistic.

# 4.3 The Lexis diagram and the stationary population




If the children are born uniformly during the year then we may expect the children aged between 0 and 1 year to be half a year old, on the average. Similarly, we find that in year 0 the expected population aged between 1 and 2 years is a fraction P(1) times B, and that they are 1.5 years old on the average. We can continue in this fashion as long as we find survivors born in the past. Suppose that the mortality of the population is such that ω is an age with the property that while there are survivors between ages ω and ω+1, there are no survivors at or above age ω+1. Denoting the population aged ω by P(ω) B, the total expected population is

$$T = B \sum_{x=0}^{\omega} P(x) \qquad (4.1)$$

with $0 < P(x) < 1$, x = 0, ... , ω. It is assumed that over time the survival fractions P(x), x = 0, ..., ω remain unchanged. Because P(x) denotes the expected fraction of survivors at the end of year 0 that were born x years ago, it is a reflection of the prevailing mortality conditions. These mortality conditions, as noted, remain unchanged over time.

Because both B and P(x) are constants, it follows that T is a constant. Moreover, for any age x, the proportion aged x, that is, P(x) B/T also remains the same. Stated otherwise, the age distribution of the population remains unchanged over time. We conclude that the yearly number of deaths is D = B; for if D > B the population would decline over time and if D < B the population would increase over time. This means that the crude birth rate (CBR) is the same as the crude death rate (CDR) or B/T = D/T. For this reason it is called a stationary population.

We shall return to a discussion of the stationary population and show that it is a special case of what is known as a stable population, a concept widely used in demographic analysis. We shall also return for a discussion of the stationary population and show that its crude birth and death rates are the inverse of its life expectancy.

# 5.0 Mortality

## 5.1 The life table

Ulpian (ca. 170-228 AD), the Roman jurist who inspired much of the writings on civil law seems to have made use of a rudimentary life table. It is uncertain how it was made and to which uses it was put. Edmund Halley (1656-1742), the astronomer, is usually credited with having estimated the first life table in modern times (the attempt made by Graunt was flawed). He made use of deaths recorded by church books in the German city of Breslau. He ordered the deaths in a manner so that they portrayed diminution due to mortality in a stationary population. About 1780, Richard Price (1771) made use of Swedish data to construct a life table correctly. More well-known than Price's historical table is the Carlisle mortality table from 1815 constructed by the actuary Joshua Milne. For many decades this table served as an important model of human survival. Above all, it was William Farr (1807-83) whose early population studies helped underpin a tradition of census taking and registration of vital events in England and Wales. It was Farr who drew attention to the excessive mortality in certain districts and within certain trades. His work became a foundation for social legislation and in the long run paved the way for the desire to improve mortality conditions, a work that continues to this very day. Dr. Farr held office during the reign of Queen Victoria (Cox, 1970, p. 301).

Life tables building on more than 90 percent complete death registration are predominantly from after World War II. While today we have (reasonably) accurate life tables for the United States, Canada, Australia, New Zealand, member states of the European Union and some other countries, the majority of today's world population is not covered by reliable life tables. A reason for this is that for the past fifty years or so there has been much more preoccupation with fertility (population growth) than with mortality. As a result, only sparse resources have been made available for upgrading vital registration systems in the majority of countries. Conflating incomplete vital registration data with census data of dubious quality necessarily results in poorly estimated life tables.

## 5.2 Age

We write $E_x$ for a number of individuals each of whom is aged exactly x years. We differentiate between (i) exact age and (ii) age last birthday. The very instant a person reaches the xth birthday the person is aged exactly x years. A person aged between exact ages x and x+1 is said to be at age x (such a person is also said to be a life aged x).

## 5.3 The central exposed to risk

The midyear population is what we would obtain were we to conduct a census on July 1. Here it is in place to mention that we operate with two different definitions of population. There is, on the one hand, what is known as the *de jure* population. This is the population that normally resides in the nation (the resident population). On the other hand, there is also the notion of the *de facto* population. This is the population that was present in the nation at the time of the census. These two population counts are not identical. In industrialized countries, the midyear population usually references the resident or *de jure* population, as determined by a population census. Whichever definition of population we make use of, it needs to be mentioned that it only affords an approximation (good or bad) to the hypothetical population we have in mind.

Persons at age x during any calendar year are assumed to be aged exactly $x+\frac{1}{2}$ years on July 1. We denote this population by $E_x^c$ where the upper-index c stands for "central part of the year[9]". We call $E_x^c$ the central exposed to risk. Denoting deaths among persons at age x during the calendar year by $D_x$ and assuming that deaths take place uniformly over time, the number of persons aged exactly x years at the beginning of the calendar year must be approximately

$$E_x = E_x^c + \frac{1}{2} D_x \qquad (5.1)$$

[9] See e.g., Benjamin and Haycocks, 1970.

## 5.4 The mortality rate and the central death rate

The probability for a person aged exactly x years to die before reaching exact age x+1 is called the mortality rate (sometimes mortality risk). The mortality rate is denoted $q_x$. Using (5.1) we have

$$q_x = \frac{D_x}{E_x} = \frac{D_x}{E_x^c + \frac{1}{2}D_x} = \frac{D_x/E_x^c}{1+\frac{1}{2}D_x/E_x^c} = \frac{m_x}{1+\frac{1}{2}m_x} \qquad (5.2)$$

where

$$m_x = \frac{D_x}{E_x^c} \qquad (5.3)$$

is called the central death rate. In many settings, (5.3) is called an age-specific mortality rate. We see that $m_x$ is the number of deaths at age x divided by an approximation to the exposure time lived by the population while at age x. In the literature, a mortality rate is not necessarily the same as a death rate. Relation (5.2) links a mortality rate with the corresponding central death rate. Numerically, this is virtually the same as (3.5).

## 5.5 The survival function

The survival function s(x) is such that s(0) = 1 and s(x) is the probability for a newborn to survive to age x. Suppose we have mortality rates $q_0, q_1, q_2, \ldots, q_r$. Letting $p_x = 1 - q_x$, we have that $p_0$ is the probability for a newborn to survive to age 1, $p_1$ is the probability for a child having reached age 1 to survive to age 2 and, generally, $p_x = 1 - q_x$ is the probability for an individual having reached age x to survive to age x+1. In consequence,

$$s(x) = p_0 \ldots p_{x-1} = \prod_{j=0}^{x-1} p_j \qquad (5.4)$$

is the probability for a newborn to survive to age x. Notice that survival in this sense is a chain process: first the child must survive from birth to age 1, then the child must survive from age 1 to age 2 and so on. At each age x, survival is a binomial trial with probability $p_x = 1 - q_x$ of reaching age x+1.

## 5.6 The number living column of the life table

Out of a cohort of $l_o$ newborns, born at the same time, we expect $l_x = s(x) l_o$ to survive to age x. We call $l_x = s(x) l_o$ the expected number of survivors at age x. In demographic literature $l_x$ is often called the number living column of the life table (Keyfitz, 1968, p. 9). We refer to $l_o$ as the radix of the life table. Usually, radix is $l_o$ = 100,000. Alternatively, we call $l_o$ a synthetic cohort. The number living column models decrementing a hypothetical birth cohort in which there are $l_x = s(x) l_o$ survivors at exact age x. In this population the number of deaths between ages x and x+1 is

$$d_x = l_x - l_{x+1}. \tag{5.5}$$

## 5.7 The number of person-years

In the hypothetical population $l_x = s(x)\ l_o$ (a population made up of expected values), the exposure time consumed by the $l_x$ survivors at age x is approximately

$$L_x = \frac{l_x + l_{x+1}}{2} \tag{5.6}$$

times one year. $L_x$ is said to be the person-years lived by the life table population between ages x and x+1. Notice that $L_x$ depends on $l_o$. It follows from (5.5) and (5.6) that the central death rate in the life table population is $m_x = d_x / L_x$.

## 5.8 The life expectancy at birth

Consider a synthetic cohort consisting of $l_o$ newborns. Between ages 0 and 1 they live $L_0$ years. Between ages 1 and 2 they live $L_1$ years and, generally, between ages x and x+1 they live $L_x$ years. Altogether, then, $l_o$ newborn babies live $L_0 + L_1 + \ldots$ years. This means that, on average, they live

$$e_o = \frac{\sum L_x}{l_o}$$

years. To determine the life expectancy it is necessary to include in the summation person-years $L_x$ until such an age ω, say, that beyond this age there are no more survivors so that the $L_x$, $x \geq \omega+1$, are zero. In consequence,

$$e_o = \frac{1}{l_o} \sum_{x=0}^{\omega} L_x$$

Using the linear approximation (5.6), we get

$$e_o = \frac{1}{2} + \frac{1}{l_o} \sum_{x=1}^{\omega} l_x \tag{5.7}$$

which is the most commonly used expression for calculating the life expectancy at birth. Occasionally (5.7) is refined to take into account that the mean age at death is less than half a year during infancy. Let $a_0$ be the mean age at death for infants, then

$$L_0 = l_1 + a_0 d_0 = l_1 + a_0 (l_0 - l_1) = a_0 l_0 + l_1 (1 - a_0) .$$

In developing nations it is customary to let $a_0 = 0.25$. Coale and Demeny (1966) provide different choices of $a_0$. Such a refinement however is only justified when the survival data are highly reliable and, at any rate, has little or no bearing on the estimated life expectancy.

## 5.9 The remaining life expectancy at age x

A person who has survived to age x has a remaining life expectancy at that age. Using the same reasoning as in the case of the life expectancy at birth, we have that

$$e_x = \frac{1}{l_x} \sum_{t=x}^{\omega} L_t \tag{5.8}$$

is the remaining life expectancy at age x.

## 5.10 The $T_x$ function

In actuarial and demographic literature, it is common to encounter the $T_x$ function. This function is the sum of the $L_x$ at and above age x, that is,

$$T_x = \sum_{t=x}^{\omega} L_t$$

using this function, we get

$$e_x = \frac{T_x}{l_x} \tag{5.9}$$

## 5.11 The life table as a stationary population

The stationary life table population is such that each year $l_o$ children are born and survival s(x) does not change over time. The total midyear population size is constant over time and equals

$$T = \sum_{x=0}^{\omega} L_x = \sum_{x=0}^{\omega} \frac{l_x + l_{x+1}}{2} =$$

$$l_o \sum_{x=0}^{\omega} \frac{s(x) + s(x+1)}{2} = l_o e_o \tag{5.10}$$

Notice also that the number of deaths per calendar year in this population is

$$D = \sum_{x=0}^{\omega} d_x = \sum_{x=0}^{\omega} l_x - l_{x+1} = l_o$$

Hence, the crude birth rate (CBR) is the same as the crude death rate (CDR), namely

$$CBR = \frac{l_o}{T} = \frac{l_o}{l_o e_o} = \frac{1}{e_o} \tag{5.11}$$

According to (5.11), we have that $1/CBR = e_o$ which tells us that the inverse of the crude birth rate in a stationary population is the life expectancy at birth. In the life table population $d_x = l_x - l_{x+1}$,

$$q_x = \frac{d_x}{l_x} \text{ and } m_x = \frac{d_x}{L_x}.$$

In the life table population, the mean age at death is the same as the life expectancy. This is easily shown. The mean age at death for those who die at age 0 is half a year so that the sum of ages for those who die at age 0 is $\frac{1}{2}d_0$. Similarly, those who die at age 1 are, on average, one and a half years old so that $1.5\,d_1$ are the sum of ages for those who die at age 1. Hence, the mean age at death is

$$\frac{\sum_{x=0}^{\omega}(x+0.5)d_x}{l_o} = \frac{1}{l_o}\sum_{x=0}^{\omega}(x+0.5)(l_x - l_{x+1}) =$$

$$\frac{1}{l_o}\left[\frac{1}{2}l_o - \frac{1}{2}l_1 + 1.5l_1 - 1.5l_2 + 2.5l_2 - \ldots\right] =$$

$$\frac{1}{l_o}\left[\frac{1}{2}l_o + l_1 + l_2 + \ldots\right] = \frac{1}{2} + \frac{\sum_{x=1}^{\omega} l_x}{l_o}$$

which we recognize as the life expectancy at birth. Here, for ease of calculation, we have assumed that individuals who die at age x, die at exact age x+0.5; a common approximation.

## 5.12 The abridged life table

It is not always possible to estimate central death rates by single-year ages. After all, this requires efficient registration practices as well as ages being recorded reliably[10]. Remember also that the observation plan usually consists of (i) a population census and (ii) a register of recorded deaths. Hence, exposures are obtained from one source (the population census) while deaths are obtained from another (the vital registration system). This is the most common way of obtaining records for estimation of life tables (it is commonly referred to as the actuarial method). It should be noted, though, that in the case of a longitudinal survey plan where household members are interviewed at the beginning of the survey and re-interviewed later on, this also yields the data for an actuarial estimation approach.

In an attempt to smooth misstatement of age (a common reporting error), life tables are often given for the broad age groups 0, 1-4, 5-9,

[10] In many nations, people do not know their exact birthday. In such cases ages are approximate and often heap at ages ending in 0 or 5 (see Spiegelman, 1980, for an excellent discussion).

10-14, … , 80-84, 85+. This means that central death rates are estimated for these ages. To this end let ${}_nD_x$ denote the deaths, during a calendar year, among persons aged between ages x and x+n. The corresponding central exposed to risk are denoted ${}_nE_x^c$ and the central death rate is

$$ {}_nm_x = \frac{{}_nD_x}{{}_nE_x^c} \tag{5.12} $$

and the mortality rate is

$$ {}_nq_x = n\frac{{}_nm_x}{\left[1+\frac{n}{2}\,{}_nm_x\right]} \tag{5.13} $$

The survival function is

$$ s(x) = (1-q_0)(1-{}_4q_1)...(1-{}_nq_{x-n}) \tag{5.14} $$

and the person-years

$$ {}_nL_x = n\,\frac{l_x + l_{x+n}}{2} \tag{5.15} $$

The life expectation at age x is

$$ e_x = \frac{T_x}{l_x},\ x = 0, \dots \tag{5.16} $$

with

$$ T_x = {}_nL_x + \dots + L_{\omega+} \tag{5.17} $$

## 5.13 The highest age group

To complete the life table we must terminate it at a certain age r, say. To estimate the life expectancy, we must know how many person-years are lived beyond age r by the $l_r$ reaching this age. Consider now the mortality rate at age r+ , namely

$$ m_{r+} = \frac{D_{r+}}{E_{r+}^c} $$

calculated for a calendar year experience. This can be interpreted as the crude death rate in a stationary population where all members are aged r and above. With this interpretation in mind, we have that

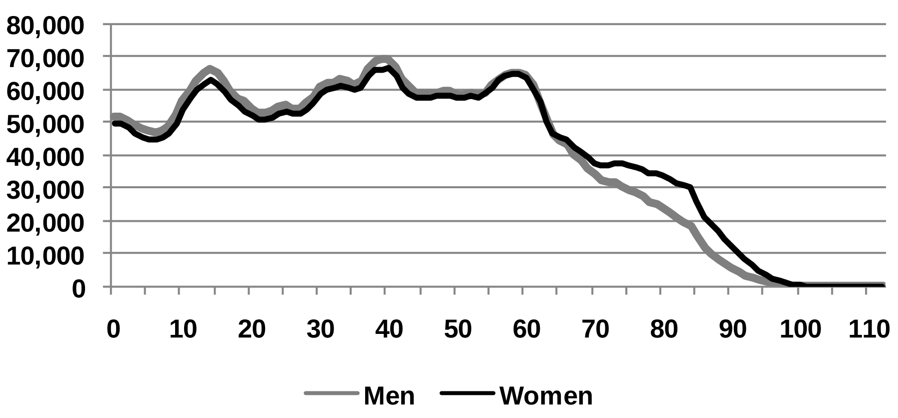
$$e_{r+} = \frac{1}{m_{r+}} \tag{5.18}$$

(see also Section 3.5 in Chapter 3). We then obtain that $L_{r+} = l_r e_{r+}$ ($l_r$ individuals reaching age r are expected to live $l_r e_{r+}$ years).

## 5.14 Graphs of the life table functions

Fig. 5.1 shows the risk populations for men and women in Sweden during 2005. These are the midyear populations. As is always the case with observed populations there are numerous undulations in the data reflecting time changes in mortality, fertility and migration.

**Fig. 5.1. Midyear populations of men and women for Sweden, 2005**




The figure also illustrates the common surplus of women at ages 65+ (the retirement ages). Figures of this nature are helpful for investigating underenumeration in censuses of women; -- we would always expect more females than males at the higher ages.

**Fig. 5.2. Male and female deaths: Sweden, 2005**

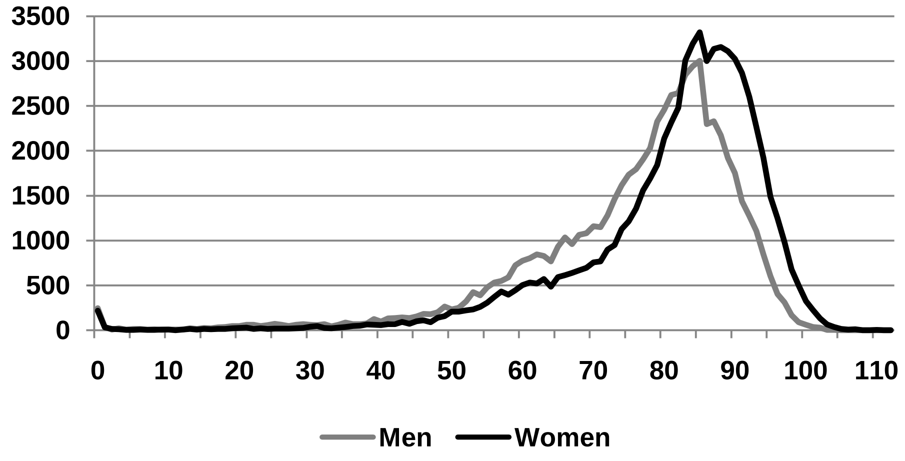



**Fig. 5.3. Male and female mortality risks: Sweden, 2005**

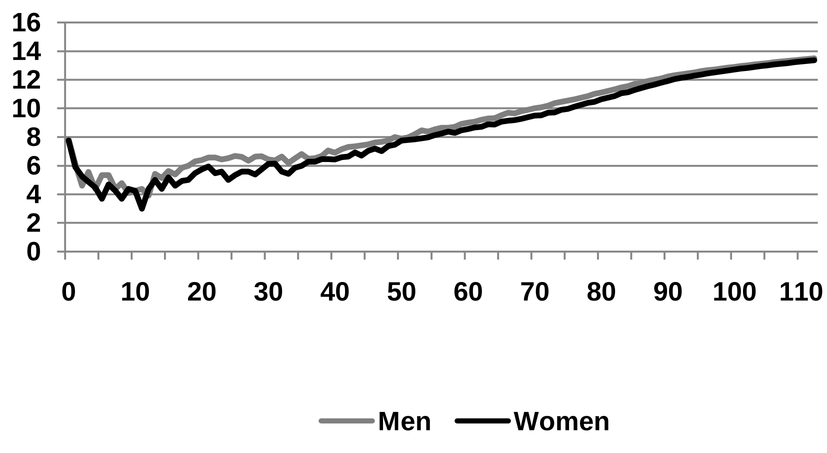



Fig. 5.2 shows deaths by age for males and females in Sweden during 2005. This figure also illustrates the common excess of male deaths over female deaths at ages beginning already at about age 50. Later, as will be seen, there is a surplus of female deaths over male deaths. It is typical of the age distribution of deaths that its peaks at a high age (here age 80) and then drops to zero at higher ages. There is also a peak at infancy. Here, too, such a graph may be helpful as a standard for gauging the completeness of death registration; -- we would expect a surplus of female deaths at the higher ages.

Fig. 5.1 and fig. 5.2 illustrate the data for estimating central death rates or risks. Fig. 5.3 shows mortality risks for men and women for Sweden in 2005. Here, to amplify the age-pattern of mortality, the transformation

$$y_x = \ln (10^6 q_x)$$

has been used. It is important to use such a transformation, as otherwise the age-pattern of child and early adult mortality creeps along the x-axis not displaying any visible variation except for adult ages.

**Fig. 5.4. Survival functions for men and women: Sweden, 2005**

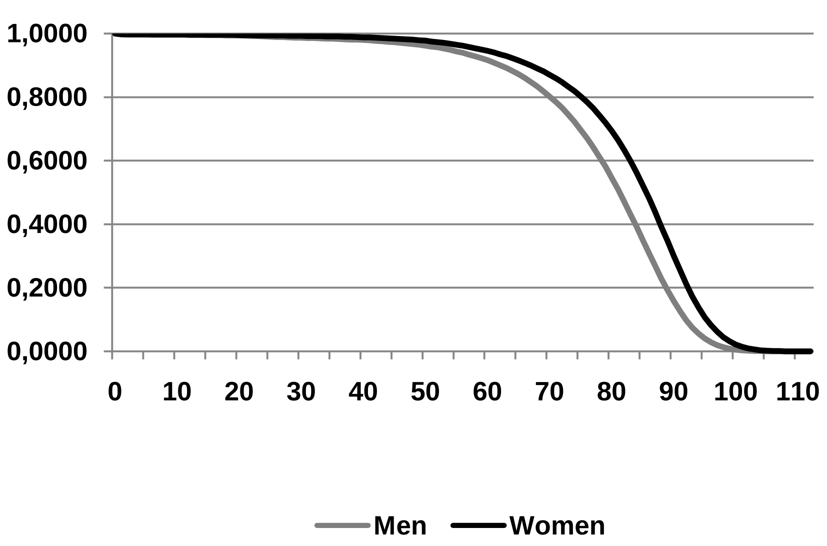
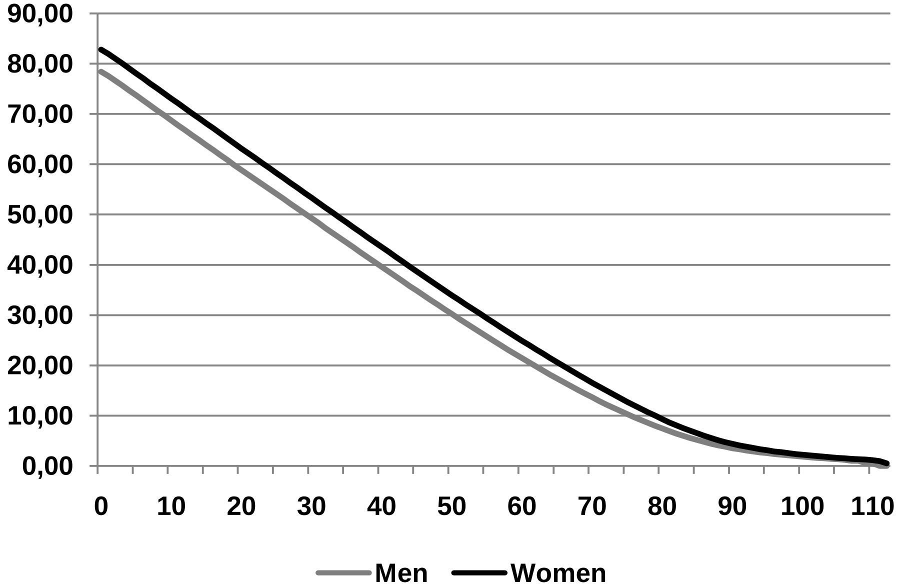



Fig. 5.4 shows the survival functions (with radix one) for men and women in Sweden, 2005. The corresponding life expectancies at age x are shown in fig. 5.5.

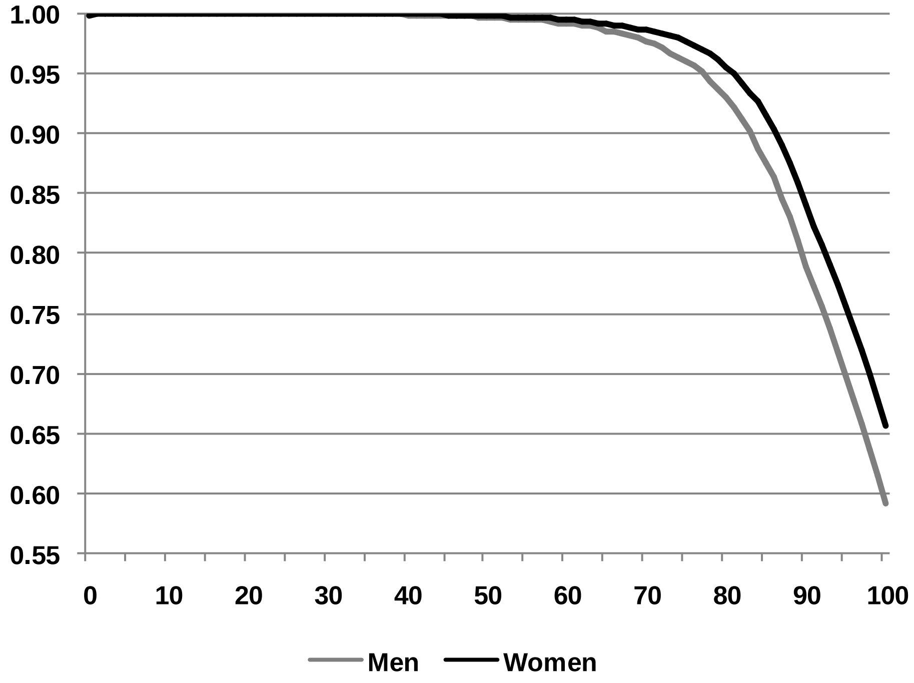
Fig. 5.6 shows the corresponding projection probabilities[11] $u_x^p = L_{x+1}/L_x$ . The projection probabilities are used to project the population from age x to age x+1. For a review of life table techniques, see Namboodiri and Suchindran (1987) and the Methods and Materials of Demography (2004).

**Fig. 5.5. Remaining life expectancies for men and women: Sweden, 2005**




[11] Because projection probabilities are of the form L(x+1)/L(x) they are somewhat insensitive to changes in central death rates m(x). This is why population projections corresponding to different life tables may be nearly the same.

**Fig. 5.6. Projection probabilities for men and women: Sweden, 2005**




To preserve space, table 5.1 gives recorded deaths (Sweden, 2000), mortality risks, survival function and life expectancies at age x for ages below 10 and for ages 90-95. Notice that in terms of the Lexis diagram, the estimation of central death rates is carried out in squares, not in parallelograms. This is the most common approach to estimating mortality rates (the actuarial method).

**Table 5.1. Life table for males: Midyear population and deaths for Sweden, 2000**

| Age | Midyear population Males | Observed Deaths | Mortality rate m(x) | s(x) | Life expectancy at age x |
|---|---|---|---|---|---|
| 0 | 46,037 | 157 | 0.00341 | 1.0000 | 77.6 |
| 1 | 45,966 | 34 | 0.00074 | 0.9966 | 76.9 |
| 2 | 46,578 | 10 | 0.00021 | 0.9959 | 75.9 |
| 3 | 47,898 | 4 | 0.00008 | 0.9956 | 74.9 |
| 4 | 51,188 | 3 | 0.00006 | 0.9956 | 74.0 |
| 5 | 55,471 | 4 | 0.00007 | 0.9955 | 73.0 |
| 6 | 58,867 | 6 | 0.00010 | 0.9954 | 72.0 |
| 7 | 61,933 | 6 | 0.00010 | 0.9953 | 71.0 |
| 8 | 64,434 | 2 | 0.00003 | 0.9952 | 70.0 |
| 9 | 65,354 | 10 | 0.00015 | 0.9952 | 69.0 |
| 10 | 63,780 | 10 | 0.00016 | 0.9951 | 68.0 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| 90 | 4,756 | 1128 | 0.23717 | 0.1327 | 3.14 |
| 91 | 3,592 | 976 | 0.27171 | 0.10468 | 2.85 |
| 92 | 2,623 | 839 | 0.31986 | 0.07977 | 2.58 |
| 93 | 1,865 | 681 | 0.36515 | 0.05794 | 2.37 |
| 94 | 1,269 | 497 | 0.39165 | 0.04021 | 2.19 |
| 95 | 845 | 363 | 0.42959 | 0.02718 | 2.01 |

## 5.15 The actuarial definition of rate

As noted, the central death rate has many different names such as age-specific mortality rate, force of mortality (the preferred name in actuarial literature), mortality intensity, instantaneous mortality rate or hazard rate (the preferred name in statistical literature). In actuarial literature it is common to define the force of mortally as

$$\mu_x = -l'_x / l_x \qquad (5.19)$$

where $\frac{d}{dx} l_x = l'_x$ (see chapter 15 for a short introduction to differentiation of a function). To explain (5.19) heuristically, let $\mu_x$ be a function of age x such that $\mu_x$ dx is the probability of death on the infinitesimal age-interval between ages x and x+dx. If in a life table population there are $l_x$ survivors at age x, then their exposure time between ages x and x+dx is $l_x$ dx for which reason the expected number of deaths between ages x and x+dx is $\mu_x l_x$ dx. It now follows that the number of survivors at age x+dx is $l_{x+dx} = l_x - \mu_x l_x dx$. Hence, $\mu_x = \frac{1}{l_x}\frac{l_{x+dx} - l_x}{dx}$. Rewriting the last expression, we get $\mu_x = \frac{1}{l_x}\frac{l_{x+dx} - l_x}{dx} = -\frac{l'_x}{l_x}$

which is (5.19). It will be noted that (5.19) is a simple first-order linear differential equation with solution

$$l_x = e^{-\int_0^x \mu_t \, dt} \tag{5.20}$$

that can be verified by the differentiation since

$$\frac{d}{dx} l_x = -\mu_x \, e^{-\int_0^x \mu_t dt} .$$

From (5.20) it will be seen that

$$q_x = \frac{l_x - l_{x+1}}{l_x} = \frac{e^{-\int_0^x \mu_t dt} - e^{-\int_0^{x+1} \mu_t dt}}{e^{-\int_0^x \mu_t dt}} = 1 - e^{-\int_x^{x+1} \mu_t dt} \approx \mu_x$$

if $\mu_t$ is constant between ages x and x+1. This also explains why $q_x$ is referred to as a rate although, in fact, it is a conditional probability.

The observed central death rate $\hat{\mu}_t$, based on exposure time $R_t$, is asymptotically normally distributed with asymptotic estimated mean

$$E(\hat{\mu}_t) = \hat{\mu}_t \tag{5.21}$$

and asymptotic estimated variance

$$Var(\hat{\mu}_t) = \hat{\mu}_t / R_t \tag{5.22}$$

In addition, central death rates $\hat{\mu}_x$ and $\hat{\mu}_z$ at different ages x and z, respectively, are asymptotically independent (insofar age intervals are non-overlapping).

It is relatively easy to understand (5.22). Since $q_x \approx \mu_x$ , $1 - q_x \approx 1$ and $E_x \approx R_x$ we have

$Var(\mu_x) \approx Var(q_x) = q_x(1-q_x)/E_x \approx q_x/E_x \approx \mu_x/R_x$. These results are often applied to data for small populations and surveys.

A result similar to (5.22) is easily obtained if one assumes that deaths are Poisson distributed with parameter λ on the age/time interval $x \le t < x+1$. Deaths (signals) take place at times $t_k$, k = 1, ... , n. Let $j_k$ be the number of cumulated deaths at time $t_k$. This means that $j_k$ is Poisson distributed with parameter $t_k \lambda$ so that its mean and variance are $E(j_k) = Var(j_k) = t_k \lambda$. The expected number of deaths at time $t_n$ is $t_n\lambda$. At time $t_n$, we let $j_n = n$.

The maximum-likelihood function[12] for the experiment is

$$L = (t_n \lambda)^n e^{-t_n \lambda}/n!$$

so that

$$\log L = n \log t_n + n \log \lambda - t_n\lambda - \log n!$$

wherefore

$$\frac{d}{d\lambda}\log L = \frac{n}{\lambda} - t_n = 0$$

yields that the maximum-likelihood estimator for λ is

$$\hat{\lambda} = j_n/t_n = n/t_n$$

[12] The maximum-likelihood method of estimation was popularized by the British statistician and evolutionary biologist Sir Ronald Fisher (1890-1962). The method however had been used earlier by Gauss, Laplace, Thiele and Edgeworth (Hald, 1998). Today this is the most commonly used method of statistical estimation.

that is, $\hat{\lambda}$ is the observed intensity with which deaths took place on the interval. Because $j_n$ is Poisson distributed with parameter $t_n\lambda$ the expected value of $\hat{\lambda}$ is $E(\hat{\lambda}) = \lambda$ with variance $Var(\hat{\lambda}) = \hat{\lambda} / t_n$.

On several occasions we have noted that a rate is the ratio between events and the exposure time elapsed for their observation. This was the intuitive position taken by the British actuary Joshua Milne (1776-1851). Later, it became clear that this choice of definition is in agreement with the exponential distribution, which is often used to model survival (life testing). We now turn to a brief outline of the exponential distribution.

A positive and continuous random variable X with probability density function

$$f(x) = \theta e^{-\theta x} \quad , x > 0 \tag{5.23}$$

is said to be exponentially distributed. The distribution function is

$$F(x) = \int_0^x f(u)du = 1 - e^{-\theta x} \tag{5.24}$$

which is the probability $P(X < x)$. The survival function is

$$S(x) = 1 - F(x) = P(X > x) = e^{-\theta x} \tag{5.25}$$

S(x) is the probability that the individual survives at least to age x (dies after age x). Application of (5.21) to (5.25) shows that $\theta$ is the piecewise constant mortality rate on the interval under consideration. Hence, underneath the definition of a piecewise constant morality rate is that deaths are exponentially distributed. We shall now show that if deaths are exponentially distributed then the number of deaths (events) divided by their corresponding exposure time is a maximum-likelihood estimator for the mortality rate.

For simplicity, we limit our attention to infancy, that is, the age interval $0 \le x < 1$, for which we write [0,1[. Assuming that deaths are exponentially distributed on [0, 1[, and that infants die independently of one another at ages $t_1,...,t_n$ on [0, 1[, the maximum-likelihood function for the experiment becomes

$$L = \theta e^{-\theta t_1} ... \theta e^{-\theta t_n} (e^{-\theta})^{N-n} dt_1...dt_n$$

(this is the probability with which we have observed the events generated by the experiment) that is,

$$L=\theta^{n} e^{-\theta(\sum_{1}^{n} t_k + N - n)} dt_1 \ldots dt_n$$

so that[13] omitting the term log $(dt_1 \ldots dt_n)$

$$\log L = n \log\theta - \theta(\sum_{1}^{n} t_k + N - n)$$

which means that

$$\frac{d}{d\theta}\log L = \frac{n}{\theta} - \sum_{1}^{n} t_k + N - n)$$

Letting $\frac{d}{d\theta}\log L = 0$ the maximum-likelihood estimator becomes

$$\hat{\theta} = \frac{n}{\sum_{1}^{n} t_k + N - n}$$

which is precisely the number of infant deaths divided by the exposure time corresponding to n infants who died, and N-n infants who survived and each contributed one year of exposure time. The above-mentioned results can be generalized to any age interval for which the rate of mortality is constant.

## 5.16 The variance of the life expectancy

Because the life expectancy is a complicated function of the central death rates it is also complicated to find an estimator for the variance of the life expectancy. Perhaps the most intuitive approach for gaining insight into the distribution of the life expectancy is to simulate life tables so that each simulation reflects the same probabilistic mechanism. This can be done by means of binomial trials. A simple approach would be like this: let $q_x$ be the probability of death between ages x and x+1 for a person aged x. Let $E_x$ be the exposed to

[13] It can be shown that maximizing L is the same as maximizing log L.

risk at age x, then the expected number of deaths at age x is $q_x\ E_x$. Draw $E_x$ random numbers on [0,1[. Let $d_j^x$ be the j:th such random number. Performing $E_x$ trials, we let $d_j^x = 1$ if $r_j^x < q_x$ and 0 otherwise. The simulated number of deaths at age x is $\tilde{D}_x = \sum d_j^x$ (summation over j). The simulated death rates become $\tilde{q}_x = \tilde{D}_x / E_x$ from which a simulated survival function and corresponding life expectancy may be calculated. Repeating the experiment a large number of times, the distribution for the life expectancy can be sketched. We shall shortly illustrate this technique.

It is however possible to follow another path for finding the variance of the life expectancy. Because an implicit approximation to the life expectancy at age α is

$$\text{Var } \hat{e}_\alpha \approx \sum_{i=\alpha}^{\omega} \left[ \frac{\delta}{\delta q_i} \hat{e}_a \right]^2 \text{Var } \hat{q}_i$$

It can be shown that, for large exposures, an approximation to the expected variance of the life expectancy at age α is

$$\text{Var}(\hat{e}_\alpha) = \sum_{i=\alpha}^{\omega} \hat{p}_{\alpha i}^2 \left[ \hat{e}_{n_i+1} + (1-a_i)\, n_i \right]^2 \frac{\hat{q}_i^2 (1-\hat{q}_i)}{D_i} \qquad (5.21)$$

In (5.21) $p_{\alpha i}$ is the probability of survival from age α to age i, $e_i$ the remaining life expectancy at age i, $n_i$ the length of age-interval beginning at age i, and $i + a_i\ n_i$ the mean age at death for those who die in age-interval i. This is known as Chiang's variance estimator (see e.g., Chiang, 1968, pp. 189-241; Irwin, 1949; Keyfitz, 1977, p. 430; Wilson, 1938).

It is assumed that $q_\omega = 1$, ω signifying the highest age at which there are survivors. Writing $E_i = D_i / q_i$ and letting $a_i = 0.5$, (5.21) for single-year ages becomes

$$\text{Var}(\hat{e}_\alpha) = \sum_{i=\alpha}^{\omega} \hat{p}_{\alpha i}^2 \left[\hat{e}_{i+1} + 0.5\right]^2 \frac{\hat{q}_i(1-\hat{q}_i)}{E_i} \tag{5.22}$$

which is more convenient for application since it refers to the exposed to risk.

As noted, the asymptotic variance (5.21) has been found with the assumption that exposures are very large. Table 5.2 shows the numerical example originally given by Chiang (1968) when he applied his estimator. He made use of the US both sexes life table for 1960 based on a population of 179,325,657 and 1,711,262 deaths; a very large population resulting in a very small standard deviation for the life expectancy at birth.

To explore what the variance of the life expectancy at birth is for a much smaller population, table 5.3 gives the same mortality risks as table 5.2 but with a population size that is scaled to 176,926. The corresponding expected deaths for an annual experience is 2,853.

**Table 5.2. US both sexes life table for 1960 and estimated standard deviations of remaining life expectancies**

| Age | Both sexes | Deaths | m(x) | $e_x$ | Sd ( $e_x$ ) |
|---|---|---|---|---|---|
| 0 | 4,126,560 | 110,873 | 0.02651 | 69.73 | 0.012 |
| 1 | 16,195,304 | 17,682 | 0.00436 | 70.62 | 0.010 |
| 5 | 18,659,141 | 9,163 | 0.00245 | 66.92 | 0.010 |
| 10 | 16,815,965 | 7,374 | 0.00219 | 62.08 | 0.010 |
| 15 | 13,287,434 | 12,185 | 0.00457 | 57.21 | 0.010 |
| 20 | 10,803,165 | 13,348 | 0.00616 | 52.46 | 0.010 |
| 25 | 10,870,386 | 14,214 | 0.00652 | 47.77 | 0.009 |
| 30 | 11,951,709 | 19,200 | 0.00800 | 43.06 | 0.009 |
| 35 | 12,508,316 | 29,161 | 0.01159 | 38.39 | 0.009 |
| 40 | 11,567,216 | 42,942 | 0.01839 | 33.81 | 0.009 |
| 45 | 10,928,878 | 64,283 | 0.02898 | 29.40 | 0.008 |
| 50 | 9,696,502 | 90,593 | 0.04564 | 25.20 | 0.008 |
| 55 | 8,595,947 | 116,753 | 0.06566 | 21.29 | 0.007 |
| 60 | 7,111,897 | 153,444 | 0.10226 | 17.61 | 0.006 |
| 65 | 6,186,763 | 196,605 | 0.14691 | 14.33 | 0.005 |
| 70 | 4,661,136 | 223,707 | 0.21335 | 11.37 | 0.004 |
| 75 | 2,977,347 | 219,978 | 0.30886 | 8.77 | 0.003 |
| 80 | 1,518,206 | 185,231 | 0.45667 | 6.57 | 0.002 |
| 85 | 648,581 | 120,366 | 0.60462 | 4.99 | 0.001 |
| 90 | 170,653 | 50,278 | 0.77079 | 3.81 | 0.001 |
| 95+ | 44,551 | 13,882 | 1.00000 | 3.21 | 0.000 |
| | 179,325,657 | 1,711,262 | | | |

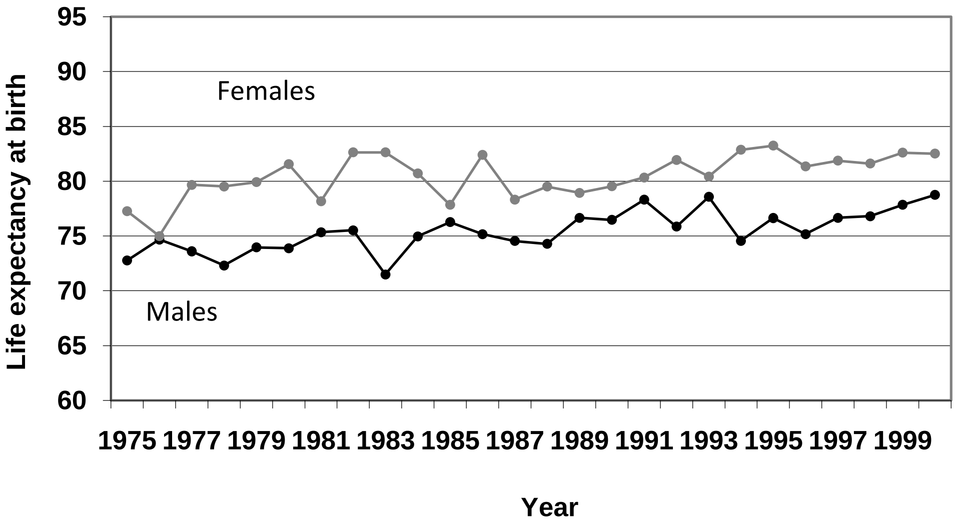
Table 5.3 also shows five simulation series each giving 20 realizations of the life expectancy together with their means and standard deviations. The mean of the five simulations is $\bar{e}_0 = 67.8$ years, and the mean of the five standard deviations is $\bar{\sigma} = 0.61$. The Chiang SD comes out at $\sigma\,(\text{Chiang}) = 0.31$. Apparently, (5.21) underestimates the variance when the risk population is small.

**Table 5.3. Five simulation series of the life expectancy**

| Simulation number | Simulated life expectancy 1 | 2 | 3 | 4 | 5 | Age | $_nq_x$ | Population | Deaths |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 69.3 | 70.0 | 68.9 | 69.3 | 68.9 | 0 | 0.0265 | 3,534 | 94 |
| 2 | 67.9 | 71.1 | 71.1 | 70.0 | 69.8 | 1 | 0.0044 | 21,910 | 96 |
| 3 | 69.3 | 69.6 | 69.3 | 69.9 | 69.3 | 5 | 0.0025 | 23,461 | 57 |
| 4 | 69.1 | 69.9 | 70.2 | 69.3 | 70.8 | 10 | 0.0022 | 22,279 | 49 |
| 5 | 69.5 | 69.7 | 70.6 | 70.2 | 69.3 | 15 | 0.0046 | 22,713 | 104 |
| 6 | 69.5 | 69.1 | 70.6 | 70.2 | 70.4 | 20 | 0.0062 | 22,231 | 137 |
| 7 | 69.8 | 69.9 | 69.5 | 68.4 | 70.1 | 25 | 0.0065 | 17,223 | 112 |
| 8 | 69.8 | 69.9 | 70.5 | 68.7 | 70.3 | 30 | 0.0080 | 11,232 | 90 |
| 9 | 69.5 | 69.6 | 70.3 | 69.6 | 70.6 | 35 | 0.0116 | 8,186 | 95 |
| 10 | 68.9 | 69.1 | 69.6 | 69.4 | 69.1 | 40 | 0.0184 | 5,638 | 104 |
| 11 | 69.9 | 69.6 | 69.5 | 68.8 | 70.4 | 45 | 0.0290 | 4,669 | 135 |
| 12 | 68.7 | 69.3 | 69.5 | 70.4 | 69.5 | 50 | 0.0456 | 3,874 | 177 |
| 13 | 69.7 | 71.1 | 70.4 | 69.6 | 69.1 | 55 | 0.0657 | 3,056 | 201 |
| 14 | 70.0 | 68.4 | 69.0 | 69.3 | 70.6 | 60 | 0.1023 | 2,414 | 247 |
| 15 | 70.5 | 70.5 | 70.8 | 69.5 | 70.0 | 65 | 0.1469 | 1,968 | 289 |
| 16 | 70.2 | 70.0 | 69.4 | 70.1 | 70.0 | 70 | 0.2134 | 1,263 | 269 |
| 17 | 70.5 | 70.7 | 70.2 | 69.4 | 70.3 | 75 | 0.3089 | 536 | 166 |
| 18 | 69.8 | 69.2 | 70.3 | 70.1 | 70.1 | 80 | 0.4567 | 392 | 179 |
| 19 | 69.0 | 69.5 | 70.1 | 69.7 | 69.5 | 85 | 0.6046 | 181 | 109 |
| 20 | 69.4 | 70.9 | 70.7 | 69.8 | 70.2 | 90 | 0.7708 | 95 | 73 |
| | | | | | | 95+ | 1.0000 | 71 | 71 |
| Mean | 69.5 | 69.9 | 70.0 | 69.6 | 69.9 | | | | |
| SD | 0.61 | 0.71 | 0.63 | 0.53 | 0.57 | | Total | 176,926 | 2,853 |

Fig. 5.8 shows life expectancies for males and females in the Swedish municipality Klippan with mean populations of about 8,000 males and females during 1975-2000. The standard deviations[14] are $\hat{\sigma}_m$ = 1.9 and $\hat{\sigma}_f$ = 2.1 years for males and females, respectively. Similar standard deviations are shown for 17 small municipalities in Sweden, 1975-2000 (table 5.4).

**Fig. 5.8. Life expectancies at birth for Klippan municipality (Sweden), 1975-2000**




It may be noted that although life expectancies increased in the municipalities during 1975-2000, nevertheless table 5.5 gives useful estimates of the sampling standard deviation of the life expectancy for small populations. It should also be mentioned that the standard deviations are rather insensitive to moderate changes in the underlying mortality risks.

[14] As calculated for the series of 26 observations.

**Table 5.4. Approximate standard deviations for life expectancies in 17 small municipalities: Sweden, 1975-2000**

| Municipality | Population | | Standard deviation | |
|---|---|---|---|---|
| | Males | Females | Males | Females |
| Arjeplog | 2,007 | 1,811 | 4.8 | 4.3 |
| Boxholm | 2,851 | 2,715 | 3.3 | 3.3 |
| Odeshög | 3,011 | 2,932 | 2.3 | 3.0 |
| Ragunda | 3,666 | 3,428 | 2.4 | 2.6 |
| Laxå | 4,021 | 3,820 | 2.3 | 2.7 |
| Torsås | 4,029 | 3,770 | 2.4 | 2.5 |
| Lessebo | 4,447 | 4,339 | 2.2 | 2.6 |
| Pajala | 4,605 | 3,993 | 2.1 | 3.1 |
| Svalöv | 6,442 | 6,169 | 1.6 | 2.1 |
| Ånge | 6,460 | 6,248 | 2.3 | 1.7 |
| Älmhult | 7,941 | 7,652 | 2.2 | 2.2 |
| Klippan | 8,053 | 8,083 | 1.9 | 2.1 |
| Sala | 10,633 | 10,676 | 2.3 | 1.7 |
| Nynäshamn | 10,755 | 10,512 | 1.7 | 1.8 |
| Laholm | 11,045 | 10,739 | 1.7 | 1.5 |
| Arvika | 13,099 | 13,502 | 2.0 | 2.0 |
| Katrineholm | 15,801 | 16,270 | 2.1 | 2.0 |

Source: M. Hartmann, 2004. Statistics Sweden.

# 6.0 The Stable Population

## 6.1 An early invention

It was the Swiss mathematician Leonard Euler (1707-83) who developed the stable population model. Euler made the assumption that all births take place on July 1, that is in the middle of the year, and that they increase exponentially over time. Survival s(x) is assumed time-invariant.

If ω years ago B births took place, then ω years later the survivors will be s(ω) B. In year ω-1 there were $B\,e^{r}$ births where r is the exponential rate of yearly increase of births. The survivors ω – 1 years later are s(ω-1) $Be^{r}$. In year ω-2 there were $B\,e^{2r}$ births of which the survivors ω - 2 years later are s(ω-2) $B\,e^{2r}$, and so on. Continuing in this fashion we find that the population size in year 0 on 1 July is

$$T_0 = s(\omega)\,B + s(\omega-1)B^{r} + \ldots + s(0)B^{\omega r} \qquad (6.1)$$

which implies that

$$T_0 = Be^{\omega r}\left[\;s(0) + s(1)e^{-r} + s(2)\,e^{-2r} + \ldots + s(\omega)\,e^{-\omega r}\right]$$

so that

$$T_0 = Be^{\omega r}\left[1 + s(1)e^{-r} + s(2)\,e^{-2r} + \ldots + s(\omega)\,e^{-\omega r}\right]$$

Dividing both sides of the equality sign by $T_0$ we obtain

$$1 = b + b\,s(1)\,e^{-r} + \ldots + b\,s(\omega)\,e^{-\omega r} \qquad (6.2)$$

where b is the crude birth rate. The right-hand side of (6.2) sums to one and thus gives the age distribution of the population. The proportion of the population aged 0 is b, the proportion aged 1 is $b\;s(1)\,e^{-r}$, the proportion aged 2 is $b\;s(2)\,e^{-2r}$ and so on. This is the stable age distribution (Keyfitz and Flieger, 1971). A stationary population is a stable population for which r = 0.

Euler's paper on the stable population model appeared in the Belgian Académie Royala des Sciences et Belles-Lettres in 1760, that is, just about a hundred years after Graunt had published his "Observations on the bills of mortality." There is a considerable amount of literature devoted to stable populations and their characteristics.

## 6.2 Application of stable age distributions

The age distribution of a stable age distribution is usually written[15]

$$c(a)\,da = b\,e^{-ra}\,s(a)\,da \tag{6.3}$$

where c(a) da is the proportion of the population aged between a and a+da. From (6.3) it follows that

$$s(a) = \frac{c(a)\,e^{ra}}{b} \tag{6.4}$$

Euler noted that if an age distribution can be found from a census and if the corresponding crude birth rate and rate of population growth are known, then the population's survivorship can be inferred. Euler, then, was an early inventor of indirect estimation. It will be noted that

$$\ln\frac{c(a)}{s(a)} = \ln b - ra \tag{6.5}$$

can be used to estimate the crude birth rate and the rate of population growth from the age-distribution c(a). Using a standard statistical package, (6.4) or (6.5) can be used for simultaneous estimation of survivorship, the crude birth rate and the natural rate of population increase (requires well-behaved data). Stable relationships like (6.5) have been used for estimation of mortality and fertility in nations with incomplete vital registration (Brass 1975).

## 6.3 An application to Abu Dhabi Emirate 2005 population census

Fig. 6.1 shows an application of (6.3) to the percent age-distribution recorded for females who were United Arab Emirate citizens in the 2005 Abu Dhabi Emirate population census. The estimated parameters are b = 0.036 and r = 0.029. The *a priori* chosen survival function is for Swedish females 1960 with a life expectancy of 74.9 years. The figure brings to light some interesting features. In the first place, it

[15] Here da signifies an infinitesimal age increment. In (6.3), the proportion of lives in the interval between ages a and a+da is c(a)da.

will be noted that, evidently, fertility began to fall rapidly some 20 years ago, that is, during the early 1980s. In the second place, it will also be seen that the accuracy of age-reporting has improved considerably after the 1980s (there is much less serration in the reported ages after 1980 than before).

The estimated crude birth rate of about 36 per 1,000, and its associated growth rate of about 3 percent per year, most likely, do not apply to the time of the census; estimates reflect more so the past than the present. Nevertheless, even if this application of the method does not yield reliable estimates for the present, it helps highlight that there must have been radical changes in reproductive behavior during the past 20 years or so.

An alternative approach is to only make use of the most recent data, for example, data going back 10 years in time (table 6.1). In this application we have used the Danish male life table for 1982 ($e_0 = 71.6$) with relation (6.5). The value at age 0 for both sexes has been modified (interpolated) so that it is -3.6 at age 0 (table 6.1). The reason for this is the very strong dip at age 0 suggesting considerable underenumeration of infants in the census. The equation for the fitted line is given in fig. 6.2. It suggests a growth rate of about 1.6 percent per year, and a crude birth rate of exp(-3.474) = 31 per 1,000 population. The results of this application can be difficult to interpret, not only because of the short range of ages but also because of the possibility of an erratic enumeration of the population.

Fig. 6.1. Stable population fitted to age-distribution for females, Abu Dhabi Emirate 2005 population census

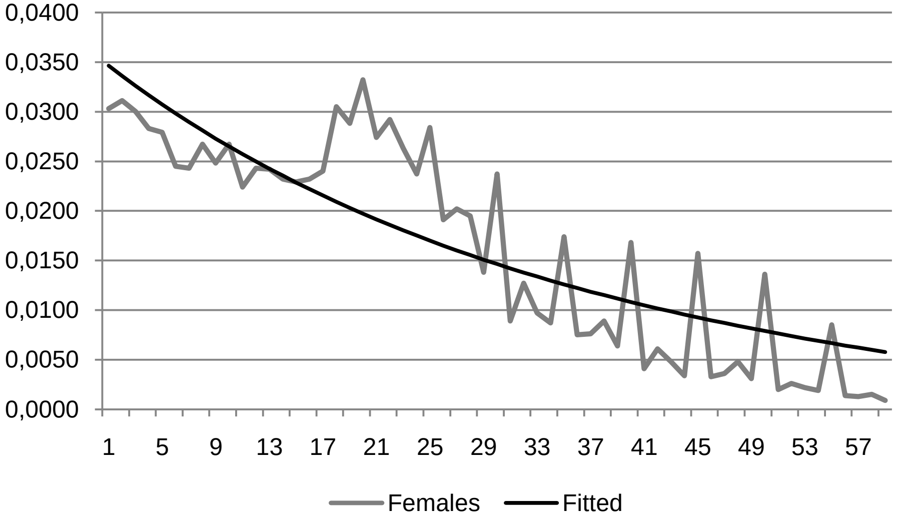
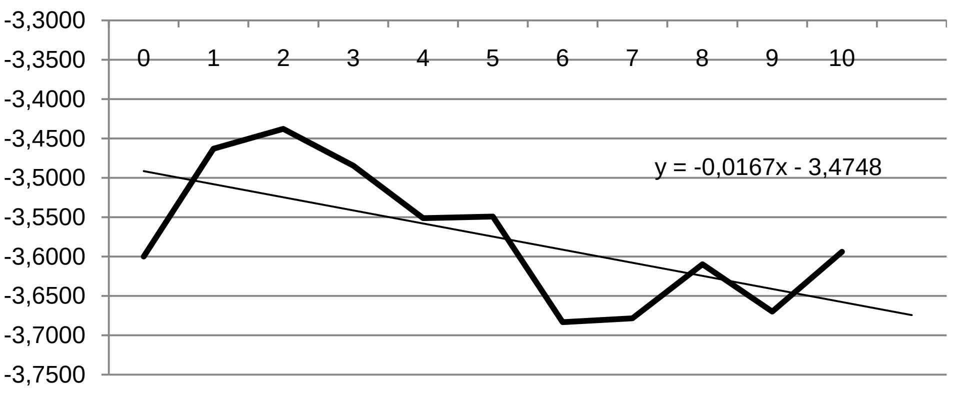


**Table 6.1. Application of stable model to age-distribution for UAE citizens as obtained from the UAE 2005 census of population**

| | ln[c(a)/s(a)] | | | ln[c(a)/s(a)] |
|---|---|---|---|---|
| Age | Males | Females | Age | Both sexes |
| 0 | -3.90869 | -3.92102 | 0 | -3.6000 |
| 1 | -3.43538 | -3.49106 | 1 | -3.4632 |
| 2 | -3.41386 | -3.46191 | 2 | -3.4379 |
| 3 | -3.46964 | -3.49935 | 3 | -3.4845 |
| 4 | -3.54446 | -3.55838 | 4 | -3.5514 |
| 5 | -3.52785 | -3.57052 | 5 | -3.5492 |
| 6 | -3.66456 | -3.70207 | 6 | -3.6833 |
| 7 | -3.64845 | -3.70869 | 7 | -3.6786 |
| 8 | -3.60322 | -3.61595 | 8 | -3.6096 |
| 9 | -3.65199 | -3.68795 | 9 | -3.6700 |
| 10 | -3.57199 | -3.61597 | 10 | -3.5940 |

Fig. 6.2. Linear application of stable population to the both sexes age distribution: Abu Dhabi 2005 population census.



## 6.4 The stable population and indirect estimation

Despite the fact that the stable population model is a pretty rigid one, it has been used on numerous occasions with success. Many years ago, the population of England and Wales had ”stable features” and was used to illustrate techniques of indirect estimation based on the stable population model (see e.g., Brass, 1974).

Roughly speaking, between the late 1950s and early 1980s much research was devoted to developing estimation methods based on the stable population model. These methods were used to estimate mortality and fertility in developing nations with incomplete vital registration and deficient censuses. Methods of this nature are still in use (see e.g., United Nations, 1968).

It was Alfred J. Lotka (1880-1949) who popularized the concept of the stable population and gave it modern statistical treatment (Shryock and Siegel, 1975). Several important contributions were also made by Ansley Coale (1917-2002).

# 7.0 Standardization

## 7.1 The crude death rate and its dependence on the age distribution

Heretofore, we have given relatively little attention to the crude death rate (3.7). We shall now see that this measure can be difficult to interpret. As noted, the crude death rate is the total number of deaths divided by the total population. Letting $m_x$ be the central death rate at age x and $E_x^c$ the corresponding central exposed to risk at age x (the midyear population aged x), we have that the total number of deaths is

$$D = \sum_{x=0}^{\omega} E_x^c \, m_x$$

The total midyear population is

$$P = \sum_{x=0}^{\omega} E_x^c$$

for which reason the crude death rate is

$$CDR = \frac{D}{P} = \sum_{x=0}^{\omega} E_x^c \, m_x \Big/ \sum_{x=0}^{\omega} E_x^c \qquad (7.1)$$

It can be shown[16] that (7.1) implies that there is an age y such that CDR = $m_y$ (depending on the $E_x$). It should be clear from (7.1) that CDR greatly depends on the age distribution of the population. Denoting the age distribution by

$$P_x = \frac{E_x^c}{\sum_0^{\omega} E_x^c}$$

[16] Mean value theorem for integrals.

(the proportion of the population aged x), the crude death rate appears as

$$CDR = \sum_{0}^{\omega} P_x m_x$$

that is as a weighted average of the central death rates. Because different populations have different age distributions, the very same set of central death rates may yield much different values of CDR. For this reason, we say that the central death rate is a mortality measure that is confounded with the age distribution (or age structure). It is in recognition of this that it may be useful to standardize mortality measures. We now turn to a discussion of how this may be achieved.

## 7.2 Standardization of mortality rates

Let A and S be two different populations. For a given calendar year, we let

$$CDR(A) = \sum_{x=0}^{\omega} E_x^c \hat{m}_x \Big/ \sum_{x=0}^{\omega} E_x^c$$

denote the crude death rate for population A. Here $\hat{m}_x$ are the estimated central death rates for population A. We refer to population S as the standard population and let

$$CDR(S) = \sum_{x=0}^{\omega} E_x^{c,s} \hat{\mu}_x \Big/ \sum_{x=0}^{\omega} E_x^{c,s}$$

Here $\hat{\mu}_x$ are estimated central death rates for the standard population S. We might ask: “What would have been the crude death rate for the standard population if we had applied to it the estimated mortality rates for population A?” The answer is

$$I(S) = \sum_{x=0}^{\omega} E_x^{c,s} \hat{m}_x \Big/ \sum_{x=0}^{\omega} E_x^{c,s} \ . \qquad (7.2)$$

We refer to this as direct standardization of mortality.

The mortality ratio

$$CMF = \sum_{x=0}^{\omega} E_x^{c,s}\,\hat{m}_x \Big/ \sum_{x=0}^{\omega} E_x^{c,s} \Big/ \sum_{x=0}^{\omega} E_x^{c,s}\,\hat{\mu}_x \Big/ \sum_{x=0}^{\omega} E_x^{c,s} =$$

$$\sum_{x=0}^{\omega} E_x^{c,s}\,\hat{\mu}_x \left(\frac{\hat{m}_x}{\hat{\mu}_x}\right) \Big/ \sum_{x=0}^{\omega} E_x^{c,s}\,\hat{\mu}_x \qquad (7.3)$$

is called the Comparative Mortality Factor (Cox 1970, p. 171). If for all x, $\hat{m}_x = \hat{\mu}_x$ then CMF = 1. The Comparative Mortality Factor expresses how much stronger or weaker the mortality of population A is relative to the mortality of the standard population.

It is in place to discuss a numerical example given by Pressat (1980, pp. 102-103). His example involves comparing mortality between white and nonwhite males in the United States in 1965. In 1965 the crude death rate was CDR = 10.85 per 1,000 for white males whereas it was CDR = 11.14 per 1,000 for nonwhite males. Comparing the crude death rates it would appear that white and nonwhite males enjoy nearly the same mortality. Pressat (1980, pp. 102-103) writes*:*

*"For the reader not forewarned, it would appear that the sanitary and health conditions are almost identical for nonwhites and whites, which would belie virtually all the age-specific rates. In order to neutralize the effect of age structure, we need to choose a standard population. No exact rule is applicable here, but one possibility is to take one of the two populations as the standard population and apply the rates of the other population to it."*

Table 7.1 gives the 1965 midyear populations by age for the nonwhite male population, the age-specific mortality rates for white males in 1965 and the corresponding expected deaths (Pressat, 1980). The total midyear population of nonwhite males is 11,190,000 and the expected deaths 93,599. As a result we obtain a standardized crude birth rate of CDR(s) = 93,599/11,190,000 = 0.00836 or 8.36 per 1,000. This shows that if the nonwhite male population had had the same age-specific mortality as white males, their crude death rate would have been 8.36 per 1,000 instead of 11.14 per 1,000. Stated otherwise, the comparative factor is CMF = 11.14/8.36 = 1.33 so that, in effect, nonwhites have 33 percent higher mortality than whites (relative to the chosen standard).

**Table 7.1. Expected number of deaths in United States nonwhite male population, based upon age-specific mortality rates for the white male population, 1965**

| | Nonwhites | White males m(x) | Expected male deaths |
|---|---|---|---|
| Total | 11,190,000 | 10.85 | 93,599 |
| Age | | | |
| 0 | 322,000 | 23.74 | 7,644 |
| 4 | 1,327,000 | 0.89 | 1,181 |
| 5 | 1,487,000 | 0.47 | 699 |
| 10 | 1,316,000 | 0.49 | 645 |
| 15 | 1,071,000 | 1.31 | 1,403 |
| 20 | 780,000 | 1.72 | 1,342 |
| 25 | 631,000 | 1.57 | 991 |
| 30 | 607,000 | 1.79 | 1,087 |
| 35 | 615,000 | 2.57 | 1,581 |
| 40 | 615,000 | 4.11 | 2,528 |
| 45 | 536,000 | 6.82 | 3,656 |
| 50 | 493,000 | 11.44 | 5,640 |
| 55 | 404,000 | 18.1 | 7,312 |
| 60 | 345,000 | 27.16 | 9,370 |
| 65 | 227,000 | 41.26 | 9,366 |
| 70 | 184,000 | 59.41 | 10,931 |
| 75 | 118,000 | 85.03 | 10,034 |
| 80 | 71,000 | 127.77 | 9,072 |
| 85+ | 41,000 | 222.43 | 9,120 |

# 8.0 Fertility

## 8.1 The crude birth rate

We have already discussed the crude birth rate (CBR) but repeat it here for convenience. The crude birth rate is the total number of births during a calendar year divided by the corresponding midyear population. It is not a “clean” rate because it relates births to both women and men and, moreover, builds on all age groups. This is why it is known as a crude rate. The crude birth rate has served as an important index of fertility or reproduction in many countries because it was the only fertility index that could be estimated easily. It is however heavily confounded with the age distribution of the population for which reason it may be misleading as an index of reproduction[17]. To this must be added that in cases where birth registration is incomplete and the census is affected by appreciable underenumeration, CBR may be materially inflated or deflated (a common problem in countries with deficient birth registration and where censuses are of poor coverage).

## 8.2 The age-specific fertility rate

If, during a calendar year, $W_x$ is the midyear female population aged x and $B_x$ is the number of their live births, then

$$f_x = \frac{B_x}{W_x} \tag{8.1}$$

is known as the age-specific fertility rate at age x. The age-specific fertility rate $f_x$ is the annual number of births divided by the number of person-years (or exposure time) lived by females while at age x. Notice that (8.1) also could be written $f_x W_x = B_x$ so that we get the result that rate times exposure time is the number of births.

[17] It is, for this reason, somewhat ironic that family planning goals often have been stated in terms of crude birth rates.

The age-specific fertility rates for all the reproductive ages

14, 15, … , 49 are known as the <u>fertility schedule</u>[18]. For five-year age groups, we write ${}_5f_x$ for the age-specific fertility rate. For five-year age groups of women

$$ {}_5f_x = \frac{{}_5B_x}{{}_5W_x} \qquad (8.2)$$

where ${}_5B_x$ are the births that took place during a calendar year for the midyear population of women aged between ages x and x+5, denoted ${}_5W_x$ .

## 8.3 The total fertility rate (TFR)

Among the many indices used for measurement of fertility, the total fertility rate (TFR) is the most commonly used. TFR is defined as the sum of the age-specific fertility rates estimated for a calendar year. Hence,

$$ TFR = \sum_{x=15}^{49} f_x \qquad (8.3)$$

where the limits of summation are ages 15 and 49, which are usually taken as the lowest and highest reproductive ages. In the event where other ages are more relevant for delimiting reproductive from non-reproductive ages, these are used in the summation for TFR. It will be appreciated that $f_{15}$ times one person-year is the number of children expected to be born by women aged 15, $f_{16}$ times one person-year is the number of children expected to be born by women aged 16, etc. Hence, TFR is the expected number of children a woman is expected to give live birth to if she has fertility $f_x$ and survives to the end of the reproductive period.

TFR is a somewhat artificial measure in that its interpretation involves a projection of the assumption that age-specific fertility rates estimated for a calendar year (or for another convenient period) will

---

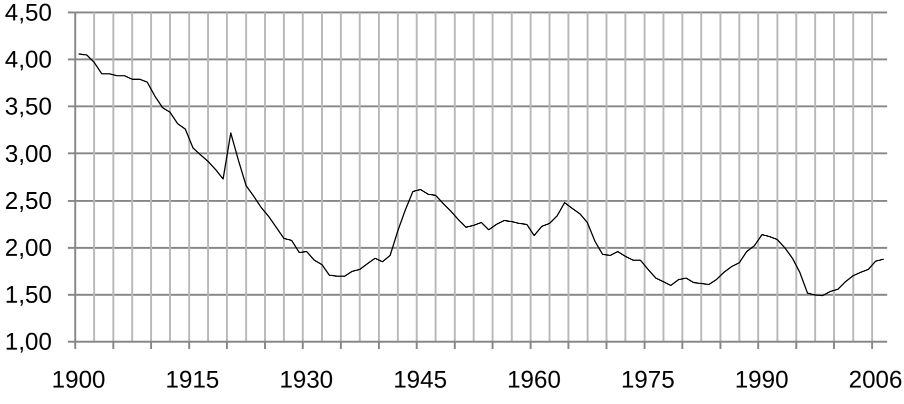
[18] In British literature there is frequent reference to fertility and mortality schedules. In American literature this is less common. Here one ordinarily speaks of age-specific mortality and fertility.

apply to a woman throughout her reproductive life. In the case of working with five-year age groups,

$$TFR = 5 \sum_{15}^{45} {}_5f_x \qquad (8.4)$$

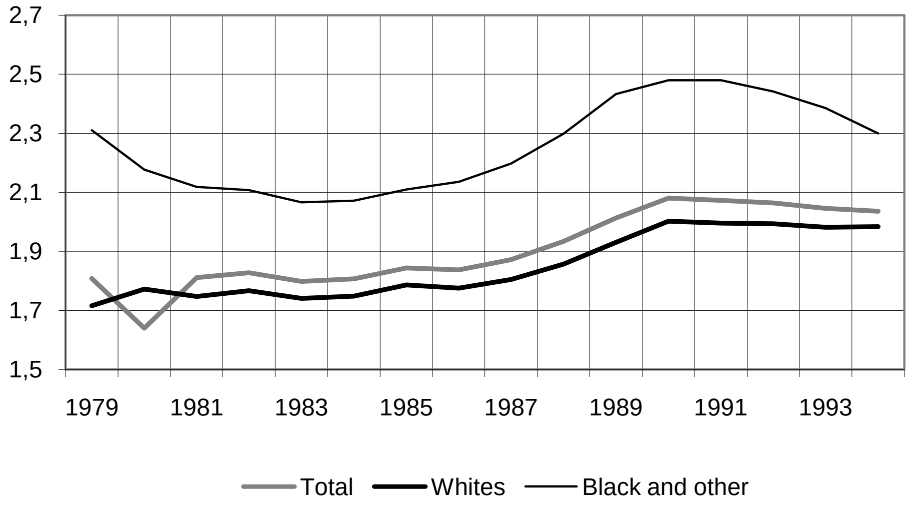
For illustrative purposes, fig. 8.1 shows the total fertility rate for Sweden during 1900-2007. It is easily appreciated that even if we have a long time-series of TFRs, yet it is of little or no aid in projecting future TFRs. Stating it differently, the history of the time-series (process) does not determine with any degree of reasonable precision its future unfolding. Because, as we shall see later, it is fertility that principally determines the future size and age-distribution of the population, it follows that population projections necessarily must be of an uncertain nature.

Fig. 8.1. Total fertility rate: Sweden, 1900-2007




Source: Statistics Sweden.

Fig. 8.2. Estimated total fertility rates for total, whites and black and others, the United States, 1979-1994




Source: Statistical Abstract of the United States 1997, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, p. 77.

Fig. 8.2 shows TFR for whites and coloreds in the United States, 1979-94. It will be noted that the TFR curves for whites and coloreds are almost parallel albeit at much different levels. There is much to be gleaned from such a diagram because it suggests very considerable social and economic differences between ethnic groups.

## 8.4 The gross reproduction rate

Because only women reproduce, a measure of fertility that perhaps is more realistic is the gross reproduction rate (GRR) which is the number of live born girl babies a woman is expected to have if she survives to the end of her reproductive period and her fertility is $f_x$. Assuming that the sex ratio at birth is 100 girls for every 105 boys

$$GRR = \frac{1}{2.05} \sum_{15}^{49} f_x \tag{8.5}$$

## 8.5 The net reproduction rate

The total fertility rate and the gross reproduction rates suffer from the disadvantage that they do not take survival of women into consideration. Denoting by $L_x$ the female life table person-years (radix one), the net reproduction rate is

$$NRR = \frac{1}{2.05}\sum_{15}^{49} L_x f_x \qquad (8.6)$$

which is the expected number of live born girl children a woman is expected to have if she has mortality $L_x$ and fertility $f_x$.

Notice, once again, that TFR, GRR and NRR are indices of fertility that build on the assumption that the chosen age-specific fertility rates will apply to women throughout their reproductive ages. Because fertility has a tendency to change relatively fast even across limited time periods, it is evident that such measures must be interpreted with caution.

## 8.6 The normalized fertility schedule

As noted, we call the set of age-specific fertility rates $f_{15},...,f_{49}$ the fertility schedule. The normalized fertility schedule consists of the age-specific fertility rates $g_{15},...,g_{49}$ where $g_x = f_x / TFR$ so that

$$\sum_{15}^{49} g_x = 1.$$

This re-scaling of the age-specific fertility rates facilitates comparison of age-patterns of fertility.

## 8.7 The mean age of the fertility schedule

The mean age of the fertility schedule is defined as

$$m = \frac{1}{TFR}\sum_{x=15}^{49}(x+0.5)\, f_x = \sum_{x=15}^{49}(x+0.5)\, g_x \qquad (8.7)$$

Basically, the mean age of the fertility schedule is its central location.

## 8.8 The variance of the fertility schedule

The variance of the fertility schedule is defined as

$$\sigma^2 = \sum_{x=15}^{49} (x + 0.5 - m)^2 \, g_x = \sum_{x=15}^{49} (x + 0.5)^2 g_x - m^2 \qquad (8.8)$$

The variance of the fertility schedule indicates how spread out it is. In populations where fertility is high (limited birth control) the variance is usually quite high (around 40 or so). In countries with high levels of fertility control it is low (around 20 or so).

## 8.9 Age-specific fertility rates for Sweden and France

Fig. 8.3 shows age-specific fertility rates for women in Sweden 2002 plotted against age. The plot is known as the age-pattern of fertility. Here it is a curve almost symmetrical about its mean. The total fertility rate calculated from these rates is TFR = 1.64. The mean age of the fertility schedule calculated using (8.7) is m = 30.5 years. The rates and calculations of mean and variance for the Swedish 2002 schedule are given in table 8.1.

**Table 8.1. Age-specific fertility rates, mean and variance: Sweden 2002**

| Age | $f_x$ | $(x+0.5)\,f_x$ | $(x+0.5)^2\,f_x$ |
|---|---|---|---|
| 15 | 0.0002 | 0.0029 | 0.0445 |
| 16 | 0.0009 | 0.0151 | 0.2490 |
| 17 | 0.0038 | 0.0674 | 1.1788 |
| 18 | 0.0068 | 0.1256 | 2.3244 |
| 19 | 0.0130 | 0.2532 | 4.9370 |
| 20 | 0.0218 | 0.4466 | 9.1561 |
| 21 | 0.0307 | 0.6595 | 14.1801 |
| 22 | 0.0429 | 0.9663 | 21.7424 |
| 23 | 0.0534 | 1.2558 | 29.5110 |
| 24 | 0.0618 | 1.5140 | 37.0929 |
| 25 | 0.0758 | 1.9322 | 49.2710 |
| 26 | 0.0876 | 2.3220 | 61.5328 |
| 27 | 0.1028 | 2.8263 | 77.7231 |
| 28 | 0.1176 | 3.3526 | 95.5483 |
| 29 | 0.1235 | 3.6420 | 107.4390 |
| 30 | 0.1311 | 3.9980 | 121.9400 |
| 31 | 0.1310 | 4.1255 | 129.9535 |
| 32 | 0.1196 | 3.8881 | 126.3632 |
| 33 | 0.1026 | 3.4369 | 115.1366 |
| 34 | 0.0903 | 3.1164 | 107.5155 |
| 35 | 0.0771 | 2.7383 | 97.2089 |
| 36 | 0.0642 | 2.3425 | 85.5000 |
| 37 | 0.0510 | 1.9126 | 71.7233 |
| 38 | 0.0405 | 1.5599 | 60.0579 |
| 39 | 0.0310 | 1.2257 | 48.4151 |
| 40 | 0.0220 | 0.8917 | 36.1140 |
| 41 | 0.0150 | 0.6221 | 25.8165 |
| 42 | 0.0098 | 0.4169 | 17.7167 |
| 43 | 0.0051 | 0.2240 | 9.7440 |
| 44 | 0.0033 | 0.1464 | 6.5168 |
| 45 | 0.0016 | 0.0734 | 3.3408 |
| 46 | 0.0006 | 0.0257 | 1.1927 |
| 47 | 0.0003 | 0.0133 | 0.6297 |
| 48 | 0.0001 | 0.0059 | 0.2869 |
| 49 | 0.0001 | 0.0068 | 0.3382 |
| Sum | 1.6391 | 50.1516 | 1577.4407 |

Calculation of mean and variance of fertility schedule:

$m = 50.1516/1.6391 = 30.6$

$\sigma^2 = 1577.4407/1.6391 - m^2 = 26.2$

**Fig. 8.3. Age-specific fertility: Sweden, 2002**


Fig. 8.4 shows the age-pattern of fertility for France 1949. Generally speaking, the age-pattern of fertility (as we usually see it) is better portrayed by the French schedule than by the Swedish which, as noted, is almost symmetrical about its mean. To get a better impression of the differences in age-pattern between the two schedules, fig. 8.5 shows the normalized schedules.

**Table 8.2. Age-specific fertility rates for France, 1949**

| Age | Women | Births | $f_x$ | Age | Women | Births | $f_x$ |
|---|---|---|---|---|---|---|---|
| 15 | 302,363 | 300 | 0.00100 | 33 | 157,734 | 18,400 | 0.11670 |
| 16 | 301,704 | 1,280 | 0.00420 | 34 | 196,636 | 20,289 | 0.10320 |
| 17 | 320,359 | 4,354 | 0.01360 | 35 | 302,120 | 28,205 | 0.09340 |
| 18 | 322,957 | 10,949 | 0.03390 | 36 | 306,002 | 25,421 | 0.08310 |
| 19 | 330,308 | 21,843 | 0.06610 | 37 | 310,195 | 22,472 | 0.07240 |
| 20 | 315,998 | 32,970 | 0.10430 | 38 | 292,986 | 18,063 | 0.06170 |
| 21 | 321,710 | 45,956 | 0.14280 | 39 | 310,944 | 16,140 | 0.05190 |
| 22 | 319,911 | 55,090 | 0.17220 | 40 | 310,984 | 12,873 | 0.04140 |
| 23 | 326,518 | 61,464 | 0.18820 | 41 | 314,115 | 10,083 | 0.03210 |
| 24 | 326,991 | 64,633 | 0.19770 | 42 | 306,096 | 7,263 | 0.02370 |
| 25 | 320,731 | 63,291 | 0.19730 | 43 | 310,655 | 5,193 | 0.01670 |
| 26 | 324,135 | 61,775 | 0.19060 | 44 | 308,358 | 3,137 | 0.01020 |
| 27 | 326,739 | 60,447 | 0.18500 | 45 | 309,479 | 1,862 | 0.00600 |
| 28 | 339,839 | 58,890 | 0.17330 | 46 | 307,179 | 953 | 0.00310 |
| 29 | 346,318 | 56,103 | 0.16200 | 47 | 311,839 | 439 | 0.00140 |
| 30 | 208,975 | 31,459 | 0.15050 | 48 | 307,399 | 176 | 0.00060 |
| 31 | 186,911 | 25,631 | 0.13710 | 49 | 296,586 | 87 | 0.00030 |
| 32 | 165,350 | 21,102 | 0.12760 | | | | |
| | | | | | | TFR | 3.0 |

**Fig. 8.4. Age-specific fertility rates: France, 1949**

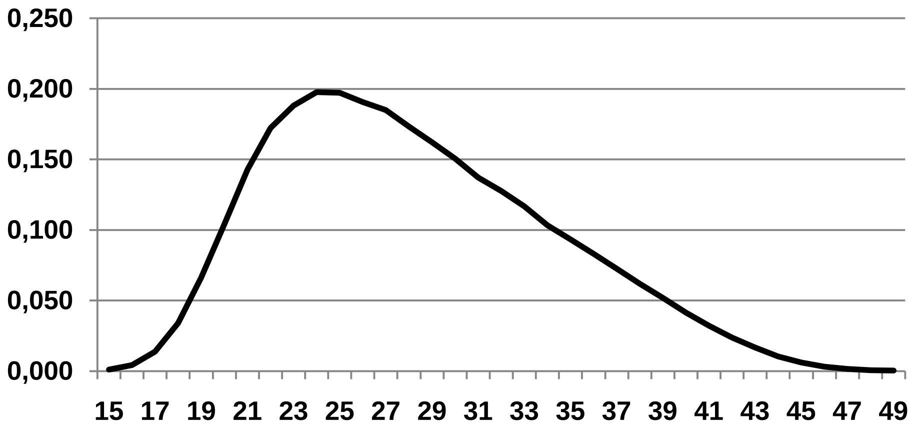
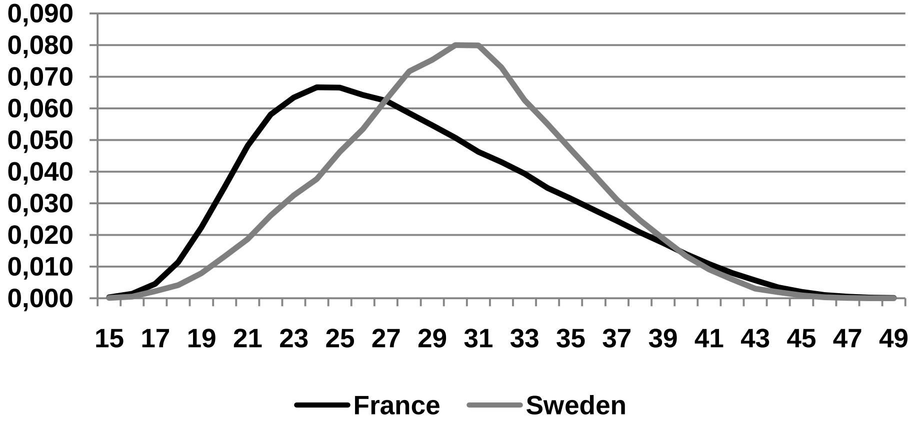


**Fig. 8.5. Normalized age-specific fertility for France 1949 and Sweden 2002**




It should be noted that the mean age of the fertility schedule is not the same as the mean age at childbearing. The reason for this is that the mean age of the fertility schedule assumes that there are equally many women in each age group (a uniform age distribution). In reality, the age distribution across fertile ages is not uniform for which reason there are slight differences between the mean age of the fertility schedule and the observed mean age at childbearing.

## 8.10 Marital, single and all fertility

The age-patterns of fertility for married, unmarried and all women are usually materially different. Table 8.3 gives age-specific fertility rates for five-year age groups of married, single and all women for Sweden, 1951-55. Fig. 8.6 illustrates the three age-patterns of fertility.

**Table 8.3. Marital, single and all age-specific fertility rates: Sweden 1951-55, per 1,000**

| Age | Marital fertility | Single fertility | All fertility |
|---|---|---|---|
| 15-19 | 543.4 | 18.9 | 37.8 |
| 20-24 | 267.6 | 30.9 | 128.2 |
| 25-29 | 167.4 | 22.6 | 128.3 |
| 30-34 | 102.5 | 16.0 | 86.8 |
| 35-39 | 55.6 | 10.1 | 47.8 |
| 40-44 | 18.9 | 3.7 | 15.9 |
| 45-49 | 1.7 | 0.2 | 1.3 |

Source: Statistics Sweden, 1985. (Befolkningsförändringar 1985, Sveriges Officiella Statistik, Del 3, Statistiska Centralbyrån, p. 87).

Fig. 8.6. Age-specific fertility rates for married, single and all women (per 1,000): Sweden, 1951-55

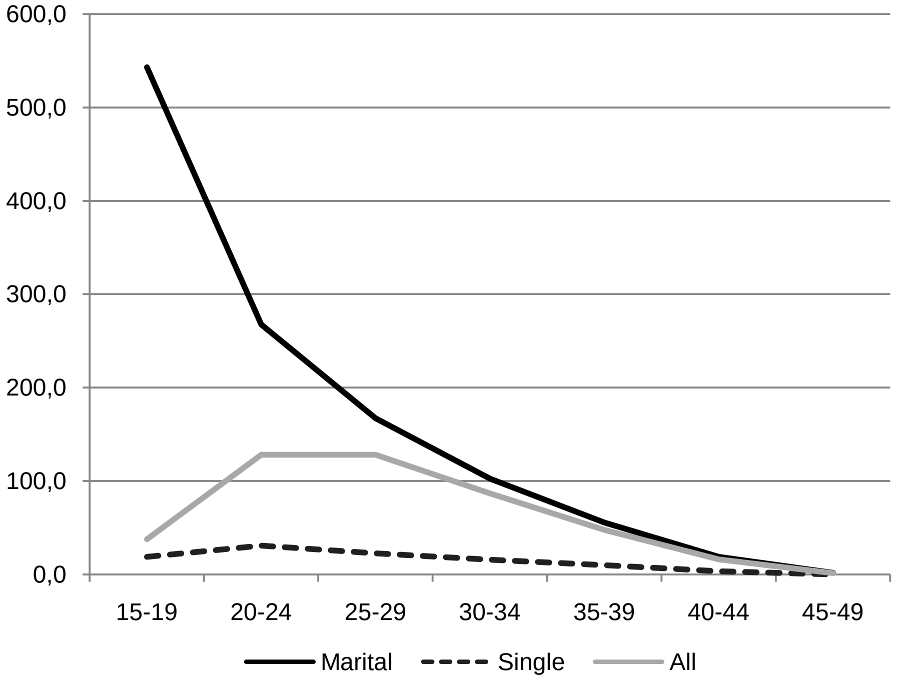



Fig. 8.6 illustrates that fertility (like so many other demographic variables) very much depends on marital status. Fertility for singles is very much different from married or steadily cohabiting couples.

Model fertility schedules representing the wide variability in the age-pattern of childbearing were developed by Coale and Trussell (1974). These have been used extensively in indirect estimation of fertility, a topic that we shall discuss later on.

## 8.11 Mathematical models of fertility

Because mathematical models have been essential for the furtherance of the physical sciences, already at the time of John Graunt it was contemplated to model demographic phenomena by means of mathematical functions. Historically the gamma probability distribution[19] has served as a popular model of fertility schedules (Keyfitz, 1968; Pressat, 1980). Here age-specific fertility is modeled

$$g(x;c,k,d) = R\,\frac{c^k}{\Gamma(k)}(x-d)^{k-1}\,e^{-c(x-d)} \qquad (8.9)$$

$x > d$. In (8.9) parameter R signifies the total fertility rate. The parameter combination $\mu = k/c + d$ is the mean age of the fertility schedule, and $\sigma^2 = k^2/c$ the variance of the fertility schedule. Parameter d does not really signify the beginning of reproduction but can often be set at zero. In (8.9), it is common to use the approximation[20]

$$\Gamma(k) \approx \sqrt{\frac{2\pi}{k}}\,k^k\,e^{(-k+\frac{1}{12k})}$$

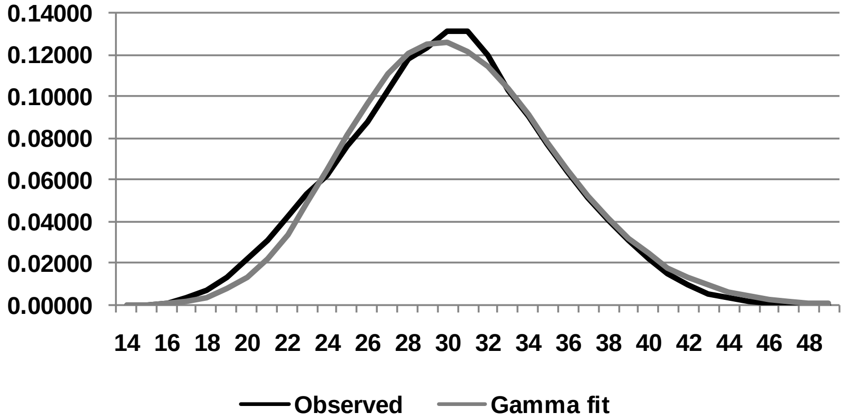
The parameters R, c, k and d can be estimated simultaneously by means of non-linear fitting using a standard statistical package. To illustrate the goodness of fit provided by the gamma function, fig. 8.7 shows the observed fertility schedule and the corresponding modeled schedule for Sweden in the year 2002. Estimated parameters are $\hat{R} = 1.64$, $\hat{c} = 1.104$ and $\hat{k} = 33.69$ with d fixed at 0.

---

[19] Apparently, it was the Swedish economist Sven D. Wicksell who first used the gamma- function to model fertility schedules (see Keyfitz, 1968, for further details).

[20] Known as Sterling's approximation.

**Fig. 8.7. Observed and Gamma-fitted age-specific fertility: Sweden, 2002**




Over time, several mathematical distribution functions have been used as models of the fertility schedule. Here we limit the discussion to the gamma distribution (8.9) and a model proposed by Brass (1968), which has been widely used by demographers working with fertility data from developing nations. The age-specific fertility rate at age x is

$$b(x; \alpha, \beta) = C\ (x-\alpha)\,(\beta-x)^2 \qquad (8.10)$$

where C, α and β are parameters. This is known as the Brass fertility polynomial. Parameter α determines the starting age at child bearing, β the end of the reproductive period, β-α the length of the reproductive period, and C the level of fertility[21]. In many cases this third-degree polynomial has given a satisfactory fit to age-specific fertility, especially in developing nations with high levels of reproduction. As in the case of (8.9) the parameters in (8.10) can be estimated using non-linear estimation in a statistical package. It is common, *a priori*, to let $\alpha = 15$ and $\beta = 50$ in (8.10). It will be noted

that since $\int_{\alpha}^{\beta} (x-\alpha)\,(\beta-x)^2\,dx = I(\alpha, \beta)$ with

[21] Parameter C determines the level of fertility but is not the same as the total fertility rate. The Brass fertility polynomial was used to develop Brass' method for estimating infant mortality from census returns from mothers on the number of children ever born and surviving children (Brass, 1968).

$I(\alpha,\beta)=$

$$\frac{1}{4}(\beta^4-\alpha^4)+(2\beta+\alpha)\frac{(\alpha^3-\beta^3)}{3}+(\beta^2+2\alpha\beta)\frac{(\beta^2-\alpha^2)}{2}-\alpha\beta^3+\alpha^2\beta^2$$

it follows that

$C = 1/\ I(\alpha, \beta)$. For $\alpha=15$ and $\beta = 50$, $C = 0.0000079967$.

Both (8.9) and (8.10) are convenient choices for graduating (or smoothing) age-specific fertility. They do <u>not</u> however provide a deeper understanding of the reproductive behavior of women. Models like (8.9) and (8.10) serve many practical purposes e.g., in population projections as well as in indirect estimation of fertility.

## 8.12 The variance of the total fertility rate

The asymptotic (large-sample) variance of the age-specific fertility rate $f_x$ is

$$\text{Var}(f_x) = f_x/W_x$$

Because fertility rates also are asymptotically independent, it follows that

$$\text{Var(TFR)} = \text{Var}(\sum f_x) \approx \sum \text{Var}(f_x) \approx \sum f_x/W_x \qquad (8.11)$$

By means of simulations it is possible to see to which extent (8.11) also applies to small populations. Table 8.4 gives the age-distribution for 4,998 women, their underlying age-specific fertility rates and expected number of children during a calendar year. The total fertility rate is TFR = 3.0 and the expected number of children is 472. The simulations involve that at ages 15-19, 130 binomial trials are conducted with probability of success (live birth) p = 0.030. Similarly, at ages 20-24, 788 binomial trials are carried out with p = 0.160 and so on. The simulated rates are the simulated number of births divided by the number of women. When simulations have been carried out for all 7 age groups, simulated rates $f_s(x)$ yield an estimated variance of TFR as given by (8.11). Given a reasonably large number of runs (simulations at ages 15-49), the variance of TFR can be estimated by

$$S^2 = (1/N) \sum_{k=1}^{N} (\text{TFR}(k) - \text{TFR (Mean)})^2 \qquad (8.12)$$

where TFR(k) references simulation k and TFR(Mean) is the mean for N simulations. This is illustrated by table 8.5 showing the results for ten simulation runs. It will be seen that each run produces its own TFR, and its own estimated standard deviation of TFR as derived from (8.11). Table 8.5 shows simulated TFRs and corresponding standard deviations provided by (8.11). The mean standard deviation for the ten runs is SD = 0.16. The standard deviation of TFR estimated by (8.12) is SD* = 0.14. Even though the number of runs is 10 the two estimates of the standard deviation are quite close.

**Table 8.4. Expected births for sample of women**

| Age | Women | Fertility rate | Expected births |
|---|---|---|---|
| 15-19 | 130 | 0.030 | 4 |
| 20-24 | 788 | 0.160 | 126 |
| 24-29 | 844 | 0.170 | 143 |
| 30-34 | 796 | 0.130 | 103 |
| 35-39 | 853 | 0.078 | 67 |
| 40-44 | 921 | 0.030 | 28 |
| 45-49 | 666 | 0.002 | 1 |
| Total | 4,998 | 0.600 | 472 |

**Table 8.5. Results for 10 TFR simulation runs**

| 10 simulations | TFR | SD |
|---|---|---|
| 1 | 3.092 | 0.154 |
| 2 | 3.031 | 0.157 |
| 3 | 3.216 | 0.168 |
| 4 | 2.999 | 0.157 |
| 5 | 3.007 | 0.160 |
| 6 | 2.926 | 0.138 |
| 7 | 2.864 | 0.162 |
| 8 | 3.181 | 0.156 |
| 9 | 2.726 | 0.147 |
| 10 | 3.050 | 0.154 |
| Mean | 3.010 | 0.160 |
| SD* | | 0.140 |

## 8.13 Population debates

Although we discuss methods of analysis in this publication it is in place to mention as an aside that in the main there have been two major population debates. At the time of the French-German war (1870-71) it was widely held that France had been disadvantaged by low fertility (too few soldiers). This debate soon spread to Germany where it was also held that reproductive levels were too low for invigorating the population. During the 1930s (with the exception of Great Britain) the European population debate built on fears that reproductive levels were too low for bringing about adequate social and economic development. For this reason, families encouraged by allowances were expected to increase childbearing. France was the first country to provide family allowances with the intention of stimulating increased reproduction. European population policies mainly build on fears that reproduction is too low (the aging society). In recent years, governments have increasingly encouraged people to continue working as long as possible in order to ease the strain on social security funds. This, presumably, also compensates for low reproduction.

After World War II a debate in the opposite direction took to the floor. Here the argument was that population growth was so high that inevitably it would lead to unprecedented misery. The terms "population explosion" and "population bomb" were widely used[22]. Population debates concerning developing nations have given much more attention to reproductive levels than to reducing the frequently high levels of infant and child mortality. Life expectancies in many developing countries have remained low in comparison with industrialized societies. Furthermore, as noted, the debates have also detracted from the desirability of improving vital registration. As a result, in developing nations, vital registration systems rarely support estimating fertility and mortality. Instead recourse is made to indirect estimation and demographic and health surveys. It may be noted that such surveys rarely are taken at short regular time intervals for which reason they only bring forth time-series of limited use. It must also be noted that indirect demographic methods provide little insight into past and ongoing demographic processes. We shall return for a discussion of this.

---

[22] The terms population bomb and population explosion were introduced by the biologist John Ehrlich during the late 1960s. The position taken by the Roman Catholic Church was that no child is "too many" and that reproduction should not be controlled by government or international agencies.

# 9.0 Migration

## 9.1 Internal and international migration

Internal migration involves moves within a country. International migration involves moves across national boundaries. With respect to internal migration (migration from one region to another within a country), we know the risk populations (provided we have reliable census counts or population registers). In the case of international migration, the risk population underlying emigration is the national population whereas immigration has a poorly defined risk population (immigrants may come from any country). In this chapter, we limit the illustrations to migration in and out of Sweden.

## 9.2 Migration in and out of Sweden

Fig. 9.1 shows emigration from Sweden between 1851 and 2002. It will be seen that emigration peaked during the late 1880s (reaching levels of about 50,000), and that from then on it declined to very low levels at the beginning of the 1940s. After World War II emigration began to increase reaching levels similar to those of the 19$^{th}$ century (table 9.1).

Fig. 9.2 shows immigration into Sweden between 1875 and 2002. Immigration during the 17$^{th}$ century was obviously modest. It was not until after World War II that immigration surged reaching its highest peak (83,598) in 1994 (table 9.1).

Fig. 9.3 shows crude rates of immigration and emigration (per 1,000) for the period 1875-2002. These are the number of immigrants (or emigrants) for a calendar year divided by the corresponding mi-dyear population.

Fig. 9.1. Out migration, both sexes: Sweden, 1875-2007

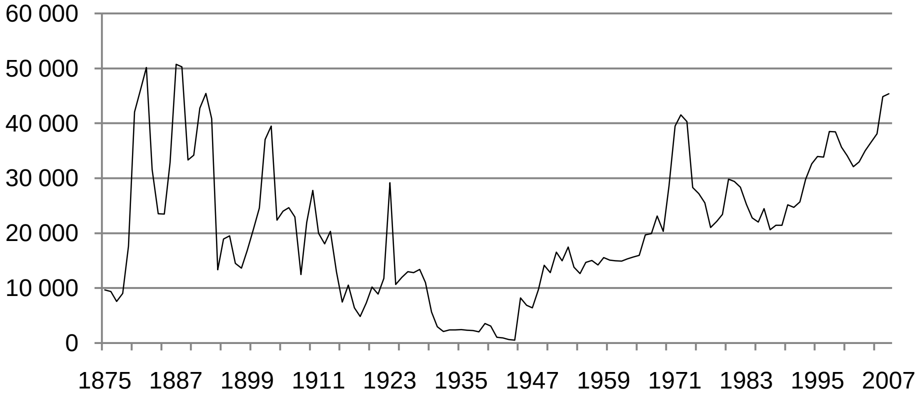



Fig. 9.2. In-migration, both sexes: Sweden, 1875-2007

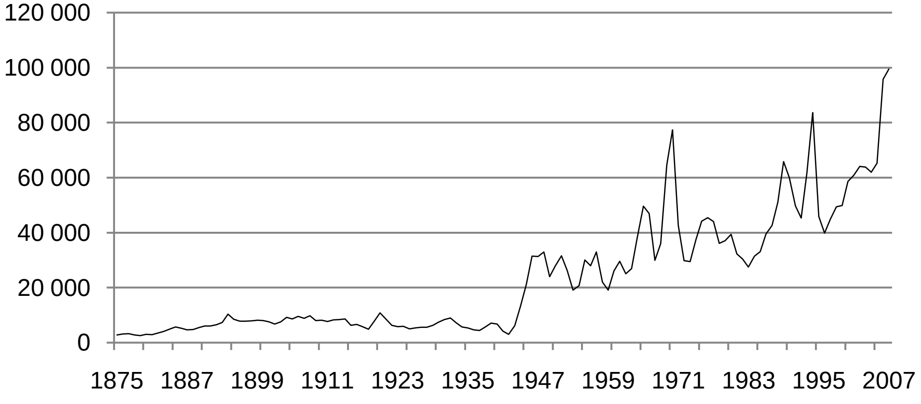
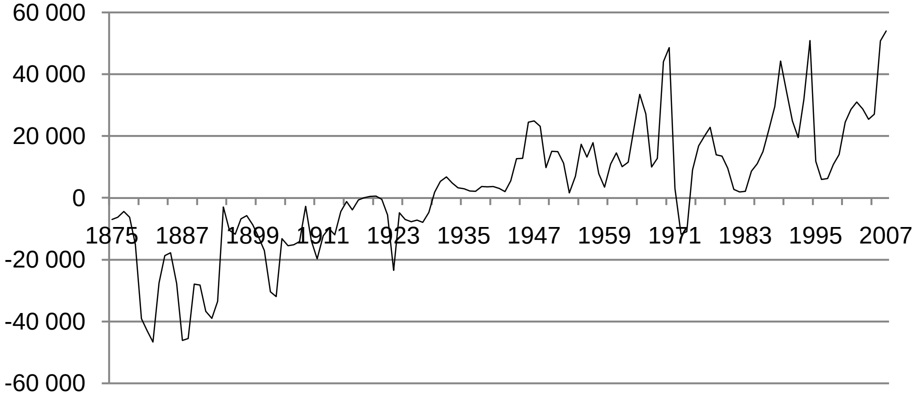


Fig. 9.3. Net migration, both sexes: Sweden, 1875-2007




The difference between emigration and immigration is net-migration. Conceptually emigration and immigration distinguish themselves from net-migration in the sense that while the former are observed processes, net-migration is a calculated difference between immigration and emigration.

Fig. 9.3 shows net-migration for the period 1875-2002. It will be seen that the time-pattern of net-migration indicates that net-migration has increased more or less steadily over the period 1875-2002. It will also be seen that the time-pattern of net-migration is more easily interpreted (almost a linear increase over time) than the similar time-patterns of immigration and emigration.

## 9.3 Statistics on migration

What distinguishes data on migration from other kinds of demographic data is that even countries that boast highly reliable data on mortality and fertility may experience poor data on migration. There are many reasons for this. Precise enumeration of passengers that cross the borders between the nation states in the European Union or cross the federal state borders in the United States is simply not possible. From a practical point of view what sometimes but certainly not always can be counted are work and residence permits issued to non-citizens. The number of such permits however usually understates the true number of persons entering a country in order to work and reside there. Moreover when foreigners who have received residence and work permits eventually leave their host country (perhaps they return home) it often happens that they are not registered as leavers. In some cases this may spuriously increase the *de jure* population. This could increase estimated life expectancies and lower estimated fertility rates for areas with high densities of migrants. It is sometimes argued that such anomalies can be resolved by means of taking population censuses, instead of relying on registers. There does not seem to be much evidence supporting this view. Yet another aspect of migration needs to be mentioned. Migration is that of the demographic processes that changes the fastest (a volatile process). This is of importance in population projections for nations that receive large numbers of migrants and where the balance between in and out migration (net migration) is decisive for whether the population grows, stagnates or declines.

# 10.0 Population Projections

## 10.1 The cohort component method

Population projections serve two purposes. First, they illustrate how mortality, fertility and migration affect the future size and composition of the population. In that sense, population projections play an important diagnostic role. Second, population projections can be used to forecast the population. Here we limit the discussion to population projections serving as an important tool for understanding structural changes in the population. The meaning of this will be explained below. In our discussion we also abstain from including migration in the projections, that is, the population to be projected is assumed closed to migration.

In the cohort component method males and females are projected independently of one another using the same projection mechanism. For this reason we let Let $P_t(x)$ be the midyear population aged x in year t for either males or females. This means that the population is made up of single-year age groups

$P_t(0)$, $P_t(1)$, $P_t(2)$, … , $P_t(h+)$

where $P_t(h+)$ is the population aged h and above in year t. We assume constant mortality and fertility[23]. $L_x$ are the person-years in the chosen life table. The probability of survival from age x to age x+1 (the projection probability) is $\pi_x = L_{x+1} / L_x$. Age-specific fertility is $f_x$.

The population $P_{t+1}(1)$ in year t+1 are the survivors from $P_t(0)$ in year t. This means that

$$\frac{L_1}{L_0} P_t(0) = \pi_0 P_t(0) = P_{t+1}(1)$$

[23] As usual, by this we mean that mortality and fertility stay the same from year to year during the projection period.

so that generally,

$$\frac{L_{x+1}}{L_x} P_t(x) = \pi_x \; P_t(x) = P_{t+1}(x+1)$$

are the survivors aged x+1 in year t+1 who were aged x in year t. For the open-ended age group h+, we have that those aged h in year t+1 are the survivors of those aged h-1 in year t, that is,

$$\frac{L_h}{L_{h-1}} P_t(h-1) = P_{t+1}(h)$$

Those aged h+ in year t+1 are those aged (h) plus those aged (h+1)+ in year t+1 whereby using the T-function

$$\frac{T_{h+1}}{T_h} P_t(h+) = P_{t+1}[(h+1)+]$$

gives

$$P_{t+1}(h+) = P_{t+1}(h) + P_{t+1}[(h+1)+]$$

Remaining is the population aged 0 in year t+1. The total number of births in year t+1 is

$$B_{t+1} = \sum_{x=15}^{49} W_{t+1}(x) \, f_x$$

Boy infants are 0.52 $B_{t+1}$ and the girl infants 0.48 $B_{t+1}$, assuming a sex-ratio at birth of 105 boys per 100 girls. Adjusting for mortality, boy infants in year t+1 are $0.52 \, s_m(0.5) \; B_{t+1}$ and girl infants $0.48 \, s_f(0.5) \; B_{t+1}$ where $s_m(x)$ is the male and $s_f(x)$ the female survival function. This illustrates the basic mechanism of the cohort component method by which a population is projected from year t to year t+1. Continuing, step-by-step, we can project the population any number of years into the future with the assumption that mortality and fertility are constant either for calendar years or for longer periods. Several program packages are available for making such calculations (see e.g., Shorter, Pasta and Sendek, 1990). The package used here is Spectrum, developed by the Futures Group International in association with the Research Triangle Institute and funded by the US Agency for International Development (this package can be downloaded for free from the Internet).

## 10.2 Illustrative projection for Argentina, 1964

In 1964, the population of Argentina was about 22 million. The life expectancy for males was about 65 and for females about 71 years. The total fertility rate was about 3.1 (Keyfitz and Flieger, 1971, pp. 362-363). To illustrate the population projection technique, the population is projected to 1974 with the fertility and mortality assumptions of table 10.1. It is assumed that TFR will drop from 3.1 in 1964 to 2.0 in 1974 (a linear decline in fertility). Mortality is assumed constant. The projection will display the effects of falling fertility on the Argentinean population. We do not take migration into account.

**Table 10.1. Assumptions for the projection**[24]

| Year | TFR | Life expectancy | |
|---|---|---|---|
| | | Males | Females |
| 1964 | 3.10 | 65 | 71 |
| 1965 | 2.99 | 65 | 71 |
| 1966 | 2.88 | 65 | 71 |
| 1967 | 2.77 | 65 | 71 |
| 1968 | 2.66 | 65 | 71 |
| 1969 | 2.55 | 65 | 71 |
| 1970 | 2.44 | 65 | 71 |
| 1971 | 2.33 | 65 | 71 |
| 1972 | 2.22 | 65 | 71 |
| 1973 | 2.11 | 65 | 71 |
| 1974 | 2.00 | 65 | 71 |

Based on the mortality and fertility assumptions in table 10.1, the program package shows that the net reproduction rate declines from NRR = 1.4 to NRR = 0.9 during the period of projection. The crude birth rate declines from CBR = 23.4 to CBR = 15.6 per 1,000. The crude death rate increases from CDR = 8.3 to CDR = 9.3 per 1,000. Not realizing that the crude death rate depends on the age distribution of the population (as well as on the underlying life table), one might think that mortality has increased. However, life expectancies for males and females are fixed during the period of projection for which reason the increase in CDR is solely a reflection of the chang-

[24] A population forecast is always based on assumptions concerning mortality, fertility and migration during the projection period.

ing age distribution. It is falling fertility that causes the change in age distribution. In 1964, the proportion aged below age 5 was 10.2 percent. In 1974, it was 8.0 percent. The proportion of elderly has increased from 6.1 percent in 1964 to 7.7 percent in 1974. Table 10.3 gives births and deaths per 1,000 during the period of projection. It will be seen that the declining total fertility rates is mirrored by a drop in the total number of births. These drop from 515,000 in 1964 to 379,000 in 1974. In contrast, the number of deaths increases from 183,000 in 1964 to 227,000 in 1974.

**Table 10.2. Demographic characteristics for Argentina during projection, 1964-74**

| Year of projection | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 |
| TFR | 3.1 | 3.0 | 2.9 | 2.8 | 2.7 | 2.6 | 2.4 | 2.3 | 2.2 | 2.1 | 2.0 |
| GRR | 1.5 | 1.5 | 1.4 | 1.4 | 1.3 | 1.2 | 1.2 | 1.1 | 1.1 | 1.0 | 1.0 |
| NR | 1.4 | 1.4 | 1.3 | 1.3 | 1.2 | 1.2 | 1.1 | 1.1 | 1.0 | 1.0 | 0.9 |
| Crude rates | | | | | | | | | | | |
| CBR | 23.4 | 22.5 | 21.6 | 20.8 | 20.0 | 19.2 | 18.4 | 17.7 | 17.0 | 16.3 | 15.6 |
| CDR | 8.3 | 8.4 | 8.5 | 8.6 | 8.7 | 8.8 | 8.9 | 9.0 | 9.1 | 9.2 | 9.3 |
| Ages | Age distribution in percent | | | | | | | | | | |
| 0-4 | 10.2 | 10.1 | 10.1 | 10.0 | 9.9 | 9.7 | 9.4 | 9.0 | 8.7 | 8.3 | 8.0 |
| 5-14 | 19.7 | 19.6 | 19.5 | 19.3 | 19.2 | 19.0 | 18.9 | 18.9 | 18.7 | 18.6 | 18.4 |
| 15-49 | 51.1 | 51.0 | 51.0 | 51.0 | 51.0 | 51.1 | 51.2 | 51.3 | 51.5 | 51.6 | 51.8 |
| 50-64 | 64.1 | 64.1 | 64.1 | 64.2 | 64.3 | 64.5 | 64.7 | 65.0 | 65.3 | 65.6 | 65.9 |
| 65+ | 6.1 | 6.2 | 6.3 | 6.5 | 6.6 | 6.8 | 7.0 | 7.2 | 7.3 | 7.5 | 7.7 |

**Table 10.3. Births and deaths per 1,000 and projected population in millions, 1964-74**

| Year | Births | Deaths | Population |
|---|---|---|---|
| 1964 | 516 | 184 | 22.04 |
| 1965 | 503 | 188 | 22.35 |
| 1966 | 489 | 192 | 22.65 |
| 1967 | 476 | 197 | 22.93 |
| 1968 | 463 | 201 | 23.19 |
| 1969 | 449 | 205 | 23.43 |
| 1970 | 436 | 210 | 23.66 |
| 1971 | 422 | 214 | 23.87 |
| 1972 | 408 | 218 | 24.06 |
| 1973 | 394 | 223 | 24.23 |
| 1974 | 379 | 227 | 24.38 |

It is in place to make a few comments concerning which population is projected. Most countries, especially the Anglo-Saxon, make use of the midyear concept[25]. This is the population that would be enumerated if a census were taken on July 1. In the European Union it has been decided to work with the population as of January 1 (which is the same as the population on December 31 the previous year). Life tables and other demographic estimates are made using these end-of-year populations (the midyear population is then the average of the populations as per December 31 in the preceding year and the population a year later). It should be borne in mind that midyear populations are used for estimation of age-specific mortality and fertility (lest indirect methods are used).

[25] The reason for this is that the midyear population is an approximation to the exposure time associated with estimating age-specific mortality and fertility rates.

**Table 10.4. Argentina and its demographic characteristics, 2003**

| Population | |
|---|---|
| Both sexes | 38,740,807 |
| Males | 5,185,548 |
| Females | 4,955,551 |
| Age distribution, percent | |
| 0-14 | 26.2 |
| 15-64 | 63.4 |
| 65+ | 10.4 |
| Rates | |
| Population growth rate, percent | 1.05 |
| Crude birth rate, per 1,000 | 17.5 |
| Crude death rate, per 1,000 | 7.6 |
| Net migration rate, per 1,000 | 0.6 |
| Infant mortality rate, per 1,000 | 16.2 |
| Life expectancy at birth | |
| Both sexes | 75.5 |
| Males | 71.7 |
| Females | 79.4 |
| TFR | 2.3 |

Source: United Nations Demographic Yearbooks.

Changes of this nature (a changing age distribution) are called structural changes. Finally it should be mentioned that population projection packages often contain model fertility and mortality tables that can be used for projecting the population (Coale and Trussell, 1974). Here we have used the model west tables for projecting the Argentinean population (Coale and Demeny, 1966).

Table 10.4 gives estimates for Argentina in 2003. It is interesting to compare table 10.2 and 10.4. Notice, for example, that the proportion of elderly has increased over time, even though the total fertility rate has remained high.

## 10.3 Midyear populations

There is no country where censuses are taken once a year. Also, there are no countries where censuses produce absolutely accurate population counts; over and underenumeration occur in all censuses. Besides that, since censuses often are taken 10 years apart, intercensal midyear populations are *estimates* that are more or less accurate. In some countries population censuses are no longer taken. Examples are Denmark and Sweden which have replaced the population census by continuous population registers. The long and the short of it is that population counts whether they derive from censuses or registers are incomplete. In some countries, this incompleteness is of no serious consequence for the reliability and usefulness of demographic estimates. In others, censuses or registers may be so incomplete that only rough estimates can be obtained; there are, it must be emphasized, many countries where it is impossible to estimate life tables with reasonable precision due to faulty midyear estimates and incomplete death registration. Because in most situations the projected population has as its starting value a census enumerated population, it is important to give some thought to how accurate this count is and especially whether it is *de facto* or *de jure.*

## 10.4 *De jure* or *de facto* populations

In most developing nations censuses are taken on a *de facto* basis. Briefly, this involves that the census office before the census is taken divides the country into census enumeration areas. Maps showing the enumeration areas are then made. These are used to guide the enumerators to the households on census night (also called the census moment) and the days following census night. Returns to the question who spent the census night in each household are then recorded in the census questionnaires. At the same time, it is also noted where the persons in each household normally reside so that returns on place of usual residence also are recorded. When censuses are taken 5 or 10 years apart, cross-tabulations of place of usual residence at the last and present census then enable an understanding of streams of migration between regions.

In contrast, a census that counts the *de jure* population inquires which persons normally reside in each housing unit. This means that persons who normally reside in a household or housing unit but who are absent at the time of the census interview are listed in the census questionnaire. This sometimes leads to persons being listed who, in fact, normally reside in another country. As a result the *de jure* population may be bigger than the *de facto* population. Such errors may implant themselves in demographic estimates of mortality and fertility, especially if these are based on births and deaths recorded by the civil registration system. Hence, demographic estimates derived from the census may reflect if the population is counted on a *de facto* or *de jure* basis.

Whether a census is taken according to the *de facto* or *de jure* principle, it is important to bear in mind that usually the returns to the census questions are provided by proxies (proxy reporting). The proxy is usually the head of the household (often an older member of the household). In other words, ordinarily the returns are not obtained by interviewing each individual in the household but by the head of the household or another member of the household. This may lead to omissions and other errors in the reporting. In countries where the census questionnaires are mailed to each household (and mailed back to the census office) errors of this nature are usually much smaller. When making population projections it is important to bear in mind which population is projected, *de jure* or *de facto*.

## 10.5 Post enumeration surveys

It should also be noted that it is desirable to a conduct post-enumeration survey after a census has been taken. The mechanism of such a survey is that after the census enumeration has taken place, a new enumeration is taken in a sample of enumeration areas. The two counts are then compared so that under or over-enumeration can be assessed. Because post-enumeration surveys add to the costing of the census they are only taken occasionally. However, when a post enumeration survey has been taken some consideration should be given whether to make use of it when projecting the population. There is also the point to be taken that a census, in all events, does not cover the total population but only a large sample of it. For this reason it might seem reasonable to always link the census operation with a survey that facilitates calculation of weights that can be used to adjust the enumerated population, so that it is in better accord with the actual population at the time of the census.

To explain the meaning of a weight, consider a five-percent sample drawn from a population. A total in the sample must then be multiplied by the weight $k = 1/0.05 = 20$ in order to estimate the similar total for the population. It will be seen that a weight is the inverse of the sampling proportion. Of course, the sampling fraction in different areas of the country may not be the same for which reason weights differ across sampling areas. The software package CsPro (Census tabulation program) issued by the US Census Bureau facilitates easy production of weighted tables.

## 10.6 The exponential growth curve

The exponential model of continuous population growth is

$$P_t = P_0 e^{r t} \qquad (10.1)$$

To estimate population growth between times 0 and t, we rewrite (10.1) to obtain

$$r = \frac{1}{t} \ln \frac{P_t}{P_0} \qquad (10.2)$$

from which we note that the time T required for the population to double its size, given a constant growth rate r, is

$$T = \frac{0.693}{r} \approx \frac{0.7}{r}$$

Application of (10.2) to the population figures in table 10.3 gives an average growth rate $r = \frac{1}{10} \ln \frac{24.38}{22.04} = 0.01$ or 1 percent per year.

Despite its mathematical simplicity, the exponential growth model is highly instructive. For example, if the growth rate r is positive, sooner or later the population explodes and becomes so big that it would exhaust all available resources required for human survival. On the other hand, if the growth rate is negative, eventually the population becomes extinct. It lies in the nature of things, then, that human populations to secure their own existence must grow during certain periods and decline during others. This is also what historical studies have shown (Cox, 1970, pp. 308-315).

# 11.0 Time Series

## 11.1 Stochastic processes

Fig. 9.1, repeated below for convenience, illustrates what is known as a time-series. Specifically, when a variable, such as out-migration, is recorded over time it forms a time-series. It is a common characteristic of observed time-series that adjacent values depend on one another. Specifically, let $x_1, x_2, ..., x_n$ be the values of a variable observed at times t = 1, … , n then it is common to find high correlations between neighboring values and, not infrequently, high correlations between variables far apart in time[26]. Correlations of this kind are called auto-correlations.

Fig. 9.1 (Chapter 9) suggests that while for any given year we can predict with reasonable precision out-migration for the following year, it would be difficult to predict out-migration several years into the future. Features of this nature are common in observed time-series. We say that a process the history of which does not uniquely determine its future is a stochastic process[27]. Out-migration, then, is an example of a stochastic process. From a formal point of view we write $y_t$ for a stochastic process. If this process is observed at times t = 0, … , n, we refer to $y_0, y_1, ..., y_n$ as a time-series.

Sometimes a stochastic process admits multiple realizations, at other times it can only be observed once. For example, the time-series of out-migration from Sweden in fig. 9.1 can only be observed once; -- we cannot live through the period 1851-2002 twice. In contrast, fig. 11.1 shows the ten temperature curves for the years 1953-62 (Chatfield, 1999, p. 247). These may be viewed as 10 realizations of the same stochastic process.

---

[26] It is typical of time-series that data depend on one another. Gottman (1981, p. 41) writes:

*"Since it is so difficult to disassociate observed events from some sort of idea of occurrence in time, it seems remarkable that most of the body of statistical methodology is devoted to observations for which the temporal sequence is of no importance. Classical statistical analysis requires independence, or at least zero correlation, among observations."*

[27] A process the future of which is entirely determined by its past history is called deterministic.

**Fig. 9.1. Out migration, both sexes: Sweden, 1875-2007**

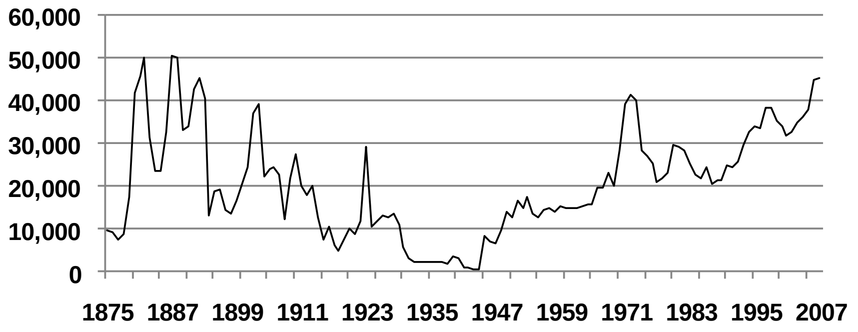
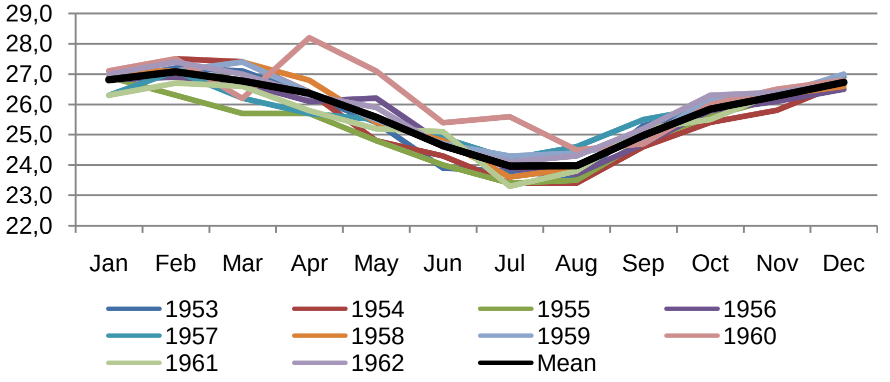



Fig. 11.1. Monthly average temperature: Recifie, 1953-62




Because many countries have reasonably long time-series of demographic data, the theory of stochastic processes has become increasingly important in the analysis of demographic phenomena, not least with respect to population projections (see e.g., Hartmann and Strandell, 2006, for important references).

## 11.2 Forecasting

We can distinguish between two different processes, namely those whose futures are completely determined by their histories (their performance in the past) and those that are not. A process the future of which is completely determined by its history is called deterministic. In contrast, as already noted, a process the future of which is not uniquely determined by its past is called stochastic (or probabilistic). Processes encountered in social science are decidedly stochastic. We cannot, for example, foretell with exactness the future total fertility rate, even if we have a very long time-series of total fertility rates.

This, of course, raises the question: What then is forecasting? In this respect, at least one aspect of sound forecasting can be mentioned, namely that we seek those features of a process that are the most time invariant, and make use of these to predict the future performance of the process. Processes that lack this property are difficult, if not impossible, to predict with reasonable precision (although exceptions exist depending on how precise we want the forecasts to be). In the second place, we need to distinguish between long-term and short-term forecasts.

## 11.3 Autoregressive time series

Mathematical modeling plays an important role in forecasting. A simple and useful model is the first-order autoregressive time-series model

$$z_t = \lambda z_{t-1} + e_t \tag{11.1}$$

In (11.1) t is time, $\lambda$ is a parameter, $|\lambda| < 1$, $z_t$ a random variable with zero mean, and $e_t$ a normally distributed error with zero mean that is independent of previous errors. Specifically, for all t, $e_t$ is independent of $e_{t-1}, e_{t-2}, \dots . e_{t-j}, \dots$ . Stated in words, every new observation is proportional to the previous one, except for the error term $e_t$ (also called an innovation). If $E(z_t) = \mu$ then (11.1) becomes

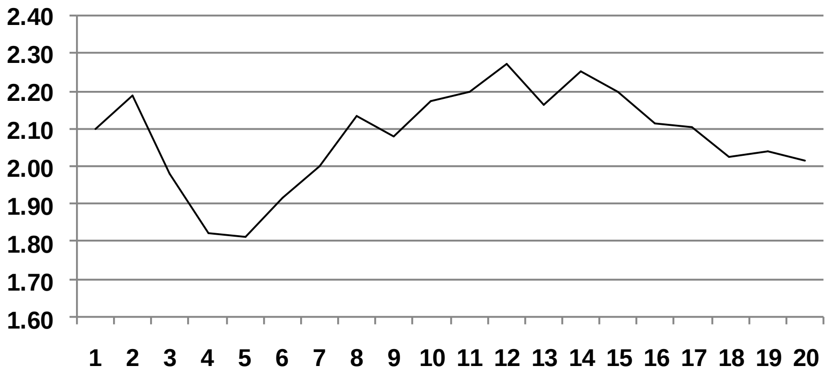
$$z_t = \lambda z_{t-1} + \mu(1-\lambda) + e_t \tag{11.2}$$

which is a more informative way of writing (11.1). In the case of (11.1), $z_t$ is expressed as $z_t - \mu$, that is, as a <u>centered variable</u>. For $\lambda = 1$, (11.1) becomes a random walk

$$z_t = z_{t-1} + + e_t$$

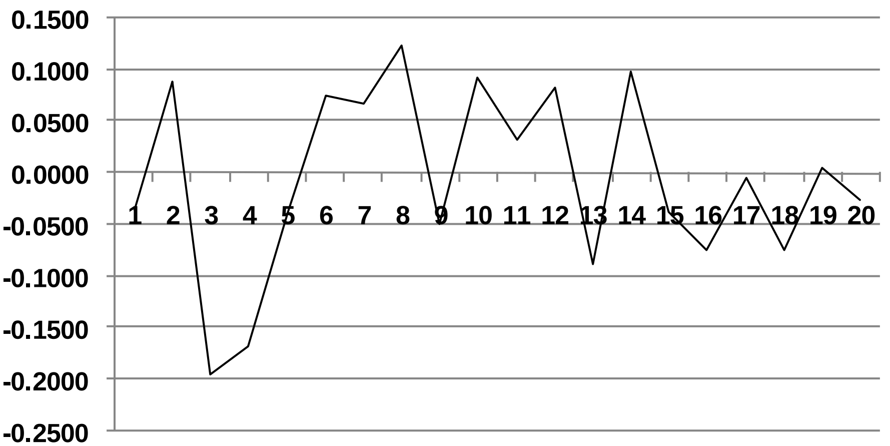
A simulation of the total fertility rate using (11.2), t = 1, … 20, is given in fig 11.2. The start value is TFR = 2.1 = μ and $\lambda = 0.9$. The innovations $e_t$ are independent and normally distributed with mean $E(e_t) = 0$ and standard deviation $\sigma_e = 0.1$ (a standard deviation similar to that for observed TFRs in Sweden, in recent years).

**Fig. 11.2. First-order autoregressive simulation of the total fertility rate (TFR)**




The graph in fig. 11.2 is surprisingly similar to observed time-series of total fertility rates (see e.g., fig. 8.1), yet it is altogether stochastic. Two important features present themselves. First, it is a tantalizing thought that if we imagine fig. 11.2 to show an observed experience then, undoubtedly, we would seek substantive explanations for the movements (trends) in the curve. Yet this would be futile since the curve merely represents correlated random behavior, -- and nothing else. Second, fig. 11.2 gives a picture of what it looks like when observations are positively correlated. In contrast, fig. 11.3 shows innovations that are independent (a time-series of the innovations underlying the graph in fig. 11.2). A time-series of independent innovations is usually called white noise (fig. 11.3).

**Fig. 11.3. White noise**




It is an important feature of (11.1 and 11.2) that both the mean and the variance of $z_t$ are time invariant. Letting $\mathrm{Var}(z_t) = \sigma_z^2$ and $\mathrm{Var}\,(e_t) = \sigma_e^2$, it follows that

$$\mathrm{Var}(z_t) = \frac{\sigma_e^2}{1-\lambda^2} \qquad (11.3)$$

The model (11.1) creates a correlated data structure. The covariance between sections[28] $z_t$ and $z_{t-1}$ is, by definition,

$$\mathrm{Cov}(z_t, z_{t-1}) = E(z_t\, z_{t-1}) = E(\lambda z_{t-1}^2 + z_{t-1} e_t) = \lambda \sigma_z^2 .$$

The covariance between sections $z_t$ and $z_{t+k}$ is

$$\mathrm{Cov}(z_t, z_{t+k}) = \lambda^k \sigma_z^2 \qquad (11.4)$$

In effect, the correlation between $z_t$ and $z_{t+k}$ is

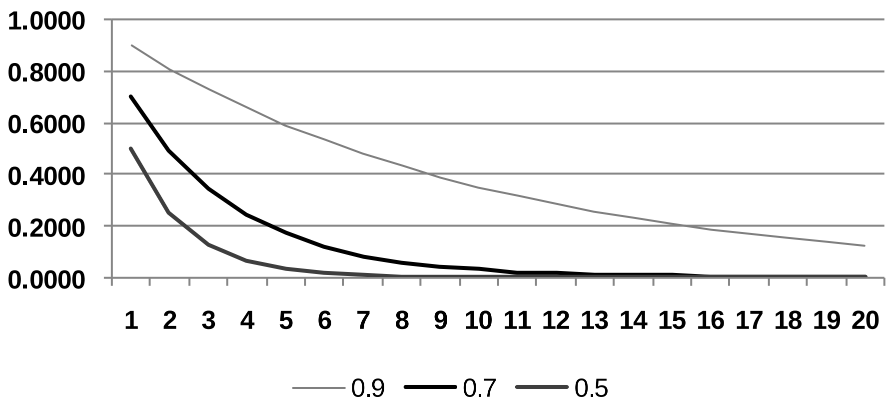
$$\rho\,(k) = \lambda^k \qquad (11.5)$$

[28] In a time series, an individual value at time t, $z_t$, is known as a section.

which is also known as the autocorrelation function. In time-series analysis and in forecasting the autocorrelation function plays an important role since it explains how closely observations across time are knitted to one another. When autocorrelations are high, what happened long ago influences current values.

Fig. 11.4 illustrates the dependence on past values in the case of the first-order autoregressive model (11.1). Time-series of total fertility rates often have autocorrelation functions similar to that in fig. 11.4 with $\lambda \approx 0.9$ so that current values are highly correlated with values some 4 or 5 years ago; a feature that enables relatively precise short-term forecasts.

**Fig. 11.4. Autocorrelation function for first-order auto-regressive process with parameters λ = 0.9, 0.8 and 0.5**




It follows from (11.1) that (the centered variable) the expected value of the series k time units into the future is

$$E(z_{t+k}) = \lambda^k z_t \tag{11.4}$$

which is called a forecast with lead time k. From (11.4) it follows that $E(z_{t+k}) \to 0$ when $k \to \infty$. This means that even if the section $z_t$ has strayed far away from its mean $\mu = 0$, a long-term forecast based on $z_t$ is the mean value $\mu = 0$. On the other hand, a short-term forecast with lead time one would be $\tilde{z}_{t+1} = E(z_{t+1}) = \lambda z_t$. At least in the case of the simple model (11.1), this illustrates the important difference between a short and a long-term forecast.

The mean square error plays an important role in forecasting. The mean square error is the squared expectation of the forecast minus the actual future value. Let $\tilde{z}_{t+k}$ be the forecast with lead time k.

For (11.1) we have $\tilde{z}_{t+k} = \lambda^k z_t$ (the forecast begins at time t+1). The mean square error is

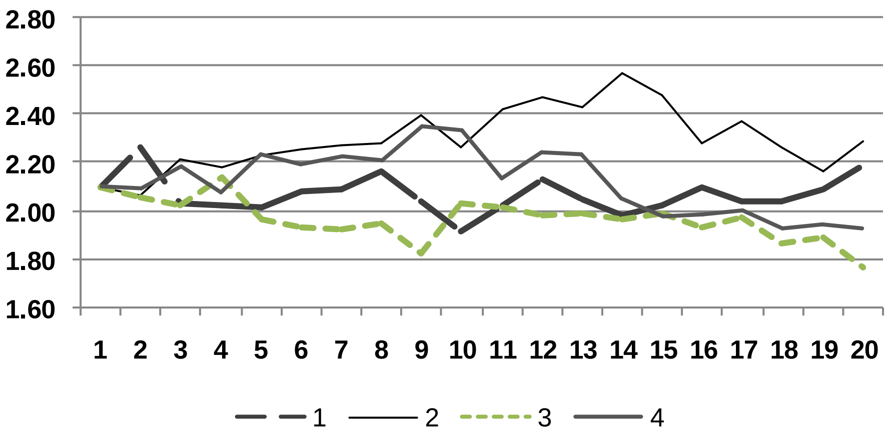
$$E(\tilde{z}_{t+k} - x_{t+k})^2 =$$

$$E(\lambda^k z_t - z_{t+k})^2 = E(\lambda^{2k} z_t^2 - 2\lambda^k z_t z_{t+k} + z_{t+k}^2) =$$

$$\lambda^{2k}\sigma_z^2 - 2\lambda^{2k}\sigma_z^2 + \sigma_z^2 = (1-\lambda^{2k})\frac{\sigma_e^2}{1-\lambda^2} \rightarrow \frac{\sigma_e^2}{1-\lambda^2} \text{ when } k \rightarrow \infty$$

which tells us that the error involved with a short-term forecast is smaller than for a long-term forecast, as indeed we would also expect. The advantage of (11.1) is that it paves the way for an uncomplicated calculation of the mean square error of a lead time k forecast.

**Fig.11.5. Four simulations of TFR for twenty-year period**




Another important concept is ergodicity. Let $z_t^j$, j = 1, … , n, be n realizations of a process. To estimate the mean at time t we would let $\mu_t = (1/n) \sum_{j=1}^{n} z_t^j$. The corresponding estimate for the variance at time t would be $\sigma_t^2 = (1/n) \sum_{j=1}^{n} (z_t^j - \mu_t)^2$. If for a large number of realizations we were to conclude that $\mu_t = \mu$ and that $\sigma_t^2 = \sigma^2$, that is, regardless of time the mean value and variance of are the same, then we would have what is known as a second-order stationary process. Such (stationary) processes share the property that the mean and variance can be estimated from just one realization. The technical term for this is that the process is *ergodic*[29] with respect to mean and variance. Ergodicity is often more or less tacitly understood to hold in projections; a topic that is beyond discussion in these notes.

Yet another aspect of (11.1) needs to be mentioned. From
$z_1 = \lambda z_0 + e_1$,
it follows that

$$z_2 = \lambda^2 z_0 + \lambda e_1 + e_2, z_3 = \lambda^3 z_0 + \lambda^2 e_1 + \lambda e_2 + e_3$$

and, generally, that

$z_n = \lambda^n z_0 + \lambda^{n-1} e_1 + ... + \lambda e_{n-1} + e_n$ which means that $z_n$, apart from a constant term, is generated by a series of n shocks or innovations. These shocks are embedded in a realization of the process so that those that took place in the remote past only have an insignificant influence on its current value (because $|\lambda| < 1$). Innovations that took place in the recent past influence the process the most.

Fig. 11.5 shows four realizations of (11.2) with μ = 2.1 and $\sigma_e = 0.1$. For each realization the start value is $z_1$ = 2.1 (a time-series model of TFR as in fig. 11.1). The process (11.2) with the above-mentioned parameter settings is stationary. Each of the four realizations can be seen as a valid forecast of the process beginning at time t = 1. Stated differently, the process has infinitely many forecasts, some of which

[29] The term ergodic is used especially by engineers. The epistemology of the word appears unknown.

would fall within *high-likelihood-event* horizons, some not. We could, of course, also assume that fig. 11.5 shows e.g., TFR for four different countries over a twenty-year period. Even though the four time series are realizations of the same underlying process, undoubtedly analysts would come up with four different theories explaining their unfolding. This raises the question: "how good are we really at explaining demographic processes?"

Seminal contributions to time series and notably autoregressive time series models were made by the Scottish statistician George Udny Yule (1871-1951) and the Russian mathematical statistician and economist Evgeny Evgenievich Slutsky (1880-1948). Markov chains, stochastic processes that are widely used in demographic research but not discussed here, were due to the Russian mathematician Andrey Andreyevich Markov (1856-1922). Another person of immense historical importance is the American mathematician Norbert Wiener (1894-1964) who introduced the theories of stochastic processes in several areas of human interest. Students who take interest in stochastic processes will run into those names time and again. Stochastic processes in demography appear among other things in the context of stochastic population projections.

# 12.0 Models in Demography

## 12.1 The Brass logit survival model

A demographic model serves the purpose of modeling something *demographic*. Over time, a plethora of models has been suggested not only with reference to an unfolding population (see Chapter 4 and 6 discussing stationary and stable populations), but also in terms of age-specific mortality and fertility (see e.g., Hartmann, 1987, for a discussion of early mortality models).

As an example, we begin by discussing a model proposed by Brass (Brass, 1968, 1971, 1975; Carrier and Goh, 1972). The reader is directed to the demographic journals for further reading on mathematical models in demography. The Brass model was developed during the 1960s in response to the need for estimating life expectancies in the absence of complete vital registration and censuses of high coverage[30]. During the 1950s it became clear that although the United Nations Organization had begun on a worldwide census program, mere census counts and age-distributions would not suffice for understanding current levels and trends of mortality and fertility. Above all, it was difficult to estimate population growth. Meanwhile, Brass had developed a method for estimating infant and childhood mortality from mothers' returns on their number of ever born and surviving children. Based on such estimates, the problem arose how to use it for estimating the life expectancy[31]. Although during the early 1950s an interesting technique had been developed by the United Nations for solving the problem[32], Brass went further and gave a solution (the Brass logit survival model) that would even produce easily calculated population projection probabilities. Hence, his model also performed as an uncomplicated aid in making population projections. As already noted, it is important to note that even today life expectancies for the majority of developing nations are estimated from data on infant and childhood mortality. Estimates of

[30] A census would have high coverage if underenumeration is less than 5 percent. In developing nations, even today, it is not uncommon that underenumeration is much higher than that.

[31] During the early 1950s, the United Nations developed the first model life tables for such estimation.

[32] Coale and Demeny (1966) expanded further on this idea when they developed the historical Princeton model life tables.

this nature, it must be emphasized, are often somewhat inauspicious. However, because this is the common approach to estimating life expectancies in developing nations (due to incomplete vital registration), a sketch of the approach is given below.

Brass noted that life table survival functions can be related in the sense that for a chosen standard survival function $l_x^s$, another survival function $l_x$ could be expressed as

$$\text{logit } l_x \approx \alpha + \beta \text{ logit } l_x^s \tag{12.1}$$

where $\text{logit } l_x = \ln \frac{1 - l_x}{l_x}$ and $\alpha$ and $\beta$ are parameters (Brass, 1971, 1974; Hill and Trussell, 1977).

Fig. 12.1 illustrates the logit relationship (12.1) using life tables for Swedish males in 1932 and 1942. It will be noted that the two curves are similar, except that they are at different levels. It is the purpose of the linear relationship (12.1) to move one curve on top of the other. Because (12.1) expresses a linear relationship, $\alpha$ and $\beta$ can be estimated by the method of least squares (ordinary linear regression estimates). Using the 1932 survival function as a standard, least squares estimates are $\hat{\alpha} = -0.323$ and $\hat{\beta} = 1.062$. This means that, on the logit scale, fitted (or modeled) survival for 1942 is

$$\text{logit } \hat{l}_x \approx -0.323 + 1.062 \text{ logit } l_x^s \tag{12.2}$$

The negative value of $\hat{\alpha} = -0.323$ transfers the 1932 logit curve down to the level of the 1942 logit curve. The parameter value $\hat{\beta} = 1.062$ increases the slope of the 1932 logit curve so that it is in harmony with the slope of the 1942 logit curve. The functionality of the parameters in (12.2) is that (i) $\alpha$ moves the standard curve up or down to the level of the logit curve to be modeled and that (ii) $\beta$ adjusts (twists) the slope of the standard curve so that it is agreement with the slope of the logit curve to be modeled. For a more accurate description of the parameters in (12.1), see chapter 14 on the logistic distribution.

**Fig. 12.1. Logit of survival for males: Sweden 1932 and 1942**

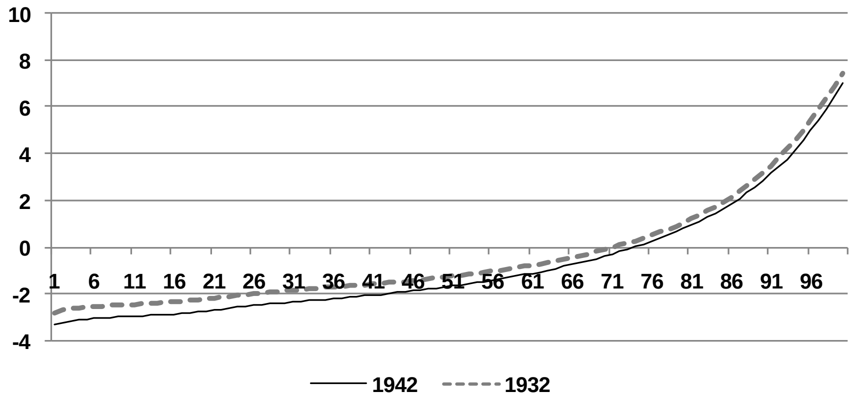



Fig. 12.2 shows that the fitted logit curve very nearly coincides with the observed one for 1942. From (12.2) it follows that modeled survival is

$$\hat{l}_x = \frac{1}{1 + e^{\alpha} \left( \frac{1 - l^s_x}{l^s_x} \right)^{\beta}} \qquad (12.3)$$

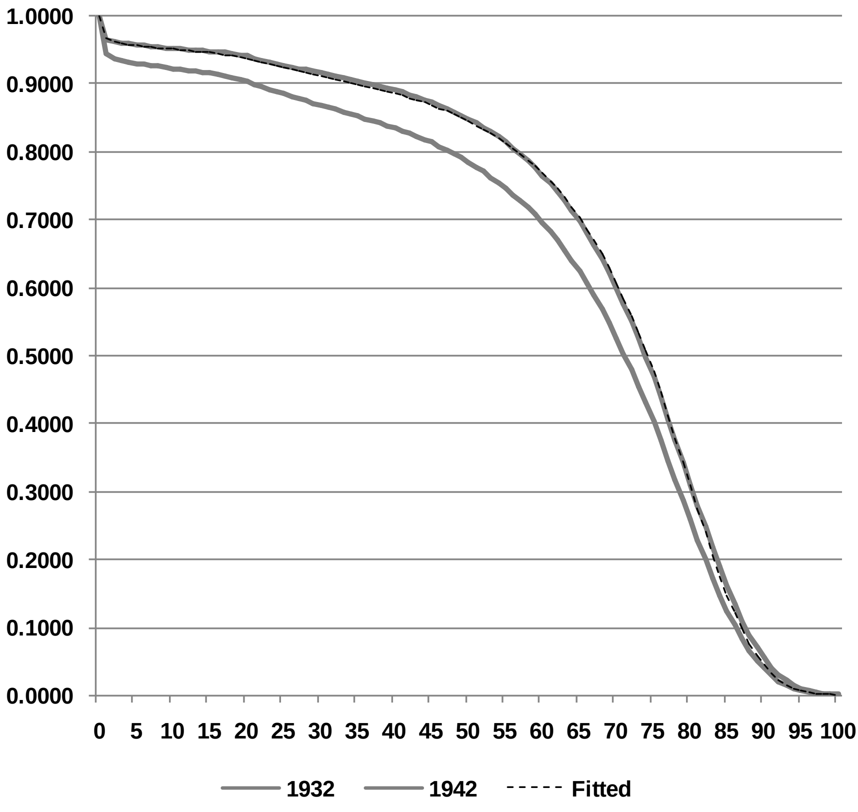
Using the above-mentioned parameter estimates, modeled survival for 1942 becomes

$$\hat{l}_x = \frac{1}{1 + e^{-0.323} \left( \frac{1 - l^s_x}{l^s_x} \right)^{1.062}} \qquad (12.4)$$

The fit to the 1942 survival curve is very close even though the standard is much different. In fact, while the life expectancy for the standard is 63.0 it is 67.6 years for 1942 survival and 67.4 years for fitted survival. In practice, this means that (12.3) is a well-chosen method of projecting survival a few years into the future. It is worth noting that this Brass model is an example of one of the finest pieces of mathematical modeling in demography.

In some situations, it is appropriate to assume that $\beta = 1$. In this event (12.1) reduces to a one-parameter model for which $\alpha$ can be estimated from estimates of infant and child mortality (a technique that is described in the demographic literature on indirect estimation).

**Fig. 12.2. Observed and modeled survival for males: Sweden 1942 (1932 male survival is standard)**



Here we give a short version of the method. If $\beta = 1$, then modeled survival is

$$\hat{l}_x = \frac{1}{1 + e^{\alpha}\left(\frac{1 - l_x^s}{l_x^s}\right)} \tag{12.5}$$

from which it follows that

$$\alpha = \ln \frac{(1 - l_x)/l_x}{(1 - l_x^s)/l_x^s} \tag{12.6}$$

(referred to as the log-odds ratio).

If infant mortality is estimated[33] at $q_0 = 1 - l_1$, and if $l_x^s$ is a conveniently chosen standard survival function thought to be similar to the one to be estimated, then it follows from (12.6) that

$$\hat{\alpha} = \ln \frac{q_0/l_1}{q_0^s/l_1^s} \tag{12.7}$$

Using the parameter estimate (12.7), estimated survival is

$$\hat{l}_x = \frac{1}{1 + e^{\hat{\alpha}}\left(\frac{1 - l_x^s}{l_x^s}\right)}$$

[33] This could be an estimate from a survey or from the population census using a variety of methods.

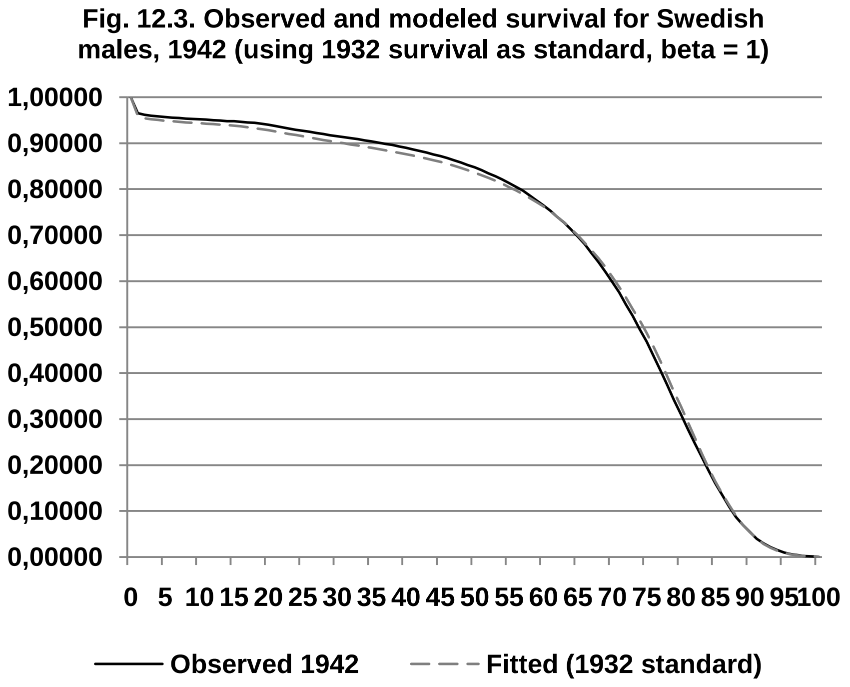



Fig. 12.3 shows the result of fitting (12.5) to 1942 survival using 1932 survival as a standard. The life expectancy for fitted survival is 67.2 years. It should be noted though that such close fits not always can be obtained. The present example capitalizes on the standard having the same essential age-pattern of mortality as 1942 male survival.

Estimating a life table from a single index, such as infant mortality, necessarily must involve a degree of error (sometimes a fairly large one), even if there is a high correlation between infant mortality and the life expectancy. Furthermore, this degree of error cannot be gauged statistically in terms of confidence limits. In the end, it is some sort of impressionistic fitting process.

## 12.2 Singular value decomposition

As a prelude to discussing the Lee-Carter mortality model (usually abbreviated LC), it is necessary first to discuss singular value decomposition. Any n-by-m matrix (a matrix with n rows and m columns) can be written as a product of three matrices. Letting A be any n-by-m matrix, the factorization involves that

$$A= U S V^T \tag{12.8}$$

U is an n-by-n orthonormal[34] matrix, S is an n-by-m matrix with non-negative numbers in its diagonal and zeroes off its diagonal, and $V^T$ denotes the transpose of an m-by-m orthonormal matrix V[35]. The orthonormal column vectors $u_k$ $(k = 1,\dots,n)$ in U, and column vectors $v_h$, h = 1, … , m in V are called left and right singular vectors, respectively. The singular values of A are the square roots of the eigenvalues of $A^T A$. The singular values $s_i$ in S (usually arranged in descending order) satisfy $A v_i = s_i u_i$, i = 1, … , m, so that each right singular vector is mapped onto the corresponding left singular vector with magnification factor $s_i$.

A major advantage of SVD is that it often (but not always) enables computing good approximations to A. This is accomplished by neglecting the smaller of the singular values in S. The approximation to A based on the first k (k < m) singular values is

$$A \approx A_k = u_1 s_1 v_1^T + \dots + u_k s_k v_k^T \tag{12.9}$$

The partial terms $u_i s_i v_i^T$ in (12.9) are called the principal images (Golub and van Loan, 1996; Hansen, 1987; Horn and Johnson, 1985; Strang, 1998). An early paper on SVD is due to Eckhardt and Young (1936). It is often adequate to make use of only the first principal image and let

[34]Two vectors $\underline{a}$ and $\underline{b}$ are orthonormal if their inner product $\underline{a}\,\underline{b} = |\underline{a}|\,|\underline{b}| \cos\theta = 0$ where $\theta$ is the angle between $\underline{a}$ and $\underline{b}$. If two vectors $\underline{a}$ and $\underline{b}$ have scalar product $\underline{a}\,\underline{b} = 0$, they are said to be independent, otherwise correlated. In a Hilbert space, an angle is an inner product.

[35] In what follows, we write V' for the transpose of V.

$$A \approx A_1 = u_1 s_1 v_1^T \qquad (12.10)$$

The LC model takes advantage of the approximation (12.10). Consider the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ 8 & 11 & 14 \end{pmatrix}$$

This matrix has singular value decomposition[1] A = U S V'

$$\text{with } U = \begin{pmatrix} 0.16775 & 0.98115 & -0.09594 \\ 0.43074 & 0.01459 & 0.90236 \\ 0.88675 & -0.19269 & -0.42017 \end{pmatrix},$$

$$S = \begin{pmatrix} 22.011 & 0 & 0 \\ 0 & 0.617 & 0 \\ 0 & 0 & 0.368 \end{pmatrix}$$

and

$$V = \begin{pmatrix} 0.40819 & -0.81419 & 0.41288 \\ 0.55624 & -0.13680 & -0.81968 \\ 0.72386 & 0.56425 & 0.39705 \end{pmatrix}.$$

The diagonal of S holds the three singular values. Letting

$$F = \begin{pmatrix} 22.011 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

contain only the first singular value of S (the two remaining diagonal elements being zero in F), the matrix

[1] Several statistical packages perform SVD. Here we have used STATA.

$$G=UFV'=\begin{pmatrix}1.507 & 2.054 & 2.673\\ 3.870 & 5.273 & 6.863\\ 7.967 & 10.856 & 14.128\end{pmatrix}\approx A$$

provides a first principal image approximation to A. Compare G and A to see how well G approximates A.

## 12.3 The LC model

The Lee-Carter (LC model) takes advantage of time-series of central deaths rates. Consider a time-series of central death rates m(x; t) where x is age and t is time, then

$$\bar{\mu}(x) = (1/m) \sum_{t=1}^{m} \log m(x;t) \tag{12.11}$$

is the mean of the logged central death rates[36] at age x across m time-periods, and

$$\mu(x;t) = \log m(x;t) - \bar{\mu}(x) \tag{12.12}$$

are the elements of the centered $\omega - by - m$ matrix

A = ( $\mu(x;t)$ ) ( $x = 0,\ldots,\omega$ and t = 1, … , m). Here $\omega$ ( $\omega \geq m$ ) denotes the highest age at which survival is considered. A can be factorized in agreement with (12.8), that is, as A = U S V'.

Using first principal image approximation (12.10), the logged rate is

$$\log m(x;t) = \bar{\mu}(x) + \rho(1)\, k(t)\, b(x) \tag{12.13}$$

$\rho(1)$ being the first singular value, k(t) (t = 1, … , m) the first column vector in V and b(x) (x = 0, … , $\omega$) the first column vector in U. Because the matrix determined by (12.12) is centered, it follows that $\sum_{t=0}^{m} k(t) = 0$ (see e.g., Wilmoth, 1993 for a discussion of estimation principles). In demographic contexts (12.13) has become known as the Lee-Carter (LC) mortality model (Booth, Maindonald and Smith, 2002; Carter and Prskawetz, 2001; Girosi and King, 2005; Lee and Carter, 1992). It should be mentioned that apparently the first speci-

[36] The logged central death rate is used in order to ensure that the modeled rate does not become negative.

fication of (12.13) as a model of age-specific mortality is due to de Gomez (1990) (Lee and Miller, 2000). Because b(x) and k(t) are unobservable, ordinary linear regression cannot be used directly. This is why ordinarily the model is estimated by means of singular value decomposition. In effect, LC is an SVD application that shares the properties of first image approximations. In effect, the LC model can only be used successfully when (a) time-series of central death rates are available and (b) over time, there is either a relatively uniform increase or decrease in age-specific mortality across all ages.

The age-series b(x) in (12.13) is often highly serrated, as illustrated for Sweden (fig. 12.4). The serration reveals that the time changes in mortality only have followed the assumption underlying the LC model to an approximate extent; nevertheless the modeled life expectancies (fig. 12.5 showing results for 1980-2005) are relatively close to the observed ones. It should be noted that b(x) may attain negative values (this sometimes happens at older ages). The proportionality factors k(t) for males and females drop almost linearly over time (fig. 12.6); a feature made use of in forecasting. Extrapolation of this type of mortality trend is usually accomplished by a random walk with drift

$$k_t = k_{t-1} + k + h_t$$

where $h_t$ is a zero mean normally distributed innovation[37] with variance $\sigma_h^2$ and k a drift parameter that determines the average speed with which $k_t$ changes, that is,

$$k = \Delta k_t / \Delta t = [k_{t(0)} - k_{t(m)}]/m$$

where t(0) and t(m) are the first and last time points of the observed $k_t$ series. It is important to realize that the variability induced by letting $k_t$ wander as a random walk with drift does <u>not</u> account for the total temporal variability in mortality!

---

[37] In a dynamic forecast, $h_t = 0$.

**Fig. 12.5. Observed and LC-fitted life expectancies by sex: 1980-2005**

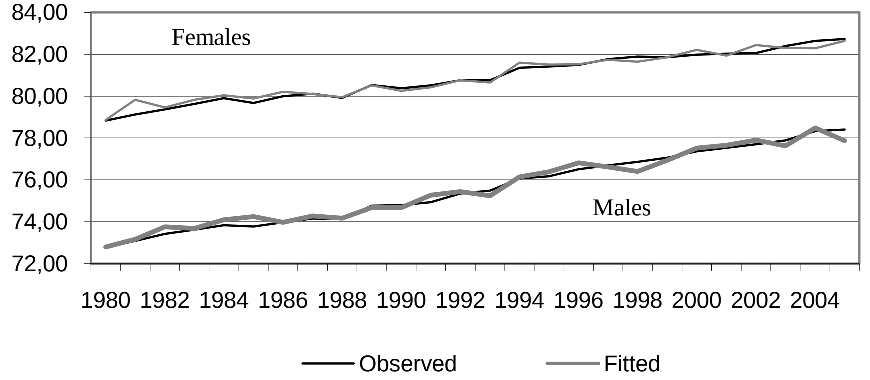



Fig. 12.4. The b(x) age-series for males and females, Sweden 1980-2005

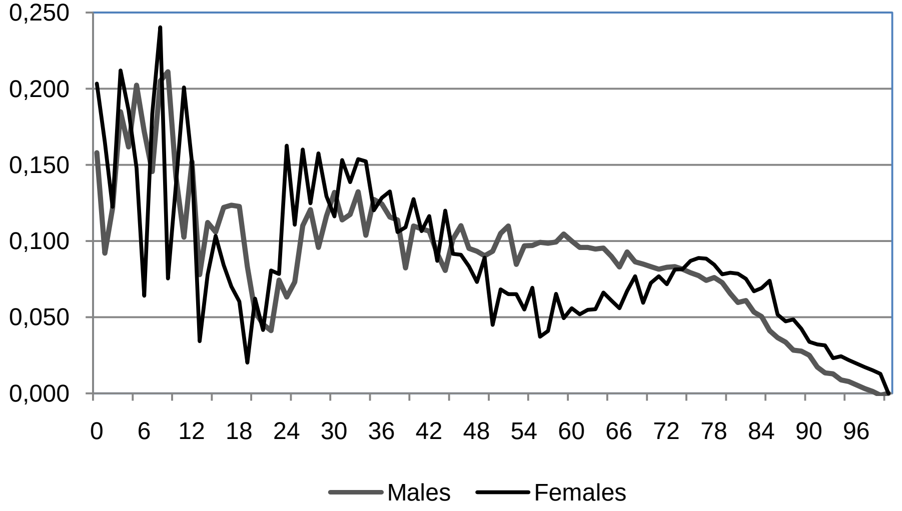


Fig. 12.6. Time-patterns for k(t), 1970-2005.


## 12.4 Models, data and documentation

No model is perfect. If indeed this were so, we would live in a world much different from the one we experience. A mathematical model, be it with few or many parameters, will never completely describe the complexities determining survival and other demographic aspects of human life. Demography is an academic discipline which, like any other, is subject to approximations. While some approximations may be better than others, nevertheless they remain approximations. This imposes the demand for sound judgment. It is always important, and indeed necessary, to appraise the results that derive from applications of various methods. No method, however ingenious, can repair incomplete and erroneous data. For this reason one must always assess how much analysis the collected data can support. This is often an initial aspect of demographic analysis, -- especially in developing nations.

# 13.0 Indirect Demographic Estimation

## 13.1 Estimating infant and child mortality

The actuarial estimation approach involves that both events and exposure time are available. In contrast, indirect demographic estimation references a situation where we wish to estimate a life table or a fertility schedule, or perhaps some other rates, although the ordinarily required event-exposure data are not available. In modern times Brass (1968) was among the first to address these problems[38]. As previously noted, his methods became known as indirect.

Over time several contributions were made to indirect estimation (see e.g., Brass, 1971; Brass et al, 1968; Carrier and Goh, 1972; Coale and Demeny, 1966; Coale and Trussell, 1974; Courbage and Fargues, 1979; Feeney, 1980; Hartmann, 1991; Palloni, 1980; Sullivan, 1972; Retherford and Cho, 1970; Trussell, 1975; Hill and Trussell, 1977; United Nations, 1968).

In the case of child mortality, the solution involved asking women in a census or survey (a) how many live births they had ever had (ever born children before the time of the census or survey) and (b) how many of those are still alive (surviving children at the time of the census or survey). The estimation method is outlined below using single-year reports from mothers on their deceased children . Given a uniform age-distribution of women, equally many women in each single-year age group,

$$D_S(x) = \frac{\int_{\alpha}^{x} f(a)\, q_S(x-a)\,da}{\int_{\alpha}^{x} f(a)\,da} \qquad (13.1)$$

is the proportion of deceased children to be reported by women aged x if between ages $\alpha$ and x they have constant fertility f(a) and their newborns have constant risk of dying before age x, denoted

[38] Interestingly enough there are many traces of indirect estimation philosophy in Graunt's works from the 17th century (Benjamin, Brass and Glass, 1963).

$q_S(x)$. Brass assumed that mortality functions q(x) = 1-l(x) (the probability of dying before age x) at different levels of mortality are proportional. With this assumption the mortality function to be estimated can be expressed as

$$q(x) = k\ q_S(x) \tag{13.2}$$

(at childhood ages this is a close approximation). If for conveniently chosen mortality $q_S(x)$ and fertility f(a), model proportions of deceased children are calculated, these can be used to estimate k in (13.2) so that estimated child mortality becomes

$$\hat{q}(x) = \hat{k}\, q_S(x) \tag{13.3}$$

As an aid for understanding the rationale of the method, it will be noted that from (13.1) and the mean value theorem for integrals it follows that there is an age $y$ with $0 < y < x - \alpha$ such that

$$D_S(y) = q_S(y) \tag{13.4}$$

for which reason the proportion of deceased children reported by women aged x is the same as the probability $q_S(y)$ for newborns to die before age y.  Given numerical specifications of $q_S(x)$ and f(a), y can be found by interpolation. The proportionality factor k in (13.2) is usually estimated as

$$\hat{k} = H(x) / D_S(x) \tag{13.5}$$

where H(x) is the observed proportion of deceased children reported by women aged about 20 years. In practice, k can be estimated by least squares from reports H(18), … , H(22). This is the *modus operandi* of the method (see also Hartmann, 1991; Sullivan, 1972 and Trussell, 1975 for variations of the original Brass method). The estimate (15.5) requires that the fertility of women also is estimated. This will be illustrated below. It will be realized that this method necessarily must give no more than a rough approximation to infant and childhood mortality.

First, it assumes that different cohorts of women have the same fertility (which is rarely the case). Second, it is assumed that children's mortality is independent of mother's age. There is however a clear tendency for children with teenage mothers to have much higher mortality than children with older more mature mothers. Third, the survival of children whose mothers are dead at the time of the census is not reported in the census or survey (a phenomenon known as left-censoring in retrospective surveys); one may suspect that such children have elevated mortality risks. Fourth, the fertility function f(a) in (13.1) can only be inferred indirectly and, of course, introduces yet another dimension of imprecision in estimated mortality. Fifth, the estimated mortality function will rarely apply to the time of the census or survey. In the case of falling mortality, the estimate would refer to a time point before the census (Brass, 1975; Feeney, 1980, 1991; Palloni, 1980). Nevertheless, whether data reference censuses or surveys this is the general method by which infant and child mortality is estimated for the majority of developing nations, -- even today.

## 13.2 Indirect estimation of fertility

Indirect estimation of fertility involves making use of the parity information obtained from a census or survey. The parity at age x is the mean number of children women have given birth to at that age. We assume that all women considered in the estimation process have the same fertility schedule[39] $\{f_a\}$. Given a uniform age-distribution of women, the mean parity for women aged 15-19 is

$$P_1 = \left[\frac{1}{2}\sum_{15}^{19} f_j + 4f_{15} + 3f_{16} + 2f_{17} + f_{18}\right]/5, \qquad (13.6)$$

for women aged 20-24

$$P_2 = \sum_{15}^{19} f_j + \left[\frac{1}{2}\sum_{20}^{24} f_j + 4f_{20} + 3f_{21} + 2f_{22} + f_{23}\right]/5, \qquad (13.7)$$

and for women aged 25-29

$$P_3 = \sum_{15}^{24} f_j + \left[\frac{1}{2}\sum_{25}^{29} f_j + 4f_{25} + 3f_{26} + 2f_{27} + f_{28}\right]/5 \qquad (13.8)$$

If women are asked if they have given live birth to a child during the 12 months before the census, the returns can be used to estimate age-specific fertility rates. Women aged x in the census are assumed

[39] It would perhaps be more appropriate to assume that the population is stationary, that is, mortality and fertility are time-invariant (and that the population is closed to migration).

to be aged x+0.5 years on the average. Because those who gave birth did so, on average, half a year earlier, age-specific fertility from the returns corresponds to exact ages 15, 16, … , 49. Adjustment for this half-year displacement however is not very important.

The general experience with retrospective reports of this nature is that they underestimate age-specific fertility (not all births are reported). However, if it is assumed that underreporting of births is independent of mother's age, then estimated age-specific fertility can be adjusted so that it is in agreement with the census reported mean parities.

To this end, let $\hat{g}_a$ be the age-specific fertility rate at age a, as estimated from births during the 12 months before the census. Fitting a Brass-fertility polynomial (or some other convenient expression) to $\hat{g}_a$ gives new graduated rates $\hat{f}_a$. Replacing $f_j$ in (13.6)–(13.8) by $\hat{f}_a$ gives estimated parities $\hat{P}_i, i = 1, 2, 3$ corresponding to the births reported for the 12 months before the census. Letting $\tilde{P}_i, i = 1, 2, 3$ be the mean parities reported in the census (from children ever born), ratio estimates

$$\hat{\gamma}_i = \tilde{P}_i / \hat{P}_i \tag{13.9}$$

i = 1, 2, 3 can be obtained.

In practice, the ratios $\hat{\gamma}_2$ and $\hat{\gamma}_3$ are used to estimate current fertility (fertility at the time of the census). This means that estimated current fertility, if based on the census reported mean parities for women aged 20-24, is

$$\hat{h}_a = \hat{\gamma}_2 \hat{f}_a \tag{13.10}$$

If based on the census reported mean parities for women aged 25-29, it is

$$\hat{h}_a = \hat{\gamma}_3 \hat{f}_a \tag{13.11}$$

Sometimes one lets

$$\hat{h}_a = \bar{\gamma} \hat{f}_a \tag{13.12}$$

with $\bar{\gamma} = (\hat{\gamma}_2 + \hat{\gamma}_3)/2$

be the estimate of current fertility. As noted, age-specific fertility estimated in this manner is based on the assumptions that women have a uniform age-distribution and share the same fertility at least up to age 30. It is also assumed that births are equally underreported by women at all ages. Nevertheless, in practice, there are always changes in fertility from one year to the next, and the age-distribution of women is never uniform. In addition, it is not very likely that births during the 12 months before the census are under-reported independently of age. In consequence, age-specific fertility estimated in this manner may be rather approximate. The above-mentioned estimation method is due to Brass and usually referred to as the P/F-method (Brass et al, 1968).

## 13.3 An application to the 2005 LAO PDR population census

Indirect estimation of child mortality and fertility is illustrated below using data from the 2005 Lao PDR population census. Table 13.1 gives children ever born and surviving children. Usually these statistics are given by five-year age groups of women. Here however are shown the returns from the census by single-year ages of women. This is advantageous because, as we shall see below, five-year age-groups may disguise important features of such data. Fig. 13.1 shows the proportions of deceased children by age of mother. As previously noted, infants and children with teenage mothers often (if not always) have much higher mortality than children with older mothers. This is exemplified by fig. 13.1. Infant mortality is estimated at 70 per 1,000 live births (the average of the proportions deceased children reported by women aged 20-24).

This is not a true Brass estimate, however data on ever born and surviving children are often somewhat incomplete and for this reason do not uphold application of a refined estimation technique, which almost always builds on assumptions that are unlikely to be met. In the case of Lao PDR the data do not seem to support a finer estimate of infant mortality. Table 13.2 shows reported births during the 12 months before the census, reporting women by age, and estimated age-specific fertility rates.

The total number of recorded births is 114,442 and the total fertility rate (TFR) is estimated at TFR = 2.62. Previous estimates were TFR = 5.5 in the 1995 Lao PDR Population Census and TFR = 5.0 from the 2000 Lao PDR Reproductive Health Survey. The resulting TFR = 2.62 is obviously too low, thus reflecting considerable underreporting of births in the census. The total census population being 5,621,982, the crude birth rate is estimated at CBR = 114,442/5,621,982 = 20.4 per 1,000 which, consequently, is also too low.

In order to calculate age-specific fertility rates corresponding to exact age x+0.5, adjusted rates can be obtained as a mean of rates at ages x and x+1. The adjusted rates facilitate calculation of mean parities corresponding to the reported births during the 12 months before the census. These mean parities are shown in table 13.3 together with the census reported mean parities as well as the ratios between the two kinds of parities.

**Table 13.1. Children ever born and surviving children, 2005 Lao PDR Population Census**

| Age | Women | Children ever born | Children surviving | Proportion surviving children | Proportion deceased children |
|---|---|---|---|---|---|
| 15 | 72,672 | 1,022 | 898 | 0.8787 | 0.1213 |
| 16 | 64,408 | 2,721 | 2,461 | 0.9044 | 0.0956 |
| 17 | 58,632 | 5,949 | 5,445 | 0.9153 | 0.0847 |
| 18 | 71,979 | 15,877 | 14,616 | 0.9206 | 0.0794 |
| 19 | 55,849 | 18,876 | 17,508 | 0.9275 | 0.0725 |
| 20 | 71,247 | 44,189 | 40,864 | 0.9248 | 0.0752 |
| 21 | 45,675 | 30,869 | 28,793 | 0.9327 | 0.0673 |
| 22 | 53,667 | 50,285 | 46,969 | 0.9341 | 0.0659 |
| 23 | 45,847 | 50,797 | 47,485 | 0.9348 | 0.0652 |
| 24 | 44,935 | 59,849 | 55,950 | 0.9349 | 0.0651 |
| 25 | 58,017 | 96,971 | 89,807 | 0.9261 | 0.0739 |
| 26 | 38,931 | 71,212 | 66,272 | 0.9306 | 0.0694 |
| 27 | 38,495 | 78,111 | 72,581 | 0.9292 | 0.0708 |
| 28 | 45,904 | 109,606 | 100,916 | 0.9207 | 0.0793 |
| 29 | 36,983 | 92,859 | 85,895 | 0.9250 | 0.0750 |
| 30 | 55,777 | 160,412 | 145,892 | 0.9095 | 0.0905 |
| 31 | 30,093 | 85,798 | 79,302 | 0.9243 | 0.0757 |
| 32 | 36,759 | 113,267 | 104,141 | 0.9194 | 0.0806 |
| 33 | 30,221 | 96,654 | 88,716 | 0.9179 | 0.0821 |
| 34 | 30,630 | 104,649 | 95,611 | 0.9136 | 0.0864 |
| 35 | 45,012 | 164,880 | 148,935 | 0.9033 | 0.0967 |
| 36 | 31,052 | 117,195 | 106,159 | 0.9058 | 0.0942 |
| 37 | 28,438 | 109,638 | 99,120 | 0.9041 | 0.0959 |
| 38 | 35,008 | 142,924 | 128,288 | 0.8976 | 0.1024 |
| 39 | 25,852 | 107,243 | 96,314 | 0.8981 | 0.1019 |
| 40 | 41,255 | 178,141 | 156,719 | 0.8797 | 0.1203 |
| 41 | 21,888 | 94,013 | 84,186 | 0.8955 | 0.1045 |
| 42 | 26,292 | 116,943 | 103,625 | 0.8861 | 0.1139 |
| 43 | 21,843 | 98,991 | 87,759 | 0.8865 | 0.1135 |
| 44 | 22,254 | 102,191 | 90,306 | 0.8837 | 0.1163 |
| 45 | 34,550 | 158,078 | 137,135 | 0.8675 | 0.1325 |
| 46 | 20,424 | 95,254 | 83,131 | 0.8727 | 0.1273 |
| 47 | 19,076 | 88,553 | 76,980 | 0.8693 | 0.1307 |
| 48 | 22,849 | 107,039 | 92,097 | 0.8604 | 0.1396 |
| 49 | 16,399 | 76,192 | 65,521 | 0.8599 | 0.1401 |

Source: Central Statistical Bureau of Lao PDR.

Fig. 13.1. Proportion deceased children by age of mothers, 2005 Lao PDR population census

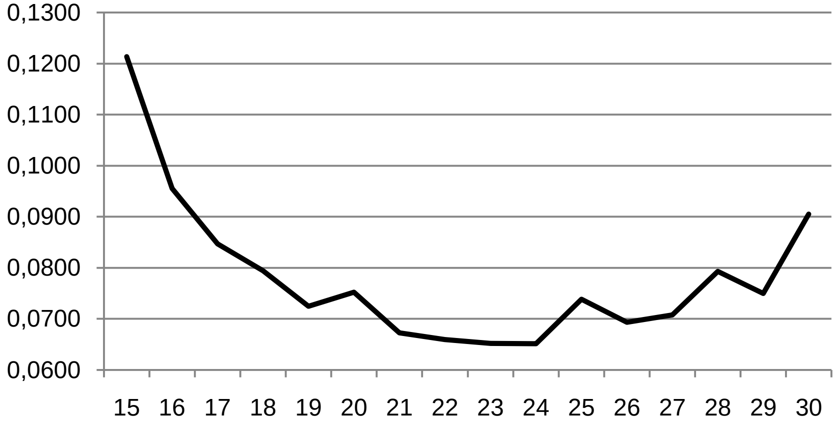



**Table 13.2. Births during the 12 months before the census and age-specific fertility, 2005 Lao PDR population census**

| Age | Women | Births | f(x) | Age | Women | Births | f(x) |
|---|---|---|---|---|---|---|---|
| 15 | 72,672 | 430 | 0.0059 | 33 | 30,221 | 2,713 | 0.0898 |
| 16 | 64,408 | 1,151 | 0.0179 | 34 | 30,630 | 2,617 | 0.0854 |
| 17 | 58,632 | 2,299 | 0.0392 | 35 | 45,012 | 3,735 | 0.0830 |
| 18 | 71,979 | 5,186 | 0.0720 | 36 | 31,052 | 2,172 | 0.0699 |
| 19 | 55,849 | 5,146 | 0.0921 | 37 | 28,438 | 1,863 | 0.0655 |
| 20 | 71,247 | 8,965 | 0.1258 | 38 | 35,008 | 2,148 | 0.0614 |
| 21 | 45,675 | 5,575 | 0.1221 | 39 | 25,852 | 1,358 | 0.0525 |
| 22 | 53,667 | 7,357 | 0.1371 | 40 | 41,255 | 1,847 | 0.0448 |
| 23 | 45,847 | 6,341 | 0.1383 | 41 | 21,888 | 846 | 0.0387 |
| 24 | 44,935 | 6,366 | 0.1417 | 42 | 26,292 | 852 | 0.0324 |
| 25 | 58,017 | 8,471 | 0.1460 | 43 | 21,843 | 630 | 0.0288 |
| 26 | 38,931 | 5,499 | 0.1412 | 44 | 22,254 | 511 | 0.0230 |
| 27 | 38,495 | 5,095 | 0.1324 | 45 | 34,550 | 696 | 0.0201 |
| 28 | 45,904 | 6,141 | 0.1338 | 46 | 20,424 | 322 | 0.0158 |
| 29 | 36,983 | 4,483 | 0.1212 | 47 | 19,076 | 227 | 0.0119 |
| 30 | 55,777 | 6,497 | 0.1165 | 48 | 22,849 | 272 | 0.0119 |
| 31 | 30,093 | 2,943 | 0.0978 | 49 | 16,399 | 161 | 0.0098 |
| 32 | 36,759 | 3,527 | 0.0959 | | | | |

Source: Central Bureau of Statistics, Lao PDR.

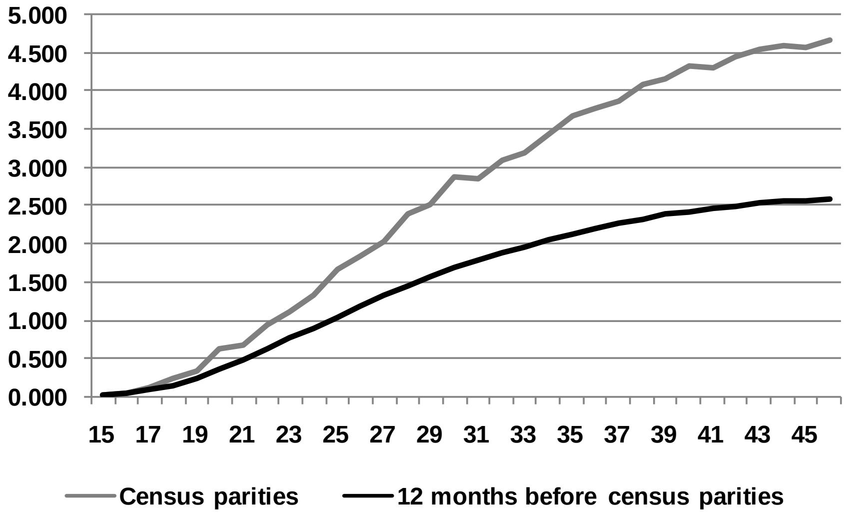
**Table 13.3. Census reported mean parities, mean parities corresponding to births 12 months before the census and their ratios, 2005 Lao PDR population census**

| Age | CEB parity | 12 month parity | Parity ratio | Age | CEB parity | 12 month parity | Parity ratio |
|---|---|---|---|---|---|---|---|
| 15 | 0.014 | 0.006 | 2.36 | 31 | 2.851 | 1.778 | 1.60 |
| 16 | 0.042 | 0.026 | 1.61 | 32 | 3.081 | 1.872 | 1.65 |
| 17 | 0.101 | 0.068 | 1.49 | 33 | 3.198 | 1.963 | 1.63 |
| 18 | 0.221 | 0.137 | 1.61 | 34 | 3.417 | 2.049 | 1.67 |
| 19 | 0.338 | 0.233 | 1.45 | 35 | 3.663 | 2.129 | 1.72 |
| 20 | 0.620 | 0.349 | 1.78 | 36 | 3.774 | 2.201 | 1.71 |
| 21 | 0.676 | 0.476 | 1.42 | 37 | 3.855 | 2.267 | 1.70 |
| 22 | 0.937 | 0.610 | 1.54 | 38 | 4.083 | 2.327 | 1.75 |
| 23 | 1.108 | 0.748 | 1.48 | 39 | 4.148 | 2.380 | 1.74 |
| 24 | 1.332 | 0.890 | 1.50 | 40 | 4.318 | 2.425 | 1.78 |
| 25 | 1.671 | 1.034 | 1.62 | 41 | 4.295 | 2.463 | 1.74 |
| 26 | 1.829 | 1.174 | 1.56 | 42 | 4.448 | 2.496 | 1.78 |
| 27 | 2.029 | 1.309 | 1.55 | 43 | 4.532 | 2.525 | 1.79 |
| 28 | 2.388 | 1.439 | 1.66 | 44 | 4.592 | 2.548 | 1.80 |
| 29 | 2.511 | 1.563 | 1.61 | 45 | 4.575 | 2.568 | 1.78 |
| 30 | 2.876 | 1.676 | 1.72 | 46 | 4.664 | 2.584 | 1.80 |

The average ratio at ages 25-30 is 1.62. Hence, upgrading the age-specific fertility rates by a factor of 1.62 adjusts them so that they are in reasonable agreement with the children ever born parities at ages 25-30. The sum of the adjusted age-specific fertility rates gives TFR = 4.2. Hence, with this method we would estimate the total fertility rate in Lao PDR around 2005 to be TFR = 4.2.

Fig. 13.2 shows the unadjusted parities obtained from births before the census (Parity12), and from children ever born (CEB).

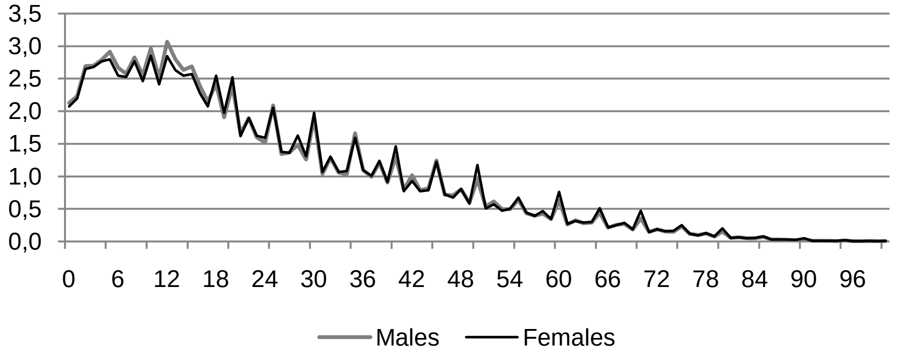
**Fig. 13.2. Parities estimated from births during the 12 months before the census and from the parities in the census**




It will be realized that since fertility obviously is falling in Lao PDR that the basic assumptions underlying the estimation method are unmet. Hence, the estimated total fertility rate as well as the infant mortality rate are approximate.

Fig. 13.3 shows the percent age-distribution for males and females in the census. It will be noted that evidently there has been a drop in fertility in the recent past. Unfortunately, there is often more underenumeration in censuses of infants and children than of the adult population, a feature that makes it difficult to assess the true magnitude of the apparent drop in fertility. In addition, it will be noted that there is considerable age-heaping.

Fig. 13.3. Percent age-distribution by sex, 2005 Lao PDR population census



# 14.0 Logistic Regression

## 14.1 The logistic distribution

The logistic distribution function has served many prominent uses in statistics and demography. It bears close resemblance to the normal distribution but is easier to work with because of its simple mathematical expression. A random variable X which can attain values between minus and plus infinity with distribution function

$$P(X < x) = F(x) = \frac{1}{1 + e^{-(x-m)/b}} \tag{14.1}$$

is logistically distributed. This distribution has mean value $\mu = m$ and variance $\sigma^2 = \frac{1}{3}\pi^2 b^2$. It follows that

$$P(X > x) = 1 - F(x) = \frac{e^{-(x-m)/b}}{1 + e^{-(x-m)/b}} \tag{14.2}$$

so that

$$\text{logit } F(x) = \ln \frac{F(x)}{1 - F(x)} = \frac{1}{b}x - \frac{m}{b} \tag{14.3}$$

This means that the log-odds ratio for the logistic distribution is a linear function of the argument x (the logistic distribution function is the only distribution with this property). If we let

$$F_S(x) = \frac{1}{1 + e^{-x}} \tag{14.4}$$

then

$$\text{logit } F(x) = -\frac{m}{b} + \frac{1}{b} \text{logit } F_S(x) \tag{14.5}$$

so that the logit of a logistic distribution function with parameters m and b can be expressed as a linear function of the logit of a standardized logistic distribution (zero mean and unit variance). Relation (14.5), of course, could also be written

$$\text{logit } F(x) = \alpha + \beta \text{ logit } F_S(x) \tag{14.6}$$

with $\alpha = -m/b$ and $\beta = 1/b$. Replacing F(x) and $F_S(x)$ in (14.6) by $l_x$ and $l_x^S$ in (12.1), respectively, we arrive at the Brass logit survival model. The survival functions $l_x$ and $l_x^S$ are now treated as if though they were logistic (a pseudo-logistic relationship).

It is now apparent that parameter $\alpha$ in (12.1) reflects not only the mean but also the variance of the distribution of deaths in $l_x$ (relative to $l_x^S$). It is also clear that $\beta$ in (12.1) is inversely proportional to the variance of the distribution of deaths in $l_x$. Hence, when the variance of the distribution of deaths in $l_x$ increases relative to $l_x^S$, then $\beta < 1$. This also leads to a decrease in the life expectancy of $l_x$. Conversely, if the variance of the distribution of deaths in $l_x$ is smaller than in $l_x^S$ then $\beta > 1$ and this increases the life expectancy; more people survive to an age in the neighborhood of the life expectancy than in the standard survival function $l_x^S$. Hence, as previously noted, $\beta$ controls the relationship between modeled child and adult mortality relative to the chosen standard survival function.

## 14.2 Regression with covariates

If the probability of an event is p, then the corresponding

$$\text{odds}(p) = p/(1-p) \qquad (14.7)$$

so that ln [odds(p)] = logit p. Odds indicate how many successes there are per failure. When the probability of success is p = 0.5, odds are one. Logits have the convenient property that they are symmetrical[40], that is, logit(p) = -logit(1-p). The probability p expressed in terms of odds is

$$p = \text{odds}(p)/(1+\text{odds}(p))$$

Hence, if the odds for an event are 2 to 1, the probability of success is p = 2/3.

[40] Probabilities are not symmetrical. Logits however are symmetrical because logit p = - logit (1-p).

The general model for logistic regression is

$$P(Y=1|\underline{\beta}) = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n}} \qquad (14.8)$$

which gives the probability that the response variable Y is 1 subject to the covariate vector $\underline{x} = (x_1, \ldots, x_n)$ and parameter vector $\underline{\beta} = (\beta_0, \ldots, \beta_n)$.

This model is easily estimated using a standard statistical package (maximum likelihood estimation). It follows from (14.8) that

$$\text{logit } P(Y=1|\underline{\beta}) = \beta_0 + \sum_{j=1}^{n} \beta_j x_j$$

The covariates need not be dichotomous, for example, continuous age could be a covariate. The dependent variable however is a zero-one variable.

Logistic regression does not predict the value of the dependent variable; rather, it gives the expected probability that the dependent variable is unity subject to the settings of the covariates and their estimated parameters.

It should be noted, incidentally, that both in the case of logistic regression as well as in the case of proportional hazards models (not discussed here), the question may arise if a covariate should be deleted from the estimated model if its corresponding coefficient is not statistically different from zero. The answer to this moot question is that it depends on the intellectual and substantive aspects of the model, an issue beyond discussion here.

Illustrative examples of logistic regression are usually given in the manuals for statistical packages. Both SPSS and Stata have explanatory texts and examples.

# 15.0 Differentiation and Integration

## 15.1 Differentiation

Adding, subtracting and multiplying numbers were the main mathematical operations until the end of the 17th century. When Galileo conducted his experiments in Pisa, a new approach to science (physics) was born. To formulate new theories it became necessary to extend the mathematical knowledge of the day. This was done by Gottfried Wilhelm Leibniz[41] (1646-1716) and Isaac Newton (1642-1728) who, independently of one another, developed the fundamentals of differential and integral calculus. This development in mathematics paved the way not only for modern physics but also for the creation of actuarial mathematics from which demography has borrowed many of its standard methods. We begin by discussing differentiation by means of a heuristic high school example.

Suppose a car has driven a distance $\Delta s$ during the time interval $t \le \tau \le t + \Delta t, \ \ \Delta t > 0$. The average speed of the car across this time interval is $v(\Delta t) = \frac{\Delta s}{\Delta t}$. However, what happens if we ask what the speed of the car is at a point in time $\tau$, $t \le \tau \le t + \Delta t$? We know how to deal with an average across a time interval, but we do not know how to deal with an average valid for a time point. We let s(t) denote the distance it has traveled at time t, t > 0. During a small time interval $\Delta t$ the car travels the distance $\Delta s = s(t+\Delta t) - s(t)$. As before, its average speed during this time interval is

$$\Delta s/\Delta t = [s(t+\Delta t) - s(t)]/\Delta t \qquad (15.1)$$

Suppose we make $\Delta t$ very small. From a practical point of view, we could then argue that (15.1) is an average speed that is valid for the time point $\tau$. It is possible for us to come closer to the answer by assuming that s(t) is a convenient function of time. As an example, suppose

$$s(t) = t^2 \qquad (15.2)$$

[41] Leibniz discovered calculus independently of Newton, and it is his notation that is used. He discovered the binary system which is used in modern electronic computers. He made major contributions to physics and technology.

Inserting (15.2) in (15.1) we get

$$\frac{s(t+\Delta t)-s(t)}{\Delta t} = \frac{t^2+2t\,\Delta t+(\Delta t)^2-t^2}{\Delta t} =$$

$$\frac{2t\,\Delta t+(\Delta t)^2}{\Delta t} = 2t+\Delta t \to 2t \text{ when } \Delta t \text{ approaches } 0 \qquad (15.3)$$

This shows that for a well-behaved choice of function s(t), we can answer the question: what is the speed of the car at time τ? In our example, the speed at time τ is the limit v(τ) = 2τ obtained when Δt approaches 0.

What we have accomplished mathematically (without knowing it) is that we have differentiated the function s(t) $= t^2$ and found the result to be $\frac{d}{dt}t^2 = 2t$.

More specifically, the derivative of a function g(x) at the point $x_0$ is

the limit $\frac{\Delta g(h)}{h} = \frac{g(x_0+h)-g(x_0)}{h}$ as $h \to 0$ (15.4)

which is its speed of change for g in a neighborhood of $x_0$. We also say that the function g is differentiable in the point $x_0$ with differential quotient $\frac{d}{dx}g(x_0) = g'(x_0)$ (15.5)

The theory of differentiation is considerable. Only a few additional comments can be made here.

First, a differential quotient, or derivative, says something about how fast a function changes in a neighborhood of its argument. Indeed, we can rewrite (15.5) so that $dg(x_0) = g'(x_0)\,dx$ which says that the amount of change for g in a neighborhood of $x_0$ is the derivative of the function in $x_0$ times a small (infinitesimal) increment dx. Second, differentiable functions change smoothly; they do not jump wildly from point to point but have a property called continuity. We leave the topic of differentiation here and refer to e.g., Penrose (2005) who gives a good description of continuity, differentiability and smoothness.

## 15.2 Integration

An integral is a sum of many small (infinitesimal) amounts. To illustrate this, suppose we wish to calculate the area underneath the curve $y = x^2$ on the open interval 0 < x < 1. This area is denoted $A = \int_0^1 x^2\,dx$. We apply a numerical approach because we do not know, as of yet, how to evaluate the area A using a mathematical expression.

Divide the x-axis between 0 and 1 into small portions of equal length w, say. This division may be denoted $x_0, x_1, \ldots, x_{n-1}$ where $x_{i+1} - x_i = w$. Calculate f(x) for the midpoint of each such interval. To this end, we use $h_i = [f(x_i) + f(x_{i+1})]/2$. Then, for each interval, calculate $h_i$ w (this is the area of a rectangle with base length w and height $h_i$) and sum all contributions $h_i$ w. Letting w = 0.025, the sum is A = 0.33344 (table 15.1). The true value is $A = \int_0^1 x^2\,dx =$ 0.33333. Approximations to integrals can be found numerically, as illustrated in table 15.1.

**Table 15.1. Numerical integration of f(x) = $x^2$ between 0 and 1**

| x | f(x) | $h_i$ | $h_i$ w | x | f(x) | $h_i$ | $h_i$ w |
|---|---|---|---|---|---|---|---|
| 0.000 | 0.00000 | 0.00031 | 0.00001 | 0.525 | 0.27563 | 0.28906 | 0.00723 |
| 0.025 | 0.00063 | 0.00156 | 0.00004 | 0.550 | 0.30250 | 0.31656 | 0.00791 |
| 0.050 | 0.00250 | 0.00406 | 0.00010 | 0.575 | 0.33063 | 0.34531 | 0.00863 |
| 0.075 | 0.00563 | 0.00781 | 0.00020 | 0.600 | 0.36000 | 0.37531 | 0.00938 |
| 0.100 | 0.01000 | 0.01281 | 0.00032 | 0.625 | 0.39063 | 0.40656 | 0.01016 |
| 0.125 | 0.01563 | 0.01906 | 0.00048 | 0.650 | 0.42250 | 0.43906 | 0.01098 |
| 0.150 | 0.02250 | 0.02656 | 0.00066 | 0.675 | 0.45563 | 0.47281 | 0.01182 |
| 0.175 | 0.03063 | 0.03531 | 0.00088 | 0.700 | 0.49000 | 0.50781 | 0.01270 |
| 0.200 | 0.04000 | 0.04531 | 0.00113 | 0.725 | 0.52563 | 0.54406 | 0.01360 |
| 0.225 | 0.05063 | 0.05656 | 0.00141 | 0.750 | 0.56250 | 0.58156 | 0.01454 |
| 0.250 | 0.06250 | 0.06906 | 0.00173 | 0.775 | 0.60063 | 0.62031 | 0.01551 |
| 0.275 | 0.07563 | 0.08281 | 0.00207 | 0.800 | 0.64000 | 0.66031 | 0.01651 |
| 0.300 | 0.09000 | 0.09781 | 0.00245 | 0.825 | 0.68063 | 0.70156 | 0.01754 |
| 0.325 | 0.10563 | 0.11406 | 0.00285 | 0.850 | 0.72250 | 0.74406 | 0.01860 |
| 0.350 | 0.12250 | 0.13156 | 0.00329 | 0.875 | 0.76563 | 0.78781 | 0.01970 |
| 0.375 | 0.14063 | 0.15031 | 0.00376 | 0.900 | 0.81000 | 0.83281 | 0.02082 |
| 0.400 | 0.16000 | 0.17031 | 0.00426 | 0.925 | 0.85563 | 0.87906 | 0.02198 |
| 0.425 | 0.18063 | 0.19156 | 0.00479 | 0.950 | 0.90250 | 0.92656 | 0.02316 |
| 0.450 | 0.20250 | 0.21406 | 0.00535 | 0.975 | 0.95063 | 0.97531 | 0.02438 |
| 0.475 | 0.22563 | 0.23781 | 0.00595 | 1.000 | 1.00000 | | |
| 0.500 | 0.25000 | 0.26281 | 0.00657 | | | | |
| | | | | | | Sum A = | 0.33344 |

We mention without proof that the derivative of $f(x) = x^n$ with respect to x is $\frac{d}{dx} x^n = n\, x^{n-1}$ (15.6)

a formula you will be using many times.

It can be shown that

$$\int_a^b x^n \, dx = \left[ \frac{x^{n+1}}{n+1} \right]_a^b = \frac{b^{n+1}}{n+1} - \frac{a^{n+1}}{n+1} \qquad (15.7)$$

which, in our example, gives $\int_0^1 x^2 \, dx = \left[ \frac{x^3}{3} \right]_0^1 = \frac{1}{3}$.

Generally, to find the integral $\int_a^b f(x)\,dx$ we seek a function F(x) such that its derivative is $\frac{d}{dx}F(x) = f(x)$. The function F(x) satisfying this requirement is called a primitive function for f(x). Hence, generally,

$$\int_a^b f(x)\,dx = \left[F(x)\right]_a^b = F(b) - F(a) \qquad (15.8)$$

High school textbooks on mathematics give the main rules for differentiation and outline a number of primitive functions used in practical situations. The interested ready may peruse Cramer (1945) which gives a very readable introduction to integration.

# References

Benjamin, B, Brass, W, and Glass, D (1963). Actuarial methods of mortality analysis; adaptation to changes in the age and cause pattern. Discussion, R. D. Clarke, R. E. Beard, W. Brass. Proceedings of the Royal Society. Series B, No. 974, Vol. 159.

Benjamin, B., and Haycocks, H. W (1970). *The Analysis of Mortality and Other Actuarial Statistics*, Cambridge.

Bogue, D. J (1969). *Principles of Demography*, Wiley.

Booth, H., Maindonald, J., and Smith, L (2002). Applying Lee-Carter Under Conditions of Variable Mortality Decline. Population Studies, Vol. 56, No. 3, pp. 325-336.

Brass, W (1971). On the Scale of Mortality in *Biological Aspects of Demography*, ed. by William Brass, pp. 69-110, Taylor and Francis.

Brass et al. (1968). *The Demography of Tropical Africa*, Princeton University Press.

Brass, W (1974). Perspective in population prediction: illustrated by the statistics of England and Wales, J.R. Statist. Soc. A 137, pp. 532-83.

Brass, W (1975) *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Laboratories for Population Statistics, The Carolina Population Center.

Carrier, N. H., and Goh, T.J (1972). The Validation of Brass's Model Life Table System. Population Studies, Vol. 26, No. 1, pp. 29-51.

Carter, L. and Prskawetz, A (2001). Examining Structural Shifts in Mortality Using the Lee-Carter Method. MPIDR Working Paper WP 2001-007, March 2001. Max Planck Institute for Demographic Research, Rostock.

Chatfield, C (1999). *The Analysis of Time-series: An Introduction*. Fifth Edition, Chapman and Hull.

Chiang, C.L (1968). *Introduction to Stochastic Processes in Biostatistics*, Wiley.

Coale, A.J., and Demeny, P.J (1966). *Regional Model Life Tables and Stable Populations*, Princeton.

Coale, A. J., and Trussell, T. J (1974). Model Fertility Schedules: Variations in the Age Structure of Childbearing in Human Populations. Population Index, Vol. 40, No. 2, pp. 185-258.

Courbage, Y and Fargues, P (1979). A method for deriving mortality estimates from incomplete vital statistics. Population Studies, Vol.. 33, No. 1, pp. 165-180.

Cox, P. R (1970). *Demography* (Fourth Edition), Cambridge.

Cramer, H (1945). *Mathematical Methods of Statistics*, Princeton.

Feeney, G (1980). Estimating infant mortality rates from child survivorship data. Population Studies Vol. 34, No. 1, pp. 102-128.

Feeney, G (1991). Child survivorship estimation: methods and data analysis. Asian and Pacific Population Forum 5(2-3), Summer-Fall 1991, 51-55 and 76-87.

Fisz, M (1963). *Probability Theory and Mathematical Statistics*, Wiley.

Girosi, F. and King, G (2005). A Reassessment of the Lee-Carter Mortality Forecasting Method. Working Paper, RAND Corporation.

Golub, G. H. and van Loan, C. F (1996). Matrix Computations, 3rd ed., Johns Hopkins University Press, Baltimore.

Gomez de Leon, C (1990). Empirical EDA Models to Fit and Project Time Series of Age-Specific Mortality Rates. Discussion Paper No. 50, Central Bureau of Statistics, Oslo, Norway.

Gottman, J. M (1981). *Time-series analysis: A comprehensive introduction for social scientists*, Cambridge, M.A.

Hald, A (1962). *Statistical Theory with Engineering Applications*, Wiley.

Hald, A (1998). *A History of Mathematical Statistics from 1750 to 1930*, New York: Wiley.

Hansen, P.R (1987). The truncated SVD as a method for regularization, BIT, 27, pp. 534-553.

Hartmann, M (1987). Past and recent experiments in modeling mortality. Journal of Official Statistics, pp. 19-36.

Hartmann, M (1991). A Parametric Method for Census Based Estimation of Child Mortality. Journal of Official Statistics. Volume 7, No. 1, pp. 45-55.

Hartmann, M., and Strandell, G (2006). *Stochastic Population Projections for Sweden*, Research and Development – Methodology Reports from Statistics Sweden, 2006, 2.

Hauser, P. M., and Duncan, O. D (1972). *The Study of Population*: An Inventory and Appraisal. (First printing 1952), University of Chicago Press.

Hill, K., and Trussell, J (1977). Further Developments in Indirect Mortality Estimation. Population Studies, Vol. 31, No. 2, pp. 313-334.

Hill, K (1977). Estimating adult mortality levels from information on widowhood. Population Studies, Vol. 31, No. 1, pp. 75-84.

Horn, R. A. and Johnson, Charles R (1985). *Matrix Analysis*, Cambridge University Press.

Irwin, J.O (1949). The standard error of an estimate of expectation of life with special reference to tumourless life in experiments with mice. Journal of Hygiene, Vol. 47, pp. 188-189.

Keyfitz, N (1968). *Introduction to the Mathematics of Population*, Addison-Wesley.

Keyfitz, N. and Flieger, W (1971). *Population: Facts and Methods of Demography*, Freeman.

Lee, R.D., and Carter, L.R (1992). Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association, Vol. 87, No. 419, pp. 659-671.

Lee, R.D., and Miller, T (2001). Evaluating the Performance of Lee-Carter Mortality Forecasts. Demography Vol. 38, No. 4, pp. 537-549.

Namboodiri, K., and Suchindran, C. M (1987). *Life Table Techniques and Their Applications*, Academic Press.

Palloni, A (1980). Estimating infant and childhood mortality under conditions of changing mortality. Population Studies, Vol. 34, No. 1, pp. 129-142.

Penrose, R (2005). *The Road to Reality*, Vintage Books.

Pressat, R (1972). *Demographic Analysis: Methods, Results, Applications*, Aldine Publishing Company, New York.

Retherford, R.D and Cho, L.J (1978). Age-parity-specific birth rates and birth probabilities from census or survey data on own children. Population Studies, Vol. 32, No. 3, pp. 567-581.

Sauvy, A (1969). *General Theory of Population*, Weidenfeld and Nicolson.

Shorter, F.C., Pasta, D. and Sendek, R (1990). *Computational Methods for Population Projections: With Particular Reference to Development Planning*. The Population Council.

Shryock, H. S., Siegel, J. S., and Associates (1976). *The Methods and Materials of Demography*. Condensed Edition by Edward G. Stockwell, US Bureau of the Census.

Siegel, J.S et al (2004). *The Methods and Materials of Demography*. Second edition. Edited by Siegel and Swanson, Elsevier Academic Press.

Smith, D., and Keyfitz, N (1977). *Mathematical Demography*, Springer-Verlag.

Spiegelman, M.A (1980). *Introduction to Demography*, Cambridge.

Strang, G (1998). *Introduction to Linear Algebra*, Wellesley-Cambridge Press.

Sullivan, J.M (1972). Models for the estimation of the probability of dying between birth and exact ages of early childhood. Population Studies, Vol. 26, No. 1, pp. 79-97.

Trussell, T.J (1975). A re-estimation of the multiplying factors for the Brass technique for determining childhood survivorship rates. Population Studies, Vol. 29, No. 1, pp. 97-107.

United Nations (1968). The Concept of a Stable Population: Application to the study of populations of countries with incomplete demographic statistics. Population Studies, No. 39.

Wenzel, E. and Ovcharov, L (1986). *Applied Problems in Probability Theory*, MIR Publishers, Moscow.

Wilmoth, J (1993). Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. Technical Report, Department of Demography, University of California, Berkeley.

Wilson, E.B (1938). The Standard Deviation of Sampling for Life Expectancy. Journal of the American Statistical Association, Vol. 33, pp. 705-708.

This book is dedicated to those who are about to begin their exploratory journey into the world of population statistics. It is written for staff in statistical offices and gives focus to basic methods used in demographic analysis.

All officiell statistik finns på: **www.scb.se**
Kundservice: tfn 08-506 948 01

All official statistics can be found at: **www.scb.se**
Customer service, phone +46 8 506 948 01