Farhat Yusuf
Jo. M. Martins
David A. Swanson

Methods of Demographic Analysis

Springer

Methods of Demographic Analysis

Farhat Yusuf • Jo. M. Martins • David A. Swanson

Methods of Demographic Analysis

Springer

Farhat Yusuf
Menzies Centre for Health Policy
Sydney School of Public Health
The University of Sydney
NSW 2006, Australia
and
Department of Marketing
and Management
Faculty of Business and Economics
Macquarie University
NSW 2109, Australia

Jo. M. Martins
Department of Marketing
and Management
Faculty of Business and Economics
Macquarie University
NSW 2109, Australia

David A. Swanson
Department of Sociology
College of Humanities, Arts and Social Sciences
University of California Riverside
92521 Riverside California, USA

ISBN 978-94-007-6783-6 ISBN 978-94-007-6784-3 (eBook)
DOI 10.1007/978-94-007-6784-3
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013951278

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

There have been many demography textbooks published over the past 50 years and we might ask: why is there a need for another? The answer is that, ideally, textbooks are tailored to the needs and capacities of the students that will be making use of the book. In my experience, this is particularly the case in the field of demography. Some prior textbooks in the field are pitched to a high, theoretical or mathematical level, probably aimed at people who plan to become academic demographers while others are pitched at a level where they can be understood by people undertaking training in human service occupations where some capacity with 'numbers' will be useful. Some texts assume relatively strong mathematical and statistical capacity on the part of the student while others try to steer around mere average capacity in this regard.

This book is written by a group that has had many years of experience teaching demography to business and economics students. Accordingly, they have pitched the book towards students that have relatively good mathematical and statistical skills without extending to matrix algebra or integral calculus. The book is comprehensive in its coverage of the field of demography and the student mastering the text will be competent in the professional application of demographic methods. It uses many worked examples making it clear what lies behind the formulae. The authors, therefore, have produced a book that will be useful at the undergraduate and master's levels where students have reasonably good quantitative skills. This is probably the mainstream of teaching in demography around the world.

Besides its wide coverage of standard demographic measures, the book includes chapters on sources of demographic data and data evaluation methods. It also has a long chapter on statistics taking the student through basic statistical measures through to correlation and regression. The final chapter of the book provides useful references to software packages that can be used to derive many of the measures described in the preceding chapters. Many of these packages are available on the internet and are freely downloadable.

I congratulate the authors on the production of this textbook as it occupies a niche in the range of available textbooks that needed to be occupied.

Director, Australian Demographic
and Social Research Institute
The Australian National University
and
President, International Union for the Scientific
Study of Population

Peter McDonald

Preface

Purpose

The study of populations is relevant to most human enterprises. As a discipline, demography is similar to statistics in that its methods are used in many other academic fields. In the case of demography, they include actuarial studies, business administration, criminal justice, geography, history, legal studies, marketing, organizational studies, planning, forecasting, political science, public policy and administration, health care and education, sociology, and urban studies. Demographic methods are used widely by practitioners in these fields. Given the wide audience and some of the recent developments in demographic methods, the contents of this book, with existing and newly developed methods, fill a crucial gap in the application of demographic methods.

The book encompasses the many facets of demographic and related methods and their applications. It introduces some statistical measures of relevance to the study of demography, related concepts and associated techniques. The book deals with basic population models and elaborates on the concepts of demographic stocks and flows. Further, it deals with the analysis of demographic phenomena in the context of cross-sectional and longitudinal/cohort analyses.

The book incorporates methods used in the computation of life and multiple decrement tables which are useful in studying demographic phenomena such as mortality, morbidity, and nuptiality, but other social events. An important application of the life table methodology is in population projections. The book examines alternative methods of population and related projections at both the national and sub-national levels.

Testing the accuracy of demographic data is an important precursor in demographic analysis. Methods of testing the accuracy of data, smoothing and adjustments that might be required are discussed.

The stable population model is a useful tool in the analysis of populations and of relevance to the estimation of demographic parameters from incomplete datasets.

These methods are of particular importance in developing countries that suffer from a paucity of demographic data.

Improved computer technology and software have enhanced the use of spreadsheets and other software in demographic analysis. The relevance of software in the public domain to demographic analysis is examined, as well as some proprietary packages.

Organization

The book is organized in 14 chapters. These chapters represent a progression going from basic concepts to more sophisticated ones. The first four chapters introduce demography as a field of study and analysis. The first chapter deals with the nature and historical context of demographic analysis. And the second provides the fundamental terms, definitions, and ideas about data that need to be mastered. In Chap. 3, some elementary statistical measures are described to enhance the introduction to demographic analysis. These include such basic measures as counts and frequencies, proportions, ratios, rates and probabilities, and measures of central tendency and dispersion, concentration, dissimilarity and relative difference. Correlation and regression methods are also examined. Chapter 4 covers fundamental demographic terms and measures. Taken together, Chaps. 3 and 4 provide the basic ideas and measurements underlying the size, distribution, and composition of human populations.

These ideas in turn, need to be mastered before proceeding to the following four chapters, which cover the components of population change, fertility, mortality, and migration. Chapter 5 covers fertility while Chap. 6 examines mortality. Cross-sectional and longitudinal approaches in demography are examined and synthetic measures of fertility and survival are considered. Chapter 7 covers a perspective so important to the study of mortality and survival that it is given a separate chapter, the Life Table. Chapter 8 examines migration.

In general, Chaps. 2, 3, 4, 5, 6, 7 and 8 look at demography in terms of *ascribed* characteristics. In Chap. 9, concepts, methods, and data that look at demography in terms of *achieved* characteristics are introduced. This chapter consists of methods of analysis and measures related to marital status and associated vital events, such as marriages and divorces, education, labour force, occupation, households and families.

Chapter 10 extends the concepts of a life table introduced in Chap. 7. The idea of death is broadened to the concept of a *decrement* and considers situations where cohorts may be subjected to multiple decrements, such as different causes of death. In addition, the chapter widens the concept to the building of multistate life tables concerned not only with decrements from life to death but also with possible movements among various active states, such as moving in and out of the labour force.

Chapter 11 on projections ties together the concepts, methods, and data discussed in Chaps. 4, 5, 6, 7 and 8 and include aspects of the discussions found in Chaps. 9 and 10. The chapter covers in detail the cohort-component method that is the most used technique of projecting populations by age and sex. It also discusses and illustrates the cohort-change method that requires less data. Further, it looks at projections of particular segments of the population such as people in the labour force and of school age.

All data are subject to errors. Chapter 12 describes some of the commonly used methods for testing the quality of demographic data, and procedures for adjusting and smoothing the data in order to improve their quality.

A major canon of mathematical demographic theory, the *Stable Population Model*, is the subject of Chap. 13. It extends the ideas found in Chap. 7 in a manner quite distinct from how these ideas were extended in Chap. 10. Some important properties and characteristics of this model are examined and its use in demographic analysis is discussed.

The book concludes with Chap. 14 that provides a survey of some demographic software available. Given the pace of technological change, this should be looked as a starting point, as new and useful products will have been added by the time the book is published.

Throughout the book numerical examples are given with information from a number of countries with some comparisons of national patterns. Where possible the Internet links (URLs) to various references are provided. All of these were tested in February 2013 to ensure that they were operational. However, the location of material on the Internet may have changed by the time the readers attempt to retrieve it.

Use

The material in this book is organised in a progressive manner that allows the user to move from an introductory level to more advanced methods of demographic analysis. The approach takes into consideration that users may have different levels of mathematical skill and that some would benefit from a step by step approach that makes the various methods of demographic analysis accessible to various levels of expertise. It follows a gradual build up from elementary to more advanced methods of demographic analysis. In addition to students of demography, the methods of analysis in the book are of relevance to practitioners of other disciplines and people in government and business.

Contents

List of Boxes

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Purpose

This book reflects the evolution of demography and its analysis over time. This chapter gives a brief sketch of this long evolution to provide context of the development of the demographic perspective and its analysis.

1.2 What Is Demography?

Demography comes from two Greek words Demos (δῆμος) meaning people and Graphos (γράφειν) meaning to write. The term was first used by a Belgian statistician, Achille Guillard, in 1855 in the title of his book *Elements de statistique humaine: Ou demographie compare* (Pressat 1985).

According to the United Nations Multilingual Demographic Dictionary 'Demography is the scientific study of human populations, primarily with respect to their size, their structure and their development' (United Nations 1958). Another dictionary emphasises some of the important components by defining Demography as 'the study of human population in relation to the changes brought about by the interplay of births, deaths and migration' (Pressat 1985).

Demography focuses on at least the following five aspects of human populations:

- size
- distribution
- composition
- components of population change, and
- determinants and consequences of population change.

Population *size* is simply the number of persons in an area at a given point in time. Population *distribution* refers to how the population is dispersed in

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_1,

geographic space at a given point in time. Population *composition* is usually defined in terms of *ascribed* characteristics that include place of birth, sex and age. These characteristics are termed ascribed because they either never change over a person's lifetime (*e.g.,* place of birth), or change in a highly predictable manner (*e.g.,* age). *Achieved* characteristics are also used and include: place of residence, occupation, marital status, and educational attainment. These characteristics are termed achieved because they can change over a person's lifetime (*e.g.,* place of residence) and can vary in ways that may not be very predictable (*e.g.,* occupation).

The *components of population change* can be defined narrowly or broadly. Defined narrowly there are three components: births, deaths, and migration. Defined more broadly, the components include the factors that affect the components of change. Births, for example, are affected by factors such as family formation and use of contraception, while deaths are affected by nutrition, the efficacy of public health measures, and medical advances. Migration is affected by personal characteristics such as age, occupation, and educational attainment, as well as by other situations which may operate as "push" or "pull" factors..

Demography consists of identifiable subfields. Broadly speaking, the two most fundamental subfields are *formal* (theoretical, pure or basic) demography and *applied* demography (population studies). The former focuses on theoretical and empirical questions of interest mostly to academic demographers, while the latter focuses on questions of interest to academics as well as others. Both formal and applied demography share a core set of methods and data. They are the focus of this book.

1.3 The Demographic Perspective: A Brief History

Concerns with population and its impact have taken many forms over time. For instance, it is reported that Confucius (502–479 BC) thought that large population growth could result in lower labour productivity. Greek philosophers such as Plato 360 BC and Aristotle 354 BC were concerned with optimum population size for the wellbeing of society. Roman preoccupation with their military activity and its financing also paid attention to population size and the tax base that it provided – first Roman census was conducted between 578–34 BC. In the middle ages, population studies were carried out by Muslim scholars such as ibn-e-Khaldun (Pollard et al. 1995; Swanson & Stephan 2004).

One of the basic tools of demography is the counting of people or the census (from the Latin to *assess*). In the past, some form of census activities were carried out in Babylon, China, India, Greece and the Roman Empire, Japan and some European countries (*e.g.,* Iceland in 1703 and Sweden in 1748). Colonial censuses were carried out, among other places, in North America (Virginia in 1620 and Quebec in 1665), Australia (New South Wales in 1828) and British India in 1871. The first United States census took place in 1790. Sampling was first used in a census in Norway (1900). Currently, almost every country carries out decennial census, and some, like Australia and Canada, do it every 5 years.

Population registers, another source of demographic information, have also been in use in a variety of modes. Ancient Egypt is said to have registered its people (around 1400 BC). Family records were kept in Japan as far back as 645 AD and were re-introduced in 1868. Population registers are now common in Scandinavian countries, and a complete enumeration of populations can now be carried out, in some countries, using these registers (*e.g.,* Finland).

In addition to the enumeration of population stocks through censuses and population registers, recording of vital events such as births, deaths and marriages has been a major source of demographic information. A well known study of mortality was John Graunt's Bills of Mortality in England (1662). Annual reports of births, marriages and deaths in Paris became available in the seventeenth century (1670), and national vital statistics registration became law in Sweden in 1748 and in England and Wales in 1874. Although vital events registration is common in most countries, comprehensive coverage of the population continues to be a challenge in many countries and their reliability may be questionable.

To overcome some of these difficulties with the vital registration system, population surveys, are carried out by countries themselves and in some cases through cooperation from international agencies such as the United Nations. Surveys have also been used to obtain information on demographic characteristics and factors associated with them. Social surveys of London were begun in 1886 and Charles Booth produced a related report in 1906. Surveys are now the most frequent source of demographic information.

Following in the tradition of examining population issues by ancient philosophers to ibn-e-Khaldun, methods of demographic analysis have evolved. The study of mortality in London by John Graunt (1662) is usually singled out as a most important development in demographic analysis. His development of the concept of the probability of dying formed a springboard for the formulation of life tables and related life expectancy by Edmund Halley (1693).

In addition to the development of analytical approaches to mortality and fertility, the discussion of the importance of population growth and perceived carrying capacity was anticipated by Giovanni Botero (1558) and taken up by Thomas Malthus (1798), in his arguments with concepts of population growth and limits proposed by William Goodwin (1793) and Nicolas Condorcet (1781). The nature of population growth and methods of its estimation were furthered by Benjamin Gompertz (1825) and the logistic growth model by Pierre Verhultz (1845, 1847). The latter was further developed by Raymond Pearl and Lowell Read (1920).

Leonard Euler's stable population model (1760) is another important tool of demographic analysis in the understanding of population dynamics. This model was later re-stated in a more complete manner by Lois Dublin and Alfred Lotka (1925). The concept of mortality as a population decrement was extended by the development of multiple decrement tables by Daniel Bernoulli to assess the impact of smallpox (1760) adjusting Halley's life tables.

Among other contributions to demographic analysis, the use of indirect standardization by William Farr (1856–1859) might be mentioned, as well as the estimation of the net reproduction rate by Richard Bockh (1884). John Snow (1854) might have been the precursor of the use of geographic information systems (GIS) through the mapping of victims of cholera to identify the source of the cause of their death.

In addition to the analysis of fertility and mortality, the analysis of migration (the other variable in population change) was also examined and Henry Carey put forward the migration gravity model (1837–1840) that was pursued by Ernst Ravenstein in his laws of migration (1889).

This progression of demographic analysis led to the development of models that project populations in the future. The cohort-component method of population projections was used by Edwin Cannan to project the population of England and Wales (1895). This projection method was further explored by Arthur Bowley in Great Britain (1924) and Pascal Whelpton (1928) in the United Sates.

1.4 Continuing Evolution of Demographic Analysis

In the 1950s the United Nations was instrumental in fostering the formulation of systematic collection of population information and methods of analysis. This involved the harnessing of efforts of many experts and the development of useful manuals. One outcome was the preparation of worldwide population statistics that showed both strengths and gaps in demographic statistics in different countries. The United Nations also provided technical support to developing countries to improve their capacity for the collection and analysis of demographic information. Indirect methods of the estimation of population and vital events were further developed to overcome the paucity of traditional tools such as censuses and vital events registration. Ansley Coale and William Brass (1963, 1966) and many others developed indirect methods of estimation of population and vital events.

Analysis of demographic phenomena continues to be given attention. The observation that countries experience different mortality and fertility levels (mooted by Warren Thompson in 1929) was developed into the concept of the demographic transition by Kingsley Davis in 1945 that helped to understand the pending population explosion in the following five decades. Better insights into the factors underlying fertility levels were provided by the work of Kingsley Davis and Judith Blake (1956). This was extended by John Bongaarts and Robert Potter who devised a framework of the proximate determinants of fertility (1983).

Economic models of family formation and fertility were explored by Gary Becker (1981). John Caldwell (1976, 1982) extended the theory of demographic transition to health transition and his theory on “wealth flows” stimulated much of the recent interest in micro-demographic research.

This inspirational work over the ages has provided current generations with the powerful tools of demographic analysis presented in the following chapters of this book.

References

Pollard, A. H., Yusuf, F., & Pollard, G. N. (1995). *Demographic techniques* (4th ed.). Sydney: A.S. Wilson Inc.

Pressat, R. (1985). *Dictionnaire de demographie* [The dictionary of demography]. English version edited by C. Wilson. Oxford: Blackwell.

Swanson, D., & Stephan, G. E. (2004). A demography time line. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 779–786). San Diego: Elsevier Academic Press.

United Nations. (1958). *Multilingual demographic dictionary*. New York: Department of Economic and Social Affairs. http://www.demopaedia.org/. Accessed Feb 2013.

Chapter 2
Demographic Information

2.1 Purpose

This chapter deals with the nature of demographic information and its main elements. It examines the range of information involved and some of their relationships. It reviews some basic demographic concepts and definitions. Further, it describes the major sources and collection methods of demographic data, including censuses, sample surveys, vital statistics registration, population registers, and administrative records.

2.2 Nature of Demographic Information

2.2.1 Components of Population Change

As discussed in the previous chapter, Demography is concerned with the size and composition of populations, as well as changes and factors influencing them. In this context, it is connected with the measurement of births, deaths, migration and other vital events that have direct impact on population changes and in turn are affected by the size and composition of populations such as age distribution.

Demography also studies other population characteristics such as sex, marital status, living arrangements and household composition, language spoken, ethnic background, health and disability, education and training, employment status and occupation, income and household consumption, population densities and urban and rural residence.

Demographic studies often deal with the assessment of *current* characteristics like the number of people in a given area and their age and sex characteristics or *trends* over time such as population growth and changes in fertility and life expectancy. Some other concerns in demographic analysis are the *relationships* between population composition such as age and vital events, for example, births or

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_2,

socio-economic factors that influence these events, such as the impact of levels of fertility on the degree of child dependence on the working population and demand for school services.

2.2.2 Concepts and Definitions

Sex and Gender

A distinction should be made between sex and gender characteristics. *Sex* could be defined as the biological characteristics of males and females – an ascribed characteristic. Male and female *gender* attributes could be seen as psychological and social characteristics arising from belief systems of what male and female behaviour is or should be – an achieved characteristic.

Age

Age is generally expressed in terms of time units after birth. The most common measure is the *number of years after birth*. The last birthday is usually the reference point to count the number of years after birth. Other age related classifications may refer to stages of the life cycle. An *infant* is generally considered to be a person *less than 1 year of age*. The infant *neonatal* period has been defined as the first *28 days after birth*. There is more than one definition of the term child. Usually *children* are people *less than 15 years of age*. However, *child mortality* usually refers to *deaths of children under 5 years of age*. On the other end of the age range, *old people* are habitually defined as those who are *65 years of age and over*. The term *adult* can also have more than one definition. It is often associated with legal responsibility that varies from country to country. It could also be seen as people who are not classified as children that is those *15 years of age and over*. There is a degree of ambiguity involved and the term *adolescen*t is frequently used to describe people in the age range of *about 13–19 years of age*. The varying perceptions lead to the need to have clear definitions in demographic analysis.

Births, Fertility and Fecundity

In demographic analysis it is useful to distinguish between fecundity and fertility. *Fecundity* refers to the *potential* or capacity *to reproduce*. The female reproductive period is from menarche to menopause, generally from the age of about 12–49 years. *Fertility* is defined in terms of the *actual number of births* that females have during their reproductive period. A related term is pregnancy. It is the state of a female who has developed an embryo or a foetus in her abdominal area after the union of a spermatozoon and an ovum. *Delivery* is the result of a pregnancy of one or more births from a pregnancy that may be born alive or stillborn foetuses (Canada 2012). *Confinement* is a pregnancy that leads in at least one live birth (Australia 2011b). *Parity* is the order of live births a woman has had. Accordingly, zero parity means that the woman has had no previous live birth, first parity means the first live birth a woman has had and second parity as the second. If in her pregnancy the woman has a set of twins born alive, and these are her first live births, then the first twin born will be the first parity and the second born the second parity (Canada 2012). *Replacement fertility* is the number of live births required for a couple to replace themselves. This is estimated at 2.1 live births during a female reproductive period.

A *birth* is the result of a pregnancy that involves the complete expulsion or extraction of the product of a pregnancy regardless of the duration of pregnancy. It may result in *live* or *stillborn* births and *single* or *multiple* births. A *live birth* is a birth that *after separation breathes or shows any evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles, whether or not the umbilical cord has been cut or the placenta is attached* (Canada 2012). *Stillbirth* (foetal death) is the death of the product of a pregnancy *prior to complete expulsion or extraction* and *is indicated by the fact that after such separation the foetus does not breathe or show any evidence of life, such as the beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles.* Registration of foetal deaths may take place if they have the *birth weight of 500 g or more or the duration of the pregnancy is 20 weeks or longer* (Canada 2012). The term *nuptial birth* may be applied to a child born to parents who are legally married. Also related to marital status is a *paternity acknowledged birth*. This is an *ex-nuptial birth* where the father acknowledges paternity (Australia 2011b). Births and other vital events may be counted either by *year of registration*, the date when they were registered, or *year of occurrence* when the event actually took place. Similarly, births and other vital events can be counted at their place of occurrence (*de facto*) or at the place of residence of the person experiencing the vital event in question (*de jure*).

Deaths, Mortality and Life Expectancy

Mortality is concerned with deaths in a given population. A *death* is the complete and permanent disappearance of all evidence of life after a live birth has taken place. It excludes deaths before a live birth (Australia 2011c). However, as will be seen, stillbirths are included in the measurement of perinatal (stillbirths and neonatal) deaths rates. Mortality measures the number of *deaths* over a period of time, usually 1 year, and relates it to the *reference population*. The importance of deaths early in life has led to the classification of deaths to include perinatal, neonatal and infant deaths. *Perinatal deaths* are the number of deaths of stillbirths (foetal deaths of at least 20 weeks or 400 g birth weight) plus neonatal deaths. *Neonatal deaths* are defined as the number of deaths of live born infants within 28 days of birth, and *infant deaths* are the number of deaths of all live births before 1 year of age

(Australia 2009). Perinatal death rates relate perinatal deaths to the number of stillbirths plus the number of live births, while neonatal and infant mortality rates relate these deaths to the number of live births. Mortality also relates deaths to age and other characteristics of the relevant population. A linked measure is survival or *life expectancy* in terms of the average *number of years* that people in a given *population are likely to live*. This may be estimated at birth or some other age. Death may affect the status of surviving relatives. *Orphans* are children with at least one parent dead. *Widowed* is the status of a male or female whose partner has died. The classification of causes of death is another dimension of the analysis of mortality. Currently, the recommended classification is the International Classification of Diseases prepared by the World Health Organization in its 10th Revision (WHO 2010). Causes of death may be *underlying* and *multiple*, with main and possibly secondary causes. Accidental and violent causes may be classified according to the *external cause* rather than the nature of the death.

Migration

Migration refers to the *movement* of people from one location to another. Migration *out of an area* is known as *emigration* and that *into* an area is *immigration. Net migration* is the *difference* between *emigration and immigration.* It can be positive or negative depending on the relative sizes of the two flows. Migration may be *temporary* or *permanent* depending on its intended purpose and or period of time involved. Different countries may apply different definitions. Migration may take place *within* a country and be denoted as *internal* (domestic migration) or *international* across country borders. It is concerned with the *origin* and *destination* of the individuals who move. In addition to the count of migration flows, demographic analysis is also concerned with relating these flows to the reference population at origin and destination.

Marital Status

Marriage is the union of two or more people who cohabit and form a family. Marriage may be *registered* according to local legal statues. Marriage may be of a male and a female, or of more than two people. Marital legal status is evolving and marriage of people of the same sex may be legal and registered. *De facto* or social marital unions may also take place and may or may not have legal status. It may be said that a *marriage exists when two people live together as husband and wife, or partners, regardless of whether the marriage is formalised through registration* (Australia 2011a). The following classification tends to be applied: *never married*, *married*, *divorced/separated* and *widowed*.

Households

The living arrangements of people are an organising framework in demography. A *household* is a group of two or more people who live in the same dwelling. A *dwelling* is a place where people live. It may be a house, an apartment or some other living place such as a caravan in a park. These dwellings may be *private* or *non-private* and may be *occupied* or *not occupied*. People in households may be *unrelated* or *related* but share their food and other basic living essentials. There are also single person households who live by themselves in a dwelling and make their own arrangements for basic living essentials including food. Households may be of *one-person*, *group* of unrelated people, *same-sex couple*, and group of *related people* either by couple union or blood. A *family* is a household of two or more people who are related by union bonds, such as a couple, or some blood relationship such as a child-parent relationship. Families include *couples only*, *couples with dependent children*, *single-parent with dependent children*, and *extended families* that may include other relatives related by union or blood. Usually, people who live in non-private dwellings such as hostels and boarding houses, hotels, hospitals and other institutions are not counted as members of households (Australia 2010).

2.2.3 Demographic Change and Implications

Although there are some models of static populations, usually populations are undergoing continuing changes in both size and composition that have an effect on and are of relevance to living conditions, government policies and business strategies. Applied demography is concerned with the application of demographic perspectives and approaches to the identification and management of social, government and business problems.

2.3 Data Collection Methods

2.3.1 Census

A census can be described as a complete enumeration of a given population. Usually, censuses are carried out on a country wide basis at a given date or period of time. According to the United Nations (2008), censuses should have the following attributes:

- enumeration of each individual in the whole population
- universal coverage of the population within a defined territory
- specified period of time of the coverage of the whole population
- periodic frequency.

This implies the identification of each individual to avoid duplication. It requires the definition of the territory to be covered and parts thereof and reasonable access to people. In view of these requirements, censuses should take place at a time when people are easily contacted in their usual place of residence. Accordingly, the timing of the census should avoid times of the year when weather may make communications difficult or when people may be away from their usual residence, as in case of seasonal work away from home or holidays.

The substantial resources involved tend to limit the periodicity of census to every 5 years, such as in the case of Australia, or ten as in the United States and many other countries.

The enumeration of individuals may follow different rules depending on residence status. For instance, short-term visitors may not be taken into consideration; the same might apply to the representatives of foreign governments or international organizations living in the country. The inverse might apply to members of the armed forces and Foreign Service personnel living in other countries. Further, residents may be classified according to their usual residence (*de jure*) or where they are at the times of the census (*de facto*).

Censuses are major undertakings requiring planning and a large organization. Often, they are the largest data collection in any country. Censuses may follow a traditional approach of enumerating each individual within a given geographical area at a given day or may use different spatial and time approaches. Frequently, population censuses are done in conjunction with housing censuses (United Nations 2008). The extent of the information collected varies from country to country. For instance, the census in the United States collects some information from all households and additional information from a sample of households. The long-form has been replaced by a survey form for the 2010 US census (Swanson & Tayman 2012).

In addition to the count and the record of the place of residence, censuses may collect information on a range of individual characteristics. The most common are age and sex. However, information on marital status, level of education, labour force status, occupation, income, country of birth, religion and previous place of residence may also be collected (Australia 2011a). Census may also be used to estimate both domestic and international migration by asking questions on the residence of individuals being enumerated at some past date such as the previous census.

The difficulties in carrying out the total enumeration of a given population tend to result in inaccuracies. There are two main kinds of errors: those arising from omission of individuals from the census (coverage) and those from misreporting of information by individuals in the census (response). Therefore, validity tests tend to be carried out to improve accuracy. These include the examination of age and sex distributions using information from previous censuses and existing knowledge of fertility, mortality and migration from vital statistics and other sources. Post-enumeration surveys of a representative sample of the population may also be used to test coverage and response errors in reporting of individual characteristics (United Nations 2008).

2.3.2 *Sample Surveys*

Sample surveys of population are another useful method of collecting demographic data. They are useful complements to demographic information obtained through censuses, and other sources.

The dated nature of censuses information also makes surveys useful tools in the compilation of demographic estimates between censuses. Surveys have a number of advantages. They tend to be less costly than censuses, can be designed with a specific purpose in mind and can dwell in depth into the subject matter. Further, they can be used in combination with demographic data obtained from other sources. After censuses are carried out, post-enumeration surveys enable the estimation of the census coverage and reliability.

When surveys include all individuals in the target population they have the attributes of a census. Here, the concern is with surveys that use a sample of the target population to assess their characteristics. There are two major types of sample surveys: those based on probability samples and those not based on probability samples. In probability samples each individual in the target population has an equal chance of being selected and thus the probability of selection can be estimated numerically. Non-probability samples take many different approaches and formats. Among them, convenience or accidental sampling is a non-random sample used because of ease of access to the individuals being sampled. Some enquiries use university students, because they are conveniently available subjects. Quota sampling involves the non-random selection of individuals with different characteristics such as sex to ensure that they are represented in the sample. This method may use a proportional approach in considering the adequacy of representation, for example, a sample in which half of the total number are males and the other half are females. Purposive or judgmental sampling involves another non-random selection of subjects. The selection is based on the knowledge of those undertaking the survey of the types of individuals that should be included in the survey. Yet another kind of non-random sampling is snowball sampling. It consists of the selection of some individuals of interest for the survey and then using their knowledge to reach others of interest for the survey. The approach tends to be used when there is difficulty in identifying individuals of interest to be surveyed.

The usefulness of non-probability sampling is limited because it may not be representative of the target population and the survey results may be biased. Therefore, it is uncertain whether findings can be generalised to the target population. A more reliable approach is the use of probability sampling even though it might be more costly (Dillman et al. 2008). Probability sampling allows the estimation of sampling errors which are used to decide the significance or otherwise of an estimate based on the sample (for further information, see any standard textbook on Statistics).

There are a number of probability sample designs that are available. Only four designs that are commonly used are described below. In many surveys more than one sampling techniques may be used.

In *simple random sampling*, individuals or households or items or events are randomly selected from the whole population to be sampled, often using random numbers. Each individual item has an equal chance of being selected and given equal value. This is the simplest of probability sampling techniques.

For the *systematic sampling*, an inventory (list) of individuals or households or items or events is compiled or is available to constitute the sampling frame. Then, say, every fifth or tenth individual or household on the list is selected for the sample. This is not strictly a random sample but is usually a reasonable practice, as long as there is no bias in the structure or order of the listing or some periodicity.

The *multistage sampling* usually involves the random selection of a higher order unit such as a region. From this, a second stage sampling may involve the random selection of households in that region. In a third stage, individuals may be randomly selected from the households selected.

The *stratified sampling* technique ensures sufficient representation in the sample of different groups in the population, the sample may be made up of strata according to the characteristics of different groups in the population. This may result in a lower cost and more representative sample. The strata size may be in proportion to its proportion in the population, otherwise relative weights need to be given to each stratum in relation to its proportion in the population.

Cluster sampling may be used when the population to be sampled is bunched and each bunch or group is close to being representative of the population or having similar characteristics. The clusters are randomly selected and every individual, or household, or item or event in the selected clusters may be included in the sample. In order to ensure representative findings, it is important that either the clusters are all of the same size or weights are assigned to overcome the difference in size of the different clusters.

Surveys can be concerned with a variety of demographic information such as fertility, health and disability status, labour force participation, training and occupation, income levels, wealth and poverty, nutrition and exercise, time use and household expenditure. They may be ad hoc in nature or be undertaken periodically and allow for the assessment of change and trends over time. They may be concerned with the whole population or defined groups within it. As is the case with a full census, the major trade-off in a sample survey is between precision, cost and time (Swanson 2012).

Sample surveys concerned with employment are carried out in most countries. Their frequency allows the study of trends over time and seasonality. Their usefulness is often enhanced by piggybacking questions concerned with other subjects as in the case of the Australian Multi-Purpose Household Survey (Australia 2007).

Housing and household sample surveys can be particularly useful in countries that take a regular census. For instance, the English Housing Survey 2009–10 collected information on age, economic status, ethnicity and marital status and income of the household reference person. It also provided information on household size and composition, and type of housing tenure (United Kingdom 2011). Household sample surveys are conducted frequently to assess household income and expenditure patterns, such as the Family Income and Expenditure Survey of Japan (Japan 2012). Yet, other sample surveys collect data on a range of human activities including how people use their time (Australia 2008).

During the period 1974–1983, the World Fertility Surveys were carried out in 45 developing countries to collect information on fertility and related demographic issues (Sprehe 1974; Cleland & Scott 1987). They were followed by the Demographic and Health Surveys, again among many developing countries to gain information on fertility, mortality and health related issues (Zuehike 2009). Other countries also conduct surveys with a focus on health status. One such survey is the United States National Health Interview Survey that has collected information periodically since 1957 (United States undated).

2.3.3 Registration of Births, Deaths, Marriages and Divorces

Usually, the compulsory registration of births, deaths, marriages and divorce can be a valuable source of demographic information. However, its degree of completion varies depending on the level of civil administration of a country and difficulties in reaching people in less accessible areas. Accordingly, some countries have to rely on sample surveys and other methods to gain a more complete view of fertility and mortality, especially in rural areas. For instance, a study in the 1990s indicated that registration of deaths in the Philippines accounted for only 68 % of estimated deaths (Fiegler & Cabigon 1994). In some countries, registration data can be more a source of confusion than usefulness, as far as demographic information is concerned, and care needs to be taken in testing its coverage and reliability; while in others these data are most valuable.

Birth registration can present problems such as the distinction between live and stillbirths and related measures of birth and perinatal death rates (Sect. 2.2.2). This can lead to both the under-enumeration of live births and infant mortality. Birth registration tends to record the date of the birth and sex of the issue, parents' details including age, usual residence, marital status and occupation.

Death registration generally records the age and sex of the individual, place of birth, marital status, occupation and cause of death. Records can suffer from various shortcomings. The first is the question of completeness of coverage due to similar issues as those of birth registration. Further, details regarding age, marital status and occupation can also pose challenges due to definitions and memory lapses. The cause of death is also an issue in terms of definitions. There can also be lack of coverage because, in some cases, deaths may take place without medical assessment of the cause of death. For instance, non-medically certified deaths constituted about 39 % of registered deaths in Malaysia in 2008 (Malaysia 2010).

The value of registration data on marriages and divorces, as a source of information on family relationships is changing considerably, as social values and partnerships evolve. Registration usually records *de jure* but not *de facto* relationships or separations. *De facto* marital status has risen considerably in some countries. This has led to some countries to offer legal status to cohabitation and even registration without legal marriage (OECD 2010). Further, there may be differences in legal statutes concerning marriage and divorce within the same country and across countries. Nevertheless, marriage registration continues to provide information on a dimension of social relationships: legal conjugal status. The recorded information usually includes the date of the event and of birth and sex of the partners, their usual residence, occupation and religion, as well as any previous legal marital status. In addition, divorce may also include the information on any children involved.

2.3.4 Population Registers

Population registers have been established mostly for administrative purposes. However, they can be useful sources of demographic data (United Nations 2011). They provide for the systematic record of the identity and characteristics of each individual resident either in a country or an area within. The data recorded may include the date of birth, sex, marital status, place of birth and nationality. In some cases, a unique number may be given to each individual on the register.

For a population register to supply reliable statistics it should have attributes such as:

- country-wide coverage
- individual as the recording unit
- legally compulsory registration
- organization that may be centralized or decentralized depending on the country
- standardized definitions
- up-to-date and complete information
- links with population and housing census
- appropriate human resources
- relevant facilities and computer capacity.

The accuracy and timeliness of the data in population registers can be affected when vital events and statistics are recorded by different agencies and the register is dependent on record linkage from more than one source. In some instances, delay in the transmission of information from local area registers to central agencies can affect their accuracy (Netherlands 2012). Yet, another issue is international and internal migration that may lead to duplication or omission due to differences in definitions to distinguish visitors from residents. Concerns have also been raised about the confidentiality of a wide range of information about a single individual in a given source. Some approaches to data handling can reduce some of the risks involved. Similar issues have been raised about confidentiality of census information.

Some Scandinavian and other countries are compiling their censuses from their population registers and complement them with other sources. For instance, Finland has established a Population Register Centre and uses an individual identity number. The Central Population Register began to function in 1971. In 1990, the population census was based on data from about 30 different administrative registers that made available demographic information such as age, sex, education, employment status, economic activity and income (Finland 2004). Population registers can also be used to estimate both international and internal migration.

2.3.5 Administrative Records

Administrative records are usually established for specific purposes such as government or private activities. This primary function may pose problems in the collection of demographic data because of:

- limited coverage
- concepts and definitions
- timeliness
- access
- consistency
- accuracy.

Marquis et al. (1996) found that an examination of records from a number of United States programs posed challenges when attempts were made to match names and residence of people registered. However, they also found that some people registered in these programs had been missed by the census. This implies that although administrative records may be, at times, of poor quality they may also contain useful information.

An advantage in the use of administrative records is that most of the large costs involved in data collection have already been incurred. The growing use of electronic data entry and classification has also improved retrieval of these data and their manipulation for statistical purposes. Greater demand for statistics, concern with data collection fatigue and sometimes budgetary constraints have contributed to efforts in the use of administrative records as a source of statistics (Brackstone 1987).

Administrative records collect a wide range of information at various levels from countrywide programs such as taxation and social security benefits to local government administrative concerns such as housing. The registration of vital events has already been discussed as a source of demographic information (Sect. 2.3.3). Statistics from schools are a major source of information on enrolments and levels of education achieved. Hospital statistics can contribute to the assessment of dimensions of health status. Administrative records of international arrivals and departures and the granting of visas are useful in the estimation of international migration. Taxation statistics can also be used to assess levels of income and also about the movement of people from year to year. The same applies to data kept by public utilities such as electricity, gas and water services.

2.4 Availability of Demographic Data from International Agencies

The major sources of demographic information are usually national agencies charged with the collection and dissemination of statistics. They may collect the data from primary sources through censuses and surveys or retrieve data from other agencies, individuals or corporations. Information may be made available in a variety of media and through several means. These include publications, special tabulations, samples of individual records in electronic media or through the Internet.

The United Nations publishes a *Demographic Year Book* with statistics from member countries. The United Nations Population Fund produce *UNFPA State of World Population*, and the United Nations Development Programme publishes the *Human Development Report*. All of these publications provide demographic information on a periodic basis. Similarly, other United Nations specialised agencies such the World Health Organization's *World Health Statistics*, the International Labour Organization's *World of Work Report*, the United Nations Educational, Scientific and Cultural Organization's *Global Education Digest*, to name a few, all publish a range of demographic information related to their responsibilities. The World Bank produces a *World Development Report* and various regional agencies also compile and publish a range of demographic information. Further, the Organization for Economic Cooperation and Development collects demographic information for industrialized countries and the EUROSTAT for the European countries. In addition to publications, most national and international agencies also provide demographic data on their websites through the Internet. Links to some of the major organisations are given below:

European Commission Statistics (EUROSTAT)
http://ec.europa.eu/index_en.htm
International Labour Organization (ILO)
http://www.ilo.org/global/lang–en/index.htm
Organization for Economic Cooperation and Development (OECD)
http://www.oecd.org/
United Nations Development Programme (UNDP)
http://www.undp.org/content/undp/en/home.html
United Nations Educational Scientific and Cultural Organization (UNESCO)
http://en.unesco.org/
United Nations Entity for Gender Equality and the Empowerment of Women
http://www.unwomen.org/
United Nations Population Division
http://www.un.org/esa/population/
United Nations Population Fund (UNFPA)
http://www.unfpa.org/public/
United Nations Statistical Division
http://unstats.un.org/unsd/default.htm

World Bank (WB)
    http://www.worldbank.org/
World Health Organization (WHO)
    http://www.who.int/en/

The following website provides the links to the central statistical agencies of all member countries of the United Nations:

http://unstats.un.org/unsd/methods/inter-natlinks/sd_natstat.asp

References

Australia. (2007). *Labour statistics: Concepts, sources and methods, Apr 2007*. Catalogue no. 6102.0.55.001. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/29F8946CF59B0168CA2572C100244BD7. Accessed Feb 2013.

Australia. (2008). *How Australians use their time, 2006*. Catalogue no. 4153.0. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/ausstats/abs@.nsf/mf/4153.0. Accessed Feb 2013.

Australia. (2009). *Perinatal deaths, Australia 2007*. Catalogue no. 3304.0. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/BDDA067A6B6A6AE1CA257705001DF4F1?opendocument. Accessed Feb 2013.

Australia. (2010). *Household and family projections, Australia, 2001 to 2026*. Catalogue no. 3236.0. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/ausstats/abs@.nsf/mf/3236.0. Accessed Feb 2013.

Australia. (2011a). *Census dictionary Australia 2011*. Catalogue no. 2901.0. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/ausstats/abs@.nsf/mf/2901.0. Accessed Feb 2013.

Australia. (2011b). *2009 births, Australia*. Catalogue no. 3301.0. Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/10BEDC49AFCACC1FCA2577CF000DF7AB/$File/33010_2009.pdf. Accessed Feb 2013.

Australia. (2011c). *2009 deaths, Australia*. Catalogue no. 3302.0. Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/CCE847BF63D7BF93CA2577D60010AC3D/$File/33020_2009.pdf. Accessed Feb 2013.

Brackstone, G. J. (1987). Issues in the use of administrative records for statistical purposes. *Survey Methodology, 13*(1), 29–43. http://www.oecd.org/std/36237567.pdf. Accessed Feb 2013.

Canada. (2012). *Births 2009*. Ottawa: Statistics Canada. http://www.statcan.gc.ca/pub/84f0210x/84f0210x2009000-eng.pdf. Accessed Feb 2013.

Cleland, J., & Scott, C. (Eds.). (1987). *The world fertility survey: An assessment of its contribution*. London: Oxford University Press.

Dillman, D., Smyth, J., & Christian, L. (2008). *Internet, mail, and mixed-mode surveys: The tailored design method*. New York: Wiley.

Fiegler, W., & Cabigon, J. (1994). *Life table estimates for the Philippines, its regions and provinces, by sex: 1970, 1980 & 1990*. Manila: Department of Health.

Finland. (2004). *Use of registers and administrative data for statistical purposes: Best practices of Statistics Finland*. Handbook 45. Helsinki: Statistics Finland.

Japan. (2012). *Annual report on the family income and expenditure survey 2011*. Tokyo: Statistics Bureau. http://www.stat.go.jp/english/data/kakei/156index.htm. Accessed Feb 2013.

Malaysia. (2010). *Statistics on the causes of death, Malaysia 2008*. Putrajaya: Department of Statistics. http://www.statistics.gov.my/portal/download_Population/files/BPD/perangkaan_sebab_kematian2006.pdf. Accessed Feb 2013.

Marquis, K., Wetrogan, S., & Palacios, H. (1996). *Towards a U.S. database from administrative records*. Washington, DC: Census Bureau. http://www.census.gov/srd/papers/pdf/km9601.pdf. Accessed Feb 2013.

Netherlands. (2012). *Population counter*. The Hague: Statistics Netherlands. http://www.cbs.nl/en-gb/menu/themas/bevolking/cijfers/extra/bevolkingsteller.hmt. Accessed Feb 2013.

OECD. (2010). *Cohabitation rate and prevalence of other forms of partnership*. Paris: Organization for Economic Cooperation and Development. http://www.oecd.org/els/soc/41920080.pdf. Accessed Feb 2013.

Sprehe, T. (1974). The World Fertility Survey: An international program of fertility research. *Studies in Family Planning, 5*(2), 35–41.

Swanson, D. A. (2012). *Learning statistics: A manual for sociology students*. San Diego: Cognella Press.

Swanson, D. A., & Tayman, J. (2012). *Subnational population estimates*. Dordrecht: Springer.

United Kingdom. (2011). *English housing survey: Household report 2009–10*. London: Department for Communities and Local Government. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6718/1937206.pdf. Accessed Feb 2013.

United Nations. (2008). *Principles and recommendations for population and housing censuses – revision 2* (Statistical papers series M no. 67/Rev.2). New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/sources/census/docs/P&R_%20Rev2.pdf. Accessed Feb 2013.

United Nations. (2011). *Population registers*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/sources/popreg/popregmethods.htm. Accessed Feb 2013.

United States. (undated). *National health interview survey*. Atlanta: Centers for Disease Control and Prevention. http://www.cdc.gov/nchs/nhis.htm. Accessed Feb 2013.

WHO. (2010). *International classification of diseases (ICD)*. Geneva: World Health Organization. http://www.who.int/classifications/icd/en/index.html. Accessed Feb 2013.

Zuehike, E. (2009). *The demographic and health surveys at 25 years and beyond*. Washington, DC: Population Reference Bureau. www.prb.org/Articles/2009/dhspolicyseminar.aspx. Accessed Feb 2013.

Chapter 3
Some Basic Statistical Measures

3.1 Purpose

The purpose of this chapter is to introduce some basic statistical measures that are commonly used in demographic analysis. The concepts are defined in general terms without going into theoretical details. Methods of calculation of various measures are described. The statistical measures discussed in this chapter consist of counts, frequencies, proportions, rates, various measures of central tendency, dispersion, comparison, correlation and regression.

3.2 Demographic Data and Analysis

Demographic data can be classified according to their *level of measurement*. This is useful because the level of measurement helps the selection of what statistical analysis is most appropriate. Several classifications can be used. This book uses the four-fold classification system proposed by Stevens (1946) that classifies data as being (1) nominal, (2) ordinal, (3) interval and (4) ratio. There are other systems such as the two-fold classification of (1) discrete and (2) continuous.

In the classification system proposed by Stevens, the *nominal* level is known as the *lowest* level of measurement. Here, the values just *name* the attribute uniquely and do not imply an ordering of cases. For example, the variable *marital* is inherently nominal. In a study it might be useful to have attributes such as never married, married, separated, divorced and widowed. These attributes are mutually exclusive and exhaustive. They could be coded *N*, *M*, *S, D* and *W* respectively, or coded as *1*, *2*, *3*, *4* and *5*. In the latter, the numbers are not numbers in a real sense since they cannot be added or subtracted. Thus, numbers assigned to serve as values for nominal level variables such as *marital status* cannot be added, subtracted, multiplied or divided in a meaningful way. An exception is the *dummy coding* of

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_3,

a nominal level variable with two values: *0* and *1*. Dummy coding allows some limited functions to be performed.

The next level of measurement is *ordinal*. Although it is at a higher level of measurement than the nominal category, it is still considered a *low level of measurement*. The attributes are rank-ordered but the distances between the attributes are not fixed. For instance, in a questionnaire, people are asked if they think that population growth is a major world problem and five possible response categories are given: (1) strongly disagree, (2) somewhat disagree, (3) neither agree nor disagree, (4) somewhat agree and (5) strongly agree. It can be seen that *strongly agree* coded as (5) indicates more agreement than a response coded as (4) *somewhat agree*, and so on. However, the distance between the five categories is not fixed. It cannot be said whether the *distance* from *strongly disagree* (1) to *somewhat disagree* (2) is the same as that from *strongly agree* (5) to *somewhat agree* (4). All it can be said is that codes (1) to (5) preserve the order of agreement. Coding ordinal levels with numbers is a *grey area* between the coding with numbers at nominal level (for which there is no meaningful addition, subtraction, multiplication and division) and interval level numbers that can be added and subtracted, and ratio level numbers for which there is meaningful addition, subtraction, multiplication and division. In some situations, numbers at ordinal level can be treated as having an interval and in others they cannot. It depends on the coding assigned to the attributes of ordinal level variables.

The *interval* level is a much higher level of measurement than the nominal level and even higher than the ordinal level. At this level, the numbers of the *distance* between attributes that can be meaningfully added and subtracted. For example, the difference in temperature between 30 and 40 °F is the same distance as between 70 and 80 °F is 10°. This means that there is a fixed distance between the temperature levels – as when a ruler is used to measure the distance between two points using the same distance scale. The distance between 2° and 4° Fahrenheit is the same as that between 7° and 9° on the same scale. However, ratios make no sense as 80° is not twice as hot as 40° Fahrenheit because there is no *true zero* on the Fahrenheit scale, where all molecular action ceases and there is complete absence of heat. This means that numbers measured at the interval level cannot be multiplied or divided in a meaningful way.

At the highest level of measurement is the *ratio*. A *true zero* exists for the numbers at this level. This means that numbers can be multiplied and divided as well as added and subtracted in a meaningful way. For instance, in the Fahrenheit system for measuring temperature, zero is not the absence of heat. Accordingly, while it can be said that the distance between 60° and 45° is the same as between 15° and 30°, it cannot be stated that 60° is twice as hot as 30°. Thus, measurement is at the interval level because the distance between points is fixed and on the same scale. However, in the Kelvin system for temperature measurement there is an absolute zero at which molecular action ceases. In that case, it can be stated that 20°

Table 3.1 Daily income of 30 persons: hypothetical data

Person	Income ($)	Person	Income ($)	Person	Income ($)
1	59.60	11	58.10	21	87.40
2	49.00	12	56.30	22	90.10
3	98.30	13	47.20	23	57.40
4	14.20	14	36.20	24	74.20
5	78.30	15	74.30	25	58.30
6	32.10	16	56.70	26	63.90
7	69.00	17	31.00	27	92.20
8	45.70	18	97.40	28	52.30
9	11.00	19	89.10	29	69.10
10	32.80	20	38.30	30	70.10

Table 3.2 Frequency distribution of daily income

Income group ($)	Width of the income group	Frequency (no. of persons)
0–29	30	2
30–49	20	8
50–69	20	10
70–89	20	6
90–99	10	4
Total	...	30

Source: Table 3.1

on the Kelvin scale are twice as hot as 10° on that scale, since not only are the distances fixed, as in the Fahrenheit scale, but there is also a true zero. This means that the Kelvin scale is at the ratio level of measurement.

Variables at higher levels of measurement can be transformed into lower levels of measurement. This is the case with *grouping* the ratio level variable of income in Table 3.1 into categories shown in Table 3.2. Here the variable measured at ratio level was transformed into a variable measured at ordinal level. This process could be continued by classifying the income data into two groups labelled *Yes* and *No*, where these two categories would be those with income below or above a certain amount of income (say the average income). This transformation would re-code income data into a nominal level variable.

The analysis of demographic information can be of a *descriptive* nature, as in the case of the estimation of percentages, measures of central tendency and dispersion discussed later. It may also be of an *inferential* nature. In that case, demographic analysis is concerned with relationships either between groups in or characteristics of population, as in the case of correlation and regression, briefly discussed later. Statistical methods should be consistent with the nature of the data, as categorical and ranked data may require different methods of analysis.

3.3 Counts and Frequencies

Counts are just one number or a series of numbers. For example, the estimated population of the world was 6,671 million people in mid-2007 (United Nations 2009). This is just one count. Other counts may relate to the population of different regions of the world, for example, that of Asia was 4,029 million and Europe 731 million.

On a smaller scale, counts may consist of data based on a few observations, such as the hypothetical data on daily income (dollars) of 30 people in Table 3.1.

The daily income of these 30 people can be summarised into income groups that cover the range of individual incomes. In making this conversion, data at the interval level is transformed into ordinal level data. In this instance, each individual is classified according to their daily income and counted as a member of the relevant income group. The result is given in Table 3.2. However, the size of each income group could be different and there could be a smaller or larger number of income groups. Nevertheless, the sum of all groups would continue to be 30. Such a distribution is called a *frequency distribution* because it counts the number of observations (frequency) in each group. Three features of each group need to be considered, when estimates are made from the frequency distribution: the lower and upper limits of the group (class) and the width (interval) of each group. In Table 3.2, the lower limits of the groups are 0, 30, 50, 70 and 90, and the upper limits are 29, 49, 69, 89 and 99. In fact, the upper limits are just less than 30, 50, 70, 90 and 100 respectively. As stated, in this example the width (interval) is the same for all, except the first, income groups but groups may have unequal intervals.

The data presented in Table 3.1 are an example of *ungrouped data*, while Table 3.2 gives an illustration of *grouped data*. In this example, the ungrouped data are measured at the ratio level while in their grouped form they are measured at the ordinal level.

3.4 Proportions and Percentages

Given a series of numbers $x_1, x_2, \ldots, x_n$, where x_1 is the first, x_2 is the second, and so on, x_n is the n^{th} number (last in the series), x_1 as a proportion of the total is defined as:

$$Proportion\ (x_1) = \frac{x_1}{\sum_{i=1}^{i=n} x_i} \tag{3.1}$$

For example, in Table 3.2 the proportion of persons earning between \$30 and \$49 is 0.266 ($= \frac{8}{30}$). This converted into a percentage is 0.266*100 = 26.6 %. It means that in this hypothetical population nearly 27 out of every 100 persons earn \$30–\$49.

Table 3.3 Calculation of proportion and percentage frequency distributions of daily income

Income group ($)	Frequency	Proportion	Percentage (%)
(1)	(2)	(3) = (2)/Total	(4) = (2)*100/Total
0–29	2	0.0667	6.7
30–49	8	0.2667	26.7
50–69	10	0.3333	33.3
70–89	6	0.2000	20.0
90–99	4	0.1333	13.3
Total	30	1.0000	100.0

Source: Table 3.2

Table 3.4 Distribution of daily income by sex: hypothetical data

	Sex of the person		
Income group ($)	Female	Male	All persons
(1)	(2)	(3)	(4) = (2) + (3)
0–29	1	1	2
30–49	5	3	8
50–69	4	6	10
70–89	2	4	6
90–99	1	3	4
Total	13	17	30

The world population in 2007 was 6,671 million and 4,029 million lived in Asia at that time (United Nations 2009). Accordingly, the proportion of the world population living in Asia was 0.604 ($= \frac{4,029}{6,671}$) or 60.4 %. In other words, there were close to 60 people living in Asia out of every 100 in the world.

Frequency distributions can be converted into proportional or percentage frequency distributions by dividing the frequency in each group by the total number, as illustrated in Table 3.3.

A hypothetical distribution of males and females by income group is shown in Table 3.4. Such a table is called a *bivariate* table as it involves two variables: income and sex of the person. It can also be called a *five by two* table since it cross-classifies five income values by two sex values. In this case, income is at the ordinal level of measurement and sex is at the nominal level.

There are three ways of calculating percentages from a bivariate table, as shown in Tables 3.5, 3.6 and 3.7.

Percentages have been calculated separately for females and males in Table 3.5. The percentage total in each column is 100. The percentages in the table indicate that the highest percentage of females is in the $30–$49 income range and that for males in the $50–$69 income group. The last column has the percentage distribution of all persons according to income. It is the same as that given in the fourth column of Table 3.3.

The percentage distribution of males and females can also be estimated for each income group as in Table 3.6. Each row adds to 100. In percentage terms, females are predominant in the $30–$49 income group, while males are relatively more numerous in the $90–$99 income range. The last row in the table shows that 43.3 % of the 30 people are female.

Table 3.5 Example of column-wise percentages

	Sex of the person		
Income group ($)	Female	Male	All persons
0–29	7.7	5.9	6.7
30–49	38.5	17.6	26.7
50–69	30.8	35.3	33.3
70–89	15.4	23.5	20.0
90–99	7.7	17.6	13.3
Total	100.0	100.0	100.0

Source: Table 3.4
Note: Totals for columns may not add due to rounding

Table 3.6 Example of row-wise percentages

	Sex of the person		
Income group ($)	Female	Male	All persons
0–29	50.0	50.0	100.0
30–49	62.5	37.5	100.0
50–69	40.0	60.0	100.0
70–89	33.3	66.7	100.0
90–99	25.0	75.0	100.0
Total	43.3	56.7	100.0

Source: Table 3.4
Note: Totals for rows may not add due to rounding

Table 3.7 Example of percentages of the total

	Sex of the person		
Income group ($)	Female	Male	All persons
0–29	3.3	3.3	6.7
30–49	16.7	10.0	26.7
50–69	13.3	20.0	33.3
70–89	6.7	13.3	20.0
90–99	3.3	10.0	13.3
Total	43.3	56.7	100.0

Source: Table 3.4
Note: Totals in columns and rows may not add due to rounding

Each cell of Table 3.4 expressed as a percentage of the total number (30) results in the percentages in Table 3.7. For example, males with incomes ranging $50–$69 constitute the highest proportion (20 %) of any female or male income group.

3.5 Ratios, Rates and Probabilities

Earlier, counts (absolute measures such as population size, births and deaths) were discussed as well as two relative measures (that focus on the relationship between two or more numbers) proportion and percentage. Relative measures can also be

Table 3.8 Calculation of the ratios of males to females in Asia and Europe, 2007

Region	No. of males (millions)	No. of females (millions)	Ratio of males to females	
			Per female	%
(1)	(2)	(3)	(4) = (2)/(3)	(5) = (4)*100
Asia	2,063	1,966	1.05	105
Europe	352	379	0.93	93

Source: United Nations (2009)

expressed as ratios, rates and probabilities. Relative measures are similar, but each has a distinct meaning.

A ratio is simply a number divided by another. However, in demographic analysis a ratio is usually the quotient of two groups of people with different characteristics in the same population or the quotient of one variable by another. As stated earlier, in 2007 there were 6,671 million people in the world; the number of males was 3,363 and females 3,308 million (United Nations 2009). The ratio of males to females was:

$$\frac{3,363}{3,308} = 1.02 \text{ males for every female} = 102 \text{ males per } 100 \text{ females.}$$

Male to female ratios for Asia and Europe are given in Table 3.8.

The number of births in China during 2007 was estimated as 15.940 million and the population in mid-2007 was 1,324.655 million (United Nations 2009). Accordingly, the rate of births in China during 2007 was:

$$\frac{15.940}{1,324.655} = 0.0120 = 12 \text{ births per } 1,000 \text{ population.}$$

In demographic analysis, a *rate* is a special type of ratio. It relates a number of events or some other demographic variable to the average population during the specified period of time, usually the mid-period population. In other words, a rate is the number of events (or some other demographic variable) occurring during a given period of time divided by the population *at risk* of the occurrence of those events (or the population to which the demographic variable relates to). For example, the birth rate is the number of births from a given population divided by the population exposed to the risk of having those births. Rates of births to population in Egypt and the United States are presented in Table 3.9. Some rates may be expressed per person, in percentages, per thousand people or even per hundred thousand depending on the size of the number of events in relation to that of the relevant population.

Although the concept of a rate is clear, it is often difficult or practically impossible to get an exact measure of the population at risk to the occurrence of the particular event. For instance, only females in their childbearing years are

Table 3.9 Calculation of birth rates in Egypt and USA, 2007

Country	No. of births (millions)	Midyear population (millions)	Birth rates	
			Per person	Per 1,000
(1)	(2)	(3)	(4) = (2)/(3)	(5) = (4)*1,000
Egypt	1.950	73.644	0.0265	26.5
USA	4.317	301.621	0.0143	14.3

Source: United Nations (2009)

subjected to the risk of giving birth. Some may die during the period – in a given area – some may move away and other move in. How can the population exposed to the risk of giving birth in that area be defined and measured? This problem is usually solved by using the midyear (year or period of time) population as an approximation. This solution is based on the assumption that births and deaths occur evenly throughout the year, so that the midyear population is a measure of the average of the reference population during the year. The resulting rate is called a *central rate*. Usually, the use of the midyear population is adequate to get reasonable results. However, if the population is growing at a substantial rate throughout the year, it might be more adequate to use the average of the population at given intervals during the year.

A distinction can also be made between central rates and *probabilities*. In a central rate the denominator is the population at the mid-point of the time period (say the middle of the year). It is meant to represent the average population during the time period. In a probability, the denominator is the population at the beginning of the time period. This is thought to correspond more closely to the population at risk of the occurrence of an event during the time period.

The distinction between central rates and probabilities is somewhat fuzzy because of the movement of people into and out of an area. For example, if age-specific rates are considered, a true single-year probability can be calculated using the population of the area at the beginning of the year and the number of deaths that took place during the year to members of the population, at the beginning of the year, for each age cohort. However, some deaths will be missed by the death registration system for the area because they relate to people who moved out of the area during the year. Further, some deaths will be improperly included because they refer to people who moved into the area during the year. Consequently, it is difficult (if not impossible) to estimate true probabilities for a variety of demographic measures.

The term *rate* is used loosely in demography, as it is elsewhere. Many measures called rates are really ratios. A growth rate, for example, is a ratio of population change over a time period to the population at the beginning of the time period. It is not a rate in the strictest sense because an area's population growth comes not only from the population of the area itself, but from other populations as well due to migration.

3.6 Measures of Central Tendency

There are many measures of central tendency. The most commonly used are the arithmetic mean (also referred to as the mean), median and mode. The mean, median and mode can be used with variables measured at the interval or ratio levels, and some times at the ordinal level. If variables are measured at the ordinal or nominal level, the mode is generally the most meaningful measure of central tendency. Two other less frequently estimated means are also discussed: the geometric and harmonic means. As their names suggest, they were designed to use variables measured at the interval and ratio levels.

3.6.1 Arithmetic Mean

The arithmetic mean (*AM*) for a series of numbers $x_1, x_2, \ldots, x_n$ is:

$$AM = \frac{\sum_{i=1}^{i=n} x_i}{n} \tag{3.2}$$

n is the total of the numbers in series (x_1 to x_n).

The estimated mean of the income data in Table 3.1 is:

$$AM = \frac{59.60 + 49.00 + \ldots + 69.10 + 70.10}{30} = \frac{1,789.60}{30} = \$59.65$$

For the grouped data, *AM* is calculated as follows:

$$AM = \frac{\sum_{i=1}^{i=n} f_i * x_i}{\sum_{i=1}^{i=n} f_i} \tag{3.3}$$

f_i represents the frequency of the i^{th} group, x_i stands for the mid-point of the interval of the i^{th} group and the summation is for all values of i. It is assumed that the individual incomes, or any other variable in the calculation of the mean, are evenly distributed within each group, and the mid-point in the group is taken as the average value of x_i for the whole group.

Equation (3.3) can also be used to calculate the *weighted mean*, when weights are used instead of frequencies.

An example of the computations involved in the estimation of the mean in grouped data (Table 3.2) is given in Table 3.10. The mid-point of each group is

Table 3.10 Calculation of the arithmetic mean for hypothetical data on daily income

Income group ($)	Frequency (f_i)	Mid-point (x_i)	Product (f_i*x_i)
(1)	(2)	(3)	(4) = (2)*(3)
0–29	2	15	30
30–49	8	40	320
50–69	10	60	600
70–89	6	80	480
90–99	4	95	380
Total	30	...	1,810
$AM = \frac{1,810}{30} = \$60.33$			

Source: Table 3.2

shown in column (3). For instance, the first group contains all incomes under $30; this means that the mid-point is 15 (***not*** 14.5). The mid-point (x_i) is multiplied by the frequency for that group in column (2), *i.e.*, (f_i). These products (f_i * x_i) in column (4) are summed and divided by the sum of frequencies to give the arithmetic mean (*AM*) of the frequency distribution in Table 3.10.

This value of *AM* ($60.33) is close to $59.65 estimated from the ungrouped data in Table 3.1. When the frequency total is relatively small, as in this case ($n = 30$), there might be some difference between the means derived from ungrouped and grouped data.

If the width of the last group is *open ended*, that is, the upper limit is not specified, then the estimation of the mean will depend upon the assumption regarding the length of that interval. For example, if in Table 3.10, the last income group was *90 and over*, the mid-point would be 105 if it is assumed that the upper limit of this group is less than $120.

The *AM* may be affected by extreme values. For instance, in a series of six numbers 23, 25, 30, 250, 20, and 24, the *AM* is 62, despite the fact that all but one values lie between 20 and 30. The following measures of central tendency are not affected by extreme values to the same extent as the *AM*.

3.6.2 Geometric and Harmonic Means

Two of the less commonly used means are the *geometric mean (GM)* and the *harmonic mean (HM)*. For a given set of numbers x_1, x_2, ..., x_n they are calculated as:

$$GM = (x_{1*}x_{2\,*}\ldots{}_*x_n)^{\frac{1}{n}} \tag{3.4}$$

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}} \tag{3.5}$$

Six numbers (23, 25, 30, 250, 20 and 24) have been chosen to illustrate the estimation of the *GM* and *HM*:

$$GM = (23 * 25 * 30 * 250 * 20 * 24)^{\frac{1}{6}}$$

$$GM = (2,070,000,000)^{\frac{1}{6}}$$

Taking the logarithm of 2,070,000,000 and following rule (3) of Box 3.1:

$$\log GM = \frac{1}{6} * \log(2,070,000,000)$$

$$\log GM = \frac{9.31597}{6}$$

$$\log GM = 1.55266$$

Taking the antilogarithm of 1.55266:

$$GM = 35.70$$

Box 3.1 Calculation of the n^{th} Root Using Logarithms

The n^{th} *root* of a number *K*, written as $(K)^{\frac{1}{n}}$ or $\sqrt[n]{K}$, is a particular number that when multiplied by itself *n* times results in a product that equals *K*. The value of *n* may be any number from 2 onwards.

For example, $(729)^{\frac{1}{6}} = 3$, because 3*3*3*3*3*3 = 729.

An easier way to calculate the n^{th} root is through *logarithms.*

Two types of logarithms are: (1) that use 10 as the base, and (2) that use a mathematical constant *e* (=2.7182828) as the base. The former are referred to as *logarithms* (log) and the latter as *natural logarithms* (LN or ln). The log of 100 is 2, because $10^2 = 10*10 = 100$. The *LN* of 100 is 4.6057, because $e^{4.6057} = 100$.

Antilogarithms enable the conversion of a logarithm back into the original number. The antilogarithm of log is 10^x and that of LN is e^x. Most calculators have log, 10^x, LN or ln and e^x keys.

The following rules apply to both *log* and *LN*:

1. Logarithm $(x*y)$ = logarithm (x) + logarithm (y)
2. Logarithm $\left(\frac{x}{y}\right)$ = logarithm (x) – logarithm (y)
3. Logarithm $(x)^y = y \, * $ logarithm (x)
4. *log(10)* and *LN(e)* are both equal to 1.

$$HM = \frac{6}{\frac{1}{23} + \frac{1}{25} + \frac{1}{30} + \frac{1}{250} + \frac{1}{20} + \frac{1}{24}}$$

$$HM = \frac{6}{0.0435 + 0.0400 + 0.0333 + 0.0040 + 0.0500 + 0.0417}$$

$$HM = \frac{6}{0.2125} = 28.24$$

One of the characteristics of these two means is that they minimise the effect of outliers, such as 250 in this example. The *AM* for the above six numbers is 62 that is substantially larger than the *GM* and *HM* values. In general, the following relationship holds for any series of numbers:

$$AM > GM > HM.$$

Usually, *GM* and *HM* are not calculated for grouped data.

3.6.3 Median

For a given set of numbers ordered in accordance with their value $x_1, x_2, \ldots, x_n$ the *median* is the figure with a value that is in the middle of the set. In other words, the median divides the ordered set into two equal parts: half of the numbers are above the value of the median and half are below that value. Accordingly, the first step in the estimation of the median for the figures in Table 3.1 is to sort the income distribution in *ascending* order:

11.0, 14.2, 31.0, 32.1, 32.8, 36.2, 38.3, 45.7, 47.2, 49.0, 52.3, 56.3, 56.7, 57.4, 58.1, 58.3, 59.6, 63.9, 69.0, 69.1, 70.1, 74.2, 74.3, 78.3, 87.4, 89.1, 90.1, 92.2, 97.4, 98.3

As the number of persons is even, the median lies between the 15th and 16th values. The median daily income is $\frac{58.1+58.3}{2} = 58.2$

If the number of persons had been odd, for example 29, then the median would have been the 15th value 58.1

The median for grouped data is estimated as:

$$Median = L + \left\{ \frac{i}{f} * [(p * n) - C] \right\} \tag{3.6}$$

The first step in the use of the above equation is to determine the group in which the median lies. The p is 0.5 in the case of the median. The median group is identified by halving the total frequency (n). This is the first group that has a

Table 3.11 Calculation of the median for hypothetical data on daily income

Income group ($)	Frequency (f_i)	Cumulative frequency (C)
0–29	2	2
30–49	8	10
50–69	10	20
70–89	6	26
90–99	4	30
Total	30	...

$Median = 50 + \frac{20}{10} * [(0.5 * 30) - 10] = 60$

Source: Table 3.2

cumulative frequency equal or larger than ($\frac{n}{2}$): the group where the median is located. The value of L is the lower limit of the median group, i is the width, and f is the frequency. C is the cumulative frequency of the group preceding the median group and n is the total of the frequencies for all groups.

In Table 3.11, $p = 0.5$, $n = 30$ and the *median group* is the one for which the value of C is just above 15 (half of 30). In Table 3.11, \$50–\$69 is the median group:

$L = 50$ (lower limit of the median group)
$i = 20$ (width of the median group)
$f = 10$ (frequency of the median group)
$n = 30$ (total number of observations)
$p = 0.5$
$C = 10$ (cumulative frequency for the group preceding the median group).

The median of \$60 is slightly different from \$58.2 calculated from the ungrouped data. Ungrouped data give more precise estimates of both *AM* and the median, but demographic data are generally available in grouped formats. Depending on the nature of the analysis the median may be preferred to the *AM*. The median is often used in place of the mean when data are highly skewed that pulls the mean way from the bulk of the observations. The median is not affected by the skewness of the distribution of observations.

3.6.4 Mode

The *mode* is the observation with the highest value in the distribution. In grouped data, it is the group that has the largest frequency (if the class intervals are about the same). In Table 3.11 the mode lies in the \$50–\$69 income group.

Some distributions have a single mode while others may have two or more modes. Figure 3.1 gives the number of live births in Australia by the age of mother (Australia 2010). In this example, the mode is 30–34 years age group, as this has the largest number of births recorded. It is an example of a distribution with one mode (*uni-modal* distribution).

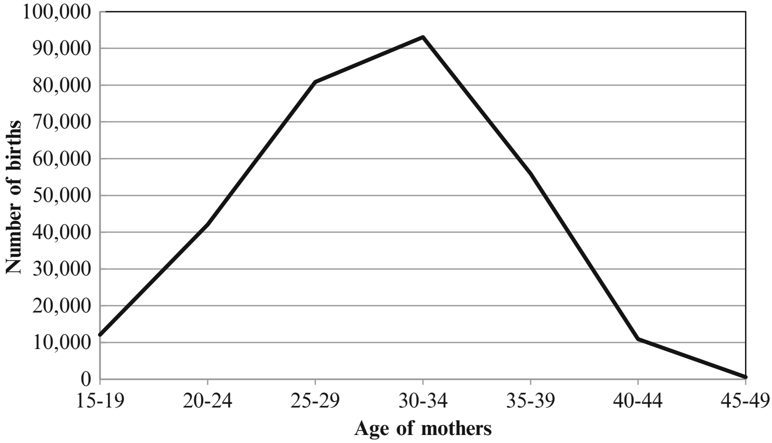


Fig. 3.1 Births in Australia by age of mother, 2009 (*Source*: Australia (2010))

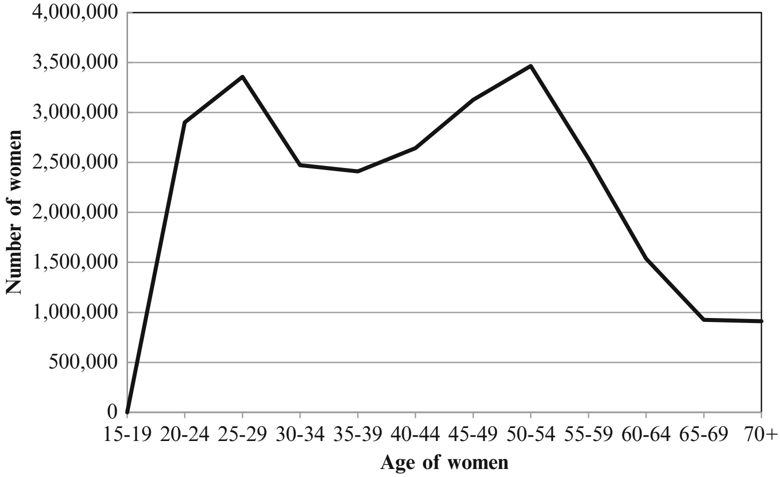


Fig. 3.2 Female labour force in Japan by age, 2000 census (*Source*: Japan (2011))

An example of a distribution with two modes (*bi-modal* distribution) is the number of Japanese females in the labour force by age (Japan 2011) in Fig. 3.2. In this instance, there were two modes: one in the 25–29 years of age group and the other in the 50–54 years group. There may be distributions with more than two modes.

3.6.5 Normal and Skewed Distributions

Figure 3.3 exhibits a bell-shaped distribution called the *normal distribution*. The value of the mode is that corresponding to the top of the hump. This is also the median that divides the distribution into two equal halves. The *AM* of this distribution also coincides

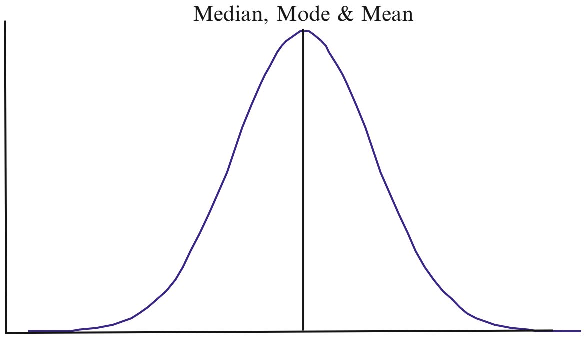


Fig. 3.3 Example of a normal distribution

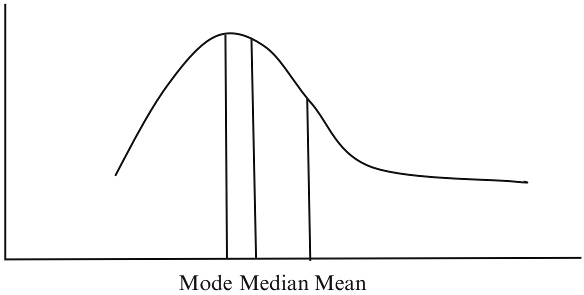


Fig. 3.4 Example of a positively skewed distribution

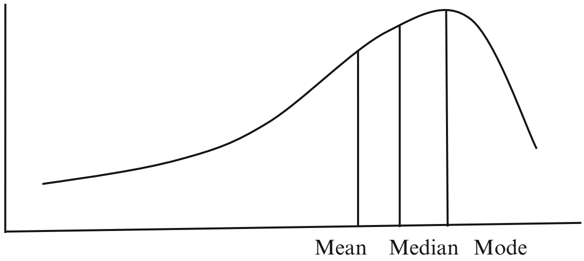


Fig. 3.5 Example of a negatively skewed distribution

with its median and mode. This distribution rarely occurs in the real world. However, it is an important component of the theory that underlies statistical inference.

A non-symmetrical distribution is called a *skewed distribution*. There are two types of skewed distributions: the *positively skewed* and the *negatively skewed*. In a positively skewed distribution the value of both the median and the mean (AM) are larger than the mode. In a negatively skewed distribution they are both smaller than the mode (Figs. 3.4 and 3.5).

The properties of these three types of distributions can be summarised as follows:

Symmetric distribution (normal): $AM = Median = Mode$
Positively skewed distribution: $AM > Median > Mode$
Negatively skewed distribution: $Mode > Median > AM.$

These properties are used in developing measures of skewness.

3.7 Measures of Dispersion

There are a number of measures of dispersion such as the variance, standard deviation and various types of quantiles. The variance and the standard deviation can be estimated for data at the interval and ratio levels and sometimes for data at the ordinal level. They should not be used with data at the nominal level unless the data have been transformed using dummy coding. Quantiles are essentially measures of data at interval and ratio levels that have been transformed into ordinal categories.

3.7.1 Variance and Standard Deviation

One of the most commonly used measures of dispersion is the *variance* or its square root the *standard deviation*. The equations to calculate the variance (*VAR*) and the standard deviation (*SD*) for a set of *n* numbers $x_1, x_2, \ldots, x_n$ that has an *AM* equal to $\bar{x}$ are:

$$VAR = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{n - 1} \tag{3.7}$$

$$SD = \sqrt{VAR} \tag{3.8}$$

Given the hypothetical series of five numbers 23, 25, 30, 20, and 24, the first step in the estimation of these measures is to find the value of $\bar{x}$. According to Eq. (3.2) the value of $\bar{x}$ (AM) is:

$$\bar{x} = \frac{23 + 25 + 30 + 20 + 24}{5} = 24.4$$

Then, the difference of each number from $\bar{x}$ is estimated, as per column (2) of Table 3.12. These differences are squared (multiplied by themselves) column (3) to convert them into positive values. They are summed and divided by the number of figures minus 1 ($5 - 1 = 4$ in this case) to give the variance (*VAR*). The standard deviation (*SD*) is the square root of the variance.

The standard deviation can be viewed as the average expected error that results from using the mean (*AM*) to predict individual values of the variable in question. For example, if *AM* of 24.4 was used to predict the values of the five numbers, the average expected error would be (*SD*) 3.65. This is an important concept to remember because it forms a basis to evaluate a regression model, a topic that is discussed later on.

Table 3.12 Calculation of the variance and standard deviation for a series of six numbers

Number (i)	Deviations from the mean $(x_i - \bar{x})$	Square of deviations from the mean
(1)	(2) = (1) − 24.4	(3) = (2)*(2)
23	−1.4	1.96
25	0.6	0.36
30	5.6	31.36
20	−4.4	19.36
24	−0.4	0.16
Total	0	53.20

$VAR = \frac{53.20}{5-1} = 13.30$

$SD = \sqrt{13.30} = 3.65$

Table 3.13 Calculation of the variance and standard deviation for hypothetical data

Income group ($)	Frequency (f_i)	Mid-points (x_i)	Deviations from mean	Square of deviations	f_i*Square of deviations
(1)	(2)	(3)	(4) = (3) − 60.33	(5) = (4)*(4)	(6) = (3)*(5)
00–29	2	15	−45.33	2,054.81	4,109.62
30–49	8	40	−20.33	413.31	3,306.48
50–69	10	60	−0.33	0.11	1.10
70–89	6	80	19.67	386.91	2,321.46
90–99	4	95	34.67	1,202.01	4,808.04
Total	30	...	...	...	14,546.70

$VAR = \frac{14,546.70}{30-1} = 501.61$

$SD = \sqrt{501.61} = 22.40$

Source: Table 3.10

The equation for the estimation in grouped data is:

$$VAR = \frac{\sum_{i=1}^{i=n} f_i * (x_i - \overline{x})^2}{\left(\sum_{i=1}^{i=n} f_i\right) - 1} \tag{3.9}$$

f_i refers to the frequency of the i^{th} group that has a mid-point of x_i, $\overline{x}$ is the AM of the whole frequency distribution, the summation is done over all values of i. The denominator is sum of frequencies (n) minus 1.

Table 3.13 gives an example of the estimation of the variance and standard deviation of grouped data. These two measures are 501.61 and 22.40 respectively.

3.7.2 Quantiles

Quantiles have been defined as the positional points of a series sorted in ascending order that divide the series into equal proportions. Although the median is usually considered a measure of central tendency, the median is a quantile that divides a series into two equal parts. In addition to the median, *quartiles, quintiles and deciles* are some of the other commonly used quantiles. They represent the partition of a series into four (quartiles), five (quintiles) and ten (deciles) equal parts. The number of positional points required is one less than the partitions, because the last positional point is the highest value of the series. Thus, there is no need to calculate the fourth quartile or the fifth quintile or the tenth decile as they are the last value of the ascending series.

Quantiles for Ungrouped Data

The identification of the *quantile points* involves first the selection of the number of partitions required. The series in Table 3.1 can be sorted in ascending order:

11.0, 14.2, 31.0, 32.1, 32.8, 36.2, 38.3, 45.7, 47.2, 49.0, 52.3, 56.3, 56.7, 57.4, 58.1
58.3, 59.6, 63.9, 69.0, 69.1, 70.1, 74.2, 74.3, 78.3, 87.4, 89.1, 90.1, 92.2, 97.4, 98.3

To determine the quartile points the series is divided into four equal parts (provided the number of observations is a multiple of 4):

11.0, 14.2, 31.0, 32.1, 32.8, 36.2, 38.3, 45.7(Q_1)
47.2, 49.0, 52.3, 56.3, 56.7, 57.4, 58.1($Median = Q_2$)
58.3, 59.6, 63.9, 69.0, 69.1, 70.1, 74.2, 74.3(Q_3)
78.3, 87.4, 89.1, 90.1, 92.2, 97.4, 98.3(Q_4).

The median that divides the series into two equal parts has already been shown as lying between 58.1 and 58.3. The median is, by definition, equal to the second quartile (Q_2). The lowest quartile point (Q_1) is positioned at 45.7. Similarly the third quartile point (Q_3) divides the second half of the series into two equal parts and is located at 74.3. The fourth quartile point (Q_4) is 98.3, the last value of the series.

To calculate quintiles the series is divided into five parts of six items each ($30 \div 5 = 6$):

11.0, 14.2, 31.0, 32.1, 32.8, 36.2(QN_1)
38.3, 45.7, 47.2, 49.0, 52.3, 56.3(QN_2)
56.7, 57.4, 58.1, 58.3, 59.6, 63.9(QN_3)
69.0, 69.1, 70.1, 74.2, 74.3, 78.3(QN_4)
87.4, 89.1, 90.1, 92.2, 97.4, 98.3(QN_5).

The five quintile points (QN_1–QN_5) are as indicated above.

Similarly, to estimate deciles, the series is divided into ten parts of three items each ($30 \div 10 = 3$):

11.0, 14.2, 31.0(D_1)
32.1, 32.8, 36.2(D_2)
38.3, 45.7, 47.2(D_3)
49.0, 52.3, 56.3(D_4)
56.7, 57.4, 58.1(D_5)
58.3, 59.6, 63.9(D_6)
69.0, 69.1, 70.1(D_7)
74.2, 74.3, 78.3(D_8)
87.4, 89.1, 90.1(D_9)
92.2, 97.4, 98.3(D_{10}).

The ten decile points (D_1–D_{10}), are noted above.

Quantiles for Grouped Data

The estimation of quantiles in frequency distributions starts by the identification of the interval that contains the quantile group. The procedure is essentially the same as that for the calculation of median described earlier. The first step involves the identification of the particular quantile group. This depends on whether quartiles, quintiles or deciles are being calculated.

Equation (3.10) is the general equation to estimate quantiles. It is the same as Eq. (3.6) to calculate the median.

$$Quantile = L + \left\{ \frac{i}{f} * [(p * n) - C] \right\} \tag{3.10}$$

L is the lower limit, i is the width, f is the frequency of the quantile group, n is the sum of frequencies, p is the value from Table 3.14 depending on which quantile is to be estimated, and C is the cumulative frequency of the group preceding the quantile group.

The value of ($p*n$) determines the quantile group. As noted earlier, this is the first group that has a cumulative frequency equal or larger than the value of ($p*n$).

Income data from Table 3.11 along with the width of each income group are given in Table 3.15.

The first step is to calculate the cumulative frequency, that is, the number of persons reporting a daily income of less than \$30, \$50, \$70, \$90 and \$100 in the fourth column of Table 3.15.

The second step is to identify the quantile group by calculating the value of ($p*n$), multiplying the appropriate value of p (Table 3.14) by n (30).

Once a particular quantile group is identified the values of L, i, f, n and C are substituted in Eq. (3.10) to calculate a particular quantile.

The following examples are concerned with the estimation of the location values for the third quartile (Q_3), the second quintile (QN_2) and the sixth decile (D_6) from Table 3.15.

Table 3.14 Proportion (p) for calculating the location points for various quantiles

Quartiles	p	*Deciles*	p
1st (Q_1)	0.25	1st (D_1)	0.10
2nd (Q_2)	0.50	2nd (D_2)	0.20
3rd (Q_3)	0.75	3rd (D_3)	0.30
4th (Q_4)	1.00	4th (D_4)	0.40
Quintiles	p	5th (D_5)	0.50
1st (QN_1)	0.20	6th (D_6)	0.60
2nd (QN_2)	0.40	7th (D_7)	0.70
3rd (QN_3)	0.60	8th (D_8)	0.80
4th (QN_4)	0.80	9th (D_9)	0.90
5th (QN_5)	1.00	10th (D_{10})	1.00

Table 3.15 Hypothetical data on daily income

Income group ($)	Width of the group (i)	Frequency (f_i)	Cumulative frequency (C_i)
00–29	30	2	2
30–49	20	8	10
50–69	20	10	20
70–89	20	6	26
90–99	10	4	30
Total	...	30	...

Source: Table 3.11

To estimate Q_3, the corresponding value of p is 0.75 (Table 3.14). Multiplying this by n (30) equals 22.5. The fourth column of Table 3.15 indicates that the Q_3 is in the $70–$89 income group, because this is the first group with a cumulative frequency equal or larger than 22.5. Accordingly:

L = 70 (lower limit of the Q_3 group)
i = 20 (width of the Q_3 group)
f = 6 (frequency of the Q_3 group)
n = 30 (total number of observations)
p = 0.75
C = 20 (cumulative frequency for the group preceding the Q_3 group).

Substituting these values in Eq. (3.10):

$$Q_3 = 70 + \left\{\frac{20}{6} * [(0.75 * 30) - 20]\right\} = 78.3$$

Q_1 and Q_2 can be estimated by substituting the values of p as 0.25 and 0.50 respectively, and obtaining the appropriate values of L, i, f, n, p and C and substituting them in Eq. (3.10).

To estimate the second quintile (QN_2), the value of p is 0.40 (Table 3.14) multiplied by n (30) equals 12. From the fourth column of Table 3.15, QN_2 is within the $50–$69 income group. Hence, the *second quintile group* is $50–$69, and the values for various components of Eq. (3.10):

$L = 50$ (lower limit of the QN_2 group)
$i = 20$ (width of the QN_2 group)
$f = 10$ (frequency of the QN_2 group)
$n = 30$ (total number of observations)
$p = 0.40$
$C = 10$ (cumulative frequency for the group preceding the QN_2 group).

Substituting the above values in Eq. (3.10):

$$QN_2 = 50 + \left\{ \frac{20}{10} * [(0.40 * 30) - 10] \right\} = 54.0$$

QN_1, QN_3 and QN_4 can be estimated using the values of p 0.20, 0.60 and 0.80 from Table 3.14 to identify the appropriate quintile group.

To estimate D_6, the value of p from Table 3.14 is 0.60. This value multiplied by n (30) equals 18. From the fourth column of Table 3.15 the D_6 is within the \$50–\$69 income group:

$L = 50$ (lower limit of the D_6 group)
$i = 20$ (width of the D_6 group)
$f = 10$ (frequency of the D_6 group)
$n = 30$ (total number of observations)
$p = 0.60$
$C = 10$ (cumulative frequency for the group preceding the QN_2 group).

Substituting the above values in Eq. (3.10):

$$D_6 = 50 + \left\{ \frac{20}{10} * [(0.60 * 30) - 10] \right\} = 66.0$$

D_1–D_5 and D_7–D_9 can be estimated by substituting the values of p as 0.10, 0.20, 0.30, 0.40, 0.50, 0.70, 0.80 and 0.90, and the identified values of L, i, f, n, p and C in Eq. (3.10).

Table 3.16 exhibits the components needed to determine quartiles, quintiles, and deciles along with their estimated values based on data in Table 3.15. The comparison of the values of the location points for various quantiles of the grouped and ungrouped data indicates that the estimated quantiles are similar.

Table 3.17 provides the interpretation of various quantiles. The first column identifies portions of the distribution (expressed as percentages) that are included in the particular quantile. The second column gives the range of location points that contain the particular portion, and the third column shows the range of values for each quantile using the hypothetical income data. This column is based on the last column of Table 3.16. It shows that in the hypothetical population, 25 % of the persons with the lowest income earned less than or equal to \$43.75 daily. The next 25 % earned from \$43.76 to \$60.00. Half of the 30 people earned more than \$60.00.

Table 3.16 Quartiles, quintiles and deciles of data in Table 3.15

Quantile location	$n * p$	Quantile group	L	i	f	C	Quantile location value
Quartiles							
1st (Q_1)	7.50	30–49	30	20	8	2	43.75
2nd (Q_2)	15.00	50–69	50	20	10	10	60.00
3rd (Q_3)	22.50	70–89	70	20	6	20	78.33
4th ($Q4$)	30.00	90–99	90	10	4	26	100.00
Quintiles							
1st (QN_1)	6.00	30–49	30	20	8	2	40.00
2nd (QN_2)	12.00	50–69	50	20	10	10	54.00
3rd (QN_3)	18.00	50–69	50	20	10	10	66.00
4th (QN_4)	24.00	70–89	70	20	6	20	83.33
5th (QN_5)	30.00	90–99	90	10	4	26	100.00
Deciles							
1st (D_1)	3.00	30–49	30	20	8	2	32.50
2nd (D_2)	6.00	30–49	30	20	8	2	40.00
3rd (D_3)	9.00	30–49	30	20	8	2	47.50
4th (D_4)	12.00	50–69	50	20	10	10	54.00
5th (D_5)	15.00	50–69	50	20	10	10	60.00
6th (D_6)	18.00	50–69	50	20	10	10	66.00
7th (D_7)	21.00	70–89	70	20	6	20	73.33
8th (D_8)	24.00	70–89	70	20	6	20	83.33
9th (D_9)	27.00	90–99	90	10	4	26	92.50
10th (D_{10})	30.00	90–99	90	10	4	26	100.00

The next quarter earned between \$60.01 and \$88.33, and the quarter of the population with the highest income earned more than \$78.33.

Similar interpretations can be given to the quintiles that divide the series into five equal parts and deciles that divide it into ten equal parts.

3.8 Measures of Relative Concentration and Inequality

As stated, quantiles divide populations into equal proportions of various sizes. A question that arises often is the degree of inequality among the quantiles. One of the most commonly used measures of inequality is the Gini Coefficient. This measure has been employed to measure inequalities in income, wealth, health and education and other characteristics. Essentially, it involves numbers at the interval or ratio levels that have been transformed into ordinal data (quantiles).

3.8.1 Gini Coefficient

In a population in which there is absolute equality of wealth each one-fifth of the population has 20 % of the total wealth. This means that 40 % of the population has

Table 3.17 Interpretation of quartiles, quintiles and deciles in Table 3.16

Percentage	Quantile range	Range of values
Quartiles		
Lowest 25 %	$\leq Q_1$	≤\$43.75
Next 25 %	$>Q_1–Q_2$	\$43.76–\$60.00
Next 25 %	$>Q_2–Q_3$	\$60.01–\$78.33
Highest 25 %	$>Q_3–Q_4$	\$78.34–\$100.00
Quintiles		
Lowest 20 %	$\leq QN_1$	≤\$40.00
Next 20 %	$>QN_1–QN_2$	\$40.01–\$54.00
Next 20 %	$>QN_2–QN_3$	\$54.01–\$66.00
Next 20 %	$>QN_3–QN_4$	\$66.01–\$83.33
Highest 20 %	$>QN_4–QN_5$	\$83.34–\$100.00
Deciles		
Lowest 10 %	$\leq D_1$	≤\$32.50
Next 10 %	$>D_1–D_2$	\$32.51–\$40.00
Next 10 %	$>D_2–D_3$	\$40.01–\$47.50
Next 10 %	$>D_3–D_4$	\$47.51–\$54.00
Next 10 %	$>D_4–D_5$	\$54.01–\$60.00
Next 10 %	$>D_5–D_6$	\$60.01–\$66.00
Next 10 %	$>D_6–D_7$	\$66.01–\$73.33
Next 10 %	$>D_7–D_8$	\$73.34–\$83.33
Next 10 %	$>D_8–D_9$	\$83.34–\$92.50
Highest 10 %	$>D_9–D_{10}$	\$92.51–\$100.0

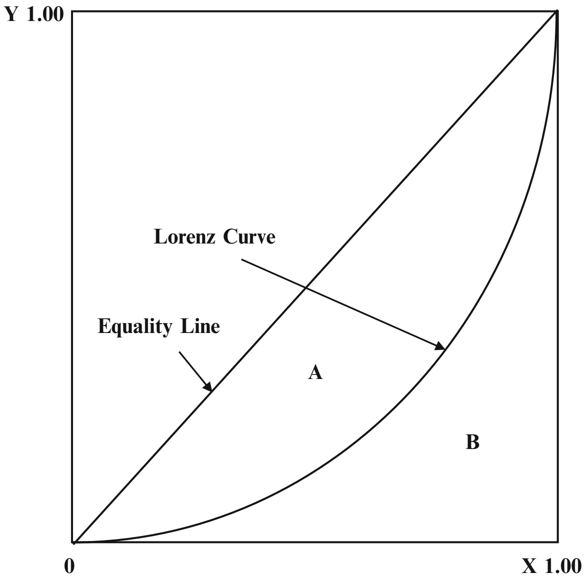
Source: The third column is based on the last column of Table 3.16

40 % of the total wealth, 60 % of the population has 60 % of the total wealth, and so on. When the proportion of population (X) measured on the horizontal axis and the proportion of wealth (Y) measured on the vertical axis are plotted, absolute equality results in a straight line with a 45° angle, referred to as the *Line of Equality* in Fig. 3.6.

However, reality is often different. If it is assumed that there is inequality of wealth such that the one-fifth of the population with lowest wealth has only 2 % of the total, the lowest 40 % have 10 %, the lowest 60 % have 30 %, the lowest 80 % have 60 %, and the remaining 20 % have 40 % of the total wealth, these values when plotted would look like the *Lorenz Curve* in Fig. 3.6.

In the square, in Fig. 3.6, the bottom side represents proportions of variable X, such as population, (adding to 1.00) and the vertical side the proportions of a related variable Y (adding to 1.00). The diagonal line divides the square in two equal triangles. Equality between equal proportions of X and Y in the distribution would follow this 45° line. The Lorenz Curve is the actual proportions of Y for the same proportions of X (Fig. 3.6). The triangle formed by the diagonal on the right side of the square consists of two parts A and B, where A comprises of the area between the

Fig. 3.6 Lorenz curve



Lorenz Curve and the Line of Equality and B is the remaining area of the triangle. The Gini Coefficient (G) is defined as:

$$G = \frac{A}{A + B}$$

The values of G lie between 0 and 1. The higher the coefficient the greater the inequality between the distributions of the variables X and Y. The Gini Coefficient is often expressed as a percentage.

An approximation of the coefficient can be estimated using the following equation:

$$G = 1 - \sum_{i=1}^{i=n} (\frown{Y}_{i+1} + \frown{Y}_i) * (\frown{X}_{i+1} - \frown{X}_i) \tag{3.11}$$

$\frown{Y}_i$ is the cumulative proportion of variable Y, $\frown{X}_i$ is the cumulative proportion of the related variable X and n is the number of proportions. For instance, $n = 5$ in the case of quintiles or $n = 10$ when dealing with deciles.

Algebraically, when X_i refers to quintiles, deciles or any other quantile, the term $(\frown{X}_{i+1} - \frown{X}_i)$ is equivalent to X_i and therefore Eq. (3.11) can be simplified as:

$$G = 1 - \left[\sum_{i=1}^{i=n} (\frown{Y}_{i+1} + \frown{Y}_i) * X_i \right] \tag{3.12}$$

Table 3.18 Calculation of the Gini coefficient for the 2007 Expenditure and Food Survey of United Kingdom

Quintiles	Households[a]	Weekly income	Income (£)[a]	Proportion X_i	Proportion Y_i	$\widehat{Y}_i$	$\widehat{Y}_{i+1}+\widehat{Y}_i$	Product
(1)	(2)	(3)	(4) = (2)*(3)	(5)	(6)	(7)	(8)	(9) = (8)*(5)
QN_1	4,950	138	683,100	0.2	0.053	0.053	0.053	0.0106
QN_2	4,960	270	1,339,200	0.2	0.104	0.157	0.210	0.0420
QN_3	4,960	427	2,117,920	0.2	0.164	0.321	0.478	0.0956
QN_4	4,960	624	3,095,040	0.2	0.240	0.560	0.881	0.1762
QN_5	4,950	1,147	5,677,650	0.2	0.440	1.000	1.560	0.3120
Total	...	...	12,912,910	1.0	1.000	...	3.182	0.6364

Source: United Kingdom (2008), Table 3.12
[a]Columns (2) and (4) are in 000s

The Food and Expenditure Survey of the United Kingdom for 2007 provides data to estimate the Gini Coefficient of the United Kingdom household weekly income, at that time. Columns (2) and (3) were taken from the survey report (United Kingdom 2008). In Table 3.18, column (2) gives the weighted estimates of the number of households in each quintile and column (3) the average weekly disposable income per household in that quintile. Total estimated income for all households in each quintile is stated in column (4). Columns (5) and (6) show the proportion of households (X_i) given in column (2), and of the estimated income (Y_i) in column (4), and column (7) is the cumulative distribution of Y_i denoted as $\widehat{Y}_i$, while columns (8) and (9) follow the procedure within the square brackets in Eq. (3.12), and the value of G is estimated as:

$$G = 1 - 0.6364 = 0.3636 \text{ or } 36.36\,\%.$$

An alternative method that gives the same answer is:

$$G = 1 - \left(2 * \left[\frac{\sum_{i=1}^{i=n} (\widehat{Y}_{i+1} + \widehat{Y}_i)}{2 * n}\right]\right) \tag{3.13}$$

Substituting the value of $\sum_{i=1}^{i=n} (\widehat{Y}_{i+1} + \widehat{Y}_i)$ as 3.182 and $n = 5$ in Eq. (3.13):

$$G = 1 - \left(2 * \left[\frac{0.3182}{2 * 5}\right]\right) = 0.3636 \text{ or } 36.36\,\%.$$

3.8.2 Quantile Concentration

The Gini Coefficient gives a measure of inequality for all the quintiles. Other inequality measures are also used. Among them, ratios of high and low quintiles are often used to measure inequality, say, between the highest and lowest quintiles.

$$\text{Inequality ratio between } QN_1 \text{ and } QN_5 = \frac{QN_5}{QN_1}$$

In Table 3.18, the average weekly disposable income was £ 138 for the lowest 20 % (QN_1) and that for the highest 20 % (QN_5) was £1,147. This indicates that the inequality ratio between Quintile 1 and 5 was:

$$\frac{QN_5}{QN_1} = \frac{1,147}{138} = 8.3$$

Hence, households in the quintile with the highest income had a gross weekly income 8.3 times higher than households in the quintile with the lowest income.

3.8.3 Indexes of Dissimilarity and Relative Difference

The *index of dissimilarity* (*ID*) compares the percentage distributions of two populations, or the same population at two points in time, to measures the extent to which the two populations are different. Conceptually, it is related to the Gini Coefficient. As such, essentially, it also measures interval and ratio level data that have been transformed into ordinal data. It is calculated as:

$$ID = \frac{\sum_{i=1}^{i=n} |A_i - B_i|}{2} \tag{3.14}$$

A_i is the percentage of population A with characteristic i, and B_i is the corresponding percentage of population B. The number of categories for the particular characteristic is n and the summation of the difference $(A_i - B_i)$ is taken irrespective whether the difference is positive or negative.

This index varies from 0, when both populations are identical, to 1, when both populations are completely dissimilar. The characteristic i may be any population characteristic such as age, education, occupation or location.

The index of dissimilarity and variants of it are found in many forms, including so-called *segregation measures* (James & Taeuber 1985; Massey & Denton 1988).

The *index of relative difference* (*IRD*) is estimated by the equation:

$$IRD = \frac{\sum_{i=1}^{i=n} \left| \left(\frac{A_i}{B_i} * 100 \right) - 100 \right|}{2 * n} \tag{3.15}$$

A_i, B_i and n are defined in the same way as in Eq. (3.14). Unlike *ID* that measures the absolute differences, the *IRD* measures the relative difference, where population in the denominator B_i is taken as the *standard*. Both indexes can be calculated for a number of populations by substituting in turn the A_i values with corresponding values for other populations.

Table 3.19 illustrates the calculation of *ID* and *IRD* for the age distributions of three countries in 2009: Japan, Spain and Egypt (United Nations 2011). Japan's age distribution is the *standard*. Both indexes for Egypt were substantially higher than those for Spain, thereby indicating that the age distribution of Spain was closer to that of Japan.

3.9 Correlation and Regression

3.9.1 Association and Correlation

Correlation and regression are designed to examine the association and relationship of two (or more) variables. Earlier, in Table 3.4, the relationship between a nominal level variable (sex) and an ordinal level variable (income in ranked groups) was examined. In this example, the age of brides and bridegrooms in 50 hypothetical marriages is given in Table 3.20. Columns (1) and (4) give the marriage number (i). The related age of the bride (y_i) is in columns (2) and (5) and that of the bridegroom (x_i) in columns (3) and (6). These data are measured at the ratio level, so both correlation and regression are appropriate analytical tools to use. Figure 3.7 shows the same data in the form of a scatter diagram.

An association is apparent in Fig. 3.7 between the ages of brides and bridegrooms: younger people tend to marry younger partners. The age of the bride (y_i) goes from its smallest value of 18 years on the horizontal axis to its highest value of 27 years along the same axis; the values of the age of the bridegroom (x_i) also becomes larger from its lowest value of 20 on the vertical axis to its highest value of 43 years of age. As the values of (x_i) get older the values of (y_i) also tend to get older: they vary in a *positive* manner. However, this *positive linear* relationship between (x_i) and (y_i) is not perfect: as the values of (x_i) become larger it is not necessarily the case that the associated values of (y_i) also get larger. If the linear relationship between (x_i) and (y_i) was *perfect* all the plotted points

Table 3.19 Calculation of the indexes of dissimilarity and relative difference for Spain and Egypt, using Japan as the *standard*

	Japan (B_i)	Spain (A_i)				Egypt (C_i)			
Age (years)	% population	% population	$\lvert A_i - B_i \rvert$	$\frac{A_i}{B_i} * 100$	$\left[\frac{A_i}{B_i} * 100\right] - 100$	% population	$\lvert C_i - B_i \rvert$	$\frac{C_i}{B_i} * 100$	$\left[\frac{C_i}{B_i} * 100\right] - 100$
(1)	(2)	(3)	(4) = (3) − (2)	(5) = (3)/(2)*100	(6) = (5) − 100	(7)	(8) = (7) − (2)	(9) = (7)/(2)*100	(10) = (9) − 100
0–14	13.4	14.8	1.5	110.4	10.4	31.7	18.3	236.6	136.6
15–24	10.3	10.7	0.5	103.9	3.9	22.5	12.3	218.4	118.4
25–29	5.9	7.6	1.6	128.8	28.8	8.8	2.9	149.1	49.1
30–34	6.8	8.8	2.0	129.4	29.4	6.5	0.3	95.6	−4.4
35–39	7.6	8.6	1.0	113.2	13.2	6.4	1.2	84.2	−15.8
40–44	6.7	8.1	1.4	106.6	6.6	5.6	1.1	83.6	−16.4
45–49	6.1	7.4	1.3	121.3	21.3	5.0	1.1	82.0	−18.0
50–54	6.1	6.5	0.4	106.6	6.6	4.2	1.9	68.9	−31.1
55–59	7.3	5.6	1.7	76.7	−23.3	3.1	4.2	42.5	−57.5
60+	29.9	21.9	7.9	73.2	−26.8	6.1	23.8	20.4	−79.6
Total (absolute value)	…	…	19.3	…	170.3	…	67.1	…	526.9
Index of dissimilarity = Total÷2=			*9.65*	…	…	…	*33.5*	…	…
Index of relative difference = Total÷(2*10)=			…	…	*8.5*	…	…	…	*26.3*

Source: United Nations (2011), Table 7a

Table 3.20 Hypothetical data of ages of brides and bridegrooms

	Age			Age	
No. (i)	Bride (y_i)	Bridegroom (x_i)	No. (i)	Bride (y_i)	Bridegroom (x_i)
(1)	(2)	(3)	(4)	(5)	(6)
1	18	20	26	24	29
2	19	20	27	24	30
3	20	22	28	24	31
4	21	22	29	24	27
5	22	24	30	24	29
6	23	25	31	24	30
7	24	26	32	24	31
8	25	29	33	25	27
9	26	31	34	25	29
10	27	26	35	25	30
11	28	35	36	25	31
12	29	37	37	25	32
13	30	39	38	25	33
14	22	34	39	25	35
15	22	24	40	25	31
16	22	25	41	25	32
17	22	26	42	25	27
18	22	34	43	25	29
19	23	20	44	25	30
20	23	27	45	25	31
21	23	29	46	25	32
22	23	20	47	25	33
23	23	27	48	25	35
24	23	29	49	38	43
25	24	27	50	27	31
			$\sum_{i=1}^{i=50} =$	1,217	1,456

Fig. 3.7 Scatter diagram of the ages of bridegrooms and brides (*Source*: Table 3.20)

Table 3.21 Further calculations based on data in Table 3.20

No. (i) (1)	y_i^2 (2)	x_i^2 (3)	$x_i * y_i$ (4)	No. (i) (5)	y_i^2 (6)	x_i^2 (7)	$x_i * y_i$ (8)
1	324	400	360	26	576	841	696
2	361	400	380	27	576	900	720
3	400	484	440	28	576	961	744
4	441	484	462	29	576	729	648
5	484	576	528	30	576	841	696
6	529	625	575	31	576	900	720
7	576	676	624	32	576	961	744
8	625	841	725	33	625	729	675
9	676	961	806	34	625	841	725
10	729	676	702	35	625	900	750
11	784	1,225	980	36	625	961	775
12	841	1,369	1,073	37	625	1,024	800
13	900	1,521	1,170	38	625	1,089	825
14	484	1,156	748	39	625	1,225	875
15	484	576	528	40	625	961	775
16	484	625	550	41	625	1,024	800
17	484	676	572	42	625	729	675
18	484	1,156	748	43	625	841	725
19	529	400	460	44	625	900	750
20	529	729	621	45	625	961	775
21	529	841	667	46	625	1,024	800
22	529	400	460	47	625	1,089	825
23	529	729	621	48	625	1,225	875
24	529	841	667	49	1,444	1,849	1,634
25	576	729	648	50	729	961	837
				$\sum_{i=1}^{i=50} =$	30,045	43,562	35,979

Source: Table 3.20

would be on a line. In statistical terms, this association is called *correlation*. An indicator of correlation is the *correlation coefficient* (r) calculated as:

$$r = \frac{n * \left(\sum_{i=1}^{i=n} x_i * y_i\right) - \left[\left(\sum_{i=1}^{i=n} x_i\right) * \left(\sum_{i=1}^{i=n} y_i\right)\right]}{\sqrt{n * \sum_{i=1}^{i=n} x_i^2 - \left(\sum_{i=1}^{i=n} x_i\right)^2} * \sqrt{n * \sum_{i=1}^{i=n} y_i^2 - \left(\sum_{i=1}^{i=n} y_i\right)^2}} \tag{3.16}$$

where (x_i) and (y_i) are the n paired observations pertaining to variables for which the correlation coefficient is calculated. The variables in Eq. (3.16) must be quantitative and may be either categorical (such as the number of children) or continuous (such as the age of a person that increases at every point in time). Data presented in Tables 3.20 and 3.21 refer to age in completed years at the time of marriage and may be considered as data at the ratio level.

There are six components in the above equation; three of these can be derived from Table 3.20 directly. They are: $n = 50$, $\sum_{i=1}^{i=n} x_i = 1,456$ and $\sum_{i=1}^{i=n} y_i = 1,217$. The three remaining components are estimated by summing x_i^2 in columns (3) and (7), y_i^2 in columns (2) and (6), $x_i * y_i$ in columns (4) and (8) in Table 3.21. Their values are 43,562 for $\sum_{i=1}^{i=n} x_i^2$, 30,045 for $\sum_{i=1}^{i=n} y_i^2$ and 35,979 for $\sum_{i=1}^{i=n} (x_i * y_i)$.

Substituting these values in Eq. (3.16):

$$\frac{(50 * 35,979) - (1,217 * 1,456)}{\sqrt{50 * 30,045 - (1,217)^2} * \sqrt{50 * 43,562 - (1,456)^2}} = 0.7695$$

A correlation coefficient varies between 0 and ± 1, where 0 means no correlation and ± 1 indicates perfect positive or negative correlation. In a positive correlation an increase in one variable is associated with an increase in the other, while in a negative correlation an increase in one corresponds to a decrease in the other variable. In the above example the correlation was positive.

Another coefficient related to r is the *coefficient of determination (R)* as:

$$R = r^2 \tag{3.17}$$

This represents the proportion of variance in one variable that could be explained by the other variable. In this example, the coefficient of determination is (0.7695^2=) 0.5922. This means that 59 % of the variance in the age distribution of brides can be explained by the age distribution of bridegrooms. This *explanation* is in terms of the mean age of the bridegrooms. Knowing the age of the bridegroom in conjunction with this regression model, the average error of the prediction of the age of the bride using the mean age of the bride can be reduced by 59 %. The coefficient of determination measures the improvement (or lack thereof) a regression model provides relative to using the mean of the variable to be predicted. It should be kept in mind that if the mean of the variable to be predicted is highly accurate (the standard deviation is very low), not much might be gained by the introduction of a new variable in the prediction process: the coefficient of determination might be very low. Conversely, the standard deviation of a variable whose values are to be predicted might be extremely high. In that case, the introduction of another variable to predict the value of the variable via regression analysis may result in a low coefficient of variation. However, such a regression may still offer a great deal of improvement over the use of the mean of that variable as a predictor. It may also be the case that when the variation of the variable to be predicted is considerably high that a regression model ends up with a high coefficient of determination; but it is still not a good predictor of the variable in question. In such cases, a transformation of the data may be needed (Swanson 2004).

It is advisable to be keenly aware of means and standard deviations of variables when considering and evaluating regression models. One way of assessing the relationship between the mean and its standard variation is the estimation of the *coefficient of variation (CV)* as:

$$CV = \frac{SD}{\bar{x}} \tag{3.18}$$

In the above equation the numerator is the standard deviation and the denominator is the arithmetic mean. The coefficient of variation can be used for purposes of comparison since it provides a measure of variation relative to the mean. It is also important to graph relationships (Anscombe 1973) and to be aware of the assumptions underlying regression, the consequences of violating them, and how to overcome these violations (Fox 1991).

Engel's Law provides a good example of negative correlation. It states that low income households spend a greater proportion of their income on food than the richer households. This correlation between household income and expenditure on food has been tested empirically.

Each correlation coefficient is associated with two *regression coefficients*, one measure the impact of variable of (x_i) on variable (y_i) and the other determines the impact of (y_i) on (x_i). These may be referred to as b_1 and b_2 respectively, and are estimated using Eqs. (3.19) and (3.20):

$$b_1 = \frac{n * \sum_{i=1}^{i=n} (x_i * y_i) - \left[\left(\sum_{i=1}^{i=n} x_i \right) * \left(\sum_{i=1}^{i=n} y_i \right) \right]}{n * \sum_{i=1}^{i=n} y_i^2 - \left(\sum_{i=1}^{i=n} y_i \right)^2} \tag{3.19}$$

$$b_2 = \frac{n * \sum_{i=1}^{i=n} (x_i * y_i) - \left[\left(\sum_{i=1}^{i=n} x_i \right) * \left(\sum_{i=1}^{i=n} y_i \right) \right]}{n * \sum_{i=1}^{i=n} x_i^2 - \left(\sum_{i=1}^{i=n} x_i \right)^2} \tag{3.20}$$

From a comparison of Eq. (3.16) with Eqs. (3.19) and (3.20), it is obvious that r is the *GM* of b_1 and b_2, that is, $\left(r = \sqrt{b_1 * b_2}\right)$.

The two regression coefficients were estimated from data in Tables 3.19 and 3.20, and were 0.46417 and 1.27584. Interpretation of the b_1 is that one unit increase in (x_i) will result in 0.46417 unit increase in (y_i), while the second regression coefficient indicates that one unit increase in (y_i) will only result in 1.27584 unit increase in (x_i).

The correlation coefficient gives an indication of association but not of cause and effect between the two variables. Regression coefficients imply that there is an independent variable (cause) and a dependent variable (effect): b_1 measures the impact of variable x_i (age of bridegroom), the independent variable, on variable y_i and in b_2 it is the variable y_i (age of the bride) that is deemed to be the independent variable.

3.9.2 Linear Regression and Multiple Regression Models

Using data in Table 3.20, it is possible to fit a *linear regression model* of the type

$$\hat{y}_i = a_1 + b_1 * x_i$$

where $\hat{y}_i$ is the predicted value of the dependent variable y_i corresponding to each value of the independent variable x_i, and a_1 and b_1 are the parameters of the linear model estimated from a given data. These parameters are also known as the *intercept* (*i.e.,* the value of $\hat{y}_i$ when x_i is zero) and *slope* of the linear model (see Box 3.2). The slope is in fact identical to the regression coefficient as indicated in Eqs. (3.19)

Box 3.2 Line Coordinates, Slope, and Intercept

Linear regression models in their basic form reflect a simple equation: $Y = \alpha + \beta{*}X$. In this equation a relationship is assumed that the value of a dependent variable Y is determined by the value of an independent variable X multiplied by a factor β plus a constant value α. The value of constant α may be zero, positive or negative. If the value of α is zero then the value of Y is entirely determined by the value of X multiplied by the value of β. In the above equation β is usually known as the slope of the line and α as the intercept.

In geometric terms, this equation can be envisaged as a line with coordinates X and Y. A line can be determined by two points A and B with coordinates (x_1, y_1) and (x_2, y_2).

The gradient or slope of the line represents the relationship between X, the independent variable, and Y the dependent valuable. This slope β is the ratio of

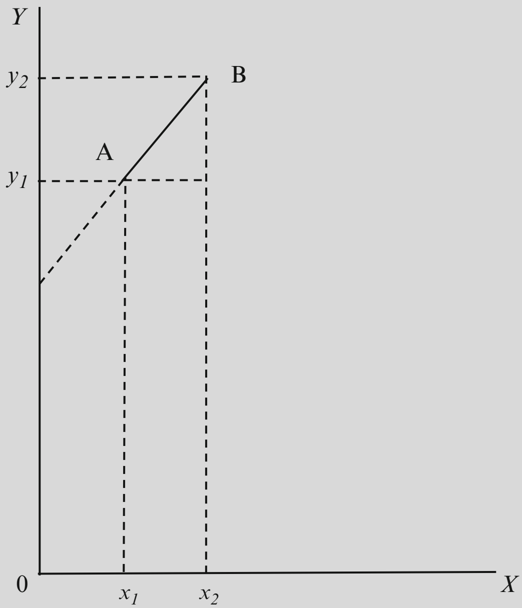
$$\beta = \frac{y_2 - y_1}{x_2 - x_1}$$

This holds as long as the value of x_2 is different from x_1 ($x_2 \neq x_1$) and y_2 is different from y_1 ($y_2 \neq y_1$). It is apparent that if x_2 has the same value as x_1 then the value of β is undetermined as ($y_2 - y_1$) divided by 0 is not defined. Similarly, if y_2 is equal to y_1 then the numerator is 0 and the value of β is also not determined. This means that the slope β can be determined as long as a line is not parallel to either the X or the Y axis. The slope β is what is known as the regression coefficient (Eqs. 3.19 and 3.20).

If the line is such that it passes through the origin where X and Y equal zero then β alone determines any value of Y given a value of X. However, if Y and X assume different values at the origin, and unless the line is parallel to Y, the line is bound to intercept the Y axis and this intercept is needed to estimate the value of Y given a value of X. The graph in this box can be used as an illustration.

(continued)

Box 3.2 (continued)



Point A is defined by (x_1, y_1) and B by (x_2, y_2). If the values for A and B in the graph are (4, 19) and (8, 24) respectively the slope $\beta = \frac{24-19}{8-4} = 1.25$

Substituting the values of x_1, y_1 for point A and β in equation $Y = \alpha + \beta * X$:

$$24 = \alpha + (1.25 * 8) = \alpha + 10 \text{ and thus } \alpha = 24 - 10 = 14.$$

The intercept is constant at 14 as it is seen when substituting the values of x_2, y_2 for point B and β in equation $Y = \alpha + \beta * X$:

$$19 = \alpha + (1.25 * 4) = \alpha + 5 \text{ and thus } \alpha = 19 - 5 = 14.$$

In this example, the value of the slope β is positive. However, the value of β can be negative if the line was defined differently in terms of *X* and *Y*, where for a given value of *X* a lower value of *Y* is obtained. The value of the intercept α can also be negative and the line would intercept the *Y* axis at a point below zero.

and (3.20). The intercepts (a_1 and a_2) for linear models corresponding with the regression coefficients b_1 and b_2 are:

$$a_1 = \frac{\sum_{i=1}^{i=n} y_i}{n} - b_1 * \frac{\sum_{i=1}^{i=n} x_i}{n} \tag{3.21}$$

$$a_2 = \frac{\sum_{i=1}^{i=n} x_i}{n} - b_2 * \frac{\sum_{i=1}^{i=n} y_i}{n} \tag{3.22}$$

The values of the two intercepts were calculated as 10.82336 and −1.93389 respectively, and the two linear regression models as:

$$\hat{y}_i = 10.82336 + 0.4617 * x_i \tag{3.23}$$

$$\hat{x}_i = -1.93389 + 1.275838 * y_i \tag{3.24}$$

The coefficient of determination (R^2) for each of the two models was 0.59221.

One of the advantages of fitting regression models is that they enable the estimation of predicted values ($\hat{y}_i$) of the dependent variable. Figure 3.8 shows the predicted values for the age of bride using the age of bridegroom as the independent variable.

A linear regression model can have more than one independent variable. *Multiple regression models* with two or more independent variables take the general form

$$\hat{y}_i = a + b_1 * x'_i + b_2 * x''_i + b_3 * x'''_i + \ldots$$

where a is the intercept (the value of $\hat{y}$ when each of the independent variables is zero) and b_1, b_2, b_3, … are the *multiple regression coefficients* for each of the independent variables (x'_i, x''_i, x''' …). In a bivariate regression model when there is only one independent variable, the regression coefficient is concerned only with the relationship between that variable and the dependent one. However, in a multiple regression model where there are more than one independent variables, the model cannot just assess the relationship between each of the independent variables and dependent variable. It needs to take cognisance of the other independent variables. The regression coefficients for each independent variable are estimated while controlling the explanatory power of the other independent variables. Accordingly, these coefficients are known as *partial regression coefficients*. In multiple regression one of the issues is the possibility that in addition to the relationship between the independent and the dependent variables, there is a also a relationship between two or more of the independent variables. This is called

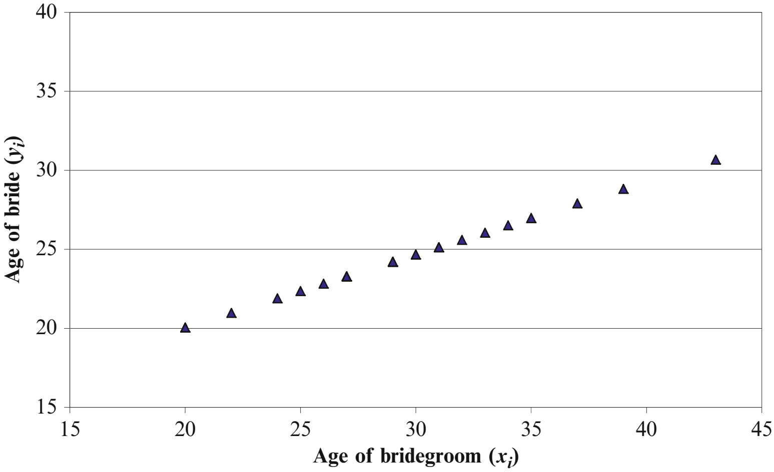


Fig. 3.8 Scatter diagram of the ages of bridegrooms with the predicted ages of brides using a linear model: $\hat{y}_i = 10.82336 + 0.4617 * x_i$

multicollinearity. In some cases, even though there is a good relationship between an independent and the dependent variable, the addition of that independent variable to the model may not add to the explanatory value of the model, because of this factor.

The *coefficient of multiple determination (R^2)* like its counterpart in the bivariate linear regression model measures the variation in the dependent variable explained by the predictive power of the independent variables:

$$R^2 = \frac{\sum_{i=1}^{i=n} (y_i - \bar{y})^2 - \sum_{i=1}^{i=n} (y_i - \hat{y})^2}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2}$$

The numerator is the *regression sum of squares*, it takes into consideration the variation not explained by the model, that is the square of the difference between the values of the dependent variable (y_i) and the predicted values by the model ($\hat{y}$) these differences are known as the residuals.

To illustrate the use of multiple regression model an additional independent variable is provided in Table 3.22 to be added to the data in Table 3.20. This additional independent variable is the age when the bride's mother herself got married.

The computations involved in the estimation of multiple regression models are complicated and can be facilitated by the use of a software package such as the Excel which has a subroutine for multiple regressions (Triola 2007).

Table 3.22 Hypothetical data on age of the mother of bride at the time of her own marriage

No. (i)	Age of mother	No. (i)	Age of mother	No. (i)	Age of mother	No. (i)	Age of mother	No. (i)	Age of mother
1	16	11	23	21	23	31	18	41	21
2	20	12	21	22	22	32	20	42	22
3	18	13	24	23	20	33	22	43	23
4	20	14	18	24	18	34	23	44	35
5	19	15	23	25	21	35	19	45	20
6	21	16	24	26	20	36	20	46	21
7	21	17	20	27	18	37	21	47	20
8	20	18	19	28	24	38	21	48	24
9	24	19	21	29	23	39	20	49	30
10	25	20	20	30	20	40	19	50	25

The multiple regression model to predict the age of bride was estimated as:

$$\hat{y}_i = 5.31050 + 0.40151 * {x'}_i + 0.34288 * x''_i \tag{3.25}$$

The age of the bridegroom was represented by ${x'}_i$ and age of the mother of bride at the time of her own marriage by x'' and as in the bivariate model the dependent variable is the age of the bride $y_{i.}$ The value of R^2 for this model is 0.71253; it was just over 20 % higher than the model using only one independent variable.

References

Anscombe, E. (1973). Graphs in statistical analysis. *The American Statistician, 27*, 17–22.

Australia. (2010). *2009 births Australia*. Catalogue no. 3301.0. Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/10BEDC49AFCACC1F-CA2577CF000DF7AB/$File/33010_2009.pdf. Accessed Feb 2013.

Fox, J. (1991). *Regression diagnostics* (Quantitative applications in the social sciences, Vol. 79). Newbury Park: Sage Publications.

James, R. D., & Taeuber, K. (1985). Measures of segregation. *Sociological Methodology, 15*, 1–32.

Japan. (2011). *e-Stat. Portal site of official statistics of Japan. 2000 population census, labour force status of population (Table 1)*. www.e-stat.go.jp/SG1/estat/ListE.do?bid=000000030097&cycode=0. Accessed Feb 2013.

Massey, D. S., & Denton, N. (1988). The dimensions of racial segregation. *Social Forces, 67*(2), 281–315.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677–680.

Swanson, D. A. (2004). Advancing methodological knowledge within state and local demography: A case study. *Population Research and Policy Review, 23*, 379–398.

Triola, M. F. (2007). *Elementary statistics using excel* (3rd ed.). Boston: Pearson Addison Wesley.

United Kingdom. (2008). *Family spending, 2007 edition*. Newport: Palgrave Macmillan. http://www.ons.gov.uk/ons/rel/family-spending/family-spending/2007-edition/index.html. Accessed Feb 2013.

United Nations. (2009). *Demographic yearbook, 2007*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dybsets/2007%20DYB.pdf. Accessed Feb 2013.
United Nations. (2011). *Demographic yearbook, 2009–2010*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dybsets/2009-2010.pdf. Accessed Feb 2013.

Chapter 4
Elements of Demographic Analysis

4.1 Purpose

This chapter discusses some of the basic concepts and methods used in demographic analysis. This follows the description of the field of demography and the nature of related data (Chaps. 1 and 2). In this chapter some of the five generic aspects of demography (a) population size, (b) geographic distribution, (c) composition, (d) components of change (births, deaths and migration), and (e) determinants of change are discussed. These include the concepts of the balancing equation, demographic stocks and flows, various types of rates and ratios, standardisation and graphical presentation of demographic data. Statistics from a variety of countries are used for illustrative purposes.

4.2 Balancing Equation

One of the most important conceptual models in Demography is that of a *balancing equation*. This model may take the following form:

$$P_{t+n} = P_t + B_{t\to t+n} - D_{t\to t+n} + I_{t\to t+n} - E_{t\to t+n} \tag{4.1}$$

P_t and P_{t+n} represent the size of a population at two points in time t and $t + n$, $B_{t\to t+n}$ is the number of births, $D_{t\to t+n}$ the number of deaths, $I_{t\to t+n}$ the number of immigrants and $E_{t\to t+n}$ the number of emigrants during the time period t to $t + n$. Usually, the value of n is 1 year, but it can be more or less in duration.

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_4,

Equation (4.1) can be re-written as:

$$(P_{t+n} - P_t) = (B_{t\to t+n} - D_{t\to t+n}) + (I_{t\to t+n} - E_{t\to t+n}) \tag{4.2}$$

$(P_{t+n} - P_t)$ is the *population growth* (*PG*), $(B_{t\to t+n} - D_{t\to t+n})$ is the *natural increase* (*NI*), and $(I_{t\to t+n} - E_{t\to t+n})$ is the *net international migration* (*NIM*). Accordingly, (*PG*) can be defined as:

$$PG = NI + NIM \tag{4.3}$$

The equalities in Eqs. (4.2) and (4.3) can be applied drawing on United States population figures for 2004–2005 (United States 2007):

$$\begin{array}{lll}
\text{Estimated population mid-2004} & = 293,657,000 & = P_{\text{mid}-2004} \\
\text{Estimated population, mid-2005} & = 296,410,000 & = P_{\text{mid}-2005} \\
\text{Births, mid-2004 to mid-2005} & = 4,129,000 & = B_{\text{mid}-2004\to\text{mid}-2005} \\
\text{Deaths, mid-2004 to mid-2005} & = 2,425,000 & = \text{D}_{\text{mid}-2004\to\text{mid}-2005} \\
\text{Immigrants, mid-2004 to mid-2005} & = 1,122,000 & = I_{\text{mid}-2004\to\text{mid}-2005} \\
\text{Emigrants, mid-2004 to mid-2005} & = 73,000 & = E_{\text{mid}-2004\to\text{mid}-2005}
\end{array}$$

As a result:

$$\begin{array}{lll}
PG_{\text{mid}-2004\to\text{mid}-2005} & = 296,410,000 - 293,657,000 & = 2,753,000 \\
NI_{\text{mid}-2004\to\text{mid}-2005} & = 4,129,000 - 2,425,000 & = 1,704,000 \\
NIM_{\text{mid}-2004\to\text{mid}-2005} & = 1,122,000 - 73,000 & = 1,049,000.
\end{array}$$

In a given population, if the number of births exceeds the number of deaths, $NI_{t\to t+n}$ is positive, as in the example given above, and negative when the number of deaths is larger than the number of births. Apart from a few countries in Europe, the general trend is for positive natural increase. Similarly, $NIM_{t\to t+n}$ is positive, zero, or negative, depending on the relative number of immigrants and emigrants. In countries such as Australia and Canada, $NIM_{t\to t+n}$ is positive. In some others, such as China and Japan, it is negligible as a proportion of the total population.

At a sub-national level – such as a province, a city, or some other segment of a country's population – two more terms are added to Eq. (4.1):

$$P_{t+n} = P_t + B_{t\to t+n} - D_{t\to t+n} + I_{t\to t+n} - E_{t\to t+n} + \hat{I}_{t\to t+n} - \hat{E}_{t\to t+n} \tag{4.4}$$

$\hat{I}_{t\to t+n}$ and $\hat{E}_{t\to t+n}$ are the number of immigrants and emigrants from and to other parts of the same country. The difference $(\hat{I}_{t\to t+n} - \hat{E}_{t\to t+n})$ is referred to as the *net domestic migration* (*NDM*). At the sub-national level equation, the counterpart of Eq. (4.3) is:

$$PG = NI + NIM + NDM \tag{4.5}$$

Table 4.1 Components of population growth: Australia, mid-2010 to mid-2011

State/Territory	Natural increase	Net migration: International	Net migration: Domestic	Population growth
(1)	(2)	(3)	(4)	(5) = (2) + (3) + (4)
New South Wales	44,441	49,607	−12,202	81,846
Victoria	32,585	44,347	3,381	80,313
Queensland	35,065	30,486	7,391	72,942
South Australia	6,724	9,364	−2,653	13,435
Western Australia	18,518	29,314	4,996	52,828
Tasmania	2,001	1,044	174	3,219
Aust. Capital Territory	2,629	699	−2,393	935
Northern Territory	3,228	1,940	1,306	6,474
Total (Australia)	145,191	166,801	0	311,992

Source: Australia (2012)

Note: The data given in this table are the sum of figures for four quarters of the year

NDM is zero at the national level because movements of population within a country cancel each other.

Urban areas, and particularly large cities, attract people from rural areas. This movement of people from rural to urban area is known as *urbanisation*.

The births component of natural increase is discussed in more detail in the chapter on fertility (Chap. 5), and the deaths component in the chapter on mortality (Chap. 6). International and domestic migration are discussed in Chap. 8.

Table 4.1 provides an illustration of the components of population growth for various States and Territories in Australia during the period mid-2010 to mid-2011 (Australia 2012). The table shows that net international migration was the biggest component of population growth in all States and Territories. Three of them (New South Wales, South Australia and the Australian Capital Territory) experienced a net loss of population through domestic migration and the others gained. Of course, the sum of the net domestic migration was zero.

4.3 Population Growth Rates

4.3.1 Annual Growth Rate

The rate of natural increase gives an approximation of the rate of population growth over a given period, if international migration (and domestic migration at sub-national level) is relatively small. A more precise measure is the net increment in the population from a given date to another. Generally, the rate of population growth is estimated on an annual basis:

$$GR_{t \to t+1} = \frac{(P_{t+1} - P_t)}{P_t} \tag{4.6}$$

$GR_{t \to t+1}$ is the annual growth rate during the period t to $t+1$, and P_t and P_{t+1} stand for the size of population at time t and $t + 1$.

An illustration is provided by United States data in Sect. 4.2:

$$\begin{aligned} GR_{mid-2004 \to \text{mid}-2005} &= \frac{(296,410,000 - 293,657,000)}{293,657,000} \\ &= \frac{2,753,000}{293,657,000} \\ &= 0.009375 \text{ or } 0.9375\,\% \end{aligned}$$

In the estimation of population growth rates, it must be ensured that the populations at both points in time are from a geographical area with the same boundaries.

4.3.2 *Exponential Population Growth Rate*

Most population enumerations, particularly the censuses, are more than 1 year apart. Equation (4.7) enables the estimation of the average annual population growth rate from such data.

$$P_{t+n} = P_t * e^{r*n} \tag{4.7}$$

P_{t+n} and P_t refer to the population size at time $t + n$ and time t, e is the mathematical constant (= 2.7182828) r is the annual *exponential population growth rate,* and n is the *number of years* between t and $t + n$.

Equation (4.7) can be written as:

$$\frac{P_{t+n}}{P_t} = e^{r*n} \tag{4.8}$$

Taking the natural logarithm (*LN*) of both sides of the equation and using rule (3) in Box 3.1, it becomes:

$$LN\left(\frac{P_{t+n}}{P_t}\right) = r * n * LN(e)$$

Since $LN(e) = 1$ according to rule (4) in Box 3.1:

$$LN\left(\frac{P_{t+n}}{P_t}\right) = r * n$$

$$r = \frac{LN\left(\frac{P_{t+n}}{P_t}\right)}{n} \tag{4.9}$$

The estimated population of India in mid-2005 and mid-2011 (India undated) offer a basis for an example of the calculation of r:

$$\begin{array}{lll} \text{Estimated population, mid-2005} & = 1,101,318,000 & = P_{\text{mid}-2005} \\ \text{Estimated population, mid-2011} & = 1,197,813,000 & = P_{\text{mid}-2011} \\ \text{Number of years} & = 6 & = n \end{array}$$

Substituting these figures in Eq. (4.9):

$$r = \frac{LN\left(\frac{1,197,813,000}{1,101,318,000}\right)}{6}$$

$$r = 0.01399829 \text{ or } 1.40\,\%\,\text{p.a.}$$

In this illustration, the given values of P_t and P_{t+n} are exactly 6 years apart. If this is not the case, it is important to ensure that that the value of n is calculated to two or more decimal places, and not just the nearest whole number.

4.3.3 Interpolation and Extrapolation

Having calculated the value of r during a given period, t to $t + n$, an estimate of the population for any date within this period is called *interpolation*, and for any date outside this period (in future or in past) is known as *extrapolation*.

As an example, the population of India on 31 December 2008 is interpolated. For this purpose, $n = 3.5$ years, $r = 0.01399829$ (as estimated in the previous section) and $P_{mid\text{-}2005} = 1{,}101{,}318{,}000$. Following Eq. (4.7):

$$\begin{aligned} P_{31/12/2008} &= 1,101,318,000 * e^{0.01399829*3.5} \\ &= 1,101,318,000 * 1.05021407 \\ &= 1,156,619,659, \text{ or } 1,156,620,000 \text{ rounded to the nearest thousand.} \end{aligned}$$

For extrapolating the population of India in the past, say, 1 April 2000, the value of n is 4 years and 275 days prior to mid-2005, that is, $n = -4\frac{275}{365} = -4.75342466$

$$\begin{aligned} P_{1/4/2000} &= 1,101,318,000 * e^{0.01399829*(-4.75342466)} \\ &= 1,101,318,000 * 0.93562566 \\ &= 1,030,421,382, \text{ or } 1,030,421,000 \end{aligned}$$

For extrapolating to a future date such as 10 October 2014, the value of n is 3 years and 102 days after mid-2011: $n = 3 + \frac{102}{365} = 3.27945205$

$$\begin{aligned} P_{10/10/2014} &= 1,197,813,000 * e^{0.01399829*3.27945205} \\ &= 1,197,813,000 * 1.04697674 \\ &= 1,254,082,356, \text{ or } 1,254,082,000 \end{aligned}$$

In the last example, the starting point to determine the value of n was mid-2011. Therefore, the calculations were based on $P_{mid\text{-}2011}$. The result would be identical, if $P_{mid\text{-}2005}$ is the starting point and n is 9.27945205

The values of r and n for interpolation and extrapolation purposes should be calculated up to at least 4 decimal places, otherwise the results could be quite off the mark. For example, if r and n are estimated to only two decimal places ($r = 0.01$ and $n = 3.28$), $P_{10/10/2014}$ will be 1,237,753,000 compared to 1,254,082,000 – an under-estimation of more than 16 million people.

4.3.4 Population Doubling Time

For a population to double in size $\left(\frac{P_{t+n}}{P_t}\right)$ must be equal to 2. Thus, Eq. (4.8) takes the form:

$$2 = e^{r*n}$$

Taking *LN* of both sides of the above equation, and following rule (4) of Box 3.1:

$$0.6931 = r * n$$

If the value of r is known, the doubling time of a population can be ascertained, without necessarily knowing the values of P_t and P_{t+n}:

$$n = \frac{0.6931}{r} \tag{4.10}$$

When r is expressed as a percentage, Eq. (4.10) can be written as:

$$n = \frac{69.31}{r\%} \tag{4.11}$$

The underlying assumption for both equations is that the value of r will remain constant throughout the period during which the population doubles itself.

Substituting the value of r for India (0.01399829) in Eq. (4.10) or (4.11), at that average annual growth rate, the population of India would double every 49.5 years.

4.3.5 Growth and Fold rates

Assuming that a population continues to increase at a particular annual growth rate of r, the number of years (n) it will take to reach z-fold its original size is:

$$n = \frac{LN(z)}{r} \tag{4.12}$$

If r is expressed in percentage form:

$$n = \frac{[LN(z)] * 100}{r\%} \tag{4.13}$$

Again, the underlying assumption for both Eqs. (4.12) and (4.13) is that the value of r will remain constant throughout the period during which the population grows z fold.

To facilitate the use of Eqs. (4.12) and (4.13) the LN values for z from 3 to 10 are given below:

z =	3	4	5	6	7	8	9	10
$LN(z)$ =	1.0986	1.3863	1.6094	1.7918	1.9459	2.0794	2.1972	2.3026

4.4 Population Dynamics and Analytical Approaches

4.4.1 Demographic Stocks and Flows

In demographic context *stocks* refer to the size and composition (characteristics) of populations at a given point in time. For instance, the population stock of the United States in mid-2005 was 296.4 million (Sect. 4.3.1) and that of India was 1,101.3 million (Sect. 4.3.2). As noted in Chap. 2, censuses, sample surveys and population registers are used to obtain such information.

Similarly demographic *flows* refer to the additions and losses in populations because of births, death and migration over a given period. These are also called *vital events*. In the United States data in Sect. 4.2, the positive flows were births (4,129,000) and immigration (1,122,000), while the negative flows were deaths (2,425,000) and emigration (73,000).

There are two types of vital events:

- those that affect only the size of a population, and
- those that influence only the composition of a population but not its size.

Births, deaths and migration (both international and domestic) belong to the first type of vital events. Marriages and divorces, and entry to and exit from the labour force are examples of the second type of vital events.

4.4.2 Cross-Sectional and Longitudinal Approaches in Demography

There are two approaches commonly used in demographic analysis: cross-sectional and longitudinal.

The *cross-sectional* approach is concerned with the collection and analysis of data concerned with population size, its composition and characteristics at a particular point, or period, of time. Most demographic data such as the census and registration of vital events are collected using this approach. Another approach involves selecting a group of people who share a common characteristic such as year of birth (referred to as the birth *cohort*) and record and analyse their experiences over time. This is called the *longitudinal* approach. Alternatively, a time-series of cross sectional data may be used to follow a particular cohort of the population.

4.4.3 Synthetic Measures

As noted earlier, data for the longitudinal analysis may take many years to capture. Often, it is not convenient to wait that long. To overcome this time-lag, synthetic measures have been developed for variables such as fertility and mortality. These measures may use current patterns of experience to assess the potential longitudinal experience, on the assumption that the current population stock, flows and characteristics stayed constant. The total fertility rate (Chap. 5) and life expectancy (Chap. 7) are examples of synthetic measures.

4.5 Crude Rates

Some demographic measures are known as *crude rates* because they do not take into consideration age, sex or other characteristics of the population. They usually involve the count of some particular vital event over a specified period of time divided by the average population during the same period. Crude rates are defined as:

$$CR_{t\to t+n} = \frac{E_{t\to t+n}}{\overline{P}_{t\to t+n}} \tag{4.14}$$

$CR_{t\to t+n}$ is the crude event rate per person during period t to $t + n$, $E_{t\to t+n}$ is the number of events during the same period and $\overline{P}_{t\to t+n}$ is the average population during the same period. Generally, the population at the mid-point of the period t to $t + n$ is taken as a satisfactory measure of the average population during that period.

The crude rates for vital events mentioned in Eq. (4.4) would involve substituting the numerator of Eq. (4.14) by $B_{t\to t+n}$, $D_{t\to t+n}$, $I_{t\to t+n}$, $E_{t\to t+n}$, $\hat{I}_{t\to t+n}$, *or* $\hat{E}_{t\to t+n}$ as appropriate. The denominator remains as $\overline{P}_{t\to t+n}$ in all cases. Similar rates can also be calculated for various components of population growth mentioned in Eq. (4.5) by replacing the numerator of Eq. (4.14) by $PG_{t\to t+n}$, $NI_{t\to t+n}$, $NIM_{t\to t+n}$, *or* $NDM_{t\to t+n}$, and keeping the denominator as $\overline{P}_{t\to t+n}$.

Table 4.2 presents the crude rates of vital events and components of population growth based on the United States data from Sect. 4.2. The denominator for each rate is the same: the average of $P_{mid\text{-}2004}$ and $P_{mid\text{-}2005}$:

$$\overline{P}_{mid-2004\to mid-2005} = \frac{293,657,000 + 296,410,000}{2} = 295,033,500$$

As noted earlier, the value of $NDM_{mid\text{-}2004\to mid\text{-}2005}$ is zero for the total population of United States, because movements of population within the country's boundaries cancel out, as a gain in one part is a loss in some other part(s).

Table 4.2 Crude rates of the United States, mid-2004 to mid-2005

Vital event/component	Numerator	Denominator	Crude rate per 1,000
(1)	(2)	(3)	(4) = (2)*1,000/(3)
Births	4,129,000	295,033,500	14.0
Deaths	2,425,000	295,033,500	8.2
Immigration	1,122,000	295,033,500	3.8
Emigration	73,000	295,033,500	0.2
Population growth	2,753,000	295,033,500	9.3
Net increase	1,704,000	295,033,500	5.8
Net migration	1,049,000	295,033,500	3.6

Source: Section 4.2

4.6 Characteristic-Specific Rates and Ratios

4.6.1 Characteristic-Specific Rates

Vital events are often associated with population characteristics such as age, sex, marital status and ethnicity. Characteristic-specific rates relate vital events to characteristic-specific groups in the population that experience such events. The general formula for the i-specific rate is:

$$R^i_{t\to t+n} = \frac{E^i_{t\to t+n}}{\overline{P}^i_{t\to t+n}} \tag{4.15}$$

$R^i_{t\to t+n}$ is the vital event rate per person of a particular characteristic i during the period t to $t + n$, $E^i_{t\to t+n}$ is the number of vital events of characteristic i during the same period, and $\overline{P}^i_{t\to t+n}$ is the average population with characteristic i during the same period.

Characteristic-specific rates can be calculated by specifying one characteristic such as age, or two characteristics such age and sex or even multiple characteristics. These rates are discussed in more detail in subsequent chapters.

4.6.2 Sex Ratios

Sex ratio is an index of balance between the number of males and females. It is sometimes referred to as the *masculinity ratio*, and is calculated by dividing the number of males by the number of females. The sex ratio can also be expressed as a *femininity ratio* by taking the number of females in the numerator and males in the denominator. The sex ratio can be estimated for the total population or for any segment of the population with characteristic i. Sex ratio SR^i_t for characteristic i at time t is defined as:

$$SR^i_i = \frac{M^i_t}{F^i_t} \tag{4.16}$$

M^i_t is the number of males of characteristic i at time t and F^i_t is the number of females of characteristic i at time t. Usually, characteristic i refers to the age, however, other characteristics can also be used.

Data for Brazil and Italy are used below to estimate the sex ratio for 0–4 population in both countries (United Nations 2009):

	Brazil (2008)	Italy (2008)
Males aged 0–4 (M_{2008}^{0-4}) =	8,226,630	1,453,690
Females aged 0–4 (F_{2008}^{0-4}) =	7,976,611	1,373,687

$$SR_{2008}^{0-4} = \frac{8,226,630}{7,976,611} = 1.031 \text{ for Brazil or 103 males per 100 females.}$$

$$SR_{2008}^{0-4} = \frac{1,453,690}{1,373,687} = 1.058 \text{ for Italy or 106 males per 100 females.}$$

Figure 4.1 shows the sex ratios for various age groups in Brazil and Italy. The declining trend in sex ratio by age is evident in both countries. It is a well-known phenomenon due to the comparatively more frequent deaths among males than females, particularly in the older ages.

4.6.3 Sex Ratio at Birth

The *sex ratio at birth* is defined as:

$$SRB_{t\to t+n} = \frac{B_{t\to t+n}^{m}}{B_{t\to t+n}^{f}} \tag{4.17}$$

$SRB_{t\to t+n}$ is the sex ratio at birth during the period t to $t + n$, and $B_{t\to t+n}^{m}$ and $B_{t\to t+n}^{f}$ are the number of male and female live births during the same period.

In most populations, the sex ratio at birth fluctuates around an average of 1.05 or 105 males per 100 females.

The number of births in Canada and Egypt (United Nations 2009) provide an example of estimation of sex ratio at birth:

	Canada (2007)	Egypt (2008)
Male live births during the year ($B_{t\to t+n}^{m}$) =	188,337	1,055,046
Female live births during the year ($B_{t\to t+n}^{f}$) =	179,527	995,658

$$SRB_{2007}^{Canada} = \frac{188,337}{179,527} = 1.049, \text{ or 105 males per 100 females.}$$

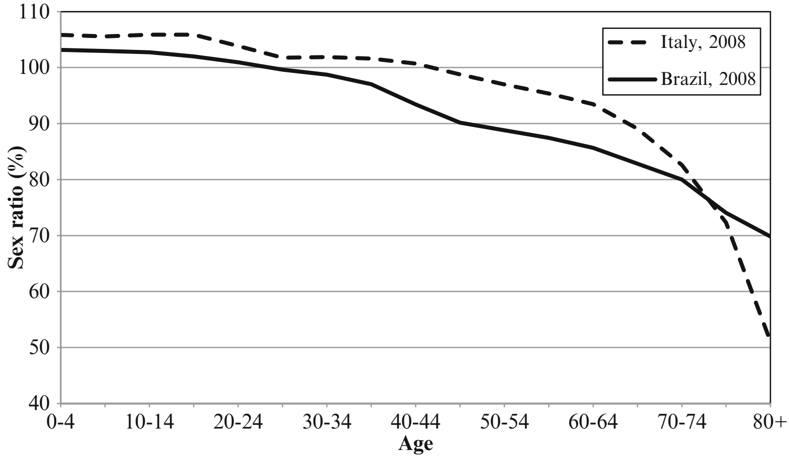


Fig. 4.1 Age-specific sex ratios: Brazil and Italy, 2008 (*Source*: United Nations 2009)

$$SRB_{2008}^{Egypt} = \frac{1,055,046}{995,658} = 1.060, \text{ or 106 males per 100 females.}$$

4.6.4 Child-Woman Ratio

Another demographic measure based on the age distribution of a population is the *child-woman ratio*. This ratio relates the number of children (both males and females added together) aged less than 5 years to women in their reproductive period at a given point in time. Usually, the reproductive period of females is considered to be either 15–44 or 15–49 years of age. This ratio is a crude indicator of the level of fertility.

$$CWR_t = \frac{P_t^{<5}}{W_t^r} \tag{4.18}$$

CWR_t is the child-woman ratio at time $t, P_t^{<5}$ is the number of children under 5 years of age at time t and W_t^r is the number of women of reproductive age group (r = 15–44 years or 15–49 years of age) at time t.

In the following example, data for Brazil and Italy supply the basis for the calculation of CWR_{2008} for both countries (United Nations 2009):

	Brazil (2008)	Italy (2008)
$P_{2008}^{<5}$ = Population under age 5 years	16,203,241	2,827,377
W_{2008}^{15-44} = Women aged 15–44 years	92,998,843	23,855,033

$$CWR_{2008}^{Brazil} = \frac{16,203,241}{92,998,843} = 0.174, \text{ or } 174 \text{ children}$$
$$0 - 4 \text{ years of age per } 1,000 \text{ women aged } 15 - 44 \text{ years.}$$

$$CWR_{2008}^{Italy} = \frac{2,827,377}{23,855,033} = 0.119, \text{ or } 119 \text{ children}$$
$$0 - 4 \text{ years of age per } 1,000 \text{ women aged } 15 - 44 \text{ years.}$$

These figures indicate that Brazilian women had, on average, more children than their Italian counterparts. These ratios are also expressed per 1,000 women. In this example, the reproductive period was defined as 15–44 years of age. Obviously, the ratios would be somewhat lower if the age group was extended to 49 years. In making comparisons of the child-woman ratios of two or more populations, it is necessary to make sure that the definition of women's reproductive period is identical.

4.6.5 Dependency Ratio

Using the age distribution of a population, it is possible to construct a measure of dependency. *Dependents* consist of two groups of people: (a) younger dependents, usually defined as people less than 15 years of age, and (b) older dependents, usually taken as people 65 years of age or older. *Independents* are considered to be the persons 15–64 years of age.

The *dependency ratio* at time t is defined as:

$$DR_t = \frac{P_t^{<15} + P_t^{65+}}{P_t^{15-64}} \tag{4.19}$$

DR_t is the dependency ratio at time $t, P_t^{<15}$ is the population less than 15 years of age at time t, P_t^{65+} is the number of persons 65 years of age or older at the same time and P_t^{15-64} is the population 15–64 years of age, also at time t. This ratio is not necessarily an index of economic dependency.

Two other related ratios are the *young dependency ratio* (YDR_t) and the *old dependency ratio* (ODR_t), both at time t:

$$YDR_t = \frac{P_t^{<15}}{P_t^{15-64}} \tag{4.20}$$

$$ODR_t = \frac{P_t^{65+}}{P_t^{15-64}} \tag{4.21}$$

These ratios are the relative contribution of the younger (<15 years) and the older (65+ years) age segments of the population to the dependency ratio. The sum of these two ratios is the equivalent to the dependency ratio.

The age distributions for Brazil and Italy are used to illustrate the calculation of dependency ratios and their components (United Nations 2009):

	Brazil (2008)	Italy (2008)
Persons aged under 15 years ($P_{2008}^{<15}$) =	50,186,610	8,405,501
Persons aged 65 years or more (P_{2008}^{65+}) =	12,377,850	12,016,006
Persons aged 15–64 years (P_{2008}^{15-64}) =	127,048,354	39,412,681

$$DR_{2008}^{Brazil} = \frac{50,186,610 + 12,377,850}{127,048,354} = 0.492 \text{ or } 49.2\,\% \text{ dependents.}$$

$$DR_{2008}^{Italy} = \frac{8,405,501 + 12,016,006}{39,412,681} = 0.518 \text{ or } 51.8\,\% \text{ dependents.}$$

$$YDR_{2008}^{Brazil} = \frac{50,186,610}{127,048,354} = 0.395 \text{ or } 39.5\,\% \text{ dependents less than 15 years of age.}$$

$$ODR_{2008}^{Brazil} = \frac{12,377,850}{127,048,354} = 0.097 \text{ or } 9.7\,\% \text{ dependents 65 years of age or older.}$$

$$YDR_{2008}^{Italy} = \frac{8,405,501}{39,412,681} = 0.213 \text{ or } 21.3\,\% \text{ dependents less than 15 years of age.}$$

$$ODR_{2008}^{Italy} = \frac{12,016,006}{39,412,681} = 0.305 \text{ or } 30.5\,\% \text{ dependents 65 years of age or older.}$$

The dependency ratios in both countries were similar. However, Italy had a larger proportion of older and a smaller proportion of younger dependents compared to Brazil.

4.6.6 Child to Old Ratios

Another demographic measure related to the age distribution is the *child to old ratio* that compares the number of children (less than 15 years of age) to the number of old people (persons 65 years of age or older):

$$COR_t = \frac{P_t^{<15}}{P_t^{65+}} \tag{4.22}$$

COR_t is the child to old ratio at time t, $P_t^{<15}$ is the number of *children* under 15 years of age at time t, and P_t^{65+} is the number of *persons* aged 65 years or older at time t. Using the data for Brazil and Italy, presented in Sect. 4.6.5, the child to old ratios in Brazil and Italy are estimated as 39.5 % and 21.3 %.

4.7 Population Density and Distribution

4.7.1 Population Density

Population density relates a geographical area to the population living in it. The demarcation of geographical areas may vary. For example, usually, lakes and waterways are included at national levels but may be excluded at sub-national level.

Population density is defined as:

$$PD_t^l = \frac{P_t^l}{S_t^l} \tag{4.23}$$

PD_t^l is the population density at time t per unit of area l, P_t^l is the population at time t within the boundaries of geographical area l, and S_t^l is the number of surface units (usually square kilometres) at time t within the boundaries of area l.

Table 4.3 presents information on population and land area in China and three of its provinces (Benewick & Donald 2009). The table states that the overall population density of China was 141 persons per square kilometres in the year 2006. However, averages for a country can mask differences in population densities between different areas of that country. This is the case in China where there are substantial disparities in population density among the three provinces.

Generally, urban areas, and in particular large cities, have higher population densities than rural areas. For example, according to the 2011 census of India, the average population density of the whole country was 382 people per square kilometres but that of the National Capital Territory (Delhi) was 11,297 (India 2011).

Table 4.3 Estimation of population density: China, 2006

Country/province	Population 2006 estimate (000's)	Land area (km^2) (000's)	Population density (persons per km^2)
(1)	(2)	(3)	(4) = (2)/(3)
CHINA	1,312,000	9,326	141
Guangdong	94,500	178	531
Sichuan	81,300	488	167
Neimongol (Inner Mongolia)	24,100	1,200	20

Source: Benewick & Donald (2009)

Table 4.4 Calculation of the index of redistribution: Australia, 1981–2011

State/territory	Population (000's) 1981	Population (000's) 2011	% distribution 1981	% distribution 2011	% difference 2011–1981
(1)	(2)	(3)	(4)	(5)	(6) = (5)–(4)
New South Wales	5,235	7,302	35.1	32.3	−2.8
Victoria	3,947	5,621	26.4	24.9	−1.6
Queensland	2,345	4,580	15.7	20.3	4.5
South Australia	1,319	1,656	8.8	7.3	−1.5
Western Australia	1,300	2,349	8.7	10.4	1.7
Tasmania	427	511	2.9	2.3	−0.6
Aust. Capital Territory	228	366	1.5	1.6	0.1
Northern Territory	123	230	0.8	1.0	0.2
Total (absolute value)	14,924	22,615	100.0	100.0	13.0
Index of redistribution = Total ÷ 2 =					6.5

Source: Australia (2012)
Note: Totals in columns may not add due to rounding

4.7.2 Index of Redistribution

Population redistribution over a given period of time within geographical areas can be examined with an index of redistribution. It is similar to the index of dissimilarity discussed in Sect. 3.8.3.

The *index of redistribution* between time t and time $t + n$ is:

$$IR_{t \to t+n} = \frac{\sum_{i=1}^{i=k} \left| P_{t+n}^{i} - P_{t}^{i} \right|}{2} \tag{4.24}$$

$IR_{t \to t+n}$ is the index of redistribution between time t and $t + n$, and P_t^i and P_{t+n}^i refer to the percentage distribution of population of geographical area i at time t and at time $t + n$. The index is simply half of the absolute sum (over all values of i from 1 to k) of the differences in the percentage distributions of population. This procedure identifies the areas which have gained or lost population.

The index varies between 0 and 100; a smaller value indicates a lower redistribution of the population during the period under review.

Table 4.4 illustrates the calculation of this index for six States and Territories in Australia over a 30-year time period 1981–2011. The overall index of 6.5 does not indicate a large-scale redistribution of the population. However, in proportional terms the two largest states of New South Wales and Victoria along with South Australia and Tasmania lost while the others gained population.

4.7.3 Index of Concentration

A similar index that measures the degree of population concentration relates the percentage distribution of the population to the percentage distribution of the

Table 4.5 Calculation of the index of concentration: Australia, 1981–2011

	Population (%)		Surface area		% diff: (Pop.-Surface)	
State/territory	1981	2011	km^2	%	1981	2011
(1)	(2)	(3)	(4)	(5)	(6) = (2)–(5)	(7) = (3)–(5)
New South Wales	35.1	32.3	801	10.4	24.7	21.9
Victoria	26.4	24.9	227	3.0	23.5	21.9
Queensland	15.7	20.3	1,731	22.5	−6.8	−2.3
South Australia	8.8	7.3	983	12.8	−3.9	−5.5
Western Australia	8.7	10.4	2,530	32.9	−24.2	−22.5
Tasmania	2.9	2.3	68	0.9	2.0	1.4
Aust. Capital Territory	1.5	1.6	2	0.0	1.5	1.6
Northern Territory	0.8	1.0	1,349	17.5	−16.7	−16.5
Total	100.0	100.0	7,691	100.0	103.3	93.6
Index of concentration = Total ÷ 2 =					51.7	46.8

Source: Table 4.4 for columns (2) and (3), and Australia (2010) for column (4)
Note: Totals in columns may not add due to rounding. Totals of columns (6) and (7) are the sum of their absolute values

relevant geographical area. It is also an indicator of the density of population within specific geographical areas. The *index of concentration* is specified as:

$$IC = \frac{\sum_{i=1}^{i=k} \left| P_t^i - S_t^i \right|}{2} \tag{4.25}$$

P_t^i and S_t^i refer to the population and surface area of a particular geographical entity i at time t, as a percentage of the total population and area of a country. This index can also be used to study the changes in concentration patterns of a country over two or more points in time. Usually, the S_t^i values do not change. However, if there have been boundary changes then appropriate adjustments will have to be made. This index also varies between 0 and 100. A value of 0 denotes that the percentage distributions of population and area are identical. A higher value of the index is an indicator of a greater concentration of population.

The index of concentration was calculated for the areas and populations of the eight States/Territories of Australia in Table 4.5. The index reveals that the extent of population concentration was reduced between 1981 and 2011.

4.8 Impact of Characteristics on Demographic Measures

In Eq. (4.15) the characteristic-specific rates are defined as:

$$R_{t\to t+n}^i = \frac{E_{t\to t+n}^i}{\overline{P}_{t\to t+n}^i}$$

Table 4.6 Estimation of the expected number of events in two hypothetical populations subjected to the same age-specific rates

Age	Population ($\overline{P}^i_{t\to t+n}$)		Rate	Expected events ($E^i_{t\to t+n}$)	
(i)	A	B	($R^i_{t\to t+n}$)	A	B
(1)	(2)	(3)	(4)	(5) = (2)*(4)	(6) = (3)*(4)
0–9	100	10	0.1	10	1
10–19	90	20	0.2	18	4
20–29	80	30	0.3	24	9
30–39	70	40	0.4	28	16
40–49	60	50	0.5	30	25
50–59	50	60	0.6	30	36
60–69	40	70	0.7	28	49
70–79	30	80	0.8	24	64
80–89	20	90	0.9	18	81
90–99	10	100	1.0	10	100
Total	550	550	…	220	385

$R^i_{t\to t+n}$ is the vital event rate per person of a particular characteristic i during the period t to $t + n$, $E^i_{t\to t+n}$ is the number of vital events of characteristic i during the same period, and $\overline{P}^i_{t\to t+n}$ is the average population with characteristic i, also during the same period.

Knowing $R^i_{t\to t+n}$ and $\overline{P}^i_{t\to t+n}$ it is possible to estimate the expected number of events $(E^i_{t\to t+n})$ in a population as:

$$E^i_{t\to t+n} = R^i_{t\to t+n} * \overline{P}^i_{t\to t+n} \tag{4.26}$$

The concept underlying these two equations is essentially the same as the concept of the weighted mean discussed previously.

Two hypothetical populations A and B in Table 4.6 have exactly the same size (550) but different age distributions. The median age of A is within the group 30–39 years of age and that of B in the group 60–69.

If both populations are subjected to the same rates ($R^i_{t\to t+n}$), the expected number of events in each can be estimated using Eq. (4.26), hence the crude rate of A was $\frac{220}{550} = 0.4$ and of B was $\frac{385}{550} = 0.7$. The crude rates could be interpreted as a substantial difference in the events between the two populations. This is contrary to the fact that both populations have been subjected to the same event rates at each age.

In the above example, the main explanation for the discrepancy lies in the different age composition of the two populations, and the crude rates do not take into consideration such differences. A procedure which overcomes this problem is *standardization*. It is commonly used in demography as well as other disciplines, such as sociology, economics and epidemiology.

4.9 Standardization

In demographic analysis, standardization is concerned with the assessment of whether differences in rates related to two or more populations (or groups within them) are the effect of disparities in the composition of those populations or attributable to other factors. In this assessment, either a population with a given composition is adopted as the *standard population* or a set of specific rates, of given categories of a characteristic, such as age, are chosen as the *standard rates*. Then, the standard population or rates for given categories are applied to the actual rates or populations in the same categories. The results allow the comparison of individual experiences of populations with different compositions. They also provide for the estimation of the magnitude of the variation in rates of fertility, mortality or any other demographic phenomena: either among countries and areas within or at different points in time. There are two methods of standardization: the *direct* and the *indirect* method.

4.9.1 Direct Standardization

The direct standardization procedure involves the selection of a population with a given composition as the *standard*. The equation to carry out the estimation procedure is:

$$DSR^A = \frac{\sum_{i=1}^{i=k} (r_i^A * P_i^s)}{\sum_{i=1}^{i=k} P_i^s} \tag{4.27}$$

where DSR^A is the *direct standardized rate* for population A, r_i^A is the specific rate of category i of a particular characteristic of population A and P_i^s is the number in the standard population in category i of the same characteristic. In the numerator, the summation involves the products of all the categories of i (from 1 to k) of the given characteristic. The denominator is the sum of the standard population for all i categories of the same characteristic.

The numerator of Eq. (4.27) represents the expected number of events in the *standard* population assuming that it is subjected to the characteristic-specific rates of category i of population A (r_i^A). The denominator is the *standard* population of category i of the same characteristic. The summations of the numerator and the denominator are of all the k values of category i of the particular characteristic. In other words, the direct standardized rate of population A is the crude rate of the *standard* population, if subjected to the *characteristic*-specific rates of population A. In Eq. (4.27), the events may refer to births, deaths, marriages or other events or variables such as income. Age is the characteristic most commonly used and i refers to various age categories. However, other characteristics, such as marital status, may also be the subject of standardization.

In essence, direct standardization is similar to the calculation of the weighted mean: the *original weights* are substituted with *new weights*. In direct standardization the *standard population weights* substitute the *original weights* of the original population.

4.9.2 Indirect Standardization

The indirect standardization procedure involves the selection of a population with a set of rates as the *standard*. This procedure has two steps:

$$\text{Step 1}: \quad SR^A = \frac{\theta^A}{\sum_{i=1}^{i=k} (R_i^s * p_i^A)} \tag{4.28}$$

where SR^A is the standardized event ratio for population A, θ^A is the sum of actual number of events in population A, R_i^s is the characteristic-specific event rate for category i in the *standard* population and p_i^A is the number in category i of population A. The summation is of all the k values of category i of the particular characteristic.

$$\text{Step 2}: \quad ISR^A = CR^s * SR^A \tag{4.29}$$

ISR^A is the indirect standardized rate for population A, CR^s is the crude event rate for the *standard* population and SR^A is the standardized event ratio for population A.

Indirect standardization is also related to the weighted mean. However, it reverses the roles of population and rates. Thus, it is the rates rather than the population that serve as the weights.

4.9.3 Examples of Standardization

Data for England and two of its regions – North East and the South East – were used to illustrate the direct and indirect methods of standardization (United Kingdom 2009). The information available consisted of (1) deaths by age and sex, and (2) death rates by age and sex. Using procedure 1 outlined in Box 4.1, the authors estimated the population of England and the two regions.

Box 4.1 A Simple Estimation Procedure

Situations may occur when the numerator or denominator of a rate or ratio is not readily available. A simple procedure to obtain the relevant figures is to estimate the information not provided:

Let R be a rate or a ratio of two numbers A and B and $R = \frac{A}{B}$
Knowing the value of R and A, the value of $B = \frac{A}{R}$ (procedure 1).
Knowing the value of R and B, the value of $A = B{*}R$ (procedure 2).
R should be expressed in relevant units. For example, if R is per 1,000 then R is divided by 1,000 before following procedures 1 or 2.

Table 4.7 Calculation of the direct standardized mortality rates for North East and South East regions of England, 2008

		Mortality rates of region (per person)		Expected deaths in England if subjected to mortality rates of	
Age group (i)	Population of England: *Standard* (P_i^s)	North East (r_i^A)	South East (r_i^B)	North East ($r_i^A * P_i^s$)	South East ($r_i^B * P_i^s$)
(1)	(2)	(3)	(4)	(5) = (2)*(3)	(6) = (2)*(4)
0–4	3,129,300	0.00098338	0.00100060	3,077	3,131
5–14	5,904,000	0.00011938	0.00008018	705	473
15–34	13,562,600	0.00063664	0.00042305	8,634	5,738
35–64	20,565,300	0.00411111	0.00299825	84,546	61,660
65+	8,285,300	0.05153265	0.04578651	426,963	379,355
Total	51,446,500	0.01063245#	0.00905952#	523,925	450,357
Direct standardized mortality rate =				$\frac{523,925}{51,446,500} = 0.0102$	$\frac{450,357}{51,446,500} = 0.0088$

Source: Table A4.1
Note: All rates refer to the 2008 calendar year and the population to mid-2008. Crude death rates for the regions are marked with a hash (#)

Direct Standardization

To illustrate this procedure, the age distribution of the population of England was taken as the *standard* and the mortality rates of two regions in England (North East and South East) as populations A and B. Table 4.7 presents these data in columns (2)–(4). Columns (5) and (6) were calculated following the numerator of Eq. (4.27). Dividing the sum of column (5) by the sum of column (2) gives the direct standardized mortality rate for the North East region. The same procedure with columns (6) and (2) gave the similar rate for the South East region.

Table 4.7 indicates that for all ages, apart from 0 to 4 age group, the North East region had higher mortality rates than the South East, and the crude death rate

Table 4.8 Calculation of the indirect standardized mortality rates for the North East and South East regions of England, 2008

Age group (years)	Mortality rates England: *Standard* (R_i^s)	Population: North East (p_i^A)	Population: South East (p_i^B)	Expected deaths if mortality rates of England prevailed: North East ($R_i^s * p_i^A$)	Expected deaths if mortality rates of England prevailed: South East ($R_i^s * p_i^B$)
(1)	(2)	(3)	(4)	(5) = (2)*(3)	(6) = (2)*(4)
0–4	0.00118205	144,400	497,700	171	588
5–14	0.00010061	284,800	985,300	29	99
15–34	0.00050455	666,000	2,075,400	336	1,047
35–64	0.00344221	1,044,000	3,419,000	3,594	11,769
65+	0.04753443	436,500	1,402,400	20,749	66,662
Total	0.00924776#	2,575,700	8,379,800	24,879	80,165
Standardized mortality ratio (SMR) =				$\frac{27,386}{24,879} = 1.10077$	$\frac{75,917}{80,165} = 0.94701$
Indirect standardized mortality rate = CDR*SMR =				0.0102	0.0088

Source: Table A4.1

Note: All rates refer to the 2008 calendar year and the population to mid-2008. Crude death rate of the standard population is marked with a hash (#)

for North East (0.01063245) was about 16.5 % higher than that for the South East (0.00905952). However, after standardizing for the effect of differences in age distributions of the two regions the mortality differential was reduced to 15.9 %.

Indirect Standardization

For this procedure the age-specific mortality rates of England were taken as the *standard* and the age distributions of the two regions in England (North East and South East) as populations A and B. Table 4.8 presents the age-specific mortality rates for England in column (2), the population of the two regions by age in columns (3) and (4). In addition, the actual number of deaths was 27,386 in the North East region and 75,917 in the South East region. The crude death rate in England was 9.24776 per 1,000 people (see Table A4.1).

Columns (5) and (6) were calculated following the denominator of Eq. (4.28). The standard rate in column (2) was multiplied by columns (3) and (4) to obtain the results in columns (5) and (6) respectively. Dividing the sums of columns (5) and (6) by the actual number of deaths in the two regions gave their standardized mortality ratios. These ratios multiplied by the crude death rate of England resulted in the indirect standardized mortality rates for the two regions as shown in Table 4.8.

The indirect standardized rates are identical to the ones estimated using the direct standardization procedure. Small differences may arise as the result of rounding.

According to Table A4.1, the crude death rates for these two regions were 10.63 and 9.06 respectively indicating that the crude rate for the South East region was 17.3 % lower than that for the North East region. However, the difference in standardized rates was 16.3 %.

Which method to use? It all depends on the availability of data. There are no hard and fast rules for the selection of the *standard*. In the comparison of populations at sub-national levels, the national population could be selected as the *standard*. Otherwise, some other relevant population may be used. When countries are compared one of them may be taken as the *standard*. Often, in cancer and other epidemiologic studies, a hypothetical world population has been used as the *standard* to calculate direct standardized rates (Ahmad et al. 2001).

4.10 Graphical Presentation of Demographic Data

Demographic data can be presented in a variety of graphical forms, such as straight-line or curve graphs, bars and pie charts. However, frequently, age and sex distributions are presented in graphical forms known as *population pyramids*. These graphs may be in terms of the age and sex distributions in absolute figures or percentages. An example of a population pyramid is given in Fig. 4.2. It is based on the age-sex distribution of the population of Egypt in 2010 in Table 4.9.

The pyramid consists of horizontal bars, each representing population in a particular age group, starting from 0 to 4 years of age, and going up to 70 years and over. Male bars are usually on the left and females on the right of the vertical line, but there is no reason why this could not be reversed. The length of the bar represents the proportion of the particular cohort of males or females and the height is 5 years, except the top-most bar that represents an open age interval of 70 years and over.

The base of a pyramid represents the population aged 0–4 years. Its length is determined by the number of births 5 years prior to the data collection and, to a lesser extent, the number of deaths and net migration of children under 5 years of age. In the 5–9 years and older age groups the pyramid bar is influenced by deaths and net migration in each age group. Overall, a pyramid reflects past trends in births, deaths and net migration in a population.

To compare two pyramids based on populations of different sizes, the number of people in each sex and age groups is converted into percentages. As mentioned in Sect. 3.4, percentages may be calculated using the total for each sex as the base, or the total for both sexes combined as the base. In this case, the latter is more appropriate, as it ensures that the sex ratios in the percentage distribution reflect the actual sex ratio for each age group in the population.

The population pyramid for Egypt is compared with that of France in Fig. 4.3. The two pyramids are based on the figures in Table 4.9 for Egypt and Table 4.10 for France. Both tables show the actual population data as well as percentage

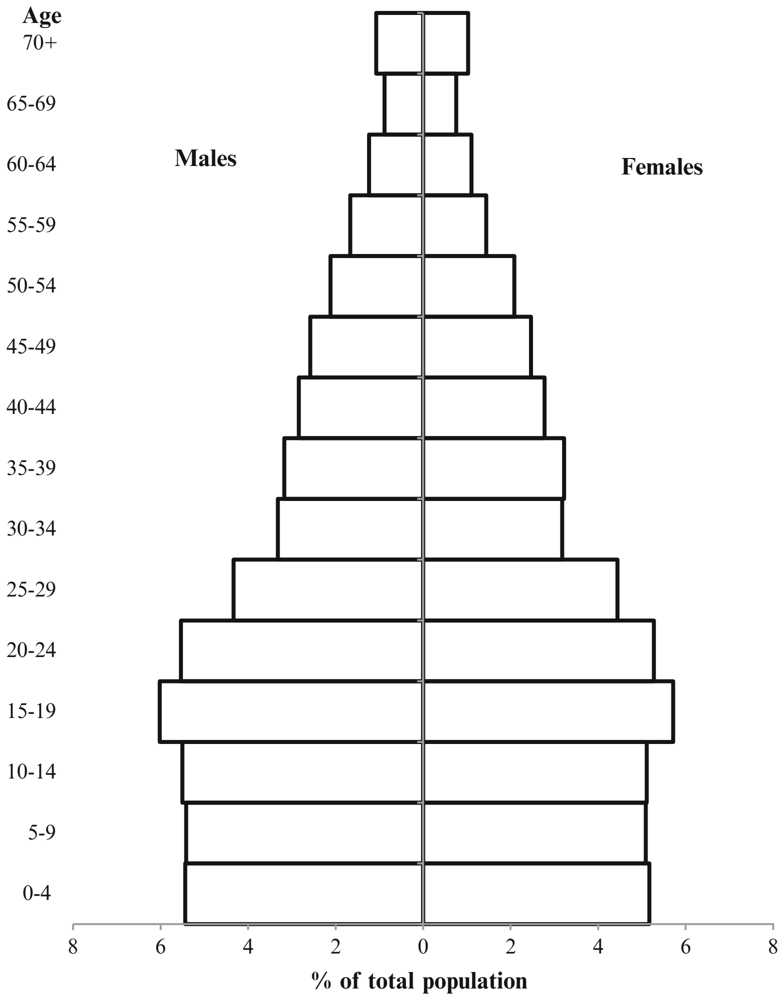


Fig. 4.2 Population pyramid for Egypt, 2010 (*Source*: Table 4.5, columns (2) and (3))

distributions. The superimposed pyramids of Egypt and France are based on the percentage distributions.

Figure 4.3 indicates that the proportion of the population under 35 years of age was much larger in Egypt than in France. The differences were substantially greater among the younger age groups. For example, the group 0–4 years of age accounted for 10.61 % of the Egyptian population (Table 4.9), and only 6.19 % of the French population (Table 4.10). Given that the population 0–4 years of age consists mainly of survivors of births five years prior to 2010, this disparity points to differences in fertility levels that were substantially higher in Egypt than in France.

Table 4.9 Sex and age distribution of the population: Egypt, 2010

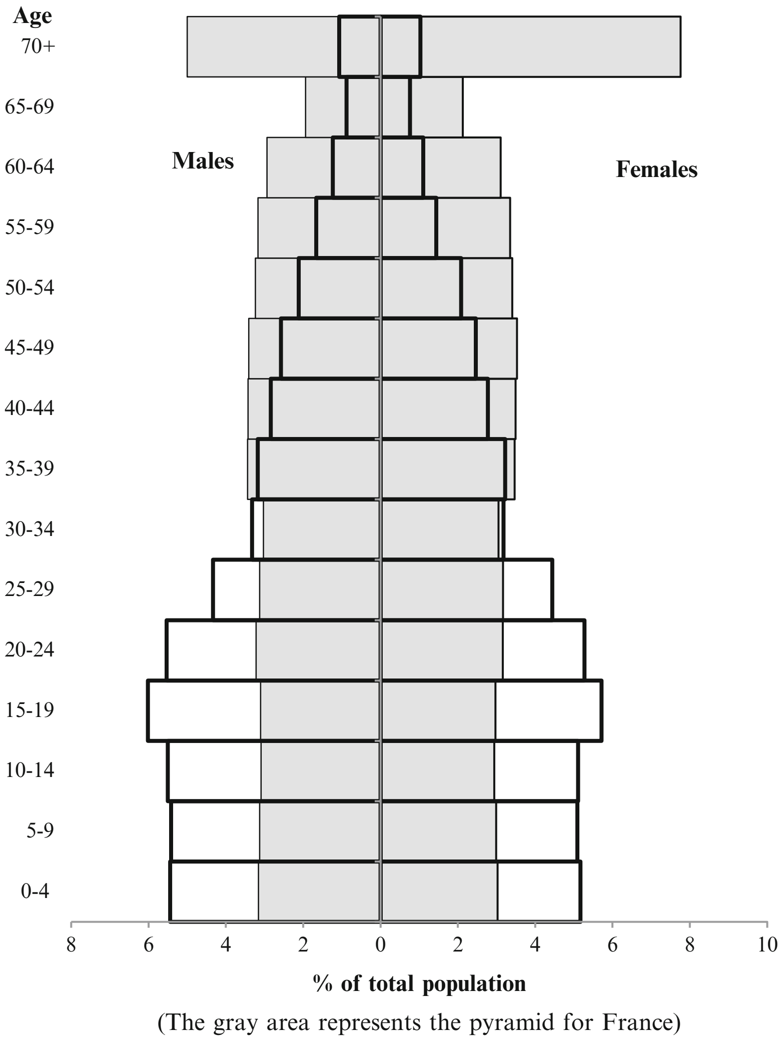
	Population		% of total population	
Age group	Males	Females	Males	Females
(1)	(2)	(3)	(4) = (2)*100/78,728,329	(5) = (3)*100/78,728,329
0–4	4,282,316	4,071,818	5.44	5.17
5–9	4,265,412	4,007,172	5.42	5.09
10–14	4,330,443	4,023,320	5.50	5.11
15–19	4,738,218	4,500,674	6.02	5.72
20–24	4,358,058	4,155,045	5.54	5.28
25–29	3,411,636	3,498,432	4.33	4.44
30–34	2,613,911	2,502,494	3.32	3.18
35–39	2,498,485	2,535,871	3.17	3.22
40–44	2,236,659	2,186,333	2.84	2.78
45–49	2,029,328	1,941,694	2.58	2.47
50–54	1,667,511	1,639,886	2.12	2.08
55–59	1,312,016	1,135,920	1.67	1.44
60–64	971,087	871,006	1.23	1.11
65–69	692,772	597,495	0.88	0.76
70+	842,588	810,729	1.07	1.03
Total[a]	40,250,440	38,477,889	51.13	48.87

Source: United Nations (2012, Table 7a)
Note: [a]Total population (both sexes) = 78,728,329

Figure 4.3 also reveals that France had a higher percentage of people aged 35 years and over than Egypt. This could be partly explained by net international migration. However, this differential is more likely to be due to French people, on average, living longer than their Egyptian counterparts. The ratio of the proportions of people aged 70 years or over in the French and Egyptian populations was 6:1. Generally, migration occurs in younger ages. Consequently, this ratio suggests dissimilar mortality levels that were considerably lower in France than in Egypt.

The age-sex distributions, hence population pyramids are good indicators of past trends in fertility, mortality and migration in a population.

Population pyramids for four selected countries – Brazil, Indonesia, Senegal and Sri Lanka – are presented in Figs. A4.1, A4.2, A4.3 and A4.4. These countries are characterised by relatively higher fertility and mortality rates. Population pyramids for four other countries – Italy, Japan, Australia and Canada – are presented in Figs. A4.5, A4.6, A4.7 and A4.8. These countries have relatively lower fertility and mortality levels. Two of these, Australia and Canada, have substantial international migration to add to their natural increases.



…

Fig. 4.3 Population pyramid for Egypt and France, 2010 (*Source*: Columns (4) and (5) in Tables 4.9 and 4.10)

Table 4.10 Age and sex and distribution of the population: France, 2010

	Population		% of total population	
Age group	Males	Females	Males	Females
(1)	(2)	(3)	(4) = (2)*100/62,793,432	(5) = (3)*100/62,793,432
0–4	1,981,984	1,902,443	3.16	3.03
5–9	1,966,731	1,874,688	3.13	2.99
10–14	1,940,517	1,843,411	3.09	2.94
15–19	1,949,381	1,866,051	3.10	2.97
20–24	2,022,307	1,986,655	3.22	3.16
25–29	1,967,866	1,988,926	3.13	3.17
30–34	1,902,811	1,918,253	3.03	3.05
35–39	2,158,654	2,181,909	3.44	3.47
40–44	2,151,121	2,194,178	3.43	3.49
45–49	2,140,122	2,213,590	3.41	3.53
50–54	2,033,465	2,141,310	3.24	3.41
55–59	1,993,592	2,104,449	3.17	3.35
60–64	1,847,480	1,955,265	2.94	3.11
65–69	1,215,547	1,338,285	1.94	2.13
70+	3,142,201	4,870,240	5.00	7.76
Total[a]	30,413,779	32,379,653	48.43	51.57

Source: United Nations (2012, Table 7a)
Note: [a]Total population (both sexes) = 62,793,432

Appendix 4.1 Input Data for Standardization and Pyramids

Table A4.1 Input data for standardization: England and two regions, 2008

Age group	England (*standard*)	North East region	South East region
	Estimated population in mid-2008 (rounded to the nearest 100)		
0–4	3,129,300	144,400	497,700
5–14	5,904,000	284,800	985,300
15–34	13,562,600	666,000	2,075,400
35–64	20,565,300	1,044,000	3,419,000
65+	8,285,300	436,500	1,402,400
All ages	51,446,500	2,575,700	8,379,800
	Deaths during 2008		
0–4	3,699	142	498
5–14	594	34	79
15–34	6,843	424	878
35–64	70,790	4,292	10,251
65+	393,837	22,494	64,211
All ages	475,763	27,386	75,917
	Age-specific mortality rates per person p.a.		
0–4	0.00118205	0.00098338	0.00100060
5–14	0.00010061	0.00011938	0.00008018
15–34	0.00050455	0.00063664	0.00042305

(continued)

Table A4.1 (continued)

Age group	England (*standard*)	North East region	South East region
35–64	0.00344221	0.00411111	0.00299825
65+	0.04753443	0.05153265	0.04578651
Crude death rates	0.00924776	0.01063245	0.00905952

Source: United Kingdom (2009)

Note: The data available from the source listed consisted of deaths and death rates by age and sex for England and the two regions as shown in the middle and the last panel of Table A4.1. Using procedure 1 outlined in Box 4.1, the authors estimated the population of England and the two regions. These estimates are given in the top panel of Table A4.1

Table A4.2 Estimated population by age and sex for Brazil, Indonesia, Senegal and Sri Lanka: 2010

	Brazil		Indonesia		Senegal		Sri Lanka	
Age group	Males	Females	Males	Females	Males	Females	Males	Females
0–4	7,016,987	6,779,172	11,658,856	11,013,204	1,040,234	1,020,301	892,000	863,000
5–9	7,624,144	7,345,231	11,970,804	11,276,366	843,592	813,300	922,000	895,000
10–14	8,725,413	8,441,348	11,659,310	11,018,180	759,199	727,242	943,000	916,000
15–19	8,558,868	8,432,002	10,610,119	10,260,967	707,601	689,538	1,015,000	988,000
20–24	8,630,227	8,614,963	9,881,969	9,996,448	611,192	617,779	974,000	968,000
25–29	8,460,995	8,643,418	10,626,458	10,673,629	498,183	530,596	799,000	832,000
30–34	7,717,657	8,026,855	9,945,211	9,876,989	382,180	409,186	779,000	801,000
35–39	6,766,665	7,121,916	9,333,720	9,163,782	307,879	345,142	758,000	780,000
40–44	6,320,570	6,688,797	8,319,453	8,199,015	242,441	283,409	707,000	718,000
45–49	5,692,013	6,141,338	7,030,168	7,005,784	191,839	228,017	615,000	635,000
50–54	4,834,995	5,305,407	5,863,756	5,693,103	163,550	175,192	554,000	572,000
55–59	3,902,344	4,373,875	4,398,805	4,046,531	135,401	142,325	390,000	427,000
60–64	3,041,034	3,468,085	2,926,073	3,130,238	95,714	115,322	297,000	312,000
65–69	2,224,065	2,616,745	2,224,273	2,467,877	68,563	78,521	225,000	260,000
70+	3,891,013	5,349,657	3,181,938	4,188,300	129,683	143,179	379,000	437,000
Total	**190,755,799**		**237,641,326**		**12,496,300**		**20,653,000**	

Source: United Nations (2012)

Table A4.3 Estimated population by age and sex for Italy, Japan, Australia and Canada: 2010

	Italy		Japan		Australia		Canada	
Age group	Males	Females	Males	Females	Males	Females	Males	Females
0–4	1,462,208	1,382,445	2,767,000	2,626,000	749,611	711,477	963,446	914,714
5–9	1,456,435	1,377,855	2,868,000	2,731,000	701,203	664,544	928,386	928,386
10–14	1,441,015	1,357,979	3,038,000	2,891,000	719,763	684,015	992,060	992,060
15–19	1,527,934	1,439,427	3,100,000	2,956,000	770,614	730,396	1,140,059	1,140,059
20–24	1,591,662	1,526,730	3,483,000	3,312,000	849,732	799,927	1,214,756	1,214,756
25–29	1,766,298	1,738,236	3,796,000	3,629,000	844,968	820,295	1,212,138	1,212,138
30–34	2,118,717	2,084,846	4,230,000	4,107,000	768,096	765,947	1,150,687	1,150,687
35–39	2,418,876	2,392,841	4,915,000	4,797,000	802,527	813,524	1,153,074	1,153,074
40–44	2,483,227	2,474,690	4,366,000	4,304,000	772,142	780,524	1,219,633	1,219,633
45–49	2,307,605	2,343,039	3,988,000	3,953,000	781,143	793,910	1,405,280	1,405,280
50–54	1,988,299	2,058,869	3,821,000	3,836,000	726,885	743,491	1,306,535	1,306,535
55–59	1,803,643	1,901,099	4,330,000	4,423,000	655,993	670,020	1,128,182	1,128,182
60–64	1,782,949	1,906,934	4,813,000	5,026,000	604,200	608,337	965,010	965,010
65–69	1,493,194	1,669,995	3,944,000	4,323,000	450,385	459,127	712,574	712,574
70+	3,645,341	5,397,940	8,600,000	12,478,000	926,992	1,172,610	1,425,462	1,748,382
Total	**60,340,328**		**127,451,000**		**22,342,398**		**34,108,752**	

Source: United Nations (2012)

Appendix 4.2 Population Pyramids

The following pyramids are based on the percentages of the total population calculated from the data shown in Table A4.2.

The authors wish to acknowledge the assistance of Lucky Tedrow of Washington State University in preparing the population pyramids included in this chapter.

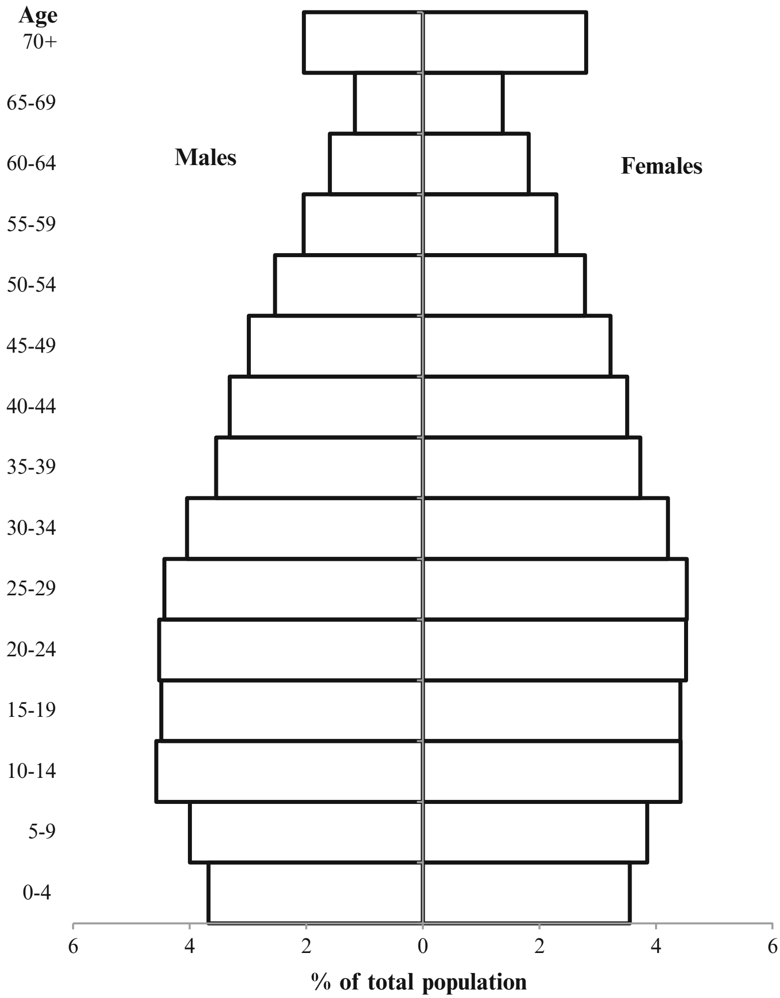


Fig. A4.1 Population pyramid for Brazil, 2010

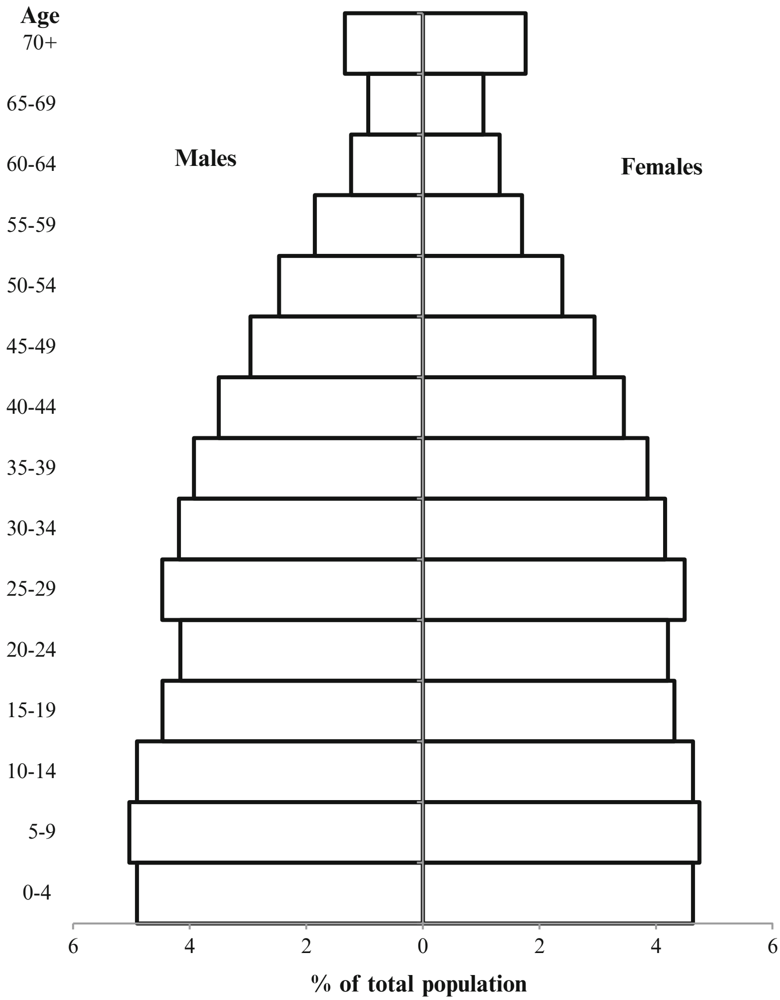


Fig. A4.2 Population pyramid for Indonesia, 2010

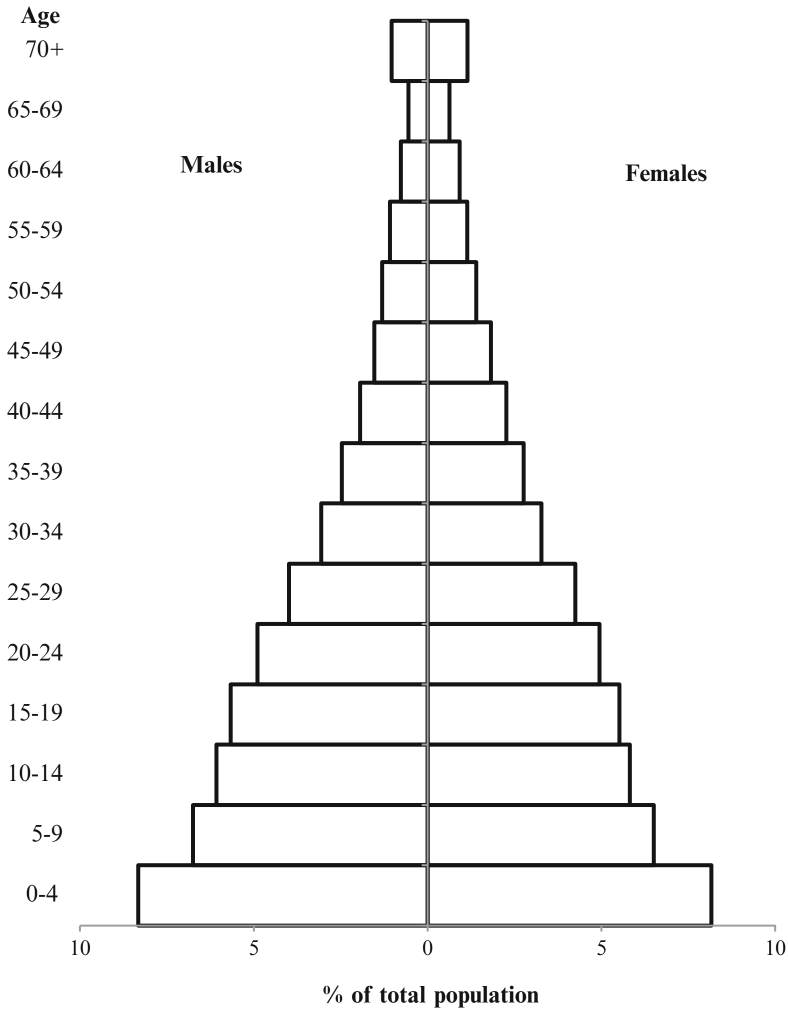


Fig. A4.3 Population pyramid for Senegal, 2010

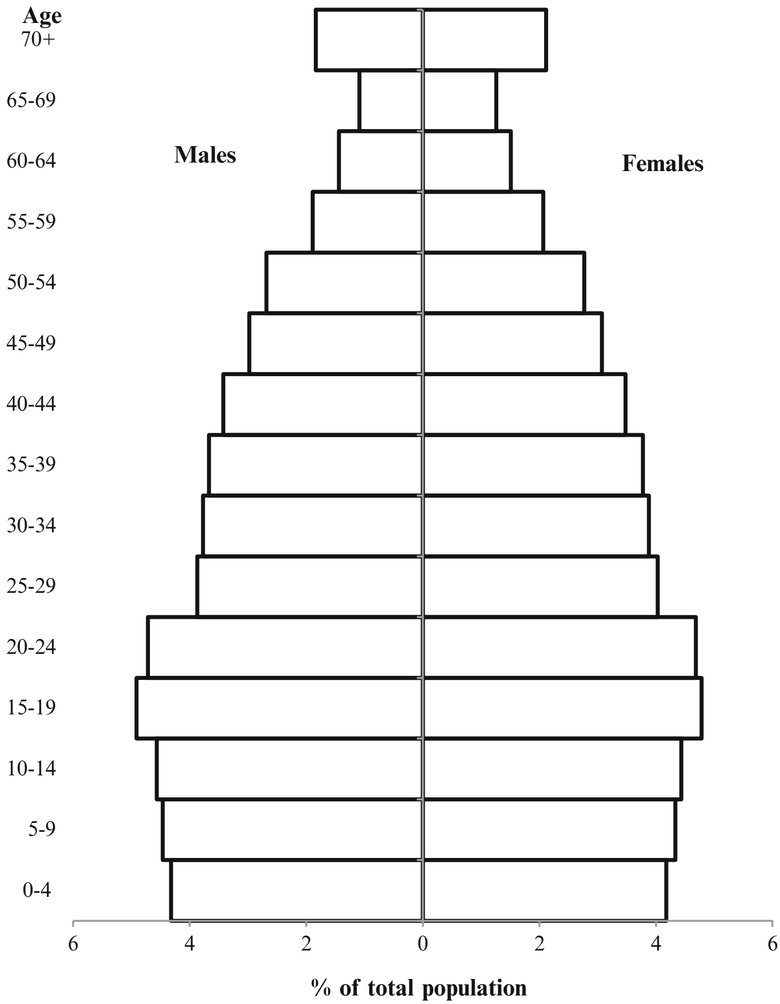


Fig. A4.4 Population pyramid for Sri Lanka, 2010

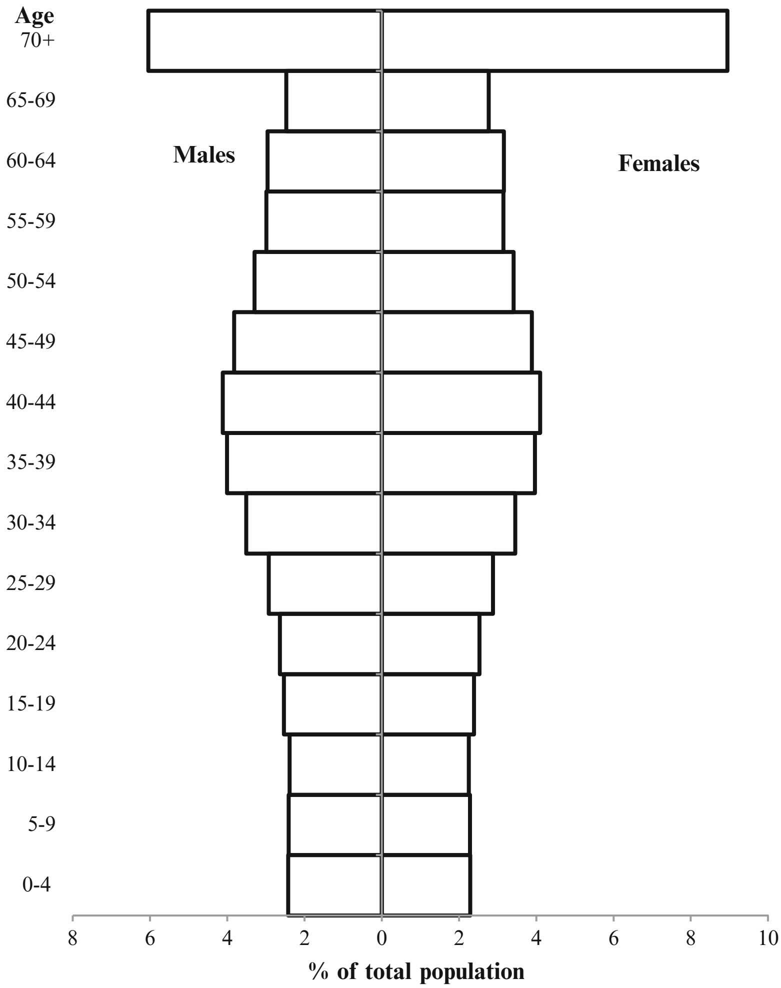


Fig. A4.5 Population pyramid for Italy, 2010

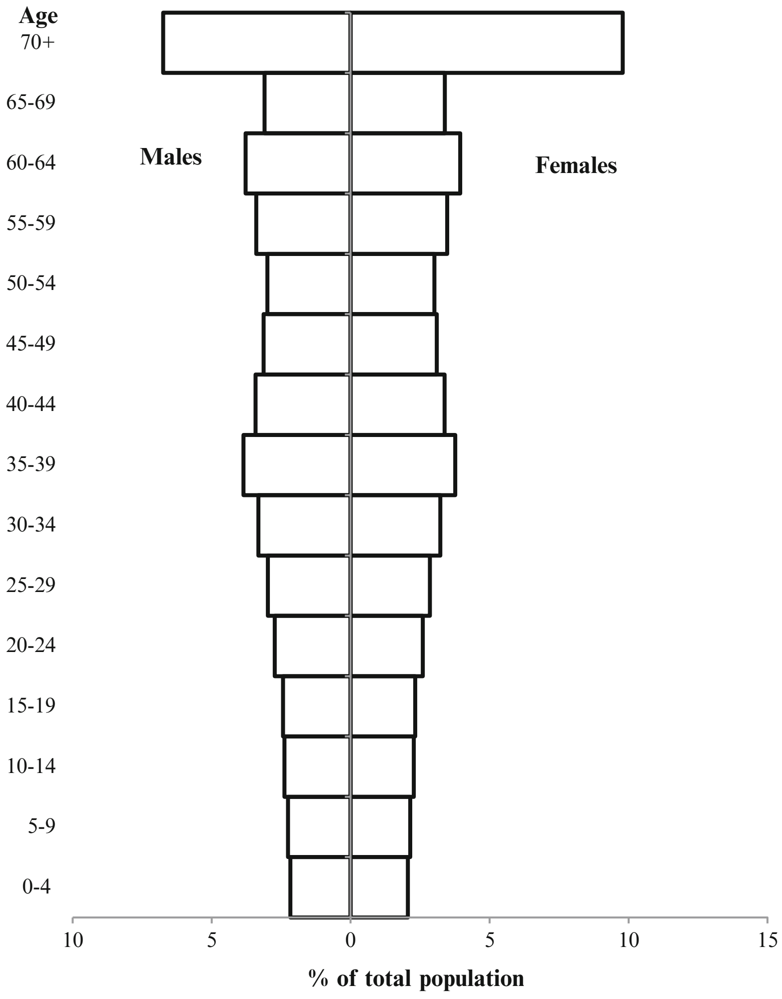


Fig. A4.6 Population pyramid for Japan, 2010

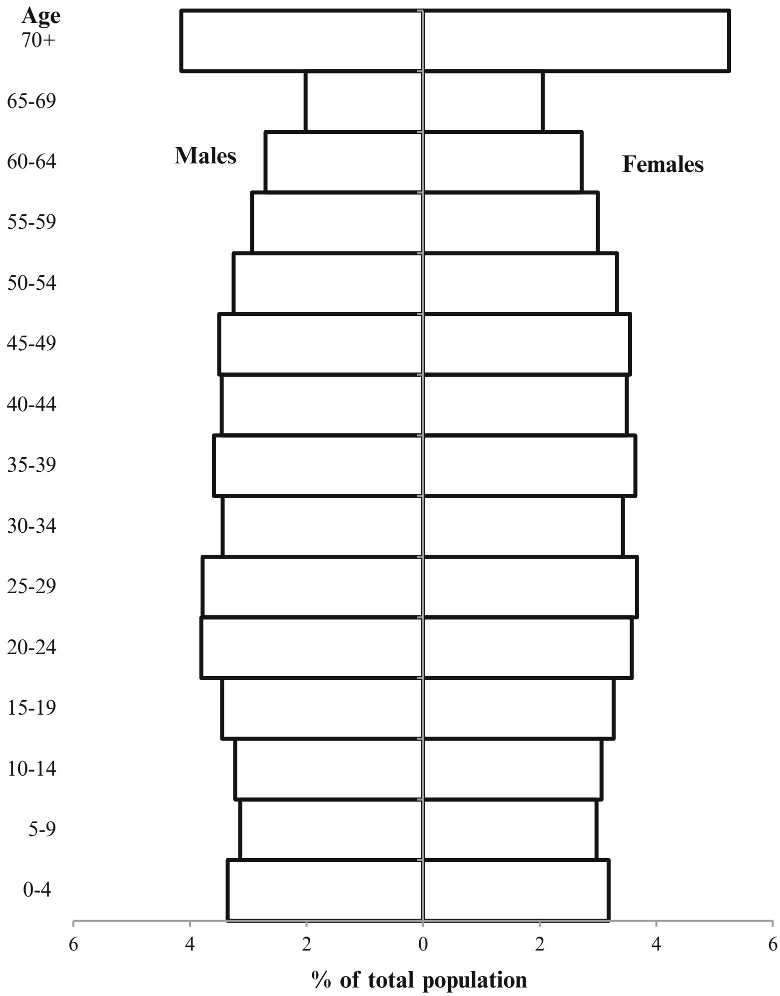


Fig. A4.7 Population pyramid for Australia, 2010

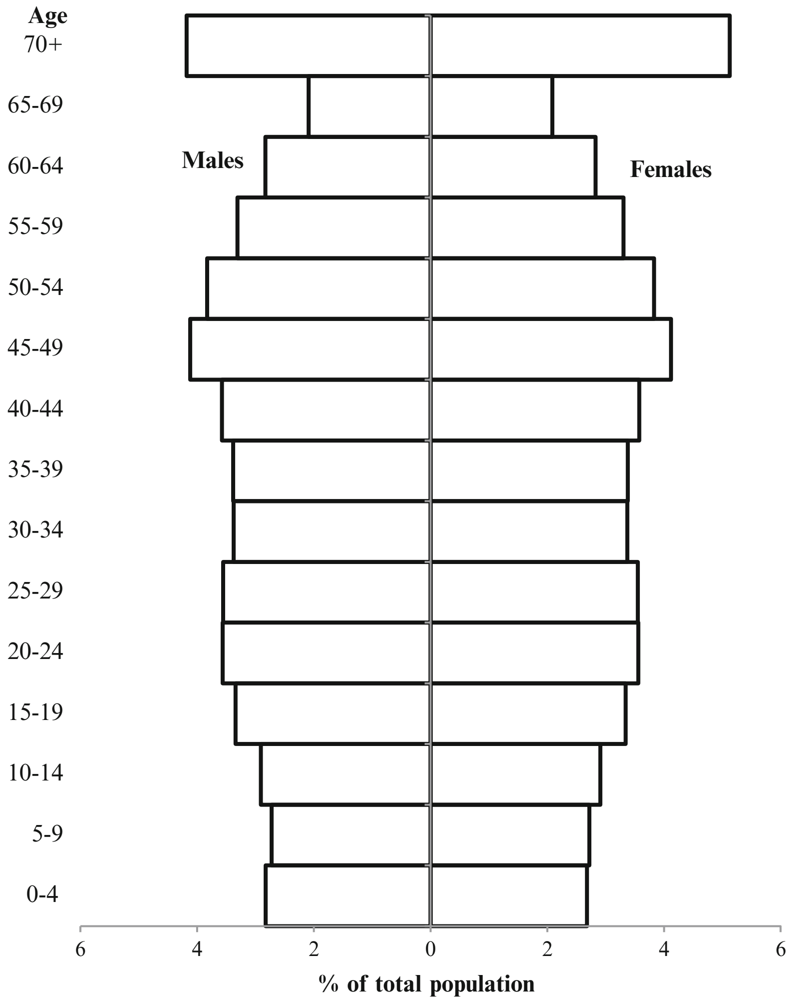


Fig. A4.8 Population pyramid for Canada, 2010

References

Ahmad, O. B., Boschi-Pinto, C., Lopez, A. D., Murray, C. J. L., Lozano, R., & Inoue, M. (2001). *Age standardization of rates: A new WHO standard* (GPE Discussion Paper Series: No. 31). Geneva: World Health Organization. http://www.who.int/healthinfo/paper31.pdf. Accessed Feb 2013.

Australia. (2010). *2009–10 year book Australia* (Catalogue No. 1301.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/AC72C92B23B6DF6DCA257737001B2BAB/$File/13010_2009_10.pdf. Accessed Feb 2013.

Australia. (2012). *September quarter 2011 Australian demographic statistics* (Catalogue No. 3101.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/FEB6A31EE56FD01ACA2579CF000F9B57/$File/31010_sep%202011.pdf. Accessed Feb 2013.

Benewick, R., & Donald, S. (2009). *The state of China atlas: Mapping the world's fastest-growing economy* (3rd ed.). Los Angeles: The University of California Press.

India. (2011). *Census 2011. Provisional population totals, paper 1 of 2011 India series 1. Chapter 7, density of population*. New Delhi: Office of the Registrar General & Census Commissioner. http://censusindia.gov.in/2011-prov-results/prov_results_paper1_india.html. Accessed Feb 2013.

India. (undated). *India statistics – India an overview. Chapter 2. Area and population*. New Delhi: Ministry of Statistics and Programme Implementation. http://mospi.nic.in/Mospi_New/site/India_Statistics.aspx?status=1&menu_id=14. Accessed Feb 2013.

Swanson, D. A., & Stephan, G. E. (2004). Glossary. In J. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 751–786). San Diego: Elsevier Academic Press.

United Kingdom. (2009). *Mortality statistics: Deaths registered in England and Wales (Series DR), 2008*. Newport: Office for National Statistics. http://www.ons.gov.uk/ons/rel/vsob1/mortality-statistics–deaths-registered-in-england-and-wales–series-dr-/2008/index.html. Accessed Feb 2013.

United Nations. (2009). *Demographic yearbook, 2008*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dyb2008.htm. Accessed Feb 2013.

United Nations. (2012). *Demographic yearbook 2009–10*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dyb2009-2010.htm. Accessed Feb 2013.

United States. (2007). *The 2007 statistical abstract*. Washington, DC: Census Bureau. www.census.gov/compendia/statab/2007/2007edition.html. Accessed Feb 2013.

Chapter 5
Fertility

5.1 Purpose

This chapter is concerned with an examination of concepts and measurement of cross sectional and longitudinal fertility. It also deals with methods to assess the influence and impact of population and female characteristics on fertility. Examples of different methods of analysis are given with data from various countries.

5.2 Perspectives on Fertility

As noted in Chap. 2, demographers like to distinguish between fecundity and fertility. *Fecundity* is the biological capacity of women to have children, while *fertility* refers to the number of children women have. The number of children of concern may be those born during a single year (period) and are often related to the reference population, in a cross-sectional approach. Or they may be the number of children born during the reproductive period of women, from a longitudinal perspective. Experience indicates that there is a close association between the age of women and the number of births they have. Therefore, much of demographic analysis of fertility tends to be related to women of varying ages during their reproductive period.

In addition to age, a number of other factors tend to be associated with fertility. There are some factors that relate to family formation and planning such as intended family size and age of first child, access to contraception, unwanted fertility, gender preferences, success with previous offspring and other factors known as the *proximate determinants of fertility* (Bongaarts 1978). While other factors tend to be associated to socio-economic (*evolutionary*) factors such as education, occupation, women participation in the formal labour force, backward intergenerational effects of parents depending on children for security in old age, trade-offs between quantity and quality of children (see for example: Becker 1960; Strassmann & Gillespie 2002). The demographic measures in this chapter are related to the first set of factors.

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_5,

5.3 Cross-Sectional Fertility Rates

5.3.1 Crude Birth Rates

A simple measure of fertility is the *crude birth rate*. It is defined as:

$$CBR_{t\to t+n} = \frac{B_{t\to t+n}}{\overline{P}_{t\to t+n}} \tag{5.1}$$

$CBR_{t\to t+n}$ is the crude birth rate during the period t to $t + n$, $B_{t\to t+n}$ and $\overline{P}_{t\to t+n}$ are the number of live births and the average population during the same period. An estimate of the average population during the period is usually the mid-period population ($P_{t+0.5*n}$). This rate is generally expressed per 1,000 people and is calculated on an annual basis, then n equals 1. Accordingly, Eq. (5.1) can be re-written as:

$$CBR_{t\to t+1} = \frac{B_{t\to t+1}}{P_{t+0.5}} \tag{5.2}$$

$CBR_{t\to t+1}$ is the crude birth rate during the 1-year period t, and $B_{t\to t+1}$ is the number of live births during the same year and $P_{t+0.5}$ is the midyear population.

All rates mentioned subsequently in this book are on an annual basis and expressed per 1,000 people, unless specified otherwise.

An example of the estimation of the crude birth rate can be given with 2007 data for Malaysia (United Nations 2008):

$$\text{Number of live births during the year 2007} = B_{2007} = 470{,}900$$

$$\text{Population in mid-2007} = P_{2007} = 27{,}720{,}000$$

$$CBR_{2007} = \frac{470{,}900}{27{,}720{,}000} = 0.0170 \text{ or } 17.0 \text{ births per } 1{,}000 \text{ population.}$$

It is apparent that this crude rate does not take into consideration population characteristics such as age and sex.

5.3.2 Age-Specific and General Fertility Rates

Age-specific fertility rates are useful in translating the relationship between age and fertility.

$$f^i_{t\to t+1} = \frac{B^i_{t\to t+1}}{W^i_{t+0.5}} \tag{5.3}$$

Table 5.1 Estimation of age-specific fertility rates: Japan, 2009

	Live births 2009			
Age (years)	Actual	Adjusted[c]	Female population mid-2009	Age-specific fertility rate per 1,000 females
(1)	(2)	(3) = (2)*k	(4)	(5) = (3)*1,000/(4)
15–19	14,687[a]	14,687	2,974,000	4.94
20–24	116,808	116,809	3,404,000	34.32
25–29	307,765	307,767	3,688,000	83.45
30–34	389,793	389,795	4,286,000	90.95
35–39	209,706	209,707	4,776,000	43.91
40–44	31,270[b]	31,270	4,243,000	7.37
Unknown	6	...	...	...
All ages 15–44	1,070,035	1,070,035	23,371,000	45.78

Source: United Nations (2008), Table 10 for column (2) and Table 7 for column (3)

Notes: [a] Includes births to females less than 15 years of age

[b] Includes births to females 45 years of age and over

[c] Births with unknown ages of the mothers were distributed on a *pro-rata* basis: $k = 1.000005607$ (Box 5.1)

$f^i_{t\to t+1}$ is the age-specific fertility rate at age i during the year t, $B^i_{t\to t+1}$ is the number of live births during the same year to females of age i, and $W^i_{t+0.5}$ is the midyear female population aged i. Although some females under 15 years of age and 45 years or older give birth, the number of such women tends to be small and the reproductive period is usually taken as between the ages of 15 and less than 45 years.

The *general fertility rate* is defined as:

$$GFR = \frac{\sum_{i=15}^{i=44} B^i_{t\to t+1}}{\sum_{i=15}^{i=44} W^i_{t+0.5}} \tag{5.4}$$

GFR is the general fertility rate, $B^i_{t\to t+1}$ is the number of live births to females aged i during the year t, and $W^i_{t+0.5}$ is the midyear female population aged i.

To illustrate the estimation of age-specific fertility rates, relevant statistics from Japan in 2004 (United Nations 2008) have been used in Table 5.1.

The general fertility rate for all females of reproductive age in Japan during 2009 was:

$$GFR = \frac{1,070,035}{23,371,000}$$
$$= 0.04578 \text{ or } 45.78 \text{ births per } 1,000 \text{ females } 15-44 \text{ years of age.}$$

Box 5.1 Distribution of Unknown Characteristics

In a population of X persons, z cases may have an unknown characteristic. To overcome this, a device is to distribute these cases among those with the known characteristic on a *pro-rata* basis. The pro-rata distribution multiplies a factor k by the number of persons in each category of known cases:

$$k = \frac{X}{X - z}$$

This method implies that z cases with unknown characteristic can be distributed on a proportional basis among those cases where characteristic is known.

5.4 Cross-Sectional and Longitudinal Fertility

The two basic approaches mentioned in Chap. 4 – cross-sectional and longitudinal – can be followed in the analysis of fertility, and are illustrated using the fertility rates for Australia, 1975–2010 (Table 5.2).

Each column of Table 5.2 refers to live births during a given year and females born in different years. For instance, the first two figures in the 1975 column (7 and 19) are based on births in that year to females who were 15 and 16 years of age and were born in 1960 and 1959. The column for the year 1975 has the births in that year to females born between 1960 and 1930. This follows the cross-sectional approach. Most demographic rates are estimated using this approach.

A longitudinal approach would follow the fertility rates diagonally. For example, in Table 5.2, rates marked with an asterisk (*) refer to the 1960 birth-cohort of women, and those marked with a hash (#) are the rates for the 1978 birth-cohort. Rates on these diagonals are those of females who were born in a given year, 1960 and 1978 respectively. The examination of cross-sectional fertility rates for a number of years (or time periods) in a diagonal way helps the study of the fertility experience of a particular cohort. This follows the longitudinal approach.

It is apparent from Table 5.2 that the 1960 birth-cohort entered the reproductive period at age 15 in 1975 and by 2004 they had reached the end of that period. Therefore, their entire fertility experience is available. This is not the case for the 1978 birth-cohort. They were only 32 years of age in 2010. They had another 12 years before they finished their reproductive life. Thus, the record of their fertility experience is incomplete.

The shapes of the curves of cross-sectional and cohort fertility age-specific rates are similar as seen in Fig. 5.1. The figure shows that the cross-sectional 2009 fertility rates (second last column of Table 5.2) were somewhat lower than those for the 1960 birth-cohort of females, but the shape of the two curves is similar. The lower age-specific fertility rates for 2009 births had a peak at an older age than those for the 1960 birth-cohort.

Table 5.2 Age-specific fertility rates, births per 1,000 females per year: Australia, 1975–2010

Age	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
15	7*	7	6	6	5	5	5	5	5	4	4	4	4	3	4	4	4	3
16	19	16*	14	13	13	12	12	12	12	10	10	9	9	9	9	9	9	9
17	38	32	30*	27	26	25	25	25	24	22	21	20	19	19	19	21	20	19
18	60	52	46	44*	40	40	40	39	38	35	34	33	31	29	29	31	32	31
19	80	73	67	61	57*	54	56	54	53	47	47	46	43	42	40	42	43	44
20	100	94	86	80	75	71*	71	70	69	61	62	59	56	55	52	54	51	53
21	118	112	107	101	94	89	89*	86	86	76	77	71	68	65	64	66	62	61
22	134	130	122	119	111	110	109	106*	103	94	95	89	82	81	78	78	75	73
23	153	146	142	134	128	128	130	121	121*	112	114	105	100	95	93	94	89	89
24	166	160	153	147	141	140	141	140	137	129*	128	123	116	111	107	108	102	103
25	170	166	160	157	153	150	151	150	151	139	142*	135	133	128	124	122	117	116
26	166	163	159	154	153	150	154	151	153	147	151	147*	140	137	136	135	130	127
27	151	152	151	148	152	150	148	148	154	148	154	148	145*	141	141	145	137	137
28	135	136	138	137	135	138	142	142	144	138	149	144	144	143*	139	147	140	140
29	124	117	120	123	123	121	127	131	133	129	134	136	137	136	138*	139	135	140
30	105	104	101	104	105	108	109	115	113	113	124	119	123	126	126	131*	126	131
31	84	85	86	83	87	87	93	94	97	97	105	106	104	109	111	117	115*	118
32	72	68	71	71	70	74	76	79	78	81	88	87	90	91	96	104	100	105*
33	58	55	56	56	56	58	62	63	67	64	71	72	75	77	80	86	87	92
34	47	44	47	46	47	47	49	51	51	51	58	58	60	63	66	69	71	75
35	39	37	35	36	37	36	39	38	40	40	45	45	48	50	52	58	56	62
36	31	28	29	28	29	29	29	31	30	31	33	35	35	38	42	43	46	47
37	24	24	23	21	21	22	22	23	22	22	26	26	27	28	30	32	33	36
38	20	17	18	17	15	16	18	17	18	17	17	19	19	21	22	24	26	26
40	15	13	13	13	13	12	12	13	12	12	13	13	14	15	16	16	18	19
41	12	10	9	9	9	9	9	9	9	8	9	9	9	10	10	12	11	14
42	8	7	6	6	6	5	6	6	6	5	5	6	6	6	7	7	7	8
43	6	5	5	4	4	4	3	4	4	4	4	3	4	4	4	4	5	5
44	4	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	3	3
45	2	2	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2

(continued)

Table 5.2 (continued)

Age	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
15	4#	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3	3
16	9	9#	9	9	8	8	8	8	7	7	7	7	7	6	6	7	7	7
17	19	19	19#	20	18	18	17	16	16	15	14	15	14	13	14	15	15	13
18	29	28	28	29#	29	27	27	25	25	25	22	23	22	22	22	24	23	21
19	42	42	41	39	40#	39	39	37	37	36	34	32	32	33	34	36	34	33
20	52	50	48	48	47	45#	46	43	44	42	40	39	39	38	42	42	41	39
21	59	58	56	53	52	51	52#	51	49	49	46	45	44	44	48	49	46	45
22	69	67	65	63	61	60	59	57#	58	55	53	52	50	51	54	57	54	52
23	82	78	77	73	71	69	68	67	66#	63	61	61	58	58	62	61	60	59
24	96	94	86	86	81	80	77	77	75	74#	72	68	68	67	71	72	68	68
25	113	108	102	96	94	92	89	89	85	82	82#	80	79	78	83	81	80	77
26	127	121	115	110	104	102	99	97	95	95	94	92#	91	90	94	95	92	89
27	132	130	126	122	118	111	109	110	105	107	102	102	103#	102	105	105	104	101
28	139	134	133	129	125	124	119	118	114	114	113	113	114	113#	120	117	114	114
29	138	136	134	131	130	128	126	122	121	122	120	121	123	123	127#	127	122	122
30	133	129	130	126	128	126	124	125	118	123	122	123	125	126	131	131#	128	125
31	120	120	120	121	119	120	120	121	119	119	123	123	127	129	136	136	131#	131
32	106	108	110	108	109	110	110	112	112	115	114	118	121	125	131	133	127	130#
33	90*	92	94	95	96	97	99	102	102	104	106	108	113	117	121	124	121	120
34	76	78*	79	81	83	84	87	88	88	93	96	99	101	106	113	113	112	111
35	61	64	66*	67	69	70	72	75	75	81	82	86	88	92	99	103	99	100
36	49	51	52	53*	56	57	58	61	61	64	67	70	75	77	81	85	84	86
37	36	40	40	42	42*	43	45	47	48	51	53	55	59	61	65	68	69	70
38	28	28	30	32	33	32*	34	37	37	39	41	43	45	47	52	55	52	55
40	20	22	21	23	23	25	24*	26	27	29	30	32	33	35	39	41	40	41
41	14	14	15	15	16	17	18	18*	18	20	21	22	23	24	27	29	29	30
42	8	9	10	10	10	11	11	12	13*	13	14	14	15	16	18	20	20	20
43	5	6	6	6	6	7	7	7	8	8*	8	9	9	10	10	12	12	13
44	3	3	4	3	4	4	4	4	4	5	4*	5	5	5	6	7	7	7
45	1	1	2	2	2	2	2	1	1	1	1	1*	1	1	2	2	2	4

Source: Australia (2011)

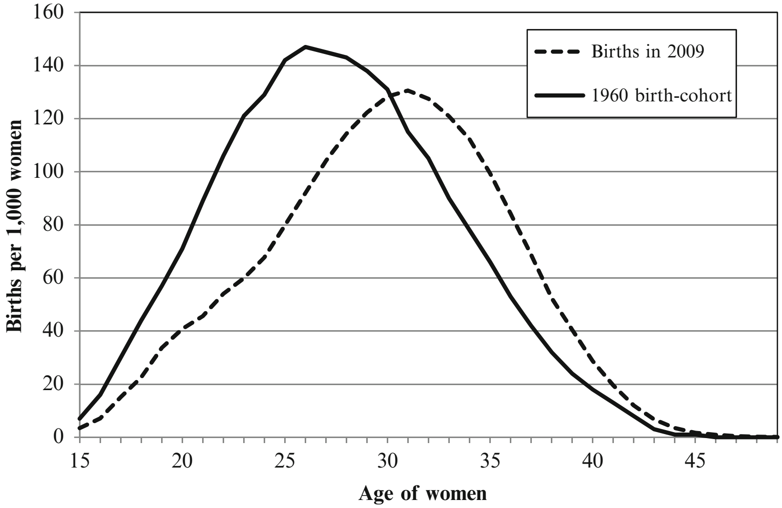


Fig. 5.1 Age-specific fertility rates in Australia, births in 2009 and of the 1960 birth-cohort of women (*Source*: Australia 2011)

5.5 Synthetic Measures of Fertility

5.5.1 Total Fertility Rate, Gross and Net Reproduction Rates

The incomplete record of fertility of the 1978 birth-cohort highlights one of the main problems with the longitudinal approach, when cohorts have not reached the end of their reproductive period. As this issue persists in the application of the longitudinal approach to demographic and non-demographic variables, some synthetic measures have been suggested. Such measures dealing with fertility are described below.

Synthetic measures of fertility provide estimates of the lifetime fertility of females, under a set of assumptions. Given a hypothetical group of 1,000 females on their 15th birthday, and assuming:

1. the group is closed to international migration
2. the group is subjected to the cross-sectional fertility rates at each age based on births occurring during a particular year (2009 in this case, as per the second last column of Table 5.2) during their reproductive period
3. no female dies before completing her reproductive period at age 45

it is posited that by the time this group of females reaches their 16th birthday they will have 3 children, 7 children during their 16th year bringing the total number of children to 10, and so on, as shown in Table 5.3.

It is apparent that this group of 1,000 females would have had, on the average, one child per female somewhere between their 31st and 32nd birthday, and that by the time

Table 5.3 Cumulative number of children by age of mother for a hypothetical group of 1,000 females aged 15 exactly subjected to the age-specific fertility rates of Australia in 2009 and assuming no mortality

By exact age	Cumulative number of children	By exact age	Cumulative number of children	By exact age	Cumulative number of children
16	3	26	431	36	1,581
17	10	27	523	37	1,665
18	25	28	627	38	1,734
19	48	29	741	39	1,786
20	82	30	863	40	1,826
21	123	31	991	41	1,855
22	169	32	1,122	42	1,875
23	223	33	1,249	43	1,887
24	283	34	1,370	44	1,894
25	351	35	1,482	45	1,896

Source: Table 5.2 (second last column)

they completed childbearing the original 1,000 females would have had 1,896 children. This estimate of lifetime fertility is called the *total fertility rate (TFR)*. Generally, the *TFR* is expressed as the average number of births per woman. In this example, the total fertility rate is 1.9 births per woman. This rate is defined as the average number of births per woman in a population subjected to a given set of age-specific fertility rates and assuming no mortality till the end of her reproductive period.

The *TFR* becomes the *gross reproduction rate (GRR)*, if only female births are considered in the estimation of age-specific fertility rates. It is defined as the average number of daughters per woman in a population subjected to a given set of age-specific fertility rates and assuming no mortality to the end of her reproductive period. Unless births are available by sex, another assumption needs to be added to the previous three assumptions. This is that the sex ratio at birth observed during the particular year continues to operate throughout the reproductive period of females.

The *GRR* can be adjusted for female mortality during the reproductive period. Such an adjusted rate is called the *net reproduction rate (NRR)*. It represents the average number of daughters per woman in a population subjected to a given set of age-specific fertility rates (daughters only) and probabilities of survival of females from birth to the average age at which mothers had their daughters. These probabilities imply a constant set of age-specific mortality rates to operate throughout the reproductive period of females. This replaces the third assumption regarding no mortality.

The gross and net reproduction rates give an indication of the replacement of the current female population in future. *NRR* of less than 1 indicates the female population is not replacing itself. In other words, the size of the daughters' generation is smaller than the size of the mothers' generation.

Two points are worth making:

- $TFR > GRR > NRR$
- GRR is approximately half the TFR.

The equations for the above fertility measures are as follows:

$$TFR_{t\to t+1} = \sum_{i=15}^{i=44} f^i_{t\to t+1} \tag{5.5}$$

$f^i_{t\to t+1}$ is the *fertility rate* for females aged *i,* during year *t*.

$$GRR_{t\to t+1} = \sum_{i=15}^{i=44} f^i_{t\to t+1} * S \tag{5.6}$$

S is the proportion of female births from all births. If $SRB_{t\to t+1}$ is the sex ratio at birth (Sect. 4.6.3), then $S = \dfrac{100}{100 + SRB_{t\to t+1}}$.

$$NRR_{t\to t+1} = R_0 = \sum_{i=15}^{i=44} f^i_{t\to t+1} * S * PS^{0\to i+0.5} \tag{5.7}$$

$PS^{0\to i}$ is the probability of survival of a female from age 0 (age at birth) to the age $i + 0.5$ being the average age at which her mother had her (see Box 5.2).

When the age of mother is given in age groups the Eqs. (5.5, 5.6, 5.7) have to include a multiplication factor equal to the width of the age intervals. The width of age intervals is usually 5 years. In that case, the multiplication factor would be 5. It would be 10 if the births related to women in 10-year age groups.

5.5.2 Mean Length of Generation

The average age of a woman at the birth of her daughter is called the *mean length of generation*. This is calculated as:

$$MLG = \frac{R_1}{R_0} = \frac{\sum_{i=15}^{i=44} \left\{ \left[f^i_{t\to t+1} * S * PS^{0\to i+0.5} \right] * (i + 0.5) \right\}}{R_0} \tag{5.8}$$

MLG is the mean length of generation during year *t*, and the remaining terms are as described in the previous section. Age *i* is from *i* to just under age *i + 1*. The average age is deemed to be the mid-point of the age group: *i* + 0.5.

Box 5.2 Probability of Survival by Age

The probability of survival of a female from exact age (a) to exact age (b) where $b > a$ is defined as $\frac{l_b^f}{l_a^f}$ where l_a^f and l_b^f stand for the survivors to exact age a and b in a female life table (see Chap. 7 for life table details). Similar probabilities can be calculated for males from a male life table.

An example of the computation of probabilities of survival is given below with an extract from the life table for Australian females based on mortality data for 2005–2007.

x	l_x	x	l_x	x	l_x	x	l_x
0	100,000	23	99,198	32	98,888	41	98,356
15	99,408	24	99,170	33	98,844	42	98,266
16	99,389	25	99,137	34	98,802	43	98,179
17	99,366	26	99,108	35	98,754	44	98,078
18	99,340	27	99,077	36	98,706	45	97,969
19	99,312	28	99,047	37	98,649	46	97,848
20	99,284	29	99,012	38	98,589	47	97,719
21	99,254	30	98,969	39	98,515	48	97,570
22	99,226	31	98,930	40	98,440	49	97,413

Source: Table A7.3

Note: x is exact age and l_x is the number of survivors at exact age x. These are out of 100,000 births, that is, survivors at exact age $x = 0$

Probability of survival (PS) for females:

- from 15 to 45 years of age $= \dfrac{97,969}{99,408} = 0.98552$
- from birth to age 30 years $= \dfrac{98,969}{100,000} = 0.98969$
- from birth to age group 25-29 years $= \dfrac{\frac{99,077+99,047}{2}}{100,000} = \dfrac{99,062}{100,000} = 0.99062$

  *This is the probability of survival from birth to age 27.5 (=25 + 0.5*5), being the average age for the 25–29 age group (as explained in Sect. 5.5.3).*
- from age group 15-19 to age 45 years $= \dfrac{97,969}{\frac{99,366+99,340}{2}} = \dfrac{97,969}{99,353} = 0.98607$

Two issues are worth noting in the above equation. First, the denominator is actually the same as Eq. (5.7) for NRR. Second, if the terms in square brackets and $(i + 0.5)$ are replaced by f_i and x_i they result in Eq. (3.2) in the estimation of the arithmetic mean in frequency distributions. Thus, Eq. (5.8) is equivalent to Eq. (3.2) and provides an estimate of the mean age of mothers at the birth of their daughters.

5.5.3 Synthetic Measures Using Grouped Data

The estimation of these synthetic measures using grouped data requires some changes to their specification. Age i is substituted by the corresponding age groups where i is the lower limit and n is the width of the age group interval. In addition, the width of the age interval is multiplied after summation in Eqs. (5.5, 5.6, 5.7). This is because all rates are annual average rates during the age interval and to compute the total rate for the age interval the average annual rate has to be multiplied by the width of the age interval. Accordingly, the equations for the estimation of grouped data become:

$$TFR_{t \to t+1} = n * \sum_{i=15}^{i=44} f_{t \to t+1}^{i \to i+n} \tag{5.5a}$$

$$GRR_{t \to t+1} = n * \sum_{i=15}^{i=44} f_{t \to t+1}^{i \to i+n} * S \tag{5.6a}$$

$$NRR_{t \to t+1} = R_0 = n * \sum_{i=15}^{i=44} f_{t \to t+1}^{i \to i+n} * S * PS^{0 \to (0.5n)} \tag{5.7a}$$

$$MLG_{t \to t+1} = \frac{R_1}{R_0} = \frac{n * \sum_{i=15}^{i=44} f_{t \to t+1}^{i \to i+n} * S * PS^{0 \to (0.5n)} * (i + 0.5n)}{R_0} \tag{5.8a}$$

The computations of the four synthetic measures are illustrated in Table 5.4 with Australian births and population in 2009 (Australia 2010a, b). Female age in 5-year age groups from 15 to 44 years is specified in Column (1). Column (2) gives the number of live births during 2009 adjusted for unknown ages. The small numbers of births before 15 years of age and after 44 are included in the 15–19 and 40–44 age groups. Column (3) states the estimated resident population in mid-2009. Column (4) has the estimated number of female births following the procedure outlined for Eq. (5.6). For example, the second figure in this column is 42,103*0.485967 = 20,461 (rounded to the nearest whole number). Columns (5) and (6) are the age-specific fertility rates for all births and female births only. These rates are expressed per 1,000 females per year. The denominator for both rates is provided by column (3). Sums of these columns are multiplied by 5 (years in each age group) to obtain the total fertility rate and the gross reproduction rate.

The total fertility rate gives the total number of live births a group of 1,000 females would have by the time they finish childbearing provided:

- the female population is closed to international migration

Table 5.4 Computation of the various synthetic measures of fertility: Australia, 2009

				Fertility rate per 1,000					
Age (years) $i \rightarrow i+5$	All births 2009[c]	Females mid-2009	Female births 2009[#]	All births $f_{2009}^{i\rightarrow i+5}$	Daughters $\hat{f}_{2009}^{i\rightarrow i+5}$	Mean age of woman i + 2.5	Prob. of survival age 0 to (x) $PS^{0\rightarrow i+2.5}$	Mortality adjusted fertility rate	Adjusted mean age of woman
(1)	(2)	(3)	(4) = S*(2)	(5) = (2)/(3)	(6) = (4)/(3)	(7)	(8)	(9) = (6)*(8)	(10) = (9)*(7)
15–19	12,130[a]	727,168	5,895	0.016681	0.008107	17.5	0.99353	0.008055	0.140963
20–24	42,103	782,583	20,461	0.053800	0.026145	22.5	0.99212	0.025939	0.583628
25–29	80,933	791,698	39,331	0.102227	0.049679	27.5	0.99062	0.049213	1.353358
30–34	93,107	751,566	45,247	0.123884	0.060204	32.5	0.98866	0.059521	1.934433
35–39	55,985	814,971	27,207	0.068696	0.033384	37.5	0.98619	0.032923	1.234613
40–44	11,480[b]	769,345	5,579	0.014922	0.007252	42.5	0.98223	0.007123	0.302728
Total	295,738	4,637,331	143,720	0.380210	0.184771	…	…	0.182774	5.549723
x 5	…	…	…	1.901050	0.923855	…	…	0.913870	27.748615
				= *TFR*	= *GRR*			= *NRR* = R_0	= R_1

Sources: Australia (2010a) for column (2) and Australia (2010b) for column 3

Notes: $\frac{R_1}{R_0} = MLG = 30.363854$ years

[a] Births to women less than 15 years of age were included

[b] Births to women 45 years of age and over were included

[c] 254 births with unknown age of mother were distributed on a *pro-rata* basis following the procedure in Box 5.1

$k = \frac{295,738}{295,738 - 254} = 1.000859607$ Column (2) shows the data after adjustment for unknown ages of mothers

[#] Births in Australia during 2009: 152,019 male and 143,719 female giving a sex ratio of 105.7752 and $S = \frac{100}{100 + 105.7752} = 0.485967$

- the female population experiences the 2009 age-specific fertility rates of Australia throughout their reproductive period, and
- no female dies before her 45^{th} birthday.

Under the above assumptions that group of 1,000 women are expected to have 1,901 children by the time they complete childbearing (*TFR*). The *GRR* reflects the number of daughters per 1,000 females under the above three assumptions plus a fourth condition that the sex ratio at birth remains constant at the 2009 level. As expected, the *GRR* of 924 daughters per 1,000 women is just under half of the *TFR*.

The average age of women in each age group is stated in column (7). The mid-point in the age group interval is assumed to be the average age. For instance, the age group 15–19 includes 5 years of life and not 4; that is why the mid-point is 17.5 and not 17. Column (8) contains the probability of survival from birth to the mid-point of each age group using the procedure outlined in Box 5.2, where the probability of survival from birth to age 25–29 is shown as an example. The reason behind the computation of these probabilities is the notion that to replace her mother a new-born daughter must survive to the age at which her mother gave birth to her. The figures in column (9) are obtained by multiplying the figures in column (6) by the corresponding ones in column (8). The sum of column (6) multiplied by 5 (the years in the class intervals) results in the *NRR*.

The *NRR* gives the total number of daughters a group of 1,000 women would have by the time they finish childbearing, under the same assumptions as those for *GRR* except that the *no mortality* assumption is replaced by the female population being subjected to the survival in the Australian female life table. In this example, the life table is based on 2005–2007 mortality. Given the low female mortality levels in Australia (98.6 % of 15 year old females survive to age 45, see Box 5.2), the use of this life table rather than the 2009 life table does not affect the results significantly. The *NRR* of 914 daughters per 1,000 women when compared with the *GRR* of 924 indicates that only 10 daughters (1.1 % of the total born) will die before reaching the age at which their mothers had them.

In column (10), the mean age of women in each age group is multiplied by the mortality adjusted fertility rates. The total of column (10) is divided by the sum of column (9). Alternatively, the sum of column (10) is multiplied by 5 and divided by *NRR*. This gives the *mean length of generation* (*MLG*). It is the average age of the mother at the birth of her daughter, under the same assumptions as those for the *NRR*. In this example, the *MLG* is estimated as 30.36 years.

5.5.4 Other Formulae for Synthetic Measures

If the value of *TFR* and the proportion of female births *S* are known the *GRR* can be estimated as:

$$GRR = S * TFR \tag{5.9}$$

If estimates of the *MLG* and *GRR* are known, the *NRR* can be computed as:

$$NRR = GRR * \frac{l^f_{MLG}}{100,000} \tag{5.10}$$

Information in Table 5.4 can be used to demonstrate the application of the above equations.

Applying Eq. (5.9):

$$GRR = 0.485967 * 1,901.05 = 923.85$$

The above estimate is the same as in Table 5.2.

To apply Eq. (5.10), the value of $l^f_{30.36}$ (*i.e.*, female survivors at age 30.36 years) is required. This can be estimated as follows:

$$l^f_{30.36} = l^f_{30} - \frac{l^f_{30} - l^f_{31}}{1} * 0.36 = 98,969 - \frac{98,969 - 98,930}{1} * 0.36 = 98,954.96$$

According to Eq. (5.10)

$$NRR = 923.85 * \frac{98,954.96}{100,000} = 914.2 \quad \text{(rounded to 914)}$$

This estimate is very close to the *NRR* estimated in Table 5.4.

5.6 Fertility Rates and Population Characteristics

Measures of fertility include fertility by the sex of the parent, parity progression ratios and children ever-born (an indicator of cohort fertility). Data on children ever-born are usually available from population censuses, these figures may be collected for all cases or on a sample basis. Other fertility measures are usually based on birth registration data or sample surveys.

Fertility rates tend to be estimated with females as the reference parents, because their reproductive period is well defined and it is only them who bear children. Fertility rates can also be computed for males, as long as the information about the father of the child is available. Figure 5.2 compares the fertility rates for females and males in Australia for the year 2009. The similarity of the two sets of rates is apparent. The median age of females was 33 years compared to 36 for males. Rates

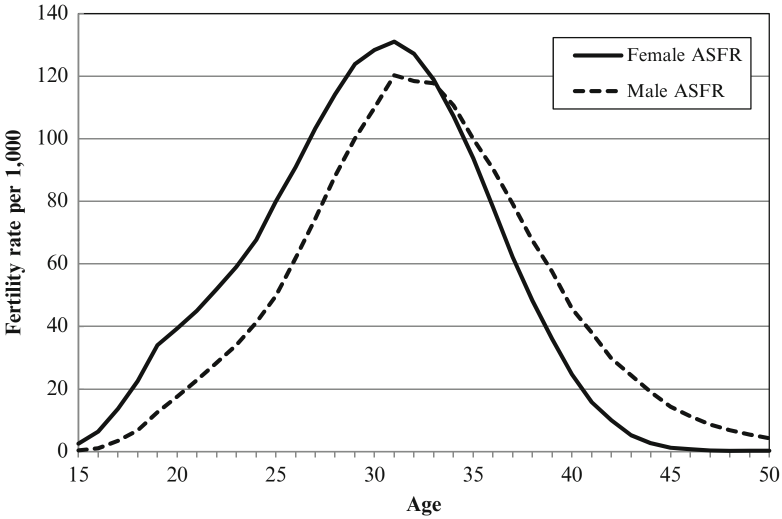


Fig. 5.2 Female and male age-specific fertility rates: Australia, 2009 (*Source*: Australia 2010a)

for males were based on 96.9 % of total births. No information about the father was known for the remaining 3.1 % of births.

The equation to calculate age-specific fertility rates by specific characteristic(s) of the population is:

$$f_{i,j} = \frac{B_{i,j}}{P_{i,j}} \tag{5.11}$$

$f_{i,j}$ is the fertility rate at age i for population with characteristic j, $B_{i,j}$ is the number of births ascribed to persons of age i and characteristic j, and $P_{i,j}$ is the number of persons of age i and characteristic j. These rates are usually computed for females and may relate to one or more characteristics such as marital status, ethnicity, religion or occupation. When data on these characteristics are not available by age, only the general fertility rates, described in Eq. (5.4), can be estimated. In some countries, the denominators for characteristic-specific rates may be available only in census years.

The number of births and women in 2010 for England and Wales (United Kingdom 2011a, b) in Table 5.5 gives an example of the estimation of female fertility rates according to their marital status characteristic. Fertility rates in column (6) are referred to as the *marital* or *nuptial fertility rates*, and those in column (7) as *non-marital* or *ex-nuptial fertility rates*.

The patterns of age-specific fertility rates of married and unmarried women in England and Wales are shown in Fig. 5.3. They reveal that the fertility rates of married women peaked at 20–24 years age and declined thereafter, while the

Table 5.5 Calculation of fertility rates of married and unmarried women in England and Wales, 2010

	Number of births to women		Estimated number of women		Fertility rates per 1,000 women		
Age group	Married	Unmarried	Married	Unmarried	Married	Unmarried	All
(1)	(2)	(3)	(4)	(5)	(6) = (2)/(4)	(7) = (3)*/(5)	(8)
15–19	1,680	38,900	5,723	1,358,040	293.6	28.6	29.8
20–24	33,809	102,945	112,200	1,743,511	301.3	59.0	73.7
25–29	107,490	90,042	457,435	1,387,343	235.0	64.9	107.1
30–34	139,869	59,858	773,514	944,611	180.8	63.4	116.2
35–39	79,315	34,444	1,057,991	824,107	75.0	41.8	60.4
40–44	16,064	9,427	1,262,836	806,978	12.7	11.7	12.3
45–49	1,058	566	1,305,431	736,264	0.8	0.8	0.8

Sources: United Kingdom (2011a) for columns (2) and (3) and United Kingdom (2011b) for columns (4) and (5)
Notes: The rates for all females in column (8) were estimated by dividing the sum of columns (2) and (3) in each age group (multiplied by 1,000) by the sum of columns (4) and (5) for the same age group. Strictly speaking, the rates given in the above table were based on *maternities* rather than live births. They are, therefore, somewhat under-estimates of the number of births as a maternity resulting in multiple births is counted as 1 irrespective of the number of births involved

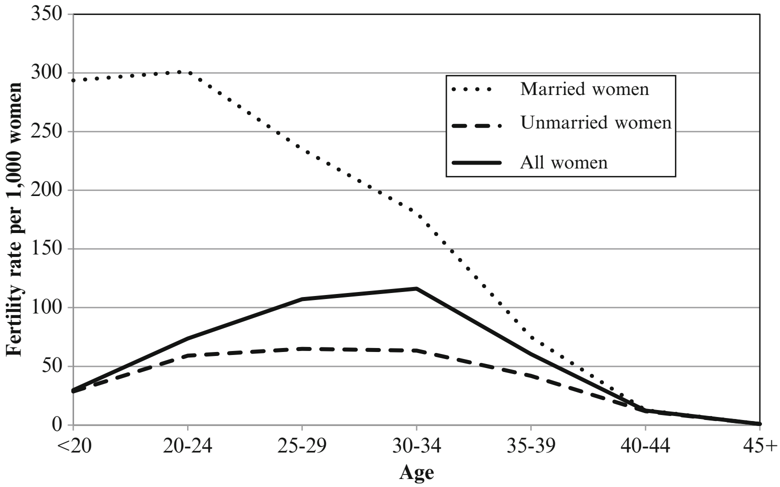


Fig. 5.3 Age-specific fertility rates for married, unmarried and all women in England and Wales, 2010 (*Source*: Table 5.5)

fertility of unmarried women rose to a plateau at 20–34 years age and then fell to lower levels. The substantial difference in the age-specific fertility rates of married and unmarried women at younger ages is an indication of the much higher levels of fertility among married women. The differences tend to diminish as fertility rates decline with age after the peak for all women at 30–34 years of age (Fig. 5.3).

5.7 Parity and Parity Progression Ratios

Information on births may be available by *parity* or the birth order defined as *equal to the previous births +1*. Accordingly, women of parity 1 are those for whom the current birth is the first, parity 2 means that the current birth is her second and so on.

Parity distributions for two countries with high fertility and another two with lower fertility in the 1990s are shown in Table 5.6. Caution needs to be taken regarding parity information, because the quality of birth registration may vary from country to country. However, if the under-registration of births is not parity specific, the relative percentages should give useful information.

Two points emerge from Table 5.6. First, the proportion of women with parities 1 and 2 was much higher in the countries with lower fertility (Canada and Germany with *TFRs* of 1.641 and 1.368). Second, the percentage of women with parities 3 and over was substantially larger in the two countries with higher fertility (Egypt and Malaysia with *TFRs* of 3.742 and 3.235) than that in the two countries with lower fertility. In general, as the fertility level of a population declines the proportion of women with lower parities (1 or 2) rises, while the proportion of higher parities (3 or over) declines. Consequently, the parity distribution of women can be seen as an indicator of their management of fertility.

Parity distributions are also available by age of women for selected years in the 1990s, in the publication used as the source for Table 5.6. However, the series is not long enough to study the parity distributions of various birth cohorts of women. This requires more detailed data on births by parity to mothers of a particular birth cohort(s). Such data are not easily available from registered births. Generally, studies of this nature are based on sample surveys in which women's detailed reproductive histories are recorded.

The *parity progression ratio* from parity n to $n + 1$ over a period of z years ($P_{t \to t+z}^{n \to n+1}$) is defined as:

$$P_{t \to t+z}^{n \to n+1} = \frac{W_{t+z}^{n+1}}{W_t^n} \tag{5.12}$$

Table 5.6 Parity distributions of women in four selected countries

			Parity (birth order) percentage distribution					
Country	Year	Births (000's)	1	2	3	4	5+	Total
Canada	1995	348.6	43.1	35.4	14.4	4.6	2.5	100.0
Germany	1997	666.4	45.6	36.9	12.3	3.3	1.8	100.0
Egypt	1995	1,604.8	36.6	23.7	16.1	10.4	13.3	100.0
Malaysia	1997	432.0	31.8	23.4	17.5	11.0	16.3	100.0

Source: United Nations (1999)

Note: Malaysia excludes data for Sabah and Sarawak

Table 5.7 Births by mother's age and parity: Canada, 1990 and 1995

Year (cohort)	Age of mother	Number of births: Parity 1	Parity 2	Parity 3	Parity 4
1990 (A)	20–24	48,555*	24,676+	6,646#	1,471
1995 (A)	25–29	53,101	42,471*	14,913+	4,270#
1990 (B)	25–29	68,647*	57,332+	20,917#	5,332
1995 (B)	30–34	35,556	45,209*	20,554+	6,408#

Source: United Nations (1999)

Notes: * Progression of each cohort from parity 1 in 1990 to parity 2 in 1995

+ Progression from parity 2 in 1990 to 3 in 1995

Progression from parity 3 in 1990 to 4 in 1995

W_t^n is the number of women of parity n at time t, and W_{t+z}^{n+1} those who continued on to have their $(n + 1)^{\text{th}}$ birth during the next z years.

The information in Table 5.6 (if the number of multiple births is ignored) indicates that of the 348,600 Canadian women who gave birth in 1995, the number of women who had their first birth in 1995 was 150,247 (=43.1 % of 348,600). Similarly, those who had their second birth in the same year was 123,404 (=35.4 % of 348,600). If it is assumed that the number of births and parity distribution in Canada remained constant over the years at the 1995 level, the estimated parity progression ratio from parity 1 to 2 is:

$$P_{1995}^{1\to 2} = \frac{123,404}{150,247} = 0.82.$$

A better approach is to look at the fertility experience of a particular age-cohort of women over time. Equation (5.12) can be used for this purpose. However, both the numerator and denominator must refer to the same cohorts, as illustrated in Table 5.7 for two age-cohorts of Canadian females: *cohort (A)* females born between 1970 and 1974, and *cohort (B)* those born between 1965 and 1969.

The first row in Table 5.7 gives the parity distribution of cohort (A) mothers who were aged 20–24 in 1990 and had a birth in that year. The second row displays the parity distribution of the same cohort of mothers 5 years later. Similar data for cohort (B) of mothers are given in the third and fourth rows. The progression of each cohort from parity 1 in 1990 to parity 2 in 1995 is indicated by (*), from parity 2 in 1990 to 3 in 1995 by (+), and from parity 3 in 1990 to 4 in 1995 by (#). Parity progression ratios for the two cohorts of women are displayed in Table 5.8.

The progression ratio of 0.875 from parity 1 to 2 for cohort (A) is given in column (2). It is the ratio of $\frac{42,471}{48,555}$ from Table 5.7. It means that 87.5 % of the women born in the period 1970–1974 who gave their first birth in 1990 went on to their second birth within the next 5 years. Other figures in columns (2)–(4) of Table 5.8 were estimated similarly.

Table 5.8 Parity progression ratios for two birth cohorts of females: Canada, 1990–1995

Mother's age-cohort: (year of birth)	Parity progression ratio		
	Parity 1–2	Parity 2–3	Parity 3–4
(1)	(2)	(3)	(4)
A: (1970–1974)	0.875	0.604	0.642
B: (1965–1969)	0.659	0.359	0.306

Source: Table 5.7

Usually, parity progression ratios are estimated from birth histories data from sample surveys, as relevant information is not available from birth registration data in many countries.

5.8 Standardized Fertility Ratios

Births by age of mother are not always available for different segments of a population, such as ethnic minorities or indigenous groups. If the female age distributions of such groups are known along with the total number of births and age-specific fertility rates of a standard population (usually the whole population to which the group belongs), it is possible to use the indirect standardization procedure described in Sect. 4.9.3 to assess differences in fertility between such groups and the population at large.

As an illustration, 2006 census data for the Indigenous female population of Australia is given in Table 5.9, along with the age-specific fertility rate for all Australian females and the actual number of Indigenous births registered in 2006 (Australia 2007). Although the census was conducted on 6 August 2006, column (3) has not been adjusted to align it with mid-2006. The Indigenous population of Australia, according to the census, consisted of all persons who identified themselves as Aboriginal or Torres Strait islanders when the census was taken.

Using the denominator of Eq. (4.28), the expected number of births to Indigenous females is estimated in column (4) assuming that they experienced the fertility rates of all Australian females. This gave a total of 6,122 expected births. The indirect standardized fertility ratio 2.041 is obtained by dividing the actual number of births (12,496) by the estimated expected number. This ratio indicates that Indigenous fertility was twice that of the national average.

Following Eq. (4.29), the indirect standardized birth rate for the Indigenous population of Australia comes to 26.1 (= 2.041*12.8) in comparison with the crude birth rate of 12.8 for the whole population of Australia. This estimate of 26.1 is close to the crude birth rate of 27.5 per thousand of the population of Indigenous people (455,016) in the 2006 census.

Table 5.9 Calculation of the standardized fertility ratio for Indigenous females using fertility rates for all females as the standard: Australia, 2006

Age (years)	Age-specific fertility rates Australia 2006 (standard)	Number of Indigenous females 2006 census	Expected number of births in 2006, if Indigenous females experienced the fertility rates for Australia in 2006
(1)	(2)	(3)	(4) = (2)*(3)
15–19	0.0154	23,626	364
20–24	0.0516	18,849	973
25–29	0.1008	15,899	1,603
30–34	0.1201	16,475	1,979
35–39	0.0633	16,480	1,043
40–44	0.0113	14,249	161

Sources: Australia (2007) for column (2) and actual Indigenous births; Australia (2013) for column (3), see Box 5.3 for details

Notes: Expected number of Indigenous births in 2006 = 6,123

Actual Indigenous births registered in 2006 = 12,496

$$\text{Standardized fertility ratio} = \frac{\textit{Actual}}{\textit{Expected}} = \frac{12,496}{6,123} = 2.041$$

Box 5.3 Obtaining Census Data From the Australian Bureau of Statistics

The link to the website is: www.abs.gov.au/websitedbs/censushome.nsf/home/Census.

This website can be used to obtain selected data from the 2006 and 2011 censuses for Australia, its States/Territories or any local government area.

- Click on *Data & analysis* in the left hand corner.
- On the right hand side the 2nd box is for *Community Profiles Search*. Select 2011 or 2006 from the drop down menu and type the name of the area. For example, typing Australia will result in 11 listings. Select the 1st one: *Australia (AUST)*. Typing the name of a locality, say Ryde, will result in 15 listings. Select the 14th one: *Ryde (C), NSW Local Government Area (LGA)*.
- After selecting the locality, click on *Go*.
- Six profiles will be offered: *Basic Community Profile*; *Aboriginal and Torres Strait Islander Peoples (Indigenous) Profile*; *Time Series Profile*; *Place of Enumeration Profile*; *Expanded Community Profile*; and *Working Population Profile*. They can be downloaded, unzipped and saved in the Excel format. Each profile has tables giving data on a variety of characteristics such as age, sex, and marital status.
- Note that the *Time Series Profiles* provide data for the selected census and the two previous ones. The three datasets in this profile use the same geographic boundaries for the area.

5.9 International Comparisons of Fertility

Age-specific fertility rates for Greece, Indonesia and Zambia (United Nations 2010) are shown in Fig. 5.4. A feature of this figure is that while the general shape of the fertility schedules is quite similar, the overall levels are substantially different leading to different total fertility rates: average of 1.3 births per woman in Greece, 2.6 in Indonesia and 6.2 in Zambia. The peak in fertility occurred at a much older age (30–34 years) in Greece with a lower *TFR* compared to the peaks in Indonesia and Zambia (at age 20–24) with higher *TFRs*.

These countries come from different parts of the world (Europe, Asia and Africa) and are at different stages of socio-economic development. There are a number of demographic, economic and social factors which influence the fertility of a population. Every year the United Nations Development Programme (UNDP) produces a report on human development that incorporates data on a variety of demographic, economic and social variables for its member countries (UNDP 2011). Two such variables in Table 5.10 indicate that the group of countries that ranked higher on the human development index in 2011 had, on average, lower fertility and higher gross national income per capita.

The correlation coefficient (r) (Sect. 3.9) was computed from the data in the UNDP report. The correlation between the *TFRs* and income per capita was -0.489 corroborating the findings of Table 5.10. However, the value of r between *TFRs* and the ranking on the human development index (a composite indicator of income and wellbeing) was substantially higher at -0.793. This indicates that factors other than income are important determinants of fertility.

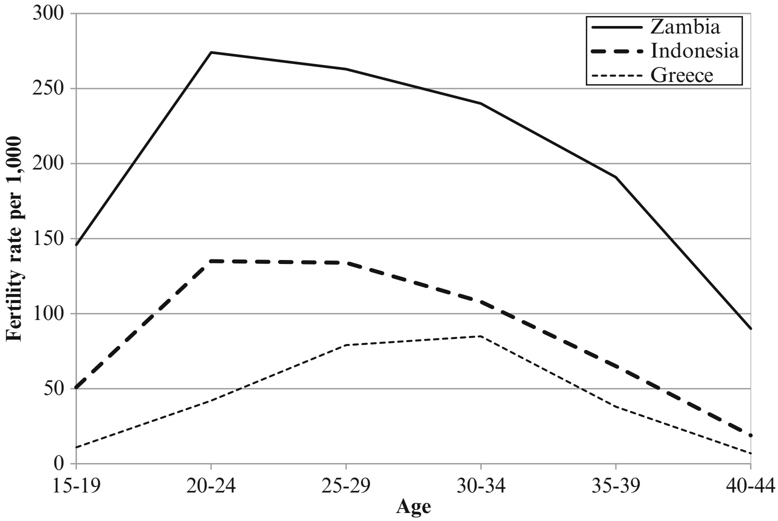


Fig. 5.4 Age-specific fertility rates for Greece, Indonesia and Zambia, 2006 (*Source*: United Nations 2010)

Table 5.10 Total fertility rates and income per capita for 180 countries by ranking in the Human Development Index, 2011

Human Development Index ranking	No. of countries	Total fertility rate 2011	Average gross national income per capita, 2011 US$ (PPP 2005)
Very high	45	1.7	$31,511
High	43	2.0	$11,584
Medium	46	2.8	$4,947
Low	46	4.6	$1,383

Source: UNDP (2011)
Note: PPP = Purchasing Power Parities are expressed in US$ for 2005. PPPs measure the domestic purchasing power of national currencies in relation to a basket of goods and services using the United States in 2005 as the standard. It is different from income expressed in currency exchange rates that reflect the value of currencies in international trade

As noted earlier, fertility is influenced by cultural, economic and social factors. Davis and Blake (1956) listed 11 intermediate variables that influence fertility. Among them the more important ones were marital status, age at which women get married, use of birth control, foetal mortality and incidence of abortion. Some of these factors will be discussed in subsequent chapters.

5.10 Maximum Biological Fertility

A population that has universal marriage and no birth control would experience close to the maximum biological fertility, also referred to as *natural fertility*. One such population was that of the Hutterites: an Anabaptist sect found mostly in upper mid-western United States and Lower Canada. During the earlier part of the twentieth century they had universal marriage and did not practice birth control (Eaton & Mayer 1953). Age-specific fertility rates for the 1921–1930 marriage cohort of Hutterite women have been considered by some demographers as indicators of natural fertility. Coale used these rates as a standard against which fertility of populations could be measured (Coale 1967).

Using Hutterite fertility rates as the standard, Coale (1967) suggested three indicators to compare the fertility of a population with the biological maximum. These were indices of general fertility (I_f), marital fertility (I_g), and proportions married (I_m). These indices are:

$$\text{Index of general fertility, } I_f = \frac{B}{\sum_{i=15}^{i=49} (w_i * G_i)} \tag{5.13}$$

B is the actual total number of births and w_i is the number of women aged i in the population, and G_i is the age-specific fertility rates of Hutterite women (standard).

$$\text{Index of marital fertility, } I_g = \frac{B_L}{\sum_{i=15}^{i=49} (m_i * G_i)} \tag{5.14}$$

B_L is the actual number of births to married women and m_i is the number of married women in the population, G_i is the age-specific fertility rates of Hutterite women (standard).

$$\text{Index of proportion married, } I_m = \frac{\sum_{i=15}^{i=49} m_i * G_i}{\sum_{i=15}^{i=49} w_i * G_i} \tag{5.15}$$

This index is the ratio of the denominators of Eqs. (5.14) and (5.13).

Equations (5.13) and (5.14) are the standardized fertility ratios for all women and married women, when the fertility rates of Hutterite women are applied to their numbers.

Algebraically, all of these indices are equal to 1 for Hutterite women: the standard. They are less than or equal to 1 for other populations indicating their proximity to natural fertility.

Australian (2006) and Indian (2001) data have been used to illustrate the estimation of the above indices in Table 5.11. The values of G_i estimated by Coale were found to be slightly different by subsequent researchers (Wetherell 2001). The total fertility rate for Hutterites estimated from Coale's figures, column (2) of Table 5.11, comes to 12.44 (=2.488*5), while Wetherell's data in column (3) results in a TFR of 14.13 (=2.826*5). The authors decided to use Wetherell's estimates of age-specific fertility rates as these were based on the actual data rather than the adjusted figures compiled by Coale.

The three indices were consistently lower for Australia. Although the Indian fertility levels are lower than those of the Hutterites, they are closer to the Hutterites than those of Australia. The difference in terms of the proportions married (I_m) between Australia and India are large (see assumptions made regarding births to married women in India in Table 5.11). Coale and Treadway (1986) have argued that in populations where fertility outside marriage is very low, the product of the index of marital fertility and the index of the proportion of married women equals the index of general fertility ($I_g * I_m = I_f$). This relationship holds more in the case of India, where births outside marriage were assumed to be 1 % of all births compared to Australia where they were 33 % of all births.

Table 5.11 Calculation of Coale's indexes for Australia 2006 and India 2001

	G_I estimate		Number of women (in 000's)				$w_i * G_i$ (in 000's)		$m_i * G_i$ (in 000's)	
			Australia, 2006		India, 2001					
Age (years)	Coale	Wetherell	All	Married	All	Married	Australia	India	Australia	India
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8) = (3)*(4)	(9) = (3)*(6)	(10) = (3)*(5)	(11) = (3)*(7)
15–19	0.300	0.623	661	4	46,276	11,316	411.8	28,829.9	2.5	7,049.9
20–24	0.550	0.549	666	63	43,443	32,899	365.6	23,850.2	34.6	18,061.6
25–29	0.502	0.502	641	226	41,865	38,578	321.8	21,016.2	113.5	19,366.2
30–34	0.447	0.447	714	397	36,912	34,781	319.2	16,499.7	177.5	15,547.1
35–39	0.406	0.406	751	470	34,535	32,183	304.9	14,021.2	190.8	13,066.3
40–44	0.222	0.236	750	484	25,859	23,288	177.0	6,102.7	114.2	5,496.0
45–49	0.061	0.063	736	481	22,541	19,591	46.4	1,420.1	30.3	1,234.2
Total	2.488	2.826	4,919	2,125	251,431	192,636	1,946.7	111,740.0	663.4	79,821.3
							All women		Married women	
							Australia	India	Australia	India
Number of births (in 000's) =							266.5	25,696.1#	179.0	25,439.1#
Index of general fertility (I_f) following Eq. (5.13) =							0.1369	0.2300	. . .	. . .
Index of marital fertility (I_g) following Eq. (5.14) =							. . .	. . .	0.2698	0.3187
Index of proportion married (I_m) following Eq. (5.15) =							0.3408	0.7143	. . .	. . .

Source: Wetherell (2001) for columns (2) and (3); Australia (2013) for columns (4) and (5); India (undated) for columns (6) and (7). Australia (2007) for births in 2006; United Nations (2011) for age-specific fertility rates in India

Notes: # The number of births for India was estimated by multiplying the reported annual age-specific fertility rates in India in 2000–2005 by the age distribution of females 15–49 in 2001. The rates were: 68.9, 228.9, 177.5, 89.5, 40.1, 12.9 and 4.9 per 1,000 for age groups 15–19 to 45–49 and downloaded from United Nations (2011). To obtain these rates select India and 2000–2005 from the drop down menu in the U.N. database and click on Apply Filters. After the estimation of total births, it was assumed that in India 99 % of births were to married women

References

Australia. (2007). *2006 births Australia* (Catalogue No. 3301.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/0C4C341C51104DC4CA2573800015C2DC/$File/33010_2006.pdf. Accessed Feb 2013.

Australia. (2010a). *2009 births Australia* (Catalogue No. 3301.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/10BEDC49AFCACC1FCA2577CF000DF7AB/$File/33010_2009.pdf. Accessed Feb 2013.

Australia. (2010b). *Population by age and sex, Australian States and Territories, June 2010* (Catalogue No. 3201.0). Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3201.0Jun%202010?OpenDocument. Accessed Feb 2013.

Australia. (2011). *2010 births Australia* (Catalogue No. 3301.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/3148B41DC3875A59CA25793300168145/$File/33010_2010.pdf. Accessed Feb 2013.

Australia. (2013). *Census home:* D*ata and analysis*. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/websitedbs/censushome.nsf/home/data?opendocument#from-banner=LN. Accessed Feb 2013.

Becker, G. S. (1960). An economic analysis of fertility. In A. J. Coale (Ed.), *Demographic and economic change in developed countries* (pp. 209–230). Princeton: Princeton University Press.

Bongaarts, J. (1978). A framework for analysing the proximate determinants of fertility. *Population and Development Review, 4*, 105–132.

Coale, A. J. (1967). Factors associated with the development of low fertility: An historic summary. *Proceedings of the World Population Conference, 2*, 205–209.

Coale, A. J., & Treadway, R. (1986). A summary of the changing distribution of overall fertility, marital fertility, and the proportions married in the provinces of Europe. In A. J. Coale & S. C. Watkins (Eds.), *The decline of fertility in Europe* (pp. 31–181). Princeton: Princeton University Press.

Davis, K., & Blake, J. (1956). Social structure and fertility: An analytical framework. *Economic Development and Cultural Change, 3*, 211–235.

Eaton, J. W., & Mayer, A. J. (1953). The social biology of very high fertility among the Hutterites: The demography of a unique population. *Human Biology, 25*, 206–264.

India. (undated). *Census data 2001. Marital status by age and sex*. www.censusindia.gov.in/Census_Data_2001/Census_data_finder/C_Series/Marital_status_by_age_and_sex.htm. Accessed Feb 2013.

Strassmann, B. I., & Gillespie, B. (2002). Life-history theory, fertility and reproductive success in humans. *Proceedings of the Royal Society B, 269*, 553–562.

UNDP. (2011). *Human development report 2011. Sustainability and equity: A better future for all*. New York: United Nations Development Programme. http://hdr.undp.org/en/reports/global/hdr2011/. Accessed Feb 2013.

United Kingdom. (2011a). *Characteristics of birth 2, England and Wales, 2010* (Excel sheet 442Kb, table 7). Newport: Office of National Statistics. http://www.ons.gov.uk/ons/search/index.html?newquery=births+to+married+women. Accessed Feb 2013.

United Kingdom. (2011b). *Population estimates by marital status – mid 2010*. Newport: Office of National Statistics. http://www.ons.gov.uk/ons/dcp171778_244768.pdf. Accessed Feb 2013.

United Nations. (1999). *Demographic yearbook. Special issues – 1999 natality statistics*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/special_issues.html. Accessed Feb 2013.

United Nations. (2008). *Demographic yearbook 2008*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dyb2008.htm. Accessed Feb 2013.

United Nations. (2010). *World fertility report 2009*. New York: Department of Economic and Social Affairs. http://www.un.org/esa/population/publications/WFR2009_Web/Main.html. Accessed Feb 2013.

United Nations. (2011). *UN data: age-specific fertility rate*. New York: Department of Economic and Social Affairs. http://data.un.org/Data.aspx?d=GenderStat&f=inID%3A13. Accessed Feb 2013.

Wetherell, C. (2001). Another look at Coale's indices of fertility, I_f and I_g. *Social Science History, 25*(4), 589–609.

Chapter 6
Mortality

6.1 Purpose

This chapter considers various methods of the analysis of mortality from cross sectional and longitudinal perspectives. It is also concerned with the examination of causes of death and the analysis of mortality related to different stages of life. The chapter also deals with other related topics such as foetal mortality, abortion and the concept and measurement of potential years of life lost.

6.2 Cross-Sectional Mortality Rates

6.2.1 Crude Death Rates

Just as birth is central to the concept of fertility, death is pivotal in mortality. A definition of death is the permanent disappearance of all evidence of the life at any time after a live birth takes place (Swanson & Stephan 2004: 757). Thus, a live birth must have occurred before a death can be recorded. Once a death is recorded, then it can be used in measurements. A mortality measure commonly used is the *crude death rate* that relates the number of deaths in a specific period of time to the average population during the same period. These crude rates are usually expressed per thousand people.

$$CDR_{t \to t+n} = \frac{D_{t \to t+n}}{\overline{P}_{t \to t+n}} \tag{6.1}$$

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_6,

$CDR_{t\to t+n}$ is the crude death rate during the period t to $t + n$, $D_{t\to t+n}$ is the number of deaths during the period, and $\overline{P}_{t\to t+n}$ is the average population during the same period. Generally the mid-period population is taken as an estimate of the average population. Thus, if deaths are over a period of 1 year, say t to $t + 1$, the population refers to $t + \frac{1}{2}$.

An example is given below using the 2007 data for Malaysia (United Nations 2008):

$$\text{Deaths during the calendar year } 2007(D_{2007}) = 123{,}300$$
$$\text{Population in mid-}2007(P_{mid-2007}) = 27{,}720{,}000$$

$$\begin{aligned}\text{Crude death rate } (CDR_{2007}) &= \frac{123{,}300}{27{,}720{,}000} = 0.00445\\ &= 4.5 \text{ deaths per } 1{,}000 \text{ population during 2007.}\end{aligned}$$

Comparisons of crude death rates need to take into consideration either changes in age distribution over time or differences in age distribution between countries, as mortality is closely associated with age.

6.2.2 Age-Specific Mortality Rates

Age-specific mortality rates are defined as:

$$ASMR^i_{t\to t+n} = \frac{D^i_{t\to t+n}}{\overline{P}^i_{t\to t+n}} \tag{6.2}$$

$ASMR^i_{t\to t+n}$ is the age-specific mortality rate for persons aged i during the period t to $t + n$, $D^i_{t\to t+n}$ is the number of deaths of persons aged i during the same period, and $\overline{P}^i_{t\to t+n}$ is the average population of persons aged i during the period. These rates may be calculated for single years of age or in age groups.

An example is given in Table 6.1 of the calculation of age-specific mortality rates in Japan during 2004. Column (2) of the table has the number of deaths during the calendar year 2004, column (3) the mid-2004 population and columns (4) and (5) give age-specific mortality rates.

Figure 6.1 presents the age-specific mortality rates for Japanese females and males. The chart shows that up to 45 years of age the mortality differences were negligible but male mortality exceeded that of females after that age.

Both the horizontal and vertical axes, also referred to as the X- and Y-axis, in Fig. 6.1 are based on the *Cartesian* scale. One or both of these axes can be changed

Table 6.1 Estimation of age-specific mortality rates: Japan, 2004

			Age-specific mortality rate	
Age (years) (*i*)	Deaths 2004	Population mid-2004	Per person	Per 1,000
(1)	(2)	(3)	(4) = (2)/(3)	(5) = (4)*1,000
0–4	4,281	5,735,000	0.00075	0.75
5–9	607	5,938,000	0.00010	0.10
10–14	589	6,060,000	0.00010	0.10
15–19	1,928	6,761,000	0.00029	0.29
20–24	3,241	7,725,000	0.00042	0.42
25–29	4,157	8,755,000	0.00047	0.47
30–34	5,969	9,819,000	0.00061	0.61
35–39	7,405	8,662,000	0.00085	0.85
40–44	10,069	7,909,000	0.00127	1.27
45–49	16,098	7,854,000	0.00205	2.05
50–54	31,307	9,300,000	0.00337	3.37
55–59	46,480	9,640,000	0.00482	4.82
60–64	61,579	8,652,000	0.00712	7.12
65–69	81,497	7,343,000	0.01110	11.10
70–74	117,114	6,466,000	0.01811	18.11
75–79	152,164	5,098,000	0.02985	29.85
80–84	160,438	3,235,000	0.04959	49.59
85–89	154,810	1,719,000	0.09006	90.06
90+	168,210	1,016,000	0.16556	165.56

Source: Japan (2012)

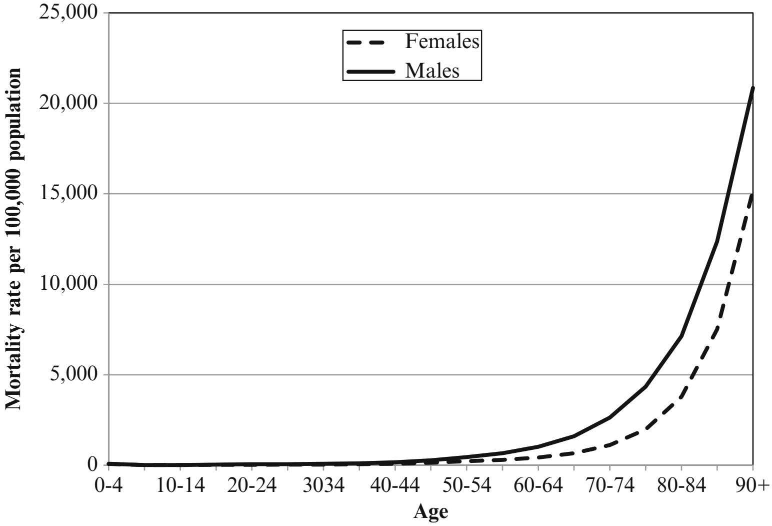


Fig. 6.1 Age-sex specific mortality rates per 100,000 population: Japan, 2004 (*Source*: Japan 2012)

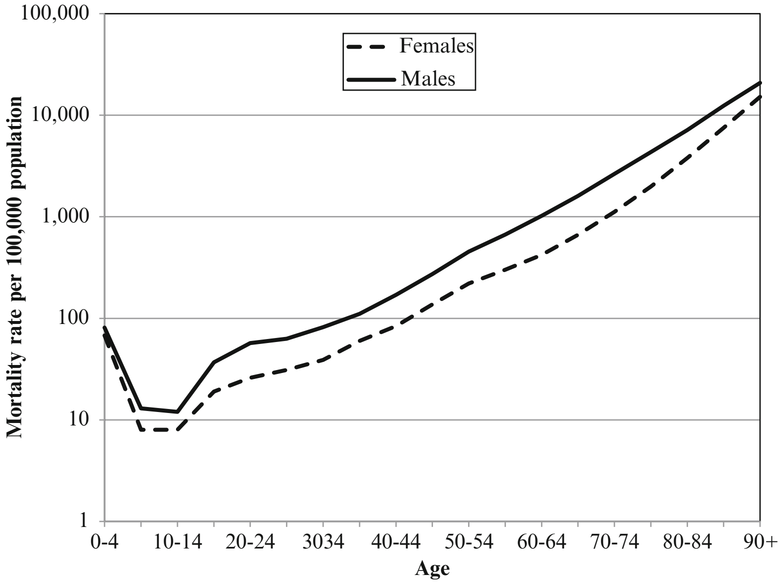


Fig. 6.2 Age-sex specific mortality rates per 100,000 population: Japan, 2004 (using logarithmic scale to base 10) (*Source*: Japan 2012)

to the logarithmic scale. In the presentation of mortality data, the logarithmic scale is used more often for the *Y*-axis. This makes the differentiation of small variations much easier.

Figure 6.2 is based on the same data as Fig. 6.1 but in this case the *Y*-axis is on a logarithmic scale with a base of 10. This means that each value shown on the *Y*-axis is 10 times that of the previous value. The sex differentials in mortality for all ages are clearer in this figure than in the previous one.

Age-specific mortality rates can be specific to other characteristics of the deceased such as marital status, occupation and ethnicity.

Age-specific mortality rates are used in the preparation of life tables (Chap. 7). In that context, usually, they are based on the average of 3 or more years of deaths to minimize random fluctuations and are referred to as the *central death rates*, denoted by m_x.

6.3 Longitudinal Mortality Rates

Mortality could also be looked at in terms of age-cohorts over time. Mortality rates by age over time for the United States are shown in Table 6.2. Each column in Table 6.2 gives mortality rates based on deaths occurring in particular years by the age of the deceased (cross sectional approach). The bold figures in the diagonal

Table 6.2 Age-specific mortality rates per 100,000 population: United States, 1900–2010

Age (years)	*1900*	*1910*	*1920*	*1930*	*1940*	*1950*
0	16,244.8	13,176.3	9,226.4	6,900.4	5,493.8	3,299.2
1–4	1,983.8	1,397.3	987.2	563.6	289.6	139.4
5–14	**386.9**#	293.5	263.9	**171.7***	103.7	60.1
15–24	585.5	**453.8**#	487.4	334.3	**204.8***	128.1
25–34	819.8	654.5	**677.5**#	465.8	305.9	**178.7***
35–44	1,023.1	898.5	811.3	**682.0**#	520.1	358.7
45–54	1,495.4	1,374.5	1,219.3	1,217.6	**1,059.9**#	853.9
55–64	2,723.6	2,624.9	2,358.8	2,402.7	2,215.5	**1,901.0**#
65–74	5,636.0	5,556.7	5,253.4	5,137.3	4,838.3	4,104.3
75–84	12,330.0	12,218.6	11,888.1	11,269.2	11,203.9	9,331.1
85+	26,088.2	25,030.8	24,825.2	22,795.6	23,565.0	20,196.9
Age (years)	*1960*	*1970*	*1980*	*1990*	*2000*	*2010*
0	2,696.4	2,142.4	1,288.3	971.9	727.4	622.4
1–4	109.1	84.5	63.9	46.8	32.8	26.6
5–14	46.6	41.3	30.6	24.0	18.6	12.8
15–24	106.3	127.7	115.4	99.2	81.5	67.7
25–34	146.4	157.4	135.5	139.2	108.0	102.8
35–44	**299.4***	314.5	227.9	223.2	199.7	170.4
45–54	756.0	**730.0***	584.0	473.4	430.7	406.7
55–64	1,735.1	1,658.8	**1,346.3***	1,196.9	1,005.4	851.2
65–74	**3,822.1**#	3,582.7	2,994.9	**2,648.6***	2,432.9	1,873.4
75–84	8,754.2	**8,004.4**#	6,692.6	6,007.2	**5,694.3***	4,785.5
85+	19,857.5	16,344.9	**15,980.3**#	15,327.4	15,324.4	**13,918.1***

Source: United States (2009), Anderson (2002), and Murphy et al. (2012)

represent the experience of a cohort of the population (longitudinal approach). The first bold figure with (#) in the diagonal (386.9) is based on deaths in 1900 of people who were born 5–14 years prior to 1900 (those born between 1886 and 1895). The second bold figure in the diagonal (453.8) relates to deaths in 1910 of people who were born 15–24 years prior to 1910. Their approximate years of birth were again between 1886 and 1895. Therefore, this diagonal represents the mortality experience of the cohort born in 1886–1895.

The second diagonal has bold figures with (*) starting from the third figure in the 1930 column (171.7) and ending in the last figure in the 2010 column (13,918.1). All of these mortality rates are for persons born between 1916 and 1925. All other diagonals starting with the mortality rates of persons 5–14 years of age in 1940 are the mortality of birth cohorts that had surviving members in 2010 and will continue to undergo additional depletion through death after 2010.

A comparison of the mortality experience of both cohorts is shown in Fig. 6.3. It indicates that the younger cohort (born between 1916 and 1925 experienced lower mortality rates compared to the 1886–1895 birth-cohort.

The higher mortality of the older birth-cohort persisted at all ages, though it was much larger up to age 45. Some of the differences might have arisen from the

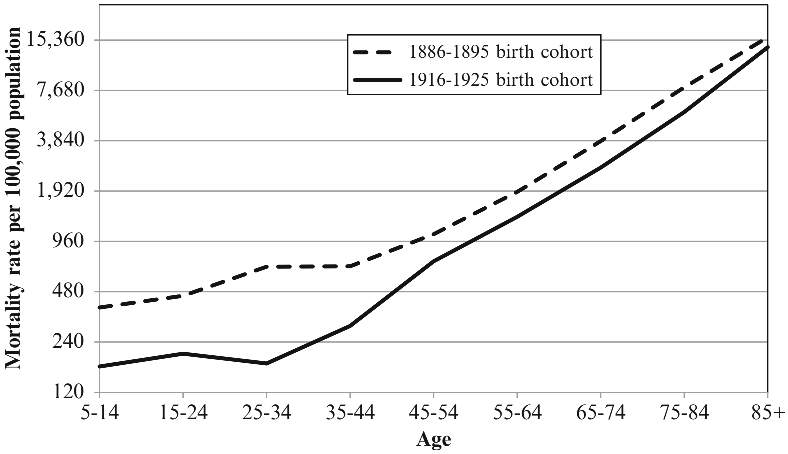


Fig. 6.3 Age-specific mortality rates per 100,000 population: United States, 1886–1895 and 1916–1925 birth cohorts (using logarithmic scale to base 2) (*Source*: Table 6.2)

quality and completeness of registration data in some states in early periods. Two peaks in age-specific mortality rates in 15–24 and 25–34 in Fig. 6.3 might have been affected by mortality at time of war.

6.4 Causes of Death

Usually, cause of death information is classified according to the International Classification of Diseases (ICD). This classification is produced by the World Health Organization (WHO) and revised every 10 years. The most current version is ICD-10 (WHO 2010). It consists of codes for individual diseases. Most countries use the detailed ICD or a modified and abridged version to code their causes of death data. In countries where registration of deaths is not adequate, mortality data are collected through sample surveys, and may not be available in the detailed ICD categories.

Information on the major causes of deaths in Netherlands is used as an illustration (Netherlands 2012a). Table 6.3 gives the percentage distribution of deaths by cause of death in column (2). These percentages are also referred to as *cause-specific mortality ratios*. The two major causes of deaths accounted for about 6 in every 10 deaths: cancer (neoplasm 32.0 %) and diseases of the circulatory system (28.7 %). Other major causes of death were diseases of the respiratory system (9.6 %), mental and behaviour disorders (5.6 %) and injury and poisoning (4.2 %)

The proportion of female deaths attributed to different causes is shown in column (3) of Table 6.3. Overall, female deaths amounted to about half of the

Table 6.3 Percentage distribution of deaths by major cause, percentage of female deaths and crude death rates for each cause: Netherlands, 2011

Causes of death	% Total	% Female	Death rate per 100,000
(1)	(2)	(3)	(4)
Infectious and parasitic diseases	1.5	52.0	12.6
Neoplasms (cancers)	32.0	45.6	261.3
Blood and blood forming organs and diseases of the immune system	0.3	58.6	2.6
Endocrine, nutritional and metabolism diseases	2.8	56.5	23.1
Mental and behaviour disorders	5.6	70.3	45.7
Diseases of the nervous system	3.5	58.4	28.2
Diseases of the circulatory system	28.7	53.2	234.2
Diseases of the respiratory system	9.6	48.8	78.1
Diseases of the digestive system	3.9	54.2	31.6
Diseases of skin and subcutaneous tissue	0.3	66.0	2.5
Diseases of the muscular system and connecting tissue	0.6	68.9	5.3
Diseases of the genitourinary system	2.6	58.9	20.8
Complications of pregnancy, childbirth and puerperium	...	100.0	...
Conditions originating in perinatal period	0.3	43.8	2.1
Congenital anomalies	0.3	49.1	2.7
Symptoms, signs and ill-defined conditions	3.8	55.1	31.4
External causes of injury and poisoning	4.2	43.6	34.5
All causes of death	100.0	51.5	816.9
Total number of deaths	136,058	70,081	...

Sources: Netherlands (2012a) for deaths and Netherlands (2012b) for population data
Notes: Figures may not add due to rounding; (...) indicates < 0.05 % or <0.005 per 100,000

deaths (51.5 %). The proportion of female deaths was substantially higher in the case of mental and behaviour disorders (70.3 %), diseases of the muscular system and connective tissue (68.9 %) and diseases of the skin and subcutaneous tissue (66.0 %). However, the proportion of female deaths were considerably lower in the cases of deaths attributed to injury and poisoning (43.6 %), conditions originating in the perinatal period (43.8 %), and cancer (45.6 %).

Cause-specific mortality rates in column (4) of Table 6.3 were obtained by dividing the number of deaths for each cause by the 2011 midyear population of Netherlands (16,655,799). As the rates are per 100,000 people the rates per person were multiplied by 100,000. The death rate due to all causes (crude death rate) was 816.9 per 100,000, or 8.2 per 1,000 people. Since age is an important determinant of disease, age and cause specific death rates are often calculated.

An international comparison of major causes of death (Table 6.4) shows that, overall, cardiovascular and communicable diseases were the two most frequently reported causes of death. The proportion of these causes varied substantially in

Table 6.4 Distribution of deaths by major cause in world regions, 2008

Cause of death (diseases and injuries)	World	World Health Organization regions: Africa	Europe	South East Asia	Americas	Western Pacific	Eastern Mediterranean
Communicable	27.5	65.0	5.8	34.7	11.7	9.9	36.3
Neoplasms	13.3	4.0	20.3	7.8	19.3	21.0	7.5
Cardiovascular	30.5	12.4	49.7	24.9	31.5	37.4	28.5
Respiratory	7.4	3.3	4.1	9.6	6.4	12.5	3.8
Injuries	9.0	6.8	7.2	10.7	9.6	9.4	10.6
All other diseases	12.3	8.5	12.9	12.3	21.5	9.8	13.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Source: The information presented in this table was compiled by the authors using the WHO mortality database (WHO 2013)
Note: Totals may not add due to rounding

different regions of the world. For example, communicable diseases accounted for most of the deaths in the Africa region but were just over one-third of deaths in South East Asia and Eastern Mediterranean regions. Cardiovascular diseases that were about half of the deaths in the Europe region were just over 1 in 10 deaths in the Africa region. Neoplasms (cancers) were less common in Africa, South East Asia and Eastern Mediterranean regions. The quality of mortality data, in particular for causes of death, tends to be poor in countries with low coverage of health and medical services.

6.5 Maternal Mortality

According to WHO, maternal mortality refers to *the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes* (WHO 2007a).

Maternal deaths may occur because of direct causes such as obstetric complications, or indirect causes such as pre-existing conditions like hypertension or diabetes which may be aggravated during pregnancy. As records of the cause of death tend to state the direct causes of death, it is necessary to investigate whether indirect causes are included. For instance, in a 2003–2005 study of maternal mortality in Australia, 29 female deaths were reported (ICD code O00-O99) as those with an underlying cause of death related to pregnancy, childbirth and the puerperium. This number was increased to 65 when indirect causes of maternal mortality were included (Sullivan et al. 2008).

The *maternal mortality ratio* is defined as:

$$MMr_{t\to t+n} = \frac{MD_{t\to t+n}}{B_{t\to t+n}} \tag{6.3}$$

$MMr_{t\to t+n}$ is the maternal mortality ratio during the period t to $t + n$, $MD_{t\to t+n}$ is the total number of maternal deaths during the same period, and $B_{t\to t+n}$ is the number of live births during the period.

As an example, Australian data (Sullivan et al. 2008) show that the number of maternal deaths was 65 in the period 2003–2005 and the number of live births during the same period was 765,198. Following Eq. (6.3):

$$\begin{aligned} MMr_{2003\to 2005} &= \frac{65}{765,198} \\ &= 0.000085 \text{ or } 8.5 \text{ maternal deaths per } 100,000 \text{ live births.} \end{aligned}$$

The measure that relates the number of maternal deaths to the number of women of reproductive age is the *maternal mortality rate* defined as:

$$MMR_{t\to t+n} = \frac{MD_{t\to t+n}}{\sum\limits_{i=15}^{i=44} W^i_{t\to t+n}} \tag{6.4}$$

$MMR_{t\to t+n}$ is the maternal mortality rate during the period t to $t + n$, $MD_{t\to t+n}$ is the number of maternal deaths in the same period and $\sum\limits_{i=1}^{i=44} W^i_{t\to t+n}$ is the sum of the number of women 15–44 (or 15–49) years of age during the period.

In Australia, the sum of the number of women 15–44 years of age in the middle of each of the 3-years 2003–2005 was 12,983,457 (Australia 2010a). Therefore, the *maternal mortality rat*e for that period was:

$$MMR_{2003\to 2005} = \frac{65}{12,983,457} = 0.000005 \text{ or } 5 \text{ per million women.}$$

Maternal mortality rates can also be calculated by age as:

$$MMR^i_{t\to t+n} = \frac{MD^i_{t\to t+n}}{W^i_{t\to t+n}} \tag{6.5}$$

$MMR^i_{t\to t+n}$ is the maternal mortality rate at age i during the period t to $t + n$, $MD^i_{t\to t+n}$ is the maternal deaths at age i during the same period and $W^i_{t\to t+n}$ is the number of women aged i during the period.

Age-specific maternal mortality rates for Australia in 2003–2005 are shown in Fig. 6.4. These rates can be standardised like any other type of demographic rates with the help of an appropriate standard.

Another useful index is the *life-time risk of maternal mortality*. In general terms, this is simply the sum of the age-specific maternal mortality rates. When the age is

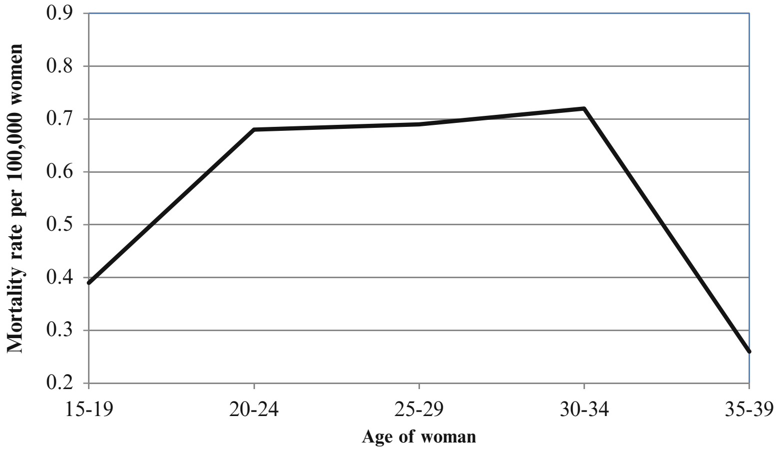


Fig. 6.4 Age-specific maternal mortality rates: Australia, 2003–2005 (*Source*: Sullivan et al. 2008)

given in 5-year or 10-year groups, the sum needs to be multiplied by 5 or 10 depending upon the number of years in the age groups.

For Australia, the sum of age-specific maternal mortality rates per 100,000 women in 5-year age groups was (0.25 + 0.39 + 0.68 + 0.89 + 0.72 + 0.26) = 3.19. Multiplying the sum by 5 resulted in 15.95. This means that for every 100,000 women 15–44 years of age, 16 died from causes related to motherhood. This assumes that no woman died of any other cause during her reproductive period. Life tables, described in the Chap. 7, can be employed to estimate the total number of women who die during their reproductive period (Box 5.2).

Maternal mortality is an indicator of women's health and also reflects the level of human development. Data compiled by the United Nations reveals that countries that are ranked very high or high on the Human Development Index (HDI) also experience lower maternal mortality ratios compared to those ranked medium or low on the HDI (UNDP 2011). Another factor influencing maternal mortality is the proportion of births attended by skilled health personnel. This proportion is negatively associated with the maternal mortality ratio (Table 6.5).

6.6 Foetal and Early Childhood Mortality

6.6.1 Perinatal, Neonatal and Infant Mortality

Pregnancy may result in a *single* or *multiple live birth*, a *still birth* (also known as a foetal death), or *abortion* (spontaneous or induced). *Infant deaths* refer to deaths within the first year of life. It is customary to divide infant deaths into two periods: *neonatal deaths* that relate to deaths during the first 4 weeks of life and *postnatal*

Table 6.5 Ranking on Human Development Index, maternal mortality ratio, and proportion of births attended by skilled health personnel

HDI ranking	Countries	Maternal mortality ratio per 100,000 births 2008	% births attended by skilled health personnel 2005–2009
Very high	47	16	99.2
High	47	51	96.1
Medium	47	135	78.1
Low	46	532	39.6

Source: UNDP (2011)

deaths that relate to deaths in the remaining 48 weeks of life. Another measurement is concerned with *perinatal deaths* defined as the sum of still births (foetal deaths) and neonatal deaths. *Spontaneous abortions* can occur at any time during the pregnancy, while *induced abortions* are usually performed in the first trimester of a pregnancy.

Caution needs to be taken regarding still birth data, as the quality and definition may vary from country to country. The definition usually involves a specification of the gestation period and or the birth weight of the foetus. A specification often used is a foetus born dead with at least 20 weeks' gestation or birth weight of at least 400 g. Information on spontaneous abortion is not routinely collected. Data on induced abortions may be available in countries where it is legal. Sample surveys of detailed pregnancy histories may be carried out to provide information on different types of pregnancy outcomes.

A demographic measure of birth outcome is the *neonatal mortality rate* defined as:

$$NMR_{t\to t+n} = \frac{D_{t\to t+n}^{first4weeks}}{B_{t\to t+n}} \tag{6.6}$$

$NMR_{t\to t+n}$ is the neonatal mortality rate for the period t to $t + n$, $D_{t\to t+n}^{first4weeks}$ is the number of infant deaths during the first 4 weeks (28 completed days) of life during the same period and $B_{t\to t+n}$ is the number of live births during the period. Generally, the period is 1 year.

Neonatal deaths can be subdivided into *early neonatal* (occurring in the first 7 days of life) in a given year. *Post-neonatal* deaths refer to the period from 29 days to less than 1-year after birth:

$$ENMR_{t\to t+n} = \frac{D_{t\to t+n}^{first7days}}{B_{t\to t+n}} \tag{6.7}$$

$$PNMR_{t\to t+n} = \frac{D_{t\to t+n}^{29^{\text{th}}day\to <1year}}{B_{t\to t+n}} \tag{6.8}$$

Table 6.6 Live and still births and infant deaths: United Kingdom, 2007

	Number of births		Number of deaths <1 year of age				
Sex	Live	Still	Neonatal	Early neonatal	Perinatal	Post-neonatal	Infant
(1)	(2)	(3)	(4)	(5)	(6) = (3) + (4)	(7)	(8) = (4) + (7)
Males	396,698	2,136	1,412	1,089	3,548	663	2,075
Females	375,547	1,891	1,105	855	2,996	479	1,584
Persons	772,245	4,027	2,517	1,944	6,544	1,142	3,659

Source: United Kingdom (2009)

$ENMR_{t\to t+n}$ and $PNMR_{t\to t+n}$ stand for the early neonatal mortality rate and the post-neonatal mortality rate, $D_{t\to t+n}^{first7days}$ and $D_{t\to t+n}^{29^{th}day\to <1year}$ for deaths in the first 7 days of life and deaths 29th day to just before the first birthday, period t to $t+n$. The denominator for both rates is the number of live births.

Stockwell & Wicks (1987) observed that in countries with low infant mortality rates, the greater management of risks in the neonatal period resulted in an increase of endogenous cases of death in the post-neonatal period. This diminished the validity of using the post-neonatal mortality rate as a proxy for endogenous mortality in infancy.

Another related demographic measure is the *perinatal mortality rate*:

$$PMR_{t\to t+n} = \frac{D_{t\to t+n}^{first4weeks} + SB_{t\to t+n}}{B_{t\to t+n}} \tag{6.9}$$

$PMR_{t\to t+n}$ is the perinatal mortality rate for a given period (year), $SB_{t\to t+n}$ is the number of still births during the period and $D_{t\to t+n}^{first4weeks}$ and $B_{t\to t+n}$ are as defined for Eq. (6.6).

The *infant mortality rate* relates to the number of infant deaths (those dying before their first birthday):

$$IMR_{t\to t+n} = \frac{D_{t\to t+n}^{<1year}}{B_{t\to t+n}} \tag{6.10}$$

$IMR_{t\to t+n}$ is the infant mortality rate for a given period (year), $D_{t\to t+n}^{<1year}$ is the deaths of infants (less than 1-year old) and $B_{t\to t+n}$ is the number of births during the period.

Table 6.6 presents 2007 data for England and Wales (United Kingdom 2009) relevant to the calculation of the five mortality rates defined in this section.

Table 6.7 Neonatal, early neonatal, perinatal, post-natal and infant mortality rates: United Kingdom, 2007

	Mortality rates per 1,000 live births				
Sex	Neonatal	Early neonatal	Perinatal	Post-neonatal	Infant
(1)	(2)	(3)	(4)	(5)	(6) = (2) + (5)
Males	3.56	2.75	8.94	1.67	5.23
Females	2.94	2.28	7.98	1.28	4.22
Persons	3.26	2.52	8.47	1.48	4.74

Source: Table 6.5

To calculate the five mortality rates defined by Eqs. (6.6), (6.7), (6.8), (6.9) and (6.10), the numerators are available in columns (4)–(8) of Table 6.6, and the denominators for all of these rates is the number of live births in column (2). The estimated rates are presented in Table 6.7.

The first set of numbers in columns (2)–(6) were calculated as:

$$\frac{1,412}{396,698}, \frac{1,089}{396,698}, \frac{3,548}{396,698}, \frac{663}{396,698}, \text{ and } \frac{2,075}{396,698}.$$

All of the quotients were multiplied by 1,000 as the rates are per 1,000 live births. This procedure was repeated in the remaining two rows of Table 6.7. The table reveals that all of the five mortality rates were higher for males than females.

Yet another demographic measure related to childhood is the *child mortality rate*. This is an age-specific mortality rate, where *i* in Eq. (6.2) stands for persons one to less than 5 years of age.

Information about perinatal, neonatal and infant mortality rates is available from birth and death registers in many countries, while in others it is collected by sample surveys. WHO has compiled estimates for the foetal, neonatal and perinatal mortality rates in 2004 for various countries. Some of these data were based on death registers, while others were obtained from various demographic surveys such as the *Demographic and Health Surveys* http://www.measuredhs.com/. International comparisons must ensure that the definitions of these types of deaths are comparable.

Neonatal, perinatal and child mortality rates for selected countries are given in Table 6.8. Neonatal and perinatal mortality rates are the WHO estimates (WHO 2007b), and the child mortality rates were calculated by the authors using UN data (United Nations 2012).

6.6.2 *Abortion Rates and Ratios*

Abortion is defined as an expulsion of a foetus from the uterus (Swanson & Stephan 2003). An induced abortion is the intentional removal of a foetus from

Table 6.8 Indicators of foetal, early childhood and infant mortality for selected countries

	Mortality rate per 1,000 births 2004				Child mortality rate	
Country	Foetal	Early neonatal	Neonatal	Perinatal	per 1,000 population	Year
Brazil	7	11	13	17	0.52	2009
South Korea	2	2	3	5	0.23	2009
Sweden	3	2	3	5	0.20	2009
Zimbabwe	19	29	36	48	12.15	2002

Source: WHO (2007b) and United Nations (2012)

the uterus. A spontaneous abortion is the premature and naturally occurring expulsion of a foetus from the uterus.

Various rates and ratios can be estimated for induced abortion data, if available. Some of these measures are similar to those described in relation to the analysis of fertility. For example, the *general abortion rate* (*GAR*), *age-specific abortion rate* (*ASAR*) and *total abortion rate* (*TAR*) are calculated by replacing the number of births by the number of abortions in the numerators of Eqs. (5.4), (5.3) and (5.5). Assumptions made in the calculation of *TFR* are implicit in the calculation of *TAR*.

In countries where female mortality during the reproductive period is low, the total abortion rate gives an estimate of the *lifetime-risk of abortion*. However, a more refined estimate is obtained by adjusting for mortality during the reproductive period as shown in Eq. (6.11):

$$ATAR = \sum_{i=15}^{i=44} \left(ASAR^{i}_{t \to t+1} * \frac{l_{45}}{l_i} \right) \tag{6.11}$$

ATAR is the mortality adjusted total abortion rate, l_i and l_{45} are the number of survivors at age i and at the end of reproductive period, 45 years of age, in an appropriate female life table (Chap. 7). If age at the time of the abortion is given in age groups, i to $i + n$, the probability of survival has to be calculated from the mid-point of the age interval to age 45.

Dividing abortions by births results in abortion ratios. In this case, the number of births is used as the number of pregnancies. This is not necessarily the best measure as a certain proportion of pregnancies result in still births, miscarriages or abortions. It has been argued that a somewhat better measure of pregnancies is births plus abortions. Again, international comparison of abortion ratios should make sure that the denominators are defined in the same way: whether only births or births plus abortions are included in the denominator.

Abortion data for South Australia (2012), where abortion is legal and registered, is given in Table 6.9 to give examples of different abortion measures. Table 6.9 reveals that abortions were equal to 26 % of the total live births in South Australia and one-in-five pregnancies resulted in abortion. Age-specific abortion rates had a similar

Table 6.9 Calculation of abortion ratios, rates and life time risk: South Australia, 2009

Age group	No. of Abortions	Births	Women	Rate per woman Abortion	Birth	Abortion ratio per Birth	Birth + abortion	Mean age (m)	Survival probability	Mortality adjusted abortion rate
(1)	(2)	(3)	(4)	(5) = (2)/(4)	(6) = (3)/(4)	(7) = (2)/(3)	(8) = (2)/[(2) + (3)]	(9)	(10)	(11) = (5)*(10)
15–19	912	837	52,385	0.01741	0.01598	1.08961	0.52144	17.5	0.98607	0.01717
20–24	1,498	3,020	56,636	0.02645	0.05332	0.49603	0.33156	22.5	0.98634	0.02609
25–29	1,105	5,721	55,785	0.01981	0.10255	0.19315	0.16188	27.5	0.98662	0.01954
30–34	726	5,978	51,165	0.01419	0.11684	0.12145	0.10829	32.5	0.98690	0.01400
35–39	588	3,443	53,685	0.01095	0.06413	0.17078	0.14587	37.5	0.98719	0.01081
40–44	225	735	57,402	0.00392	0.01280	0.30612	0.23438	42.5	0.98747	0.00387
Total	5,054	19,734	327,058	0.09273	0.36562	0.25611	0.20389	…	…	0.09148
×5	…	…	…	0.46365	1.82810	…	…	…	…	0.45740
				= *TAR*	= *TFR*					= *ATAR*

Sources: South Australia (2012) for column (2); Australia (2010b) for column (3); and Australia (2010a) for column (4)

Notes: 1. Births and abortions data refer to the calendar year 2009, and the number of women refers to mid-2009

2. Births and abortions to women under 15 years of age were included in the 15–19 age group, and for those 45 years of age or older in the 45–49 age group

3. For the estimation of column (10), it is assumed that the mortality of females in South Australia in 2009 was similar to that for the whole country as implied in the Australian female life table (see extract in Box 5.2). The average number of survivors for each age group, say 15–19, was estimated as the average of l_{17} and l_{18}. Survival probability refers to the probability of survival from the mean age (m) to age 45 taken as the end of the reproductive period. Thus, the values in column (10) were estimated by dividing l_{45} by the l_x for the mean age of each age group (see Box 5.2, last dot point)

4. *TAR* total abortion rate, *TFR* total fertility rate, *ATAR* mortality adjusted abortion rate

pattern to age-specific fertility rates. However, age-specific ratios had a U-shaped age-pattern. The total abortion rate came to 464 lifetime abortions per 1,000 women of reproductive age. This may be compared with the total fertility rate for South Australia of 1,828 lifetime births per 1,000 women. Adjustment for mortality resulted in a reduction of just over 1.3 % in the total abortion rate.

Countries, where abortions are legal, tend to report data on abortions. For example, abortion statistics are available for England and Wales (United Kingdom 2010). The estimated general abortion rate for England and Wales in 2009 was 17 abortions per 1,000 women 15–44 years of age. The authors used the published age-specific abortion rates to calculate the 2009 *TAR* for England and Wales: 525 abortions per 1,000 women. Both of these figures indicate that abortion rates in England and Wales were somewhat higher than those for South Australia.

6.7 Characteristic Specific Mortality Rates

In addition to age and sex other characteristics of the deceased can be used in the analysis of mortality. Such analyses require the availability of deaths by age, sex and the selected characteristic. In addition to age-sex-characteristic-specific mortality rates, standardization procedures (Sect. 4.9) can be used. In this section, standardization is applied to two characteristics: birthplace and marital status.

6.7.1 Mortality and Place of Birth

In countries where migration plays an important role, mortality information might be available by birthplace of the deceased. In others, information may be collected for various ethnic groups.

Deaths registered in Australia in 2010 offer an example in Table 6.10. This table shows that the crude death rates of foreign-born people were higher than those of Australian-born. When mortality was standardized for age, a different pattern emerges: foreign-born people had lower age-standardized mortality rates. There were sizeable differences in the crude death rates of foreign-born people; the differences were reduced substantially when the direct standardization procedure was applied.

6.7.2 Mortality and Marital Status

Figure 6.5 exhibits the age standardized death rates by marital status at the time of death for males and females in England and Wales. The data available consisted of age-sex-marital status specific rates for 2005 and the mid-2005 population (United Kingdom 2007). The direct age-standardized death rates by marital status were calculated by the authors using Eq. (4.27).

Table 6.10 Crude and standardized mortality rates by sex and birthplace: Australia, 2010

	Morality rate per 1,000 population			
	Crude		Standardized	
Birthplace	Males	Females	Males	Females
All countries	6.6	6.2	6.7	4.9
Australia-born	6.1	6.1	7.0	5.1
Foreign-born	7.8	6.6	6.0	4.3
North-West Europe	12.9	12.0	6.6	4.9
Southern and Eastern Europe	17.9	13.0	6.0	3.9
North Africa and the Middle East	5.2	4.4	5.9	4.3
South-East Asia	2.5	2.1	4.1	3.1
North East Asia	2.4	2.0	3.9	3.0
Southern and Central Asia	2.0	2.5	4.5	3.7
Americas	4.3	3.0	6.6	4.3
Sub-Saharan Africa	2.7	2.7	4.7	3.9

Source: Australia (2011, Table 9.1)

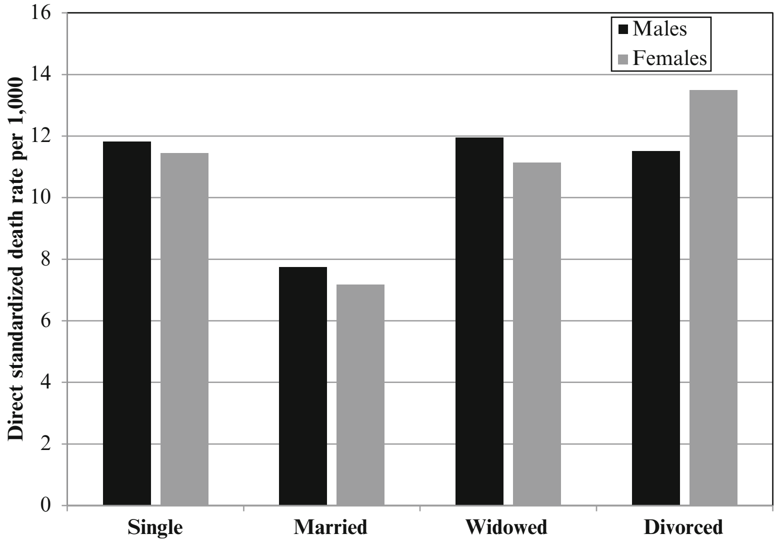


Fig. 6.5 Direct standardized death rates per 1,000 by sex and marital status: England and Wales, 2005 (*Source*: United Kingdom 2007)

Three points emerge from Fig. 6.5. First, the rates for married people were substantially lower than the other groups. Second, while male death rates were generally higher than female rates, the rates for divorced female were higher than those of divorced males. Third, the male rates for single, widowed and divorced were of similar order of magnitude.

Various socio-economic and psychological factors influence the relative mortality levels of married and unmarried persons. Their influence can be measured by further analysis, provided appropriate data are available.

6.8 Potential Years of Life Lost

Another index is the *potential years of life lost* (*PYLL*) concerned with premature deaths, sometimes related to specific causes of death. For example, if a disease such as HIV-AIDS is considered, it affects mostly young people. If a 32 year old person dies due to this disease in a country with a life expectancy of 80 years, the potential years of life lost are 48 years (80–32). Alternatively, a disease like prostate cancer which affects mostly older men, a death from this cause at the age of 72 years means 8 years of life lost. In the calculation of this index, the average life expectancy at birth, an age close to it or some other age standard needs to be chosen. The equation for the calculation of this index is:

$$PYLL = \sum_{i=0}^{i=n} (E - x_i) \times d_i \tag{6.12}$$

PYLL is the potential years of life lost before a set standard age, *E*, that people are expected to reach. This might be the life expectancy at birth or some other age

Table 6.11 Potential years of life lost for male deaths: Australia, 2005–2007

Age group	Mid-point (x_i)	$E–x_i$	Male deaths (d_i)	Potential years of life lost
(1)	(2)	(3)	(4)	(5) = (3)*(4)
0	0.5	78.5	2,107	165,400
1–4	2.0	77.0	393	30,261
5–9	7.5	71.5	229	16,374
10–14	12.5	66.5	259	17,224
15–19	17.5	61.5	1,078	66,297
20–24	22.5	56.5	1,776	100,344
25–29	27.5	51.5	1,824	93,936
30–34	32.5	46.5	2,284	106,206
35–39	37.5	41.5	2,762	114,623
40–44	42.5	36.5	3,736	136,364
45–49	47.5	31.5	5,355	168,683
50–54	52.5	26.5	7,147	189,396
55–59	57.5	21.5	10,264	220,676
60–64	62.5	16.5	13,151	216,992
65–69	67.5	11.5	16,921	194,592
70–74	72.5	6.5	21,680	140,920
75–79	77.5	1.5	31,894	47,841
Total	...	...	122,860	2,026,129

Source: Chapter 7, Table A7.1

(say, 79 years), x_i is the age at death, and d_i is the number of deaths at that age. If deaths data are in age groups, then x_i is the mid-point of each age group.

Australian data for male deaths in 2005–2007 given in Table A7.1 and stated in Table 6.11 provides an example of the estimation of potential years of life lost.

The analysis shows that 122,860 people died before their 80th birthday and 2,026,129 potential years of life were lost – an average of 16.5 years. This index can be used as an indicator of the relative impact of premature deaths from different diseases on society and the economy.

According to recent Australian statistics (Australia 2012), cardiovascular diseases that made up just over one-third of all deaths in 2009, accounted for 17 % of the total potential years of life lost. On the other hand, external causes of mortality and morbidity which accounted for just over 6 % of the total deaths represented 24 % of the total potential years of life lost. External causes include, among other things, various types of accidents. Their disproportionate contribution to the potential years of life lost arises from younger people being more prone to die from external causes.

References

Anderson, R. N. (2002). Deaths: leading causes for 2000. *National Vital Statistics Reports, 50*(16), 1–85. http://www.cdc.gov/nchs/data/nvsr/nvsr50/nvsr50_16.pdf. Accessed Feb 2013.

Australia. (2010a). *Population by age and sex, Australian States and Territories, June 2010* (Catalogue No. 3201.0). Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/ausstats/abs@.nsf/mf/3201.0. Accessed Feb 2013.

Australia. (2010b). *2009 births Australia* (Catalogue No. 3301.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/10BEDC49AFCACC1FCA2577CF000DF7AB/$File/33010_2009.pdf. Accessed Feb 2013.

Australia. (2011). *2010 deaths Australia* (Catalogue No. 3302.0). Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3302.02010. Accessed Feb 2013.

Australia. (2012). *Causes of death Australia* (Catalogue No. 3303.0). Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3303.02010?OpenDocument.

Japan. (2012). Population and households. In *Historical statistics of Japan* (Chapter 2, Population (table 2.3), and vital statistics (table 2.30a)). Tokyo: Statistics Bureau. http://www.stat.go.jp/english/data/chouki/02.htm. Accessed Feb 2013.

Murphy, S. L., Xu, J., & Kochanek, K. D. (2012). Deaths: Preliminary data for 2010. *National Vital Statistics Reports, 60*(4), 1–51. http://www.cdc.gov/nchs/data/nvsr/nvsr60/nvsr60_04.pdf. Accessed Feb 2013.

Netherlands. (2012a). *StatLine – Causes of death; main primary causes of death, sex and age, 2011*. Den Haag/Heerlen: Centraal Bureau voor de Statistiek. http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLEN&PA=7052eng&LA=EN. Accessed Feb 2013.

Netherlands. (2012b). *StatLine – Population; key figures*. Den Haag/Heerlen: Centraal Bureau voor de Statistiek. http://statline.cbs.nl/StatWeb/publication/?DM=SLEN&PA=37296eng&D1=0-51,56-68&D2=0,10,20,30,40,50,(l-1)-l&LA=EN&VW=T. Accessed Feb 2013.

South Australia. (2012). *Seventh annual report of the South Australian Abortion Reporting Committee, 2009*. First session, Fifty-Second Parliament. Adelaide: Parliament of South Australia.

Stockwell, E. G., & Wicks, J. (1987). The age-cause proxy relationship in infant mortality. *Social Biology, 34*, 249–253.

Sullivan, E. A., Hall, B., & King, J. F. (2008). *Maternal deaths in Australia 2003–2005* (Maternal deaths series no. 3). Canberra: Australian Institute of Health and Welfare. http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737421514. Accessed Feb 2013.

Swanson, D. A., & Stephan, G. E. (2004). Glossary. In J. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 751–786). San Diego: Elsevier Academic Press.

UNDP. (2011). *Human development report 2011. Sustainability and equity: A better future for all*. New York: United Nations Development Programme. http://hdr.undp.org/en/reports/global/hdr2011/. Accessed Feb 2013.

United Kingdom. (2007). *Mortality statistics: General, England and Wales (Series DH1: discontinued) – No. 38, 2005*. Newport: Office of National Statistics. http://www.ons.gov.uk/ons/search/index.html?newquery=dh1%2C+no.+38%2C+2005. Accessed Feb 2013.

United Kingdom. (2009). *Mortality statistics: Childhood, infant and perinatal, England and Wales (Series DH 3) No. 40, 2007*. Newport: Office of National Statistics. http://www.ons.gov.uk/ons/search/index.html?newquery=DH3%2C+No.+40%2C+2007. Accessed Feb 2013.

United Kingdom. (2010). *Abortion statistics, England and Wales, 2009*. London: Department of Health. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_116039. Accessed Feb 2013.

United Nations. (2008). *Demographic yearbook 2008*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dyb2008.htm. Accessed Feb 2013.

United Nations. (2012). *Demographic yearbook 2009–10*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dyb2009-2010.htm. Accessed Feb 2013.

United States. (2009). *National vital statistics system, HIST290. Death rates for selected causes by 10-year age groups, race, and sex: Death registration states, 1900–32, and United States, 1933–98*. Atlanta: Centers for Disease Control and Prevention. http://www.cdc.gov/nchs/nvss/mortality/hist290.htm. Accessed Feb 2013.

WHO. (2007a). *Maternal mortality in 2005: Estimates developed by WHO, UNICEF, UNFPA and The World Bank*. Geneva: World Health Organization. www.who.int/maternal_child_adolescent/documents/9789241596213/en/index.html. Accessed Feb 2013.

WHO. (2007b). *Neonatal and perinatal mortality: Country, regional and global estimates*. Geneva: World Health Organization. http://whqlibdoc.who.int/publications/2007/9789241596145_eng.pdf. Accessed Feb 2013.

WHO. (2010). *International classification of diseases (ICD)*. Geneva: World Health Organization. http://www.who.int/classifications/icd/en/index.html. Accessed Feb 2013.

WHO. (2013). *WHO mortality database: Tables*. Geneva: World Health Organization. http://www.who.int/healthinfo/mortality_data/en/. Accessed Feb 2013.

Chapter 7
Life Tables

7.1 Purpose

The subject of this chapter is the estimation of the average life expectancy of populations and expectation of life at different ages through the preparation of life tables. The chapter discusses assumptions made in the building of life tables and data requirements. The sequential elements in the construction of life tables are described and defined. The text distinguishes between complete and abridged life tables and between life tables that reflect actual mortality experience over time and those of a synthetic nature that estimate life expectancy based on current mortality rates. Further, the chapter examines different methods of estimation of life tables and compares results from these methods. In addition, it describes some uses of life tables and compares life expectancy in a number of countries in different parts of the world.

7.2 Longevity and Life Tables

One of the concerns of demographic analysis is the measurement of the longevity of populations. In other words, how long people live, on average. The length of life has differed over time and from one population to another for a variety of reasons. Mortality rates give an indication of the chances of survival of populations. However, as age-specific mortality rates show, the chances of survival vary from one age to another at different stages of the life span. In general terms, there are also differences in mortality between males and females. Therefore, crude deaths rates are poor indicators of survival and the length of life, on average, because of disparities in the age composition of different populations. A measure of how long people are expected to live, on average, can be obtained by following a cohort born at a given point in time and see how long they live up to different stages of their life. Thus, people born in a given geographical location and year could be followed to see how many survived on average to a given age until all died and then

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_7,

the average number of years lived by the cohort could be ascertained: number of years lived in total divided by the number of people born in the given year and geographical area. This data could also be used to estimate the average life expectancy of these people at given ages. However, this would take a considerable time and be of limited use in the assessment of expected average length of life and any changes that might take place over time or among populations in different geographical areas.

Methods have been devised to overcome these constraints, with a number of caveats. Life tables have been designed to estimate expected years of life using survival rates at a given age, or age range, of a given population. Life tables rely on the number of deaths for a given age and the number of people with the same age in the population. These data allow for the estimation of the chances that people will survive from one age to another, and how long they are likely to live on average. As the computation of single-year survival rates is rather cumbersome, the estimation may rely on 5 or 10-year age groups. Estimates of life expectancy are often made on the assumption that the age-specific mortality rates during a given period will prevail throughout the life of the population. For that reason, they have been labelled as synthetic measures of life expectancy.

7.3 Assumptions Made in Constructing a Life Table

It is important to maintain the distinction between a given life table, which is a model of mortality for a *synthetic* population, and the *real* population to which it is meant to apply. In constructing a life table, assumptions need to be made about the life table as a model of mortality for a synthetic population, as well as a model for a real population to which it is meant to apply.

Three major assumptions underlie the life table as a model of mortality for a synthetic population. These assumptions are that the synthetic population experiences:

Assumption 1: no migration
Assumption 2: annual age-specific death rates that do not change over time
Assumption 3: annual number of births that remains constant over time; the annual number of births chosen is usually 100,000 and accordingly the synthetic population has 100,000 deaths annually over time; thus, the synthetic population is *stationary* in that it never changes in size.

Such life table describes the experience of a synthetic cohort of 100,000 newborn in a population as they live from birth to the highest attained age. Realistically, the above conditions are unlikely to happen in a real population. A life table should be taken as an indication of what would happen if its major assumptions prevailed.

However, these requirements can be substantially relaxed in practice and the assumptions underlying the use of a given life table for a real population can be stated as follows:

Assumption 1(a): the real population to which a life table is applied can experience migration, but the migrants as a whole have the same mortality regime specified by the life table

Assumption 2(a): the real population to which a life table is applied has the same annual age-specific death rates used in constructing the life table and this mortality regime remains constant over the period of time for which the life table is meant to apply

Assumption 3(a): the annual number of births in the real population to which a life table is applied may be quite different from the (constant) number of births used in the life table and this number may change over time; however, those born are subjected to the same mortality regime underlying the life table over the period of time for which the life table is meant to apply.

In using a life table, one must be cognizant of both sets of assumptions. If the mortality rates in a real population are changing rapidly then the length of time for which a given life table can be applicable is limited. This may require adjustments to the underlying mortality schedule on which the life table is based.

Life tables are usually constructed separately for males and females. However, they can also be prepared for both sexes combined.

7.4 Data Requirements

The basic data required are the number of:

- deaths by the age of the deceased over a period of time
- people classified by age who were alive in the middle of the same year.

Generally, the information on deaths is obtained from death registers while population figures are taken from population censuses or population registers. In countries where the death registration is incomplete, sample surveys are conducted to estimate the data required for the construction of life tables.

7.5 Constructing a Complete Life Table

A complete life table depicts the mortality and survival experience of a hypothetical population of 100,000 newborn as they progress in life in single years of age, *i.e.,* from birth to age 1, age 2 and so on, till every one of them has died. The concepts and formulae used in the computation of complete life tables are given below.

7.5.1 Age Specific Death Rates

As discussed in Sect. 6.6.2, the age specific mortality rate at each age x, (m_x), is derived by taking in the numerator the number of deaths of persons aged x, (d_x), usually over a period of 1 year, and dividing it by the number of people aged x, (P_x), in the middle of the same year:

$$m_x = \frac{d_x}{P_x} \tag{7.1}$$

In the above equation two points are worth noting: (a) age x may take values 0, 1, 2 and so on, and (b) x refers to age x to $< x + 1$.

7.5.2 Probability of Dying

The probability of dying (q_x) between age x and $x + 1$ is defined as the number of deaths (d_x) at that age in a given year divided by the population *exposed-to-the risk* of dying at the same age within the same year. Obviously, the midyear population (P_x) has already survived the risk of dying during the first half of that year. Therefore, it is the population aged x at the beginning of the year which is the true population *exposed-to-the-risk* of dying within that year. To estimate this, it is assumed that deaths at any age are uniformly distributed within a given year.

Under the above assumption half of the deaths at any age x occurs in the first 6 months of a given year and the remaining half in the second 6 months. Therefore, the population *exposed-to-the-risk* of dying between age x and $x + 1$ at the beginning of a given year is the population aged x in the middle of the same year plus half of the deaths aged x in the given year.

Algebraically q_x may be expressed as:

$$q_x = \frac{d_x}{P_x + \frac{1}{2} d_x} \tag{7.2}$$

Dividing both the numerator and denominator of the above equation by (P_x) will result in Eq. (7.3):

$$q_x = \frac{\frac{d_x}{P_x}}{\frac{P_x}{P_x} + \frac{1}{2}\left(\frac{d_x}{P_x}\right)} \tag{7.3}$$

Using Eq. (7.1), $(\frac{d_x}{P_x})$ can be replaced by (m_x), therefore Eq. (7.3) can be re-written as:

$$q_x = \frac{m_x}{1 + \frac{1}{2}m_x} \quad \text{simplified as :}$$

$$q_x = \frac{2m_x}{2 + m_x} \tag{7.4}$$

This is the general equation to convert the age-specific death rate (m_x) to the corresponding probability of dying (q_x). This equation is used only for ages 1 *and over*. For age 0, the following equation is often used:

$$q_0 = \frac{d_0}{B} \tag{7.5}$$

where (d_0) is the number of deaths between ages 0 and 1 (*i.e.*, infant deaths) in a given year and (B) refers to the number of live births in that year. While more complicated formulae involving *separation factors* can also be used for converting m_0 to q_0 (see for example Kintner 2004), it is quite common to take m_0 as a reasonable approximation of q_0.

On the other hand, at the oldest attainable age ω (usually considered to be 100+), everybody must die. Thus:

$$q_\omega = 1 \tag{7.6}$$

Given that an x year old person will either die before reaching age $x + 1$, survive to the next age $x + 1$, it is certain that $(q_x + p_x) = 1$, and

$$p_x = 1 - q_x \tag{7.7}$$

while (q_x) is defined as the probability of dying between age x and $x + 1$, (p_x) is the probability of surviving to age $x + 1$.

7.5.3 *Number of Deaths and Survivors*

Given a set of (q_x) values and starting with, say, 100,000 births in a hypothetical population, the number of deaths (d_x) between age x and $x + 1$ can be estimated, and survivors (l_x) at each age x, (where $x = 0, 1, 2, 3\ldots$) till all of the 100,000 people have died. For this purpose, the following equations are used:

$$d_x = q_x * l_x \tag{7.8}$$

$$l_{x+1} = l_x - d_x \tag{7.9}$$

Equation (7.9) can be re-written as:

$$d_x = l_x - l_{x+1} \tag{7.10}$$

The value of (l_0), the number of survivors at age 0, is the number of births started with. This number is called the *radix* of the life table.

7.5.4 Person-Years Lived and Life Expectancy

Person-years lived during a given year by (l_x) people alive at the beginning of that year is denoted by (L_x). This can be estimated as shown below.

Let (l_x) be the number of survivors aged x at the beginning of a given year say t. Of these (d_x) will die within that year, so that (l_{x+1}) would be alive aged $x + 1$ at the beginning of the next year $t + 1$. Obviously, (l_{x+1}) lived for 1 whole year during the year t. To this, one must add the average period lived by the people who died aged x (d_x) in year t. Using the assumption regarding the uniformity of the distribution of deaths (Sect. 7.5.2), it can be argued that, on the average, those who died at age x (d_x) lived for half a year each during the year t. Thus:

$$L_x = l_{x+1} + \frac{1}{2} d_x$$

Substituting the value of (d_x) from Eq. (7.10)

$$L_x = l_{x+1} + \frac{1}{2}(l_x - l_{x+1})$$

$$L_x = \frac{1}{2}\ (l_x + l_{x+1}) \tag{7.11}$$

$$L_x = 0.5 l_x + 0.5 l_{x+1} \tag{7.12}$$

If the assumption in Sect. 7.5.2 holds then:

$$L_x = l_{x+0.5} \tag{7.13}$$

Because deaths are not uniformly distributed in early years of life (particularly the first year), (L_0) may be calculated as follows:

$$L_0 = 0.3 l_0 + 0.7 l_1 \tag{7.14}$$

The coefficients 0.3 and 0.7 may vary from population to population depending upon the relative age distribution of infant deaths within a given year. In many cases, these coefficients may be assumed to be 0.5 each, and Eq. (7.12) is used.

If age (x+) is an open interval, such as age interval (100+), Eqs. (7.11) or (7.12) cannot be used for this age. In such cases, the estimate of (L_{x+}) is obtained by the following equation:

$$L_{x+} = \frac{d_{x+}}{m_{x+}} \tag{7.15}$$

In the absence of data on (d_{x+}) and (m_{x+}) an alternative, though less accurate, estimate can be obtained as follows:

$$L_{x+} = l_x * log(l_x) \tag{7.16}$$

Using (L_x), it is a simple calculation to obtain (T_x) that is defined as the total person-years lived beyond age x

$$T_x = L_x + L_{x+1} + L_{x+2} + \ldots + L_\omega \tag{7.17}$$

where ω is the highest attainable age. This can be done rather easily by cumulating (L_x) from age ω to age x.

Since (T_x) is the total person-years lived beyond age x, and (l_x) is the number of persons alive at age x, the average number of person-years lived beyond age x, also referred to as the life expectancy at age x (e_x), is calculated as follows:

$$e_x = \frac{T_x}{l_x} \tag{7.18}$$

7.5.5 Characteristics of Life Table Populations

As mentioned earlier, life tables represent hypothetical populations that will evolve as a result of the following conditions affecting populations over long periods of time:

- population will remain closed to migration
- every year it will be augmented by a constant number of births
- it will experience a constant schedule of mortality (q_x) every year
- deaths will generally occur uniformly within each year.

Under these conditions, the characteristics of such a hypothetical population are as follows:

- L_x is the number of persons at each age
- T_0 is the total size of the population
- l_0 is the number of births
- d_x is the number of deaths at each age
- the crude birth rate $= \frac{l_0}{T_0}$

To calculate the crude death rate, the total number of deaths needs to be estimated.

$$\sum_{x=0}^{x=\omega} d_x = d_0 + d_1 + d_2 + \cdots + d_{\omega-1}$$

where ($d_{\omega-1}$) is the number of deaths in the last year of life before every one has died.

Using Eq. (7.10) and substituting the values of d_x results in the following:

$$\sum_{x=0}^{x=\omega} d_x = (l_0 - l_1) + (l_1 - l_2) + (l_2 - l_3) + \cdots = l_0$$

It is obvious that (l_1), (l_2), (l_3) and the remaining terms will all cancel out as for each of them there is a positive and a negative value. Thus,

$$\text{the crude death rate} = \frac{l_0}{T_0}$$

The growth rate of a life table population will always be zero because of the equality of its birth and death rates and the population being closed to migration. Given the zero growth of the life table population it is also referred to as a *stationary* population.

The age specific death rate at any age x in a life table population is equal to $\frac{d_x}{l_x}$. This also remains constant. Theoretically, it is not appropriate to calculate the age specific birth rates in a life table population. However, as noted before, the total number of births in such a population remains constant.

The average age at death for persons of a given age x can be estimated as $(x + e_x)$.

7.5.6 Graphs of q_x, l_x, d_x and e_x

Figures 7.1, 7.2, 7.3, and 7.4 show the general shapes of the life table columns based on the life tables presented in Table A7.2 for males and Table A7.3 where females data are shown. These general shapes are applicable to most life tables irrespective

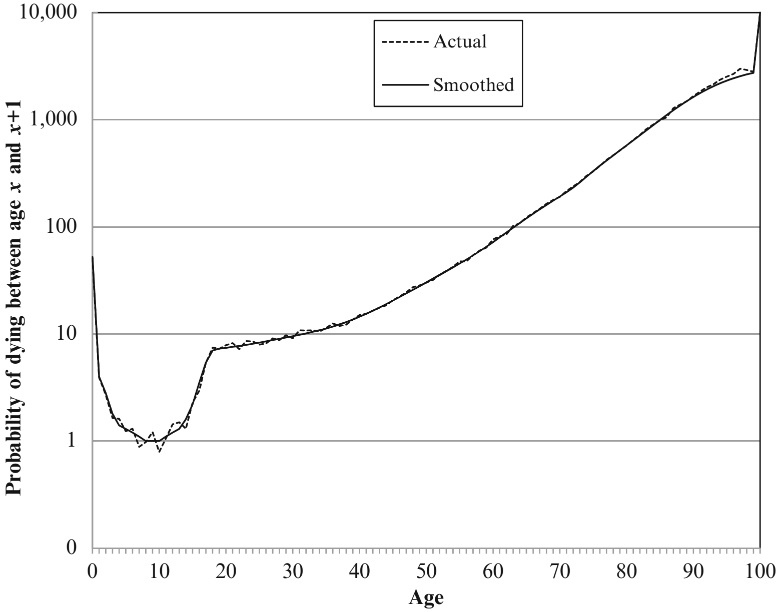


Fig. 7.1 Actual and smoothed values of age-specific probabilities of dying between age x and $x + 1$ per 10,000 on a logarithmic scale for Australian males, 2005–2007 (*Source*: Table A7.2 for actual and Australia (2009) for smoothed data)

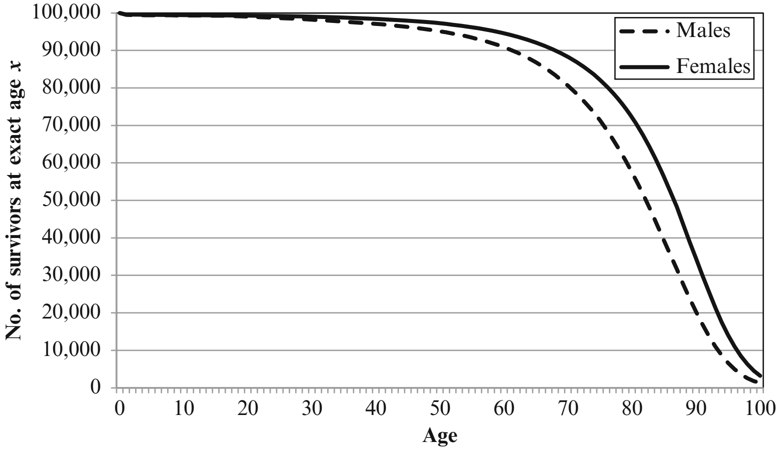


Fig. 7.2 No. of survivors at exact age x (l_x) for Australian males and females, 2005–2007 (*Source*: Tables A7.2 and A7.3)

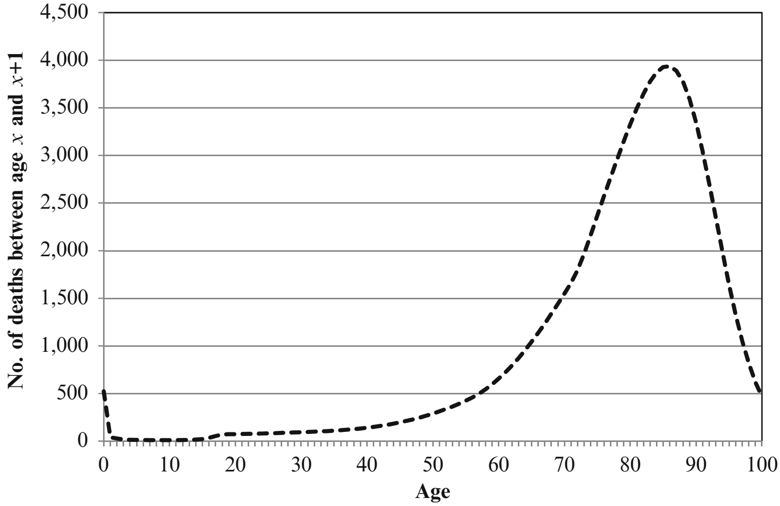


Fig. 7.3 No. of deaths between age x and $x + 1$ (d_x) for Australian males, 2005–2007 (*Source*: Table A7.2)

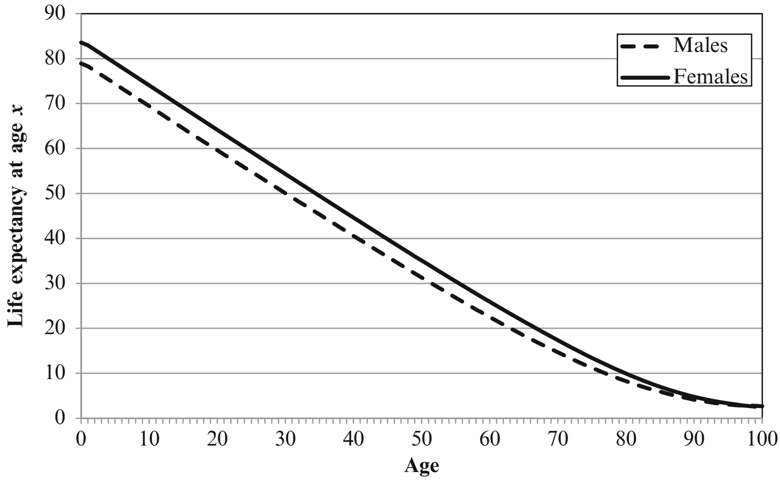


Fig. 7.4 Life expectancy at age x (e_x) for Australian males and females, 2005–2007 (*Source*: Tables A7.2 and A7.3)

of gender, the population/country or period of time. The general shape of the (L_x) column is very similar to (l_x), while that of (T_x) is very similar to (e_x) column. Therefore, graphs for these life table columns are not presented here.

The female life table columns are very similar in shape compared to those for the male life tables. However, female (q_x) and (d_x) values are lower than those for males. Consequently, (L_x), (T_x) and (e_x) values for female life tables are generally higher.

Table 7.1 Extract from the complete life table for Australian males, 2005–2007

x	m_x	q_x	d_x	l_x	L_x	T_x	e_x
0	0.00512	0.00506	506	100,000	99,646	7,890,093	78.9
1	0.00039	0.00039	39	99,494	99,475	7,790,447	78.3
2	0.00027	0.00027	27	99,455	99,442	7,690,972	77.3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	0.34386	0.29341	739	2,517	2,148	7,088	2.8
99	0.32830	0.28201	501	1,778	1,528	4,940	2.8
100+	0.37428	1.00000	1,277	1,277	3,412	3,412	2.7

Source: Table A7.2

7.5.7 Smoothing of Data

Depending on the quality of data, it is common practice to smooth the irregularities in the data by using mathematical or other models, as was done in the Australian life tables (Australia 2009). Figure 7.1 shows the differences in the actual values estimated by the authors and the smoothed values from the official life tables. The life expectancies at birth calculated in the official life tables were 79.0 for males and 83.7 for females. These figures were only 0.1 year larger than the corresponding figures of 78.9 and 83.6 for males and females respectively, as given in Tables A7.2 and A7.3.

7.5.8 Example of a Complete Life Table

An extract from a complete life table for Australian males is shown in Table 7.1. The full life tables for males and females are given in Tables A7.2 and A7.3.

The input data were the registered deaths in Australia by age and sex during the 3-year period January 1, 2005 to December 31, 2007 and the 2006 census population. Although the census was conducted as on August 6, 2006, the population was assumed to be the mid-2006 population.

The average annual death rates (m_x) for males were calculated by dividing the annual average of 3-year deaths by the midyear population. Using Eq. (7.1) and data from Table A7.1, the (m_x) values for males at various ages were calculated as:

$$m_0 = \frac{2,107 \div 3}{137,142} = 0.00512$$

$$m_1 = \frac{1,577 \div 3}{133,215} = 0.00039$$

$$\cdots \qquad \cdots \qquad \cdots$$

$$\cdots \qquad \cdots \qquad \cdots$$

$$m_{99} = \frac{326 \div 3}{331} = 0.32830$$

$$m_{100} = \frac{521 \div 3}{464} = 0.37428$$

The (m_x) values were converted to the corresponding (q_x) values as shown below:

$$q_0 = \frac{2,107 \div 3}{416,576 \div 3} = 0.00506 \quad \text{Eq. (7.5)}$$

$$q_1 = \frac{2 * 0.00039}{2 + 0.00039} = 0.00039 \quad \text{Eq. (7.4)}$$

$$\ldots \quad \ldots \quad \ldots$$

$$\ldots \quad \ldots \quad \ldots$$

$$q_{99} = \frac{2 * 0.32830}{2 + 0.32830} = 0.28201 \quad \text{Eq. (7.4)}$$

$$q_{100+} = 1.00000 \quad \text{Eq. (7.6)}$$

To calculate the number of deaths (d_x) and survivors (l_x) at each age x, Eqs. (7.8) and (7.9) were applied. Starting with 100,000 newborn (l_0):

$$l_0 = 100{,}000 \qquad d_0 = 0.00506 * 100{,}000 = 506$$

$$l_1 = 100{,}000 - 506 = 99{,}494 \qquad d_1 = 0.00039 * 99{,}494 = 39$$

$$\ldots \quad \ldots \quad \ldots$$

$$\ldots \quad \ldots \quad \ldots$$

$$L_{99} = 2{,}517 - 739 = 1{,}778 \qquad d_{99} = 0.28201 * 1{,}778 = 501$$

$$L_{100+} = 1{,}778 - 501 = 1{,}277 \qquad d_{100+} = 1.00000 * 1{,}277 = 1{,}277$$

Equations (7.12), (7.14) and (7.15) were used for calculating the (L_x) values as:

$$L_0 = (0.3 * 100{,}000) + (0.7 * 99{,}494) = 99{,}646 \quad \text{Eq. (7.14)}$$

$$L_1 = 0.5 * (99{,}494 + 99{,}455) = 99{,}475 \quad \text{Eq. (7.12)}$$

$$\ldots \quad \ldots \quad \ldots$$

$$\ldots \quad \ldots \quad \ldots$$

$$L_{99} = 0.5 * (1{,}778 + 1{,}277) = 1{,}528 \quad \text{Eq. (7.12)}$$

$$L_{100+} = 1{,}277 \div 0.37428 = 3{,}412 \quad \text{Eq. (7.15)}$$

The (T_x) values were obtained by cumulating (L_x) values starting from age 100+ going upwards towards age 0 as shown below:

$$T_{100+} = 3{,}412$$

$$T_{99} = 3{,}412 + 1{,}528 = 4{,}940$$

$$T_{98} = 4{,}940 + 2{,}148 = 7{,}088$$

$$\ldots \quad \ldots \quad \ldots$$

$$\ldots \quad \ldots \quad \ldots$$

$$T_1 = 7{,}691{,}198 + 99{,}475 = 7{,}790{,}673$$

$$T_0 = 7{,}790{,}673 + 99{,}646 = 7{,}890{,}319$$

Finally, the (e_x) values were estimated for each age x using Eq. (7.18) as follows:

$$\begin{aligned}
\text{Age } 0 &= \frac{7,890,319}{100,000} = 78.9 \\
\text{Age } 1 &= \frac{7,790,673}{99,494} = 78.3 \\
\ldots & \quad \ldots \quad \ldots \\
\ldots & \quad \ldots \quad \ldots \\
\text{Age } 99 &= \frac{4,940}{1,778} = 2.8 \\
\text{Age } 100 &= \frac{3,412}{1,277} = 2.7
\end{aligned}$$

7.6 Constructing an Abridged Life Table

An abridged life table, like a complete life table, depicts the mortality and survival experience of a hypothetical population of 100,000 newborn as they progress in life in age groups of different size, *i.e.*, 5 year age groups, 10 year age groups or any other age groups. The first year of life is usually kept separate. As in the case of other life tables, generally, abridged life tables are calculated for males and females separately, they can also be calculated for both sexes combined.

Basically, the concepts and equations used in the computation of abridged life tables are the same as those for the complete life tables as shown below:

Complete life table	Abridged life table	
m_x	${}_nm_x$	Death rate between age x and $x + n$
q_x	${}_nq_x$	Probability of dying between age x and $x + n$
p_x	p_x	Probability of survival to exact age $x + n$
l_x	l_x	Survivors at age x
d_x	${}_nd_x$	Deaths between age x and $x + n$
L_x	${}_nL_x$	Person-years lived between age x and $x + n$
T_x	T_x	Person-years lived beyond age x
e_x	e_x	Life expectancy beyond age x

Given below are the formulae for the abridged life tables that are different from the complete life tables:

$$ {}_nm_x = \frac{{}_nd_x}{{}_nP_x} \tag{7.1a}$$

$$ {}_nq_x = \frac{2 * n*{}_nm_x}{2 + (n*{}_nm_x)} \tag{7.4a} $$

$$ {}_nd_x = {}_nq_x * l_x \tag{7.8a} $$

$$ l_{x+n} = l_x - {}_nd_x \tag{7.9a} $$

$$ {}_nL_x = \frac{n}{2}(l_x + l_{x+n}) \tag{7.12a} $$

7.6.1 Graphs of the Abridged Life Table Columns

The shape of abridged life table columns is the same as shown and described in Sect. 7.5.6. The comments made in Sect. 7.5.7 regarding the smoothing of data also apply to the abridged life tables.

7.6.2 Example of an Abridged Life Table

Table 7.2 shows an extract from the abridged life table for Australian males (Table A7.4). The corresponding abridged life table for females is given in Table A7.5.

The input data were the registered deaths in Australia by age and sex during the 3-year period January 1, 2005 to December 31, 2007 and the 2006 census population. As stated in Sect. 7.5.8, although the census was conducted on August 6, 2006, the population was assumed to be the mid-2006 population (see Table A7.1).

The average annual death rates for males were calculated by dividing the annual average of 3-year deaths by the midyear population. Using the data from Table A7.1 and Eq. (7.1a), the (${}_nm_x$) values for males at various ages were calculated as:

$$
\begin{aligned}
m_0 &= \frac{2,107 \div 3}{137,142} = 0.00512 \\
{}_4m_1 &= \frac{393 \div 3}{528,707} = 0.00025 \\
\ldots &\quad \ldots \quad \ldots \\
\ldots &\quad \ldots \quad \ldots \\
{}_5m_{95} &= \frac{4,082 \div 3}{4,322} = 0.31482 \\
m_{100+} &= \frac{521 \div 3}{464} = 0.37428
\end{aligned}
$$

Table 7.2 Extract from the abridged life table for Australian males, 2005–2007

x	n	${}_nm_x$	${}_nq_x$	${}_nd_x$	l_x	${}_nL_x$	T_x	e_x
0	1	0.00512	0.00506	506	100,000	99,646	7,892,971	78.9
1	4	0.00025	0.00100	99	99,494	397,778	7,793,325	78.3
5	5	0.00011	0.00055	55	99,395	496,838	7,395,547	74.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	5	0.21163	0.69202	14,357	20,746	67,838	87,746	4.2
95	5	0.31482	0.88084	5,628	6,389	17,875	19,908	3.1
100+	5+	0.37428	1.00000	761	761	2,033	2,033	2.7

Source: Table A7.4

The (${}_nm_x$) values were converted to the corresponding (${}_nq_x$) values as shown below:

$$q_0 = \frac{2,107 \div 3}{416,576 \div 3} = 0.00506 \quad \text{Eq. (7.5)}$$
$${}_4q_1 = \frac{2 * 4 * 0.00025}{2 + (4 * 0.00035)} = 0.00100 \quad \text{Eq. (7.4a)}$$
$$\ldots \qquad \ldots \qquad \ldots$$
$$\ldots \qquad \ldots \qquad \ldots$$
$${}_5q_{95} = \frac{2 * 5 * 0.31482}{2 + (5 * 0.31482)} = 0.88084 \quad \text{Eq. (7.4a)}$$
$$q_{100+} = 1.00000 \quad \text{Eq. (7.6)}$$

To calculate the number of survivors (l_x) at each exact age x, and the number of deaths (${}_nd_x$) between age x and $x + n$, Eqs. (7.9a) and (7.8a) were applied respectively. Starting with 100,000 newborn (l_0):

$$l_0 = 100{,}000 \qquad d_0 = 0.00506*100{,}000 = 506$$
$$l_1 = 100{,}000 - 506 = 99{,}494 \qquad {}_4d_1 = 0.00100*99{,}494 = 99$$
$$\ldots \qquad \ldots \qquad \ldots$$
$$\ldots \qquad \ldots \qquad \ldots$$
$$l_{90} = 20{,}744 - 14{,}355 = 6{,}389 \qquad {}_5d_{90} = 0.88084*6{,}389 = 5{,}628$$
$$l_{100+} = 6{,}389 - 5{,}628 = 761 \qquad d_{100+} = 1.00000*761 = 761$$

Eqs. (7.14), (7.12a) and (7.16) were used for calculating the (L_x) values as follows:

$$L_0 = (0.3 * 100{,}000) + (0.7*99{,}494) = 99{,}646 \quad \text{Eq. (7.14)}$$
$${}_4L_1 = (4 \div 2) * (99{,}494 + 99{,}395) = 397{,}778 \quad \text{Eq. (7.12a)}$$
$$\ldots \qquad \ldots \qquad \ldots$$
$$\ldots \qquad \ldots \qquad \ldots$$
$${}_5L_{95} = (5 \div 2) * (6{,}389 + 761) = 17{,}875 \quad \text{Eq. (7.12a)}$$
$$L_{100+} = 761 \div 0.37428 = 2{,}033 \quad \text{Eq. (7.15)}$$

Values of the (T_x) and (e_x) columns were obtained using Eqs. (7.17) and (7.18) respectively.

7.7 Other Methods of Constructing Abridged Life Tables

Greville's method provides another way of building an abridged life table. This method uses a more complicated formula to convert (${}_nm_x$) to (${}_nq_x$) values (Greville 1943; Kintner 2004). The formula to convert (${}_nm_x$) to (${}_nq_x$) values is:

$$ {}_nq_x = \frac{{}_nm_x}{\frac{1}{n} + \left[{}_nm_x\left(0.5 + \frac{n}{12}\right) * \{{}_nm_x - 0.95\}\right]} \tag{7.19}$$

where n is the width of the age interval. This formula is based on the assumption that the mortality pattern of a population follows the Gompertz model. This model is described in Sect. 11.3.3.

For purposes of illustration, the value of (${}_5\, m_5$) in Table 7.2, 0.00011, has been converted to (${}_5q_5$) using Greville's method as shown below:

$$ {}_5q_5 = \frac{0.00011}{\frac{1}{5} + \left[0.00011\left(0.5 + \frac{5}{12}\right) * \{0.00011 - 0.95\}\right]} = 0.00055.$$

Fergany (1971) suggested another somewhat simpler method of converting (${}_nm_x$) to (${}_nq_x$) values:

$$ {}_nq_x = 1 - e^{(-n * {}_nm_x)} \tag{7.20}$$

where n is the width of the age interval. Again, the value of (${}_5\, m_5$) in Table 7.2, 0.00011, provides an example in applying Eq. (7.22):

$$ {}_5q_5 = 1 - e^{(-5*0.00011)} = 0.00055.$$

Both of these methods give an estimate of (${}_5q_5$) identical to 0.00055 given in Table 7.2. A comparison of (${}_nq_x$) values estimated by using the Fergany and Greville methods *vis-a-vis* the one proposed in Eq. (7.4a) did not reveal any significant differences between the results obtained from the three methods, and the life expectancy at birth (e_0) estimated through them were practically identical (Table 7.3).

Table 7.3 Comparison of ${}_nq_x$ values from abridged life tables for Australia, 2005–2007 using three alternative methods of computation

	${}_nq_x$ values for males				${}_nq_x$ values for females		
x	This book	Fergany	Greville	x	This book	Fergany	Greville
5	0.00055	0.00044	0.00044	5	0.00045	0.00036	0.00036
10	0.00060	0.00060	0.00060	10	0.00045	0.00045	0.00045
20	0.00404	0.00404	0.00404	20	0.00150	0.00150	0.00150
30	0.00519	0.00519	0.00519	30	0.00215	0.00215	0.00215
40	0.00827	0.00827	0.00827	40	0.00479	0.00479	0.00479
50	0.01769	0.01769	0.01770	50	0.01099	0.01099	0.01099
60	0.04401	0.04400	0.04408	60	0.02684	0.02683	0.02686
70	0.11392	0.11379	0.11430	70	0.06922	0.06919	0.06938
80	0.31259	0.30961	0.31345	80	0.21814	0.21718	0.21906
90	0.69202	0.65290	0.66950	90	0.59841	0.57425	0.58731
e_0 (males)	78.9	78.9	79.0	e_0 (females)	83.6	83.6	83.6

7.8 Uses of Life Tables

Life tables represent one of the most commonly used tools, not only in demography but other areas of scientific research. Life tables can be prepared not only for human populations as well as for other living organisms, as well as non-living entities.

7.8.1 Population Projections

For population projections, the proportion of people of a given age who will survive over a specific period of time needs to be estimated. These proportions, called survival ratios, are generally estimated by using (L_x) column of the life table. For example, the survival ratio of persons aged (x to $x + n$) over a five-year period can be estimated by taking (L_x) values for the age group ($x + 5$ to $x + n + 5$) and dividing it by (L_x) values for the age group (x to $x + n$). Section 11.4 describes the use of life tables in population projections in more detail.

7.8.2 Other Types of Life Tables

Life tables discussed in this chapter involve only one decrement: death. Life tables that involve more than one decrement can be prepared. An example is nuptiality tables, where the experience of, say, a group of 20-year old single people may be followed. As they go to the next age, some will die while of those who survive some will remain single and others will get married. In this case, there are two decrements: death and change of marital status. Another example of multiple decrement tables is working life tables and cause of death tables. This topic is addressed in Chap. 10.

Cohort life tables can be compiled, if data on the mortality experience of a particular birth cohort is available. However, this type of data is not easily available in some countries. In addition, for populations experiencing negligible migration, one can use *cohort survivorship rates* to construct applicable life tables (Swanson & Tedrow 2011; United Nations 1982).

7.8.3 Insurance

Insurance companies prepare life tables to decide how much premium they will charge their customers for life and other insurance products. The people who do this work are usually actuaries. More recently, they have incorporated data on life-style risk factors such as smoking to adjust their premiums. Life tables used by insurance companies are generally based on the mortality experience of policy holders rather than the general population. As discussed in Chap. 1, the idea of a life table can be traced back to John Graunt, who was not an actuary. Historically, much of the development and refinement of life tables was done by people who worked in the insurance industry (Elderton & Fippard 1914; Hull 1899).

7.8.4 Epidemiological Research

Life table methodology is used in epidemiological research when testing the efficacy of drugs and treatments of different types of medical conditions. The assessment of the relative importance of mortality and disability from specific causes involves life table methodology. This also applies to the impact of the possible eradication of some diseases such as polio (see Chap. 10).

7.8.5 Other Industrial and Commercial Uses

Life table methodology is employed to estimate the usefulness and quality of manufactured goods. For example, a particular type of light bulb may last only 2,000 hours. This is similar to life expectancy and its estimation involves life table methodology. As noted by Keyfitz (1985), a watch can run only as long as its individual parts operate and each of the parts has a working life expectancy, for which an individual life table can be constructed. Life tables can also be applied in the development of maintenance schedules for machinery, such as aircraft engines and other parts (Altman & Goor 1946).

7.9 Availability of International Life Tables

Life tables for countries, and in some cases even particular provinces/regions of countries, are usually calculated by the national statistical agencies. In addition, many international organisations such as the World Health Organization (WHO 2013) produce life tables for its member countries.

7.10 International Comparisons

Sex-specific life expectancies at selected ages for four countries belonging to different regions of the world give examples of the variation in life expectancy at different ages. These life tables were compiled by the United Nations based on information provided by the statistical agencies of the member countries. Quality of mortality data varies from country to country. The figures indicate that the life expectancies of males are generally lower than those of females. Although the selected countries are not necessarily representative of their region (for instance, Japan in Asia is the country with the longest average life expectancy in the world), it appears that mortality levels in Africa, Asia and South America are generally higher than those in Europe (Table 7.4).

Table 7.4 Life expectancy at various ages in selected countries

	Life expectancy at age								
Country, year and (region)	0	10	20	30	40	50	60	70	80
Botswana, 2006 (Africa)									
Males	54.0	54.0	45.6	37.9	30.0	22.4	15.4	9.4	4.8
Females	66.0	61.4	52.3	43.4	34.7	26.2	18.2	11.2	5.8
Brazil, 2009 (South America)									
Males	69.4	61.8	52.3	43.6	35.0	26.9	19.5	13.4	9.0
Females	77.0	68.9	59.1	49.5	40.0	31.1	22.8	15.6	10.0
France, 2006–2008 (Europe)									
Males	77.4	67.8	58.0	48.5	39.0	30.0	21.9	14.5	8.2
Females	84.3	74.7	64.8	55.0	45.2	35.8	26.8	18.2	10.5
Myanmar, 2008 (Asia)									
Males	64.3	56.9	46.9	38.2	30.9	24.0	17.4	10.9	5.9
Females	68.3	60.4	51.0	42.2	33.9	25.8	18.1	10.9	5.3

Source: United Nations (2012), Table 21

Appendices

Appendix 7.1: Input Data for Life Tables

Table A7.1 Deaths, population, m_x and q_x by age and sex for Australia

Age	Deaths, 2005–2007		Population, 2006		m_x		q_x	
x	Males	Females	Males	Females	Males	Females	Males	Females
0	*2,107*	*1,667*	*137,142*	*130,510*	*0.00039*	*0.00031*	*0.00506*	*0.00423*
1	157	117	133,215	125,413	0.00027	0.00020	0.00027	0.00020
2	107	73	131,863	124,538	0.00016	0.00016	0.00016	0.00016
3	65	59	131,550	124,761	0.00016	0.00011	0.00016	0.00011
4	64	43	132,079	125,691	0.00039	0.00031	0.00039	0.00031
1–4	*393*	*292*	*528,707*	*500,403*	*0.00025*	*0.00019*	*0.00125*	*0.00095*
5	50	41	134,757	127,887	0.00012	0.00011	0.00012	0.00011
6	53	43	135,688	129,309	0.00013	0.00011	0.00013	0.00011
7	36	32	135,644	128,760	0.00009	0.00008	0.00009	0.00008
8	40	32	136,510	128,600	0.00010	0.00008	0.00010	0.00008
9	50	33	137,324	130,009	0.00012	0.00008	0.00012	0.00008
5–9	*229*	*181*	*679,923*	*644,565*	*0.00011*	*0.00009*	*0.00055*	*0.00045*
10	33	22	138,726	130,985	0.00008	0.00006	0.00008	0.00006
11	45	29	141,998	134,268	0.00011	0.00007	0.00011	0.00007
12	61	27	141,896	134,837	0.00014	0.00007	0.00014	0.00007
13	64	48	142,339	135,354	0.00015	0.00012	0.00015	0.00012
14	56	51	143,270	135,055	0.00013	0.00013	0.00013	0.00013
10–14	*259*	*177*	*708,229*	*670,499*	*0.00012*	*0.00009*	*0.00060*	*0.00045*
15	99	77	144,257	135,432	0.00023	0.00019	0.00023	0.00019
16	128	94	144,591	135,537	0.00030	0.00023	0.00030	0.00023
17	228	106	141,421	133,676	0.00054	0.00026	0.00054	0.00026
18	313	111	139,874	132,252	0.00075	0.00028	0.00075	0.00028
19	310	113	141,924	134,690	0.00073	0.00028	0.00073	0.00028
15–19	*1,078*	*501*	*712,067*	*671,587*	*0.00050*	*0.00025*	*0.00250*	*0.00125*
20	342	125	145,217	138,317	0.00079	0.00030	0.00079	0.00030
21	359	121	145,795	141,713	0.00082	0.00028	0.00082	0.00028
22	315	120	145,722	141,470	0.00072	0.00028	0.00072	0.00028
23	382	122	148,494	144,190	0.00086	0.00028	0.00086	0.00028
24	378	142	147,772	142,113	0.00085	0.00033	0.00085	0.00033
20–24	*1,776*	*630*	*733,000*	*707,803*	*0.00081*	*0.00030*	*0.00404*	*0.00150*
25	346	120	145,090	139,326	0.00079	0.00029	0.00079	0.00029
26	345	127	140,700	137,402	0.00082	0.00031	0.00082	0.00031
27	374	124	136,920	136,037	0.00091	0.00030	0.00091	0.00030
28	359	144	136,577	136,554	0.00088	0.00035	0.00088	0.00035
29	400	176	136,913	137,954	0.00097	0.00043	0.00097	0.00043
25–29	*1,824*	*691*	*696,200*	*687,273*	*0.00087*	*0.00034*	*0.00434*	*0.00170*

(continued)

Table A7.1 (continued)

Age	Deaths, 2005–2007		Population, 2006		m_x		q_x	
x	Males	Females	Males	Females	Males	Females	Males	Females
30	378	160	138,033	138,300	0.00091	0.00039	0.00091	0.00039
31	457	178	140,798	141,827	0.00108	0.00042	0.00108	0.00042
32	468	194	144,884	146,492	0.00108	0.00044	0.00108	0.00044
33	486	191	149,262	151,549	0.00109	0.00042	0.00109	0.00042
34	495	234	156,961	158,659	0.00105	0.00049	0.00105	0.00049
30–34	*2,284*	*957*	*729,938*	*736,827*	*0.00104*	*0.00043*	*0.00519*	*0.00215*
35	543	238	158,929	160,395	0.00114	0.00049	0.00114	0.00049
36	572	265	152,278	152,353	0.00125	0.00058	0.00125	0.00058
37	532	275	149,190	151,315	0.00119	0.00061	0.00119	0.00061
38	530	326	144,668	145,683	0.00122	0.00075	0.00122	0.00075
39	585	329	142,998	144,099	0.00136	0.00076	0.00136	0.00076
35–39	*2,762*	*1,433*	*748,063*	*753,845*	*0.00123*	*0.00063*	*0.00613*	*0.00315*
40	646	369	143,664	145,076	0.00150	0.00085	0.00150	0.00085
41	674	408	145,606	147,990	0.00154	0.00092	0.00154	0.00092
42	754	407	152,324	152,849	0.00165	0.00089	0.00165	0.00089
43	810	482	153,253	156,110	0.00176	0.00103	0.00176	0.00103
44	852	515	153,474	155,162	0.00185	0.00111	0.00185	0.00111
40–44	*3,736*	*2,181*	*748,321*	*757,187*	*0.00166*	*0.00096*	*0.00827*	*0.00479*
45	940	575	151,534	155,746	0.00207	0.00123	0.00207	0.00123
46	997	602	148,662	151,727	0.00224	0.00132	0.00224	0.00132
47	1,064	670	145,121	146,697	0.00244	0.00152	0.00244	0.00152
48	1,160	701	141,342	144,902	0.00274	0.00161	0.00274	0.00161
49	1,194	756	140,060	141,505	0.00284	0.00178	0.00284	0.00178
45–49	*5,355*	*3,304*	*726,719*	*740,577*	*0.00246*	*0.00149*	*0.01222*	*0.00742*
50	1,263	788	139,951	140,443	0.00301	0.00187	0.00301	0.00187
51	1,299	818	135,295	136,196	0.00320	0.00200	0.00319	0.00200
52	1,412	869	131,617	132,851	0.00358	0.00218	0.00357	0.00218
53	1,531	965	131,552	132,926	0.00388	0.00242	0.00387	0.00242
54	1,642	1,011	128,806	128,996	0.00425	0.00261	0.00424	0.00261
50–54	*7,147*	*4,451*	*667,221*	*671,412*	*0.00357*	*0.00221*	*0.01769*	*0.01099*
55	1,810	1,034	126,318	128,064	0.00478	0.00269	0.00477	0.00269
56	1,795	1,130	125,839	126,182	0.00475	0.00299	0.00474	0.00299
57	1,970	1,199	120,511	120,428	0.00545	0.00332	0.00544	0.00331
58	2,221	1,335	122,914	122,128	0.00602	0.00364	0.00600	0.00363
59	2,468	1,367	128,140	126,479	0.00642	0.00360	0.00640	0.00359
55–59	*10,264*	*6,065*	*623,722*	*623,281*	*0.00549*	*0.00324*	*0.02708*	*0.01607*
60	2,487	1,587	108,036	106,259	0.00767	0.00498	0.00764	0.00497
61	2,540	1,430	103,539	102,125	0.00818	0.00467	0.00815	0.00466
62	2,500	1,634	97,683	98,361	0.00853	0.00554	0.00849	0.00552
63	2,752	1,592	89,588	89,555	0.01024	0.00593	0.01019	0.00591
64	2,872	1,647	88,021	87,581	0.01088	0.00627	0.01082	0.00625
60–64	*13,151*	*7,890*	*486,867*	*483,881*	*0.00900*	*0.00544*	*0.04401*	*0.02684*
65	3,027	1,689	82,127	82,839	0.01229	0.00680	0.01221	0.00678
66	3,216	1,890	79,287	80,433	0.01352	0.00783	0.01343	0.00780
67	3,382	1,895	75,814	78,941	0.01487	0.00800	0.01476	0.00797
68	3,567	2,089	71,875	74,811	0.01654	0.00931	0.01640	0.00927
69	3,729	2,251	69,111	73,074	0.01799	0.01027	0.01783	0.01022
65–69	*1,6921*	*9,814*	*378,214*	*390,098*	*0.01491*	*0.00839*	*0.07187*	*0.04109*

(continued)

Table A7.1 (continued)

Age	Deaths, 2005–2007		Population, 2006		m_x		q_x	
x	Males	Females	Males	Females	Males	Females	Males	Females
70	3,825	2,332	66,313	69,944	0.01923	0.01111	0.01905	0.01105
71	3,991	2,553	61,145	66,218	0.02176	0.01285	0.02153	0.01277
72	4,257	2,704	58,941	63,666	0.02407	0.01416	0.02378	0.01406
73	4,559	2,956	57,185	62,892	0.02657	0.01567	0.02622	0.01555
74	5,048	3,401	55,513	61,449	0.03031	0.01845	0.02986	0.01828
70–74	*21,680*	*13,946*	*299,097*	*324,169*	*0.02416*	*0.01434*	*0.11392*	*0.06922*
75	5,554	3,711	55,271	62,549	0.03350	0.01978	0.03295	0.01959
76	6,059	4,145	53,469	61,965	0.03777	0.02230	0.03707	0.02205
77	6,430	4,534	49,760	59,520	0.04307	0.02539	0.04216	0.02507
78	6,805	5,089	47,596	58,196	0.04766	0.02915	0.04655	0.02873
79	7,046	5,444	44,296	55,629	0.05302	0.03262	0.05165	0.03210
75–79	*31,894*	*22,923*	*250,392*	*297,859*	*0.04246*	*0.02565*	*0.19193*	*0.12052*
80	7,255	6,019	40,960	53,834	0.05904	0.03727	0.05735	0.03659
81	7,345	6,551	36,590	51,571	0.06691	0.04234	0.06474	0.04146
82	7,335	7,109	32,686	47,593	0.07480	0.04979	0.07210	0.04858
83	7,426	7,401	28,931	44,291	0.08556	0.05570	0.08205	0.05419
84	7,339	7,973	25,918	41,305	0.09439	0.06434	0.09014	0.06233
80–84	*36,700*	*35,053*	*165,085*	*238,594*	*0.07410*	*0.04897*	*0.31259*	*0.21814*
85	6,930	8,128	22,212	37,638	0.10400	0.07198	0.09886	0.06948
86	6,295	7,971	18,884	32,957	0.11112	0.08062	0.10527	0.07750
87	5,643	7,834	13,723	25,681	0.13707	0.10168	0.12828	0.09676
88	4,925	7,602	11,052	22,456	0.14854	0.11284	0.13827	0.10681
89	4,548	7,339	9,453	20,088	0.16037	0.12178	0.14847	0.11479
85–89	*28,341*	*38,874*	*75,324*	*138,820*	*0.12542*	*0.09334*	*0.47741*	*0.37840*
90	4,168	7,373	7,658	17,495	0.18142	0.14048	0.16633	0.13126
91	3,676	7,100	6,058	15,240	0.20227	0.15529	0.18369	0.14410
92	3,103	6,624	4,628	12,245	0.22349	0.18032	0.20103	0.16541
93	2,440	5,705	3,379	9,330	0.24070	0.20382	0.21484	0.18497
94	1,898	4,752	2,352	7,279	0.26899	0.21761	0.23710	0.19626
90–94	*15,285*	*31,554*	*24,075*	*61,589*	*0.21163*	*0.17078*	*0.69202*	*0.59841*
95	1,452	3,924	1,667	5,390	0.29034	0.24267	0.25353	0.21641
96	1,086	3,118	1,165	3,946	0.31073	0.26339	0.26895	0.23274
97	728	2,369	684	2,691	0.35478	0.29345	0.30133	0.25590
98	490	1,792	475	1,868	0.34386	0.31977	0.29341	0.27569
99	326	1,268	331	1,209	0.32830	0.34960	0.28201	0.29758
95–99	*4,082*	*12,471*	*4,322*	*15,104*	*0.31482*	*0.27523*	*0.88084*	*0.81522*
100+	521	2,466	464	1,995	0.37428	0.41203	1.00000	1.00000

Sources: Australia (2008) for births; Australia (2009) for deaths, 2005–2007 and population, 2006

Notes: (1) Births, 2005–2007 were: 416,576 for males and 394,377 for females

(2) $m_x = \dfrac{Deaths, 2005-2007}{3 * Population, 2006}$. The q_x values were estimated by using Eq. (7.4) for ages 1–99 years, and Eq. (7.5) for age 0. The value of q_{100+} was assumed to be 1 for both males and females. The underlined values refer to ${}_nm_x$ and ${}_nq_x$ estimated using Eqs. (7.1a) and (7.4a) respectively

Appendix 7.2: Complete Life Tables by Sex: Australia, 2005–2007

Table A7.2 Complete life table for Australian males, 2005–2007

x	m_x	q_x	d_x	l_x	L_x	T_x	e_x
0	0.00512	0.00506	506	100,000	99,646	7,890,093	78.9
1	0.00039	0.00039	39	99,494	99,475	7,790,447	78.3
2	0.00027	0.00027	27	99,455	99,442	7,690,972	77.3
3	0.00016	0.00016	16	99,428	99,420	7,591,530	76.4
4	0.00016	0.00016	16	99,412	99,404	7,492,110	75.4
5	0.00012	0.00012	12	99,396	99,390	7,392,706	74.4
6	0.00013	0.00013	13	99,384	99,378	7,293,316	73.4
7	0.00009	0.00009	9	99,371	99,367	7,193,938	72.4
8	0.00010	0.00010	10	99,362	99,357	7,094,571	71.4
9	0.00012	0.00012	12	99,352	99,346	6,995,214	70.4
10	0.00008	0.00008	8	99,340	99,336	6,895,868	69.4
11	0.00011	0.00011	11	99,332	99,327	6,796,532	68.4
12	0.00014	0.00014	14	99,321	99,314	6,697,205	67.4
13	0.00015	0.00015	15	99,307	99,300	6,597,891	66.4
14	0.00013	0.00013	13	99,292	99,286	6,498,591	65.4
15	0.00023	0.00023	23	99,279	99,268	6,399,305	64.5
16	0.00030	0.00030	30	99,256	99,241	6,300,037	63.5
17	0.00054	0.00054	54	99,226	99,199	6,200,796	62.5
18	0.00075	0.00075	74	99,172	99,135	6,101,597	61.5
19	0.00073	0.00073	72	99,098	99,062	6,002,462	60.6
20	0.00079	0.00079	78	99,026	98,987	5,903,400	59.6
21	0.00082	0.00082	81	98,948	98,908	5,804,413	58.7
22	0.00072	0.00072	71	98,867	98,832	5,705,505	57.7
23	0.00086	0.00086	85	98,796	98,754	5,606,673	56.8
24	0.00085	0.00085	84	98,711	98,669	5,507,919	55.8
25	0.00079	0.00079	78	98,627	98,588	5,409,250	54.8
26	0.00082	0.00082	81	98,549	98,509	5,310,662	53.9
27	0.00091	0.00091	90	98,468	98,423	5,212,153	52.9
28	0.00088	0.00088	87	98,378	98,335	5,113,730	52.0
29	0.00097	0.00097	95	98,291	98,244	5,015,395	51.0
30	0.00091	0.00091	89	98,196	98,152	4,917,151	50.1
31	0.00108	0.00108	106	98,107	98,054	4,818,999	49.1
32	0.00108	0.00108	106	98,001	97,948	4,720,945	48.2
33	0.00109	0.00109	107	97,895	97,842	4,622,997	47.2
34	0.00105	0.00105	103	97,788	97,737	4,525,155	46.3
35	0.00114	0.00114	111	97,685	97,630	4,427,418	45.3
36	0.00125	0.00125	122	97,574	97,513	4,329,788	44.4
37	0.00119	0.00119	116	97,452	97,394	4,232,275	43.4
38	0.00122	0.00122	119	97,336	97,277	4,134,881	42.5

(continued)

Table A7.2 (continued)

x	m_x	q_x	d_x	l_x	L_x	T_x	e_x
39	0.00136	0.00136	132	97,217	97,151	4,037,604	41.5
40	0.00150	0.00150	146	97,085	97,012	3,940,453	40.6
41	0.00154	0.00154	149	96,939	96,865	3,843,441	39.6
42	0.00165	0.00165	160	96,790	96,710	3,746,576	38.7
43	0.00176	0.00176	170	96,630	96,545	3,649,866	37.8
44	0.00185	0.00185	178	96,460	96,371	3,553,321	36.8
45	0.00207	0.00207	199	96,282	96,183	3,456,950	35.9
46	0.00224	0.00224	215	96,083	95,976	3,360,767	35.0
47	0.00244	0.00244	234	95,868	95,751	3,264,791	34.1
48	0.00274	0.00274	262	95,634	95,503	3,169,040	33.1
49	0.00284	0.00284	271	95,372	95,237	3,073,537	32.2
50	0.00301	0.00301	286	95,101	94,958	2,978,300	31.3
51	0.00320	0.00319	302	94,815	94,664	2,883,342	30.4
52	0.00358	0.00357	337	94,513	94,345	2,788,678	29.5
53	0.00388	0.00387	364	94,176	93,994	2,694,333	28.6
54	0.00425	0.00424	398	93,812	93,613	2,600,339	27.7
55	0.00478	0.00477	446	93,414	93,191	2,506,726	26.8
56	0.00475	0.00474	441	92,968	92,748	2,413,535	26.0
57	0.00545	0.00544	503	92,527	92,276	2,320,787	25.1
58	0.00602	0.00600	552	92,024	91,748	2,228,511	24.2
59	0.00642	0.00640	585	91,472	91,180	2,136,763	23.4
60	0.00767	0.00764	694	90,887	90,540	2,045,583	22.5
61	0.00818	0.00815	735	90,193	89,826	1,955,043	21.7
62	0.00853	0.00849	759	89,458	89,079	1,865,217	20.9
63	0.01024	0.01019	904	88,699	88,247	1,776,138	20.0
64	0.01088	0.01082	950	87,795	87,320	1,687,891	19.2
65	0.01229	0.01221	1,060	86,845	86,315	1,600,571	18.4
66	0.01352	0.01343	1,152	85,785	85,209	1,514,256	17.7
67	0.01487	0.01476	1,249	84,633	84,009	1,429,047	16.9
68	0.01654	0.01640	1,367	83,384	82,701	1,345,038	16.1
69	0.01799	0.01783	1,462	82,017	81,286	1,262,337	15.4
70	0.01923	0.01905	1,535	80,555	79,788	1,181,051	14.7
71	0.02176	0.02153	1,701	79,020	78,170	1,101,263	13.9
72	0.02407	0.02378	1,839	77,319	76,400	1,023,093	13.2
73	0.02657	0.02622	1,979	75,480	74,491	946,693	12.5
74	0.03031	0.02986	2,195	73,501	72,404	872,202	11.9
75	0.03350	0.03295	2,350	71,306	70,131	799,798	11.2
76	0.03777	0.03707	2,556	68,956	67,678	729,667	10.6
77	0.04307	0.04216	2,799	66,400	65,001	661,989	10.0
78	0.04766	0.04655	2,961	63,601	62,121	596,988	9.4
79	0.05302	0.05165	3,132	60,640	59,074	534,867	8.8
80	0.05904	0.05735	3,298	57,508	55,859	475,793	8.3
81	0.06691	0.06474	3,510	54,210	52,455	419,934	7.7
82	0.07480	0.07210	3,655	50,700	48,873	367,479	7.2
83	0.08556	0.08205	3,860	47,045	45,115	318,606	6.8
84	0.09439	0.09014	3,893	43,185	41,239	273,491	6.3

(continued)

Table A7.2 (continued)

x	m_x	q_x	d_x	l_x	L_x	T_x	e_x
85	0.10400	0.09886	3,884	39,292	37,350	232,252	5.9
86	0.11112	0.10527	3,727	35,408	33,545	194,902	5.5
87	0.13707	0.12828	4,064	31,681	29,649	161,357	5.1
88	0.14854	0.13827	3,819	27,617	25,708	131,708	4.8
89	0.16037	0.14847	3,533	23,798	22,032	106,000	4.5
90	0.18142	0.16633	3,371	20,265	18,580	83,968	4.1
91	0.20227	0.18369	3,103	16,894	15,343	65,388	3.9
92	0.22349	0.20103	2,772	13,791	12,405	50,045	3.6
93	0.24070	0.21484	2,367	11,019	9,836	37,640	3.4
94	0.26899	0.23710	2,051	8,652	7,627	27,804	3.2
95	0.29034	0.25353	1,674	6,601	5,764	20,177	3.1
96	0.31073	0.26895	1,325	4,927	4,265	14,413	2.9
97	0.35478	0.30133	1,085	3,602	3,060	10,148	2.8
98	0.34386	0.29341	739	2,517	2,148	7,088	2.8
99	0.32830	0.28201	501	1,778	1,528	4,940	2.8
100+	0.37428	1.00000	1,277	1,277	3,412	3,412	2.7

Notes: (1) Values of m_x and q_x were copied from Table A7.1

(2) Values of the remaining five columns were estimated using Eqs. (7.8), (7.9), (7.12), (7.17) and (7.18) respectively for all ages except for L_0 and L_{100+} for which Eqs. (7.14) and (7.15) were used

Table A7.3 Complete life table for Australian females, 2005–2007

x	m_x	q_x	d_x	l_x	L_x	T_x	e_x
0	0.00426	0.00423	423	100,000	99,704	8,358,643	83.6
1	0.00031	0.00031	31	99,577	99,562	8,258,939	82.9
2	0.00020	0.00020	20	99,546	99,536	8,159,377	82.0
3	0.00016	0.00016	16	99,526	99,518	8,059,841	81.0
4	0.00011	0.00011	11	99,510	99,505	7,960,323	80.0
5	0.00011	0.00011	11	99,499	99,494	7,860,818	79.0
6	0.00011	0.00011	11	99,488	99,483	7,761,324	78.0
7	0.00008	0.00008	8	99,477	99,473	7,661,841	77.0
8	0.00008	0.00008	8	99,469	99,465	7,562,368	76.0
9	0.00008	0.00008	8	99,461	99,457	7,462,903	75.0
10	0.00006	0.00006	6	99,453	99,450	7,363,446	74.0
11	0.00007	0.00007	7	99,447	99,444	7,263,996	73.0
12	0.00007	0.00007	7	99,440	99,437	7,164,552	72.0
13	0.00012	0.00012	12	99,433	99,427	7,065,115	71.1
14	0.00013	0.00013	13	99,421	99,415	6,965,688	70.1
15	0.00019	0.00019	19	99,408	99,399	6,866,273	69.1
16	0.00023	0.00023	23	99,389	99,378	6,766,874	68.1
17	0.00026	0.00026	26	99,366	99,353	6,667,496	67.1
18	0.00028	0.00028	28	99,340	99,326	6,568,143	66.1
19	0.00028	0.00028	28	99,312	99,298	6,468,817	65.1
20	0.00030	0.00030	30	99,284	99,269	6,369,519	64.2

(continued)

Table A7.3 (continued)

x	m_x	q_x	d_x	l_x	L_x	T_x	e_x
21	0.00028	0.00028	28	99,254	99,240	6,270,250	63.2
22	0.00028	0.00028	28	99,226	99,212	6,171,010	62.2
23	0.00028	0.00028	28	99,198	99,184	6,071,798	61.2
24	0.00033	0.00033	33	99,170	99,154	5,972,614	60.2
25	0.00029	0.00029	29	99,137	99,123	5,873,460	59.2
26	0.00031	0.00031	31	99,108	99,093	5,774,337	58.3
27	0.00030	0.00030	30	99,077	99,062	5,675,244	57.3
28	0.00035	0.00035	35	99,047	99,030	5,576,182	56.3
29	0.00043	0.00043	43	99,012	98,991	5,477,152	55.3
30	0.00039	0.00039	39	98,969	98,950	5,378,161	54.3
31	0.00042	0.00042	42	98,930	98,909	5,279,211	53.4
32	0.00044	0.00044	44	98,888	98,866	5,180,302	52.4
33	0.00042	0.00042	42	98,844	98,823	5,081,436	51.4
34	0.00049	0.00049	48	98,802	98,778	4,982,613	50.4
35	0.00049	0.00049	48	98,754	98,730	4,883,835	49.5
36	0.00058	0.00058	57	98,706	98,678	4,785,105	48.5
37	0.00061	0.00061	60	98,649	98,619	4,686,427	47.5
38	0.00075	0.00075	74	98,589	98,552	4,587,808	46.5
39	0.00076	0.00076	75	98,515	98,478	4,489,256	45.6
40	0.00085	0.00085	84	98,440	98,398	4,390,778	44.6
41	0.00092	0.00092	90	98,356	98,311	4,292,380	43.6
42	0.00089	0.00089	87	98,266	98,223	4,194,069	42.7
43	0.00103	0.00103	101	98,179	98,129	4,095,846	41.7
44	0.00111	0.00111	109	98,078	98,024	3,997,717	40.8
45	0.00123	0.00123	121	97,969	97,909	3,899,693	39.8
46	0.00132	0.00132	129	97,848	97,784	3,801,784	38.9
47	0.00152	0.00152	149	97,719	97,645	3,704,000	37.9
48	0.00161	0.00161	157	97,570	97,492	3,606,355	37.0
49	0.00178	0.00178	173	97,413	97,327	3,508,863	36.0
50	0.00187	0.00187	182	97,240	97,149	3,411,536	35.1
51	0.00200	0.00200	194	97,058	96,961	3,314,387	34.1
52	0.00218	0.00218	211	96,864	96,759	3,217,426	33.2
53	0.00242	0.00242	234	96,653	96,536	3,120,667	32.3
54	0.00261	0.00261	252	96,419	96,293	3,024,131	31.4
55	0.00269	0.00269	259	96,167	96,038	2,927,838	30.4
56	0.00299	0.00299	287	95,908	95,765	2,831,800	29.5
57	0.00332	0.00331	317	95,621	95,463	2,736,035	28.6
58	0.00364	0.00363	346	95,304	95,131	2,640,572	27.7
59	0.00360	0.00359	341	94,958	94,788	2,545,441	26.8
60	0.00498	0.00497	470	94,617	94,382	2,450,653	25.9
61	0.00467	0.00466	439	94,147	93,928	2,356,271	25.0
62	0.00554	0.00552	517	93,708	93,450	2,262,343	24.1
63	0.00593	0.00591	551	93,191	92,916	2,168,893	23.3
64	0.00627	0.00625	579	92,640	92,351	2,075,977	22.4
65	0.00680	0.00678	624	92,061	91,749	1,983,626	21.5
66	0.00783	0.00780	713	91,437	91,081	1,891,877	20.7

(continued)

Table A7.3 (continued)

x	m_x	q_x	d_x	l_x	L_x	T_x	e_x
67	0.00800	0.00797	723	90,724	90,363	1,800,796	19.8
68	0.00931	0.00927	834	90,001	89,584	1,710,433	19.0
69	0.01027	0.01022	911	89,167	88,712	1,620,849	18.2
70	0.01111	0.01105	975	88,256	87,769	1,532,137	17.4
71	0.01285	0.01277	1,115	87,281	86,724	1,444,368	16.5
72	0.01416	0.01406	1,211	86,166	85,561	1,357,644	15.8
73	0.01567	0.01555	1,321	84,955	84,295	1,272,083	15.0
74	0.01845	0.01828	1,529	83,634	82,870	1,187,788	14.2
75	0.01978	0.01959	1,608	82,105	81,301	1,104,918	13.5
76	0.02230	0.02205	1,775	80,497	79,610	1,023,617	12.7
77	0.02539	0.02507	1,974	78,722	77,735	944,007	12.0
78	0.02915	0.02873	2,205	76,748	75,646	866,272	11.3
79	0.03262	0.03210	2,393	74,543	73,347	790,626	10.6
80	0.03727	0.03659	2,640	72,150	70,830	717,279	9.9
81	0.04234	0.04146	2,882	69,510	68,069	646,449	9.3
82	0.04979	0.04858	3,237	66,628	65,010	578,380	8.7
83	0.05570	0.05419	3,435	63,391	61,674	513,370	8.1
84	0.06434	0.06233	3,737	59,956	58,088	451,696	7.5
85	0.07198	0.06948	3,906	56,219	54,266	393,608	7.0
86	0.08062	0.07750	4,054	52,313	50,286	339,342	6.5
87	0.10168	0.09676	4,670	48,259	45,924	289,056	6.0
88	0.11284	0.10681	4,656	43,589	41,261	243,132	5.6
89	0.12178	0.11479	4,469	38,933	36,699	201,871	5.2
90	0.14048	0.13126	4,524	34,464	32,202	165,172	4.8
91	0.15529	0.14410	4,314	29,940	27,783	132,970	4.4
92	0.18032	0.16541	4,239	25,626	23,507	105,187	4.1
93	0.20382	0.18497	3,956	21,387	19,409	81,680	3.8
94	0.21761	0.19626	3,421	17,431	15,721	62,271	3.6
95	0.24267	0.21641	3,032	14,010	12,494	46,550	3.3
96	0.26339	0.23274	2,555	10,978	9,701	34,056	3.1
97	0.29345	0.25590	2,155	8,423	7,346	24,355	2.9
98	0.31977	0.27569	1,728	6,268	5,404	17,009	2.7
99	0.34960	0.29758	1,351	4,540	3,865	11,605	2.6
100+	0.41203	1.00000	3,189	3,189	7,740	7,740	2.4

Notes: (1) Values of m_x and q_x were copied from Table A7.1

(2) Values of the remaining five columns were estimated using Eqs. (7.8), (7.9), (7.12), (7.17) and (7.18) respectively for all ages except for L_0 and L_{100+} for which Eqs. (7.14) and (7.15) were used

Appendix 7.3: Abridged Life Tables by Sex: Australia, 2005–2007

Table A7.4 Abridged life table for Australian males, 2005–2007

x	n	${}_nm_x$	${}_nq_x$	${}_nd_x$	l_x	${}_nL_x$	T_x	e_x
0	1	0.00512	0.00506	506	100,000	99,646	7,892,971	78.9
1	4	0.00025	0.00100	99	99,494	397,778	7,793,325	78.3
5	5	0.00011	0.00055	55	99,395	496,838	7,395,547	74.4
10	5	0.00012	0.00060	60	99,340	496,550	6,898,709	69.4
15	5	0.00050	0.00250	248	99,280	495,780	6,402,159	64.5
20	5	0.00081	0.00404	400	99,032	494,160	5,906,379	59.6
25	5	0.00087	0.00434	428	98,632	492,090	5,412,219	54.9
30	5	0.00104	0.00519	510	98,204	489,745	4,920,129	50.1
35	5	0.00123	0.00613	599	97,694	486,973	4,430,384	45.3
40	5	0.00166	0.00827	803	97,095	483,468	3,943,411	40.6
45	5	0.00246	0.01222	1,177	96,292	478,518	3,459,943	35.9
50	5	0.00357	0.01769	1,683	95,115	471,368	2,981,425	31.3
55	5	0.00549	0.02708	2,530	93,432	460,835	2,510,057	26.9
60	5	0.00900	0.04401	4,001	90,902	444,508	2,049,222	22.5
65	5	0.01491	0.07187	6,246	86,901	418,890	1,604,714	18.5
70	5	0.02416	0.11392	9,188	80,655	380,305	1,185,824	14.7
75	5	0.04246	0.19193	13,717	71,467	323,043	805,519	11.3
80	5	0.07410	0.31259	18,052	57,750	243,620	482,476	8.4
85	5	0.12542	0.47741	18,952	39,698	151,110	238,856	6.0
90	5	0.21163	0.69202	14,357	20,746	67,838	87,746	4.2
95	5	0.31482	0.88084	5,628	6,389	17,875	19,908	3.1
100+	5+	0.37428	1.00000	761	761	2,033	2,033	2.7

Note: In the above table x refers to the beginning of an age interval and n its width. Values of ${}_nm_x$ and ${}_nq_x$ were copied from Table A7.1. The remaining five columns were estimated using Eqs. (7.8a), (7.9a), (7.12a), (7.17) and (7.18). Eq. (7.15) was used for the estimation of L_{100+}

Table A7.5 Abridged life table for Australian females, 2005–2007

x	n	${}_nm_x$	${}_nq_x$	${}_nd_x$	l_x	${}_nL_x$	T_x	e_x
0	1	0.00426	0.00423	423	100,000	99,704	8,360,108	83.6
1	4	0.00019	0.00076	76	99,577	398,156	8,260,404	83.0
5	5	0.00009	0.00045	45	99,501	497,393	7,862,248	79.0
10	5	0.00009	0.00045	45	99,456	497,168	7,364,855	74.1
15	5	0.00025	0.00125	124	99,411	496,745	6,867,687	69.1
20	5	0.00030	0.00150	149	99,287	496,063	6,370,942	64.2
25	5	0.00034	0.00170	169	99,138	495,268	5,874,879	59.3
30	5	0.00043	0.00215	213	98,969	494,313	5,379,611	54.4
35	5	0.00063	0.00315	311	98,756	493,003	4,885,298	49.5
40	5	0.00096	0.00479	472	98,445	491,045	4,392,295	44.6
45	5	0.00149	0.00742	727	97,973	488,048	3,901,250	39.8
50	5	0.00221	0.01099	1,069	97,246	483,558	3,413,202	35.1
55	5	0.00324	0.01607	1,546	96,177	477,020	2,929,644	30.5
60	5	0.00544	0.02684	2,540	94,631	466,805	2,452,624	25.9
65	5	0.00839	0.04109	3,784	92,091	450,995	1,985,819	21.6
70	5	0.01434	0.06922	6,113	88,307	426,253	1,534,824	17.4
75	5	0.02565	0.12052	9,906	82,194	386,205	1,108,571	13.5
80	5	0.04897	0.21814	15,769	72,288	322,018	722,366	10.0
85	5	0.09334	0.37840	21,387	56,519	229,128	400,348	7.1
90	5	0.17078	0.59841	21,023	35,132	123,103	171,220	4.9
95	5	0.27523	0.81522	11,502	14,109	41,790	48,117	3.4
100+	5+	0.41203	1.00000	2,607	2,607	6,327	6,327	2.4

Note: In the above table x refers to the beginning of an age interval and n its width. Values of ${}_nm_x$ and ${}_nq_x$ were copied from Table A7.1. The remaining five columns were estimated using Eqs. (7.8a), (7.9a), (7.12a), (7.17) and (7.18). Eq. (7.15) was used for the estimation of L_{100+}

References

Altman, O., & Goor, C. (1946). Operating analysis of B-29 aircraft engines. *Journal of the American Statistical Association, 41*(334), 190–203.

Australia. (2008). *2007 births Australia* (Catalogue No. 3301.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/DC32A0611500BAA0CA2574EF00142139/$File/33010_2007.pdf. Accessed Feb 2013.

Australia. (2009). *Australian life tables, 2005–07*. Canberra: Australian Government Actuary. http://www.aga.gov.au/publications/life_tables_2005-07/downloads/Australian_Life_Tables_2005-07.pdf. Accessed Feb 2013.

Elderton, W., & Fippard, R. (1914). *The construction of mortality and sickness tables: a primer*. London: Adam and Charles Black Publishers.

Fergany, N. (1971). On the human survivorship function and life table construction. *Demography, 8*(3), 331–334.

Greville, T. N. E. (1943). Short method of constructing abridged life tables. *The Record of the American Institute of Actuaries, 32*(1), 29–35.

Hull, C. (1899). *The economic writings of Sir William Petty together with the Observations Upon the Bills of Mortality more probably by Captain Jon Graunt*. London: Cambridge University Press.

Keyfitz, N. (1985). *Applied mathematical demography*. New York: Wiley.

Kintner, J. (2004). The life table. In J. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 301–340). San Diego: Elsevier Academic Press.

Swanson, D., & Tedrow, L. (2011). Using cohort change ratios to estimate life expectancy in populations closed to migration: a new approach. *Actuarial Research Clearing House Newsletter, 1*, 1–9. http://www.soa.org/search.aspx?searchterm=swanson. Accessed Feb 2013.

United Nations. (1982). *Model life tables for developing countries*. New York: Department of Economic and Social Affairs. http://www.un.org/esa/population/techcoop/DemMod/model_lifetabs/model_lifetabs.html. Accessed Feb 2013.

United Nations. (2012). *Demographic yearbook 2009–10*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dyb2009-2010.htm. Accessed Feb 2013.

WHO. (2013). *Life tables for WHO Member States*. Geneva: World Health Organization. http://www.who.int/healthinfo/statistics/mortality_life_tables/en/. Accessed 15 Feb 2013.

Chapter 8
Migration

8.1 Purpose

This chapter discusses some of the basic concepts and methods used in the study of migration within and between countries. Examples of different methods of analysis are given with data from various countries.

8.2 Basic Concepts

In general, the conceptual difficulties in defining a migrant are greater than in defining a birth or a death, the other two components of population change. These difficulties are associated with the variety of definitions in different countries and with the assortment of collection systems. Unlike birth and death, migration necessarily involves two areas: origin and destination. In addition, people may move more than one time, while death happens only once and there is a limit to the number of births a woman can have.

Migration is broadly thought of as the movement by individuals, groups or populations seeking to make relatively permanent changes of residence (Swanson & Stephan 2004). Supplementing this definition is one for *migrant* as a person who makes a relatively permanent change of residence from one area (origin), to another (destination) during a specified period of time.

There are two types of migrations. International migration is the movement of people between countries and domestic or internal migration relates to population movements within a country. In both types of migration short-term movements of, say, less than 1 year, are usually not included, even if they involve a change of residence. It is easier to measure international migration than internal migration. In the latter case, it is essential to define which of the address changes are to be considered migration. For example, changes within the same suburb or town may

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_8,

not be construed as migration. Generally speaking, internal migration involves a change of address across administrative or statistically-defined boundaries.

A migrant is a subset of the category *mover* that is defined as a person who reports in a census, survey or records in a population register that he or she lived at a different address, nationally or other country, at an earlier date. The way a population is defined affects how migration is defined. For example, if the *de jure* definition is used, a change in one's usual place of residence over specified boundaries determines the migration status. However, under the *de facto* definition of population, migration is a difficult concept to measure without some idea of where a person originated. Hence, the concept of permanent residence is introduced. In addition, migration might also be defined in terms of the place of birth.

For a specific area, whether it is a country or a particular geographic or administrative area within a country, migration flows consist of two components: the in-migrants (immigrants) and the out-migrants (emigrants). Gross or total migration refers to the sum of these two and the net migration is the difference between these two components. Thus, while gross migration will always be a positive number, the net migration may be positive, indicating a gain of population, or negative meaning a loss of population. In this context, it is important to keep in mind that there is no such person as a *net migrant*. The concept is akin to the concept of surplus in that like surplus is the result of subtracting outlays from revenues.

8.3 Migration Rates and Ratios

8.3.1 Crude Migration Rates

Migration is measured separately for in-migration and out-migration as shown below:

$$CIMR_{t\to t+n} = \frac{IM_{t\to t+n}}{\overline{P}_{t\to t+n}} \tag{8.1}$$

$CIMR_{t\to t+n}$ is the crude in-migration rate during the period t to $t + n$, $IM_{t\to t+n}$ is the number of in-migrants (immigrants) during the same period and $\overline{P}_{t\to t+n}$ is the average population during the period. Usually, the period is 1 year and the mid-period population is taken as an estimate of the average population. Thus, if the number of in-migrants (immigrants) over a 1-year period, say t to $t + 1$, the population refers to $t + \frac{1}{2}$.

Similarly, the crude out-migration rate, $COMR_{t\to t+n}$, is calculated by dividing the number of out-migrants (emigrants) ($OM_{t\to t+n}$) by $\overline{P}_{t\to t+n}$.

$$COMR_{t\rightarrow t+n} = \frac{OM_{t\rightarrow t+n}}{\overline{P}_{t\rightarrow t+n}} \tag{8.2}$$

Other commonly used migration measures include:

$$GMR_{t\rightarrow t+n} = \frac{IM_{t\rightarrow t+n} + OM_{t\rightarrow t+n}}{\overline{P}_{t\rightarrow t+n}} \tag{8.3}$$

$$NMR_{t\rightarrow t+n} = \frac{IM_{t\rightarrow t+n} - OM_{t\rightarrow t+n}}{\overline{P}_{t\rightarrow t+n}} \tag{8.4}$$

$GMR_{t\rightarrow t+n}$ is the *gross migration rate* (also referred to as the *total migration rate*), and $NMR_{t\rightarrow t+n}$ is the *net migration rate*. The net migration rate may be positive or negative depending upon whether in-migrants (immigrants) were more or less than the out-migrants (emigrants).

$$MTR_{t\rightarrow t+n} = \frac{GMR_{t\rightarrow t+n}}{NMR_{t\rightarrow t+n}} \tag{8.5}$$

$$MER_{t\rightarrow t+n} = \frac{NMR_{t\rightarrow t+n}}{GMR_{t\rightarrow t+n}} \tag{8.6}$$

$MTR_{t\rightarrow t+n}$ and $MER_{t\rightarrow t+n}$ are the *migration turnover rate* and *migration effectiveness rate* respectively.

These rates can be used to measure international as well as the internal migration data. They are generally expressed per 1,000 population.

Data for Estonia provide an illustration (see Box 8.1). The estimated average population of Estonia in 2010 was 1,340,160, and a total of 2,810 persons migrated to Estonia and 5,294 migrated out of the country during the calendar year 2010. The four migration rates are:

$$\begin{aligned}\text{Crude in-migration rate } (CIMR_{2010}) &= \frac{2,810}{1,340,160} = 0.00210\\ &= 2.10 \text{ per } 1,000\end{aligned}$$

$$\begin{aligned}\text{Crude out-migration rate } (COMR_{2010}) &= \frac{5,294}{1,340,160} = 0.00395\\ &= 3.95 \text{ per } 1,000\end{aligned}$$

$$\begin{aligned}\text{Gross (total) migration rate } (GMR_{2010}) &= \frac{2,810 + 5,294}{1,340,160} = 0.00605\\ &= 6.05 \text{ per } 1,000\end{aligned}$$

$$\text{Net migration rate } (NMR_{2010}) = \frac{2,810 - 5,294}{1,340,160} = -0.00185$$
$$= -1.85 \text{ per } 1,000$$

The negative value of the net migration rate indicates that as a result of the international migratory movements Estonia lost population during 2010.

$$\text{Migration turnover rate } (MTR_{t \to t+n}) = \frac{6.05}{-1.85} = -3.27 \text{ per } 1,000$$

$$\text{Migration effectiveness rate } (MER_{t \to t+n}) = \frac{-1.85}{6.05} = -0.31 \text{ per } 1,000.$$

Box 8.1 Obtaining the Demographic Data from Statistics Estonia

The link to the website is: http://pub.stat.ee/px-web.2001/Dialog/statfile1.asp

- Select *Population* then *Population indictors and composition*. Then select *Population figure and composition* (Table PO021 for population in 2010. In the table *Sex*, *Year* and *Age group(s)* have to be specified).
- Select *Population* then select *Vital events*. This will give access to data on births, deaths, migration and other events. Select a particular vital event and then select the appropriate table (Table POR05 for in- and out-migration. In the table *Year*, *Country*, Indicator (*immigration/emigration/net migration*) and *Sex* have to be specified).
- Click on *Population and Housing Census* to get access to the 2000 and 2011 census tables. For example, to obtain the age-sex distributions for the 2000 census, select *Age* and then select, for example, Table PC201. In the table one could specify the *Sex*, *Place of residence*, *Age* and *Ethnic nationality*.

The website allows the user to select appropriate categories within the table and the output can be saved and downloaded in the Excel format.

8.3.2 Characteristic-Specific Migration Rates

Specifying one or more characteristics in the numerators and the same characteristics in the denominators of Eqs. (8.1), (8.2), (8.3), (8.4), (8.5) and (8.6) would result in the corresponding characteristic-specific migration rates. Age and/or sex are the most commonly used characteristics.

8.3.3 Migration Ratios

Two types of migration ratios can also be calculated: one compares net migration with natural increase (Sect. 4.2) and the other relates net migration to total population growth (Sect. 4.3).

$$MNIR_{t\to t+n} = \frac{IM_{t\to t+n} - OM_{t\to t+n}}{B_{t\to t+n} - D_{t\to t+n}} \tag{8.7}$$

$MNIR_{t\to t+n}$ is the net migration to natural increase ratio during the period t to $t + n$, $B_{t\to t+n}$ and $D_{t\to t+n}$ stand for the births and deaths during the same period, and $IM_{t\to t+n}$ and $OM_{t\to t+n}$ as defined above.

$$MPGR_{t\to t+n} = \frac{IM_{t\to t+n} - OM_{t\to t+n}}{(B_{t\to t+n} - D_{t\to t+n}) + (IM_{t\to t+n} - OM_{t\to t+n})} \tag{8.8}$$

$MPGR_{t\to t+n}$ denotes the ratio of net migration to population growth during the period t to $t + n$, and the remaining terms are as defined in Eqs. (8.1), (8.2) and (8.7).

Again, these ratios can be applied to international as well as internal migration data. They are generally expressed on a percentage basis.

The number of live births in Estonia during 2010 was 15,825 and that of deaths 15,790. These figures with those on migration provide a basis for the estimation of these two ratios for Estonia:

$$MNIR_{2010} = \frac{2,810 - 5,294}{15,825 - 15,790} = \frac{-2,484}{35} = -71$$

This means that for every natural increase of one person the net out-migration consisted of 71 persons.

$$MPGR_{2010} = \frac{2,810 - 5,294}{(15,825 - 15,790) + (2,810 - 5,294)} = \frac{-2,484}{-2,449} = 1.01$$

The above ratio indicates that net migration accounted for the whole of the population loss in Estonia during 2010.

8.4 Direct Methods of Estimation

Countries may compile data on international migration using the arrival and departure cards that passengers in international travel fill-in, they may also use administrative records on the issuance of visas, other migration records, population

Table 8.1 Schematic presentation of the place of usual residence at two points in time

	Population at time $t + n$					
Population at time t	*Area 1*	*Area 2*	*Area 3*	...	*Area k*	Total at time t
Area 1	$\mathbf{p_{11}}$	p_{12}	p_{13}	...	p_{1k}	p_1^t
Area 2	p_{21}	$\mathbf{p_{22}}$	p_{23}	...	p_{2k}	p_2^t
Area 3	p_{31}	p_{32}	$\mathbf{p_{33}}$	...	p_{3k}	p_3^t
...	...	...	...	...	...	...
Area k	p_{k1}	p_{k2}	p_{k3}	...	$\mathbf{p_{kk}}$	p_k^t
Total at time $t + n$	p_1^{t+n}	p_2^{t+n}	p_3^{t+n}		p_k^{t+n}	...

registers and surveys (Chap. 2). Movements within national boundaries are more difficult to measure and information is usually collected through censuses and surveys, unless there are reliable population registers that can be used for this purpose. Questions on place of residence at specific points in time, such as last year, or at the time of a previous census or survey are useful in deriving estimates of both internal and international migration.

8.4.1 Place of Residence at Specific Times in the Past

In some censuses and population surveys questions may be asked as to where a person was living a certain number of years prior to the current residence. Cross-classifying the current place of usual residence with that n years ago would result in a matrix such as Table 8.1. The matrix would exclude both (a) people born during the period and (b) those whose residence was one of the designated areas at $t + n$ but were not living in any of the designated areas n years ago, that is those who moved from another country during the reference period. Accordingly, the matrix is concerned with movements from one area to another during the period t to $t + n$. It does not involve the actual total populations in each area at t or $t + n$ only those who were alive at the two points in time and lived in one of the specified areas at both points in time.

People in the diagonal of the matrix (shown in bold) represent those people who reported their place of usual residence in the same area at both points in time. These are the non-movers during the period t to $t + n$. The sum of each column less the non-movers represents the in-migration to that area from the other areas, and the sum of each row less the non-movers, stands for the out-migration from the area to the other areas. Accordingly,

$$IM_i = (p_{1i} + p_{2i} + \ldots + p_{ki}) - p_{ii} \tag{8.9}$$

IM_i is the in-migration to any area, say i, during the period t to $t + n, p_{1i}, p_{2i}, \ldots, p_{ki}$ are the persons who were living in that area at time $t + n$ but living in area 1, area

2, . . . area k, at time t, and p_{ii} are the people who were living in area i at both times. Similarly, out-migration (OM_i) from area i is defined as:

$$OM_i = (p_{i1} + p_{i2} + \ldots + p_{ik}) - p_{ii} \tag{8.10}$$

$p_{i1}, p_{i2}, \ldots, p_{ik}$ are the people who were living in area i at time t but were in another designated area at time $t + n$.

The population figures in Table 8.1 refer to those people who were alive and lived in one of the areas at time t as well as at time $t + n$. Obviously, births and deaths occurring during this period were not included. The same applies to people who lived in none of the areas at time t (*i.e.,* the international migrants). Each cell of the matrix, except those in the diagonal, represent internal *migration streams* and *counter-streams*. For example, p_{12} is the migration stream from area 1 to area 2 and p_{21} is the migration counter-stream from area 2 to area 1.

To estimate the in-migration and out-migration rates for area i, the denominators for both would be the average population of area i, between time t and $t + n$. The in-migration rate for area i ($INMR_i$) and the out-migration rate (OMR_i) for area i during the time period t to $t + n$ are calculated as:

$$INMR_i = \frac{IM_i}{\frac{p_i^t + p_i^{t+n}}{2}} = \frac{2 * IM_i}{p_i^t + p_i^{t+n}} \tag{8.11}$$

$$OMR_i = \frac{OM_i}{\frac{p_i^t + p_i^{t+n}}{2}} = \frac{2 * OM_i}{p_i^t + p_i^{t+n}} \tag{8.12}$$

In Eqs. (8.11) and (8.12), p_i^t and p_i^{t+n} denote the average population of area i at time t and $t + n$ respectively, with the caveats mentioned above.

The net internal migration rate (NMR_i) for area i during the time period t to $t + n$ is calculated as:

$$NMR_i = INMR_i - OMR_i \tag{8.13}$$

Australian data in Table 8.2 shows the State/Territory of usual residence of people at the time of the 2006 census and 5 years prior to the census. These people were alive at the time of the 2006 census and 5 years earlier. Their place of residence was also one of the states/territories 5 years earlier.

The **bold** numbers in the diagonal represent those who were living in the same State/Territory at the time of both censuses. From the point of view of measuring interstate migration, they represented the non-migrants or non-movers during the 5 -year period. For each location the numbers in each column less the number in bold were those people who were in that state/territory in the 2006 census but were located in another state/territory 5 years earlier. These were in-migrants to that

Table 8.2 Place of usual residence as reported at the 2006 census and 5 years prior to the census: Australia, 2006 census

	Place of usual residence at the 2006 census								
Place of usual residence 5 years ago	NSW	VIC	QLD	SA	WA	TAS	NT	ACT	Australia
New South Wales (NSW)	**5,224,678**	53,456	141,266	14,003	18,864	9,467	5,884	24,845	5,492,463
Victoria (VIC)	43,970	**3,965,394**	52,113	13,992	15,221	8,192	4,976	5,539	4,109,397
Queensland (QLD)	60,134	32,243	**2,936,555**	9,326	13,761	7,831	8,432	6,164	3,074,446
South Australia (SA)	11,418	16,145	16,409	**1,242,211**	6,365	2,331	4,454	1,996	1,301,329
Western Australia (WA)	13,900	15,353	17,945	5,194	**1,512,988**	3,175	4,291	2,206	1,575,052
Tasmania (TAS)	4,779	8,233	8,785	1,683	3,012	**378,813**	699	827	406,831
Northern Territory (NT)	5,466	4,811	13,765	5,670	4,710	845	**119,850**	1,207	156,324
A. C. Territory (ACT)[a]	22,710	5,677	9,559	1,649	1,942	897	811	**229,101**	272,346
Australia	5,387,055	4,101,312	3,196,397	1,293,728	1,576,863	411,551	149,397	271,885	16,388,188

Source: Australia (2013)

Notes: [a]Australian Capital Territory. The figures for each state/territory exclude people <5 years of age, and international migrants arriving during the inter-censual period, and those who did not state their residence 5 years before the census state/territory. The numbers in each row less the number in bold represent the out-migrants from the particular state/territory.

The information presented in Table 8.2 is summarized in Table 8.3 in terms of internal in-migration, out-migration and net-migration for each state/territory during the 5-year period, 2001–2006. Obviously, the table includes only those people who were alive (aged 5 years or older) and were living in Australia at both points in time. Births and deaths during the intervening 5-years were not counted.

Equations (8.9) and (8.10) were used to calculate the in- and out-migration in columns (2) and (3), column (4) is the net migration, the difference between columns (2) and (3), and column (5) is the average population for each area. These served as the denominators for the estimation of migration rates as per Eqs. (8.11), (8.12) and (8.13) in columns (7) and (8). The net migration rate is stated in column (9) that is the difference between columns (7) and (8).

8.4.2 Place of Birth

In some countries, information on the place of birth is also used instead of or in addition to determining the place of residence sometime in the past. The principle in measurement is the same as described above, except instead of place of usual residence in the past place of birth is asked. Persons who are enumerated at a place other than their birthplace are termed as *lifetime migrants*.

Both methods mentioned above have limitations because they do not allow the estimation of the number of migratory moves within the reference period. This information can only be collected by directly asking a relevant question in the census or survey or through reliable population registers.

8.5 Indirect Methods of Estimation

8.5.1 Life Table Method

If a population age distribution is known at two points in time t and $t + n$, it is possible to estimate the number of survivors at time $t + n$ using an appropriate life table. The difference between the actual number of persons at time $t + n$ and survivors of the original population at time t would give an indirect estimate of net migration experienced by the population provided that consideration is taken of births during the period that have added to the population and any related mortality. It is also assumed that the mortality experience in the life table is close to that of the subject population.

Table 8.3 Interstate migration between 2001 and 2006, population in 2001, and migration rates for Australian States and Territories

	Migration			Average population	Migration rate per 1,000 population		
(State/Territory)	In	Out	Net	2001–2006	In	Out	Net
(1)	(2)	(3)	(4)	(5)	(6) = (2)/(5)	(7) = (3)/(5)	(8) = (6) − (7)
New South Wales	162,377	267,785	−105,408	5,439,759	29.85	49.23	−19.38
Victoria	135,918	144,003	−8,085	4,105,355	33.11	35.08	−1.97
Queensland	259,842	137,891	121,951	3,135,421	82.87	43.98	38.89
South Australia	51,517	59,118	−7,601	1,297,529	39.70	45.56	−5.86
Western Australia	63,875	62,064	1,811	1,575,957	40.53	39.38	1.15
Tasmania	32,738	28,018	4,720	409,191	80.01	68.47	11.53
Northern Territory	29,547	36,474	−6,927	152,861	193.29	238.61	−45.32
A. C. Territory[a]	42,784	43,245	−461	272,115	157.23	158.92	−1.69
Australia	778,598	778,598	0	16,388,188	47.51	47.51	0.00

Source: Table 8.2
Note: [a]Australian Capital Territory

Let p_x^t be the population aged x at time t, and L_x be the stationary population aged x from an appropriate life table. In this context, an appropriate life table may be the one that represents the mortality experience of the area for which net-migration is to be estimated. For example, if the area is part of a province or a state the life table for that province/state or even the whole country may be used.

Assuming no migration in and out of a population, which is one of the basic conditions of life tables, the proportion of x year olds in the life table population who will survive for n years would be $\frac{L_{x+n}}{L_x}$. This is called the *survival ratio*, or the *probability of survival* of an x year old person to age $x + n$ in a life table population. Multiplying this probability by p_x^t would give the expected population aged $x + n$ at time $t + n$. The difference between this number and the actual population of $x + n$ year olds at time $t + n, p_{x+n}^{t+n}$, would give an estimate of the net migration ($M_x^{t\to t+n}$) of x year olds to age $x + n$. This can be written algebraically as:

$$M_x^{t\to t+n} = p_{x+n}^{t+n} - \left(p_x^t * \frac{L_{x+n}}{L_x}\right) \tag{8.14}$$

If the age distribution is given in age intervals of width, say, 5 years, Eq. (8.14) can be re-written as:

$$M_{x\to x+5}^{t\to t+5} = {}_5p_{x+5}^{t+5} - \left({}_5p_x^t * \frac{{}_5L_{x+5}}{{}_5L_x}\right) \tag{8.15}$$

The terminology of abridged life table has been used in the above equation, so that the subscript on the left hand side stands for the width of the age interval, 5 in this case. It can, however, be an interval of any width.

As an example, Table 8.4 provides estimates of net migration by age for Inyo County, California during the period 1990–2000. Age distributions of the population given in columns (2) and (4) were obtained from the United States (1992) and United States (undated). ${}_5L_x$ values in column (5) were taken from the life table for California (United States 1999).

As pointed out in Chap. 7, in a life table population (l_0) births occur every year and are equal to 100,000. Therefore, the survival ratios in column (6) were calculated as follows:

- From birth to age group 0–4: $\frac{{}_5L_0}{5 * l_0} = \frac{496,798}{5 * 100,000} = 0.993596$
- From birth to age group 5–9: $\frac{{}_5L_5}{5 * l_0} = \frac{496,167}{5 * 100,000} = 0.992334$
- Survival ratios for ages 0–4 in 1990 to 10–14 in 2000 and so on till 60–64 in 1990 to 70–74 in 2000 were estimated as $\frac{{}_5L_{x+10}}{{}_5L_x}$ where x varied from 0 to 60.
- Survival ratio for age 65+ in 1990 to age 75+ in 2000 was: $\frac{L_{75+}}{L_{65+}} = \frac{785,060}{398,465 + 356,460 + 785,060} = 0.509784$.

To estimate the number of survivors in 2000 shown in column (7) for age groups 0–4 and 5–9, the number of registered births in Inyo county during the 5 calendar

Table 8.4 Estimation of net migration using the forward survival ratio method: Inyo County, California, USA, 1990–2000

Age in 1990 years	Population in 1990 census	Age in 2000 years	Population in 2000 census	Life table population $_5L_x$	Survival ratio 1990–2000	Number of survivors in 2000	Net migration 1990–2000
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8) = (4) − (7)
0–4	1,196	0–4	961	496,798	0.993596	1,037	−76
5–9	1,350	5–9	1,184	496,167	0.992334	1,137	47
10–14	1,261	10–14	1,360	495,759	0.997909	1,193	167
15–19	977	15–19	1,236	494,652	0.996947	1,346	−110
20–24	687	20–24	673	492,334	0.993091	1,252	−579
25–29	1,071	25–29	644	490,018	0.990632	968	−324
30–34	1,367	30–34	849	487,323	0.989822	680	169
35–39	1,531	35–39	1,232	483,746	0.987200	1,057	175
40–44	1,378	40–44	1,482	478,877	0.982669	1,343	139
45–49	1,109	45–49	1,597	472,082	0.975888	1,494	103
50–54	954	50–54	1,314	462,720	0.966261	1,332	−18
55–59	929	55–59	1,101	449,044	0.951199	1,055	46
60–64	1,080	60–64	883	428,560	0.926176	884	−1
65+	3,391	65–69	892	398,465	0.887363	824	68
		70–74	898	356,460	0.831762	898	0
		75+	1,639	785,060	0.509784	1,729	−90
Total	18,281		17,945	7,768,065	...	18,229	−284

Sources: United States (1992) for column (2); United States (undated) for column (4); United States (1999) for column (5) years prior to 2000, *i.e.,* 1995–1999 and during the period 1990–1994 were obtained (United States 2013). These were 1,044 and 1,146 respectively. Multiplying these births by the relevant survival ratios (0.993596 and 0.992334) the first two figures in column (7) were obtained. The third figure was derived by multiplying population 0–4 in 1990 by the survival ratio from 0–4 to 10–14, *i.e.,* 0.997909 and so on for the remainder of the column.

The difference between columns (4) and (7) is the estimated net migration. The result indicates that Inyo County lost 284 during the 10-year period (Table 8.4).

The net migration rate can be estimated dividing the net migration figure by the average population of the county in 1990 and 2000 (*i.e.,* 18,113):

$$NMR = \frac{-284}{18,113} = -0.0157 \text{ or } -15.7 \text{ per } 1,000.$$

8.5.2 Census Survival Ratios

The estimation of internal migration for a given area may use age-specific census survival ratios instead of life table survival ratios, if a relevant life table is not available. The census survival ratios used may be from a country wide census or of some other population that contains the area for which internal migration is to be estimated. *Census survival ratio* is defined as:

$$CSR_{x \to x+n}^{t \to t+n} = \frac{P_{x+n}^{t+n}}{P_x^t} \tag{8.16}$$

$CSR_{x \to x+n}^{t \to t+n}$ is the ratio of persons aged x in a census conducted at time t and persons aged $x + n$ in a census conducted at time $t + n$, and P_x^t and P_{x+n}^{t+n} are the number of persons aged x and $x + n$ in the reference population. The estimate of net internal migration is obtained as:

$$M_x^{t \to t+n} = p_{x+n}^{t+n} - \left(p_x^t * \frac{P_{x+n}^{t+n}}{P_x^t} \right) \tag{8.17}$$

$M_x^{t \to t+n}$ is the value of internal migration of people aged x from time t to $t + n$, p_x^t is the population at time t of age x in the area, p_{x+n}^{t+n} is the number of people of age $x + n$ in the area at $t + n$, and the other terms are as defined for Eq. (8.16).

It is apparent that the age-specific survival ratios are a combination of the impact of both mortality and net migration in the reference population during the inter-census period. If migration in the reference population is small, this approach would give approximate estimates of net migration for the area. However, if immigration or emigration during the inter-census period is large in relation to the reference population care needs to be used in the interpretation of results.

8.5.3 Other Administrative Records

Records of utility connections and disconnections, electoral rolls and other administrative records can be used to estimate net migration. However, none of these records are easily available and there are no standard methods which could be described here. Again, population registers can be used for this purpose, if they are reliable.

8.6 Determinants of Migration

8.6.1 Push and Pull Factors

Two types of factors affect the level of migration in and out of an area. *Push* factors influence a person to move away from an area, while *pull* factors attract a person to move into an area. Differences in socio-economic and political conditions are among the most important determinants of migration, both internal and international. They may act as pull factors for areas/countries where they are favourable and push factors where they are not favourable. There are a number of migration theories that deal with various push and pull factors (see for example: Ravenstein 1889; Lee 1966).

An important factor in the decision to move is the distance between the places of origin and destination (Morrison et al. 2004). This factor has largely been dealt through a set of models, one of which is based on Newton's Law of Gravity.

8.6.2 Gravity Model

A popular gravity model is that proposed by George Zipf that provides an estimate of the expected net number of migrants between two areas based on their distance from each other and their respective population sizes (Zipf 1946). Zipf's version of the gravity model is defined as

$$M_{ij} = \frac{p_i * p_j}{D_{ij}} \tag{8.18}$$

M_{ij} is the migration stream between area i and area j, p_i and p_j are the populations of area i and j respectively and D_{ij} is the distance between the two areas measured from a centroid or a central city.

According to this model, given the equality of populations, the greater the distance between area i and area j, the less likely it is for people to move from one to the other. Given the equality of distances, the higher the respective populations in areas i and j, the more likely it is for people to move from one area to the other.

This concept was tested using data on the number of high school graduates in Washington State counties in the Spring of 1990 and enrolments in Washington State University from these counties in the Fall of 1990 (Table 8.5).

Counties are listed in alphabetic order in columns (1) and (5), and distances (D_{ij}) in miles from the seat of each county to the seat of King county, where Washington State University is located, are shown in columns (2) and (6). The number of high school graduates in the Spring of 1990 in each county (H_i) is presented in columns (3) and (7). The number of graduates from each county enrolled at Washington State University in the Fall of 1990 (S_i) are given in columns (4) and (8).

Following the gravity model, two hypotheses are to be tested:

1. the number of students from each county who enrolled in Washington State University in Fall 1990 (S_i) was negatively correlated with the distance (D_{ij}) from the county seat to Washington State University (located in King County).
2. the number of students from each county who enrolled in Washington State University in Fall 1990 (S_i) was positively correlated with the total number of students in each county who graduated from high school in the Spring of 1990.

A model can be built to test these two hypotheses using multiple regression analysis (Sect. 3.9). In this model, the number of students enrolled in Washington State University (S_i) is the dependent variable. The number of high school graduates in each county (H_i) is one of the independent variables and the distance in miles between the each county seat and Washington State University (D_{ij}) is another independent variable.

To use the linear multiple regression model, the natural logarithm (LN) values for all variables were calculated. In the case of King County the distance was amended from zero to 0.00001 as it is not possible to calculate the natural logarithm of zero.

The proposed model was:

$$LN(\hat{S}_i) = a + b_1 * LN(D_{ij}) + b_2 * LN(H_i)$$

where $\hat{S}_i$ is the expected value of S_i based on the model, a is the intercept and b_1 and b_2 are the two multiple regression coefficients.

Table 8.5 Number of high school graduates in Washington State counties in the Spring of 1990 and enrolments in Washington State University from these counties in the Fall of 1990

County (i)	D_{ij}	H_i	S_i	County	D_{ij}	H_i	S_i
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Adams	225	177	18	Lewis	90	682	101
Asotin	321	171	21	Lincoln	262	130	23
Benton	201	1,391	323	Mason	84	341	44
Chelan	157	527	130	Okanogan	244	364	55
Clallam	109	499	128	Pacific	141	171	17
Clark	165	2,484	519	Pend Orielle	334	115	11
Columbia	276	31	2	Pierce	35	5,072	1,837
Cowlitz	126	919	182	San Juan	114	85	37
Douglas	175	279	40	Skagit	63	750	199
Ferry	312	59	5	Skamania	206	64	13
Franklin	235	403	43	Snohomish	27	3,934	1,937
Garfield	285	23	2	Spokane	285	3,913	951
Grant	177	665	67	Stevens	349	365	25
Grays Harbor	103	627	94	Thurston	64	1,714	535
Island	76	457	131	Wahkiakum	152	27	6
Jefferson	86	173	41	Walla Walla	279	418	105
King	0	12,463	15,998	Whatcom	89	1,137	314
Kitsap	60	1,964	699	Whitman	281	308	136
Kittitas	117	256	59	Yakima	149	1,807	356
Klickitat	222	238	19	Total	6,676	45,203	25,223

Source: Miyashiro (1990)
Notes: (H_i) Number of high school graduates in Washington State counties in the Spring of 1990. (S_i) Number of students from each Washington State county enrolled in Washington State University in Fall of 1990. (D_{ij}) Distance in miles from each county seat to King County where Washington State University is located. The authors are grateful to Lucky Tedrow at Western Washington University for supplying these data

Using a statistical package the model was estimated as:

$$LN(\hat{S}_i) = -2.22674 - 0.09744 * LN(D_{ij}) + 1.167174 * LN(H_i)$$

The results from this multiple regression model support the two hypotheses. Thus, the number of high school graduates in each county (H_i) is positively and the distance from each county is negatively related to the number of students enrolled in Washington State University (S_i). In other words, the larger the number of graduates (H_i) in each county the larger the number of students enrolled from that county (S_i), and the greater the distance from each county to Washington State University (D_{ij}) the lower the number enrolled (S_i).

The coefficient of determination (R^2) for the model was 0.946157. It indicates that the two independent variables in the model explain nearly 95 % of the variation

Table 8.6 Actual and expected values of enrolments in Washington State University by the county from where the students came from

County (i)	S_i	$LN(\hat{S}_i)$	$\hat{S}_i$	County	S_i	$LN(\hat{S}_i)$	$\hat{S}_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Adams	18	3.2870	27	Lewis	101	4.9507	141
Asotin	21	3.2121	25	Lincoln	23	2.9120	18
Benton	323	5.7043	300	Mason	44	4.1484	63
Chelan	130	4.5955	99	Okanogan	55	4.1206	62
Clallam	128	4.5673	96	Pacific	17	3.2923	27
Clark	519	6.4003	602	Pend Orielle	11	2.7452	16
Columbia	2	1.2337	3	Pierce	1,837	7.3846	1,611
Cowlitz	182	5.2660	194	San Juan	37	2.4971	12
Douglas	40	3.8426	47	Skagit	199	5.0963	163
Ferry	5	1.9729	7	Skamania	13	2.1083	8
Franklin	43	4.2431	70	Snohomish	1,937	7.1133	1,228
Garfield	2	0.8822	2	Spokane	951	6.8774	970
Grant	67	4.8553	128	Stevens	25	4.0890	60
Grays Harbor	94	4.8394	126	Thurston	535	6.0595	428
Island	131	4.4999	90	Wahkiakum	6	1.1306	3
Jefferson	41	3.3540	29	Walla Walla	105	4.2690	71
King	15,998	9.9021	19,972	Whatcom	314	5.5483	257
Kitsap	699	6.2247	505	Whitman	136	3.9119	50
Kittitas	59	3.7814	44	Yakima	356	6.0388	419
Klickitat	19	3.6339	38	Total	25,223		28,011

in the number of students enrolled in Washington State University coming from the counties within that state.

In the context of Zipf's model Eq. (8.18), variable (S_i) is the equivalent of the dependent variable (M_{ij}), (H_i) is one of independent variables equivalent of (p_i), (D_{ij}) is the other independent variable and (p_j) drops out because in this case it has a constant value.

Substituting the values of LN (D_i) and LN (H_i) for each county in the model, the corresponding values of LN (S_i) were estimated. Taking the antilog (e^x) resulted in the estimated value of $\hat{S}_i$ for each county (Table 8.6).

8.7 Centre of Gravity of Population

This concept is similar to that used in Physics. Consider a conglomerate of population, such as a country or a province or a city that consists of n small units within. These may be census collection districts or a suburb or any other identifiable unit for which three pieces of information are available: its population size, longitude and latitude. Based on these n observations it is possible to calculate a weighted average of the longitudes and latitudes using population of each unit as the weight. Thus, for an area i, population is p_i, longitude is y_i and latitude is x_i, weighted

Table 8.7 Example of the calculation of the centre of gravity of population

Unit	Population	Latitude	Longitude	Latitude product	Longitude product
(1)	(2)	(3)	(4)	(5) = (2)*(3)	(6) = (2)*(4)
Alt	39	46.16 N	123.39 W	1,800.24	4,812.21
Cath	1,023	46.12 N	123.23 W	47,180.76	126,064.29
Deep	204	46.21 N	123.41 W	9,426.84	25,175.64
Gray	263	46.12 N	123.37 W	12,129.56	32,446.31
Pug	831	46.10 N	123.23 W	38,309.10	102,404.13
Ska	401	46.16 N	123.27 W	18,510.16	49,431.27
Total	2,761	…	…	127,356.66	340,333.85
$\overline{X}$ (weighted average of latitude) $= \frac{127{,}356.66}{2{,}761} =$				46.13 N	
$\overline{Y}$ (weighted average of longitude) $= \frac{340{,}333.85}{2{,}761} =$					123.26 W

average latitude ($\overline{X}$) and longitude ($\overline{Y}$) for the whole population (consisting of n units) could be calculated using Eq. (3.3) as follows:

$$\overline{X} = \frac{\sum_{i=1}^{i=n} p_i * x_i}{\sum_{i=1}^{i=n} p_i} \tag{8.19}$$

$$\overline{Y} = \frac{\sum_{i=1}^{i=n} p_i * y_i}{\sum_{i=1}^{i=n} p_i} \tag{8.20}$$

The intersection of the $\overline{X}$ longitude and the $\overline{Y}$ latitude would determine the location of the demographic centre of gravity for that particular population.

The population of a hypothetical county in Washington State provides an example for the calculation of the demographic centre of gravity following Eqs. (8.19) and (8.20) (Table 8.7).

References

Australia. (2013). *Census home: data and analysis*. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/websitedbs/censushome.nsf/home/data?opendocument#from-banner=LN. Accessed Feb 2013.

Lee, E. (1966). A theory of migration. *Demography, 3*, 47–57.

Miyashiro, T. (1990). *Migration of college students in the state of Washington, 1990* (Student Paper Series, Technical Report S.8.16.91-3). Bellingham: Demographic Research Laboratory, University of Washington.

Morrison, P., Bryan, T., & Swanson, D. A. (2004). Internal migration and short-distance mobility. In J. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 493–521). San Diego: Elsevier Academic Press.

Ravenstein, E. (1889). The laws of migration. *Journal of the Royal Statistical Society, 52*, 241–301.

Swanson, D. A., & Stephan, G. E. (2004). Glossary. In J. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 751–786). San Diego: Elsevier Academic Press.

United States. (1992). *1990 Census of population: general population characteristics, California, table 54*. Washington, DC: Bureau of Census. http://www.census.gov/prod/cen1990/cp1/cp-1-6-1.pdf. Accessed Feb 2013.

United States. (1999). *California life expectancy: abridged life tables by race/ethnicity for California, 1995–97*. Sacramento: State of California Department of Health Services. http://www.cdph.ca.gov/pubsforms/Pubs/OHIRLifetables1995-1997.pdf. Accessed Feb 2013.

United States. (2013). *Historical and projected State and County births, 1970–2021, with actual and projected fertility rates by mother's age and race/ethnicity, 2000–2021*. Sacramento: California Department of Finance. http://www.dof.ca.gov/research/demographic/reports/projections/births/. Accessed Feb 2013.

United States. (undated). *American fact finder*. Washington, DC: Bureau of Census. http://factfinder2.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t#. Accessed Feb 2013 (To obtain the data required in column (4) of Table 8.4, type Table QT-P1 in the first box and Inyo County, California in the second box.)

Zipf, G. K. (1946). The P_1P_2/D hypothesis: on the intercity movement of persons. *American Sociological Review, 11*, 677–686.

Chapter 9
Some Demographic Events and Characteristics Analysis

9.1 Purpose

Although some demographic events do not change the size of populations they have an impact on its characteristics and reflect and or have an effect on what people do. This chapter deals with methods of analysis and measures related to marital status and associated vital events, such as marriages and divorces, education, labour force and occupation, households and families. Examples are given with information from a number of countries with some comparisons of national patterns.

9.2 Marriages and Divorces

Marriage and divorce are examples of vital events that affect the composition but not the size of a population. Customs related to marriage vary depending on culture and evolving social norms. Most countries have laws that allow marriage only after a person has attained a certain age. In some societies, only monogamous marriages are legal, while in others polygamous marriages are lawful. In other countries, marriages can only take place between a male and a female. Marriage conventions and laws have been undergoing change in many societies. For example, in some, couples may cohabit without getting married, in others cohabitation may be followed by marriage. These unions may or may not have legal status. Further, in some societies, same-sex marriages are happening and such marriages have even been legally recognised. Social conventions and the laws about divorce and separation also vary from country to country. This section deals with indices related to the study the marriage and divorce patterns.

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_9,

9.2.1 Marriage Rates

The *crude marriage rate* is defined as:

$$CMR_{t\to t+n} = \frac{M_{t\to t+n}}{\overline{P}_{t\to t+n}} \tag{9.1}$$

$CMR_{t\to t+n}$ is the crude marriage rate during the period t to $t + n$, $M_{t\to t+n}$ is the number of marriages during the period, and $\overline{P}_{t\to t+n}$ is the average population during the same period. Generally the mid-period population is taken as an estimate of the average population. Thus, if marriages are over a period of 1 year, say t to $t + 1$, the population refers to $t + \frac{1}{2}$. This equation is similar to that for other vital events (Sect. 4.5) such as births (Sect. 5.3.1) and deaths (Sect. 6.2.1).

Usually, data on marriages are available on annual basis by age of partners and other characteristics, such as previous marital status and place of birth. A *characteristic-specific marriage rate* is calculated as follows:

$$MR^x_{t\to t+1} = \frac{M^x_{t\to 1}}{\overline{P}^x_t} \tag{9.2}$$

$MR^x_{t\to t+1}$ is the characteristic x-specific marriage rate during year t, $M^x_{t\to t+1}$ is the number of marriages of persons with characteristic x during the same year t, and $\overline{P}^x_t$ is the population with characteristic x in the middle of year t. As a marriage involves two people, these rates are invariably calculated for each partner separately.

Again, this equation follows Eq. (4.15) concerned with characteristic-specific rates (Sect. 4.6.1).

The minimum age at which people can get married is specified in marriage laws. This may be the same for both partners or may be different for males and females. The registration of marriages may not be recorded consistently in some countries. In such cases, sample surveys are used to obtain these data.

Spanish data on marriages registered in 2009 provide an illustration. Given that the marriages registered in 2009 amounted to 174,062, and the estimated population in mid-2009 was 45.929 million (United Nations 2011), the crude marriage rate for Spain in 2009 was:

$$CMR = \frac{174,062}{45,929,000} = 0.0038 \text{ or } 3.8 \text{ marriages per } 1{,}000 \text{ population.}$$

A comparable rate for Egypt was 9.9 per 1,000 in 2009 and for Japan 4.6 (United Nations 2011).

The estimation of age-sex specific marriage rates for Spain in 2009 is shown in Table 9.1. The table indicates that up to 30 years of age, rates for females were higher than those of males, while for persons 30 years of age or older the rates for males were higher than those for females. Marriage rates for females peaked in the 25–29 age group. Those for males peaked 5 years later.

Table 9.1 Marriages, population, and age-specific marriage rates: Spain, 2009

	Marriages (2009)		Population (Mid-2009)		Marriage rate (Per 1,000 population)	
Age (years)	Males	Females	Males	Females	Males	Females
(1)	(2)	(3)	(4)	(5)	(6) = (2)/(4)	(7) = (3)/(5)
15–24	7,691	17,896	2,527,369	2,405,173	3.04	7.44
25–29	44,556	58,627	1,780,458	1,689,346	25.03	34.70
30–34	60,326	53,912	2,101,710	1,956,070	28.70	27.56
35–39	29,290	21,970	2,027,105	1,902,267	14.45	11.55
40–44	12,951	9,860	1,879,716	1,818,521	6.89	5.42
45–49	7,426	5,641	1,701,846	1,695,119	4.36	3.33
50–54	4,757	3,275	1,470,154	1,501,308	3.24	2.18
55–59	2,915	1,605	1,260,998	1,317,563	2.31	1.22
60+	4,150	1,276	4,418,967	5,660,706	0.94	0.23

Source: United Nations (2011)
Note: Although the minimum legal age at marriage was 18 years, 336 males and 1,670 females below that age were married and included in the 15–19 age group

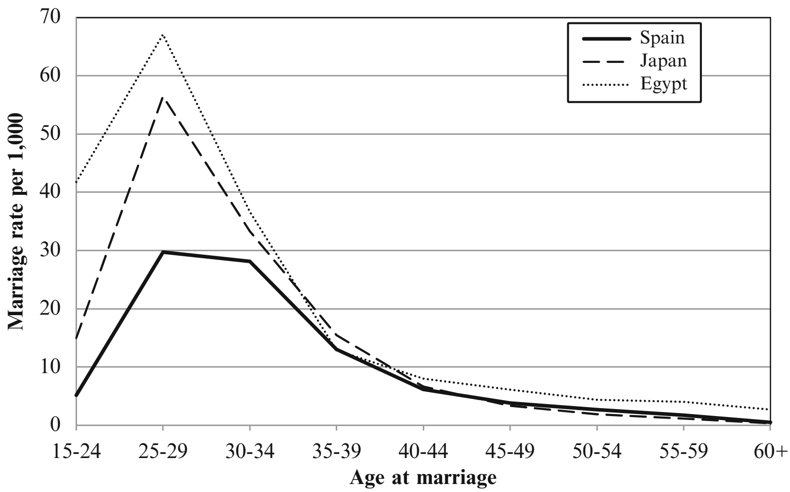


Fig. 9.1 Age-specific marriage rates per 1,000 persons for Spain, Japan and Egypt, 2009 (*Source*: United Nations 2011)

The median age at marriage for Spanish males was 32.9 years compared to 31.0 years for females. Corresponding figures were 27.9 and 22.2 for Egyptian males and females, and 30.7 and 28.9 for their Japanese counterparts. The trend of males marrying somewhat younger females tends to take place in most populations. Age-specific marriage rates (sexes combined) for Spain, Japan and Egypt are presented in Fig. 9.1. Although crude marriage rates in these countries are different, the age patterns are similar.

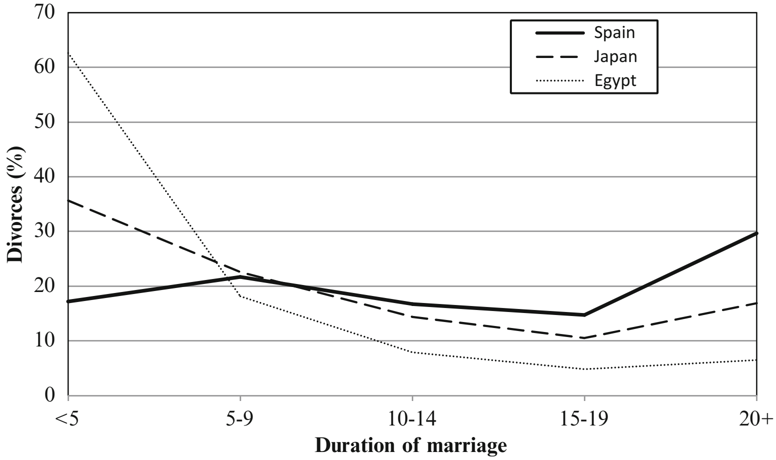


Fig. 9.2 Percentage distribution of divorces by duration of marriage in years: Spain, Japan and Egypt, 2009 (*Source*: United Nations 2011)

9.2.2 Divorce Rates

Analysis of data on divorces is done in a similar manner to that of marriages. Accordingly, crude divorce rates could be estimated for any other characteristic such as age by simply substituting marriages by divorces in the numerators of Eqs. (9.1) and (9.2).

In Spain, a total of 98,207 divorces were recorded in 2009 (United Nations 2011), consequently the crude divorce rate (*CDiR*) for that year was:

$$CDiR = \frac{98,207}{45,929,000} = 0.0021 \text{ or } 2.1 \text{ divorces per } 1,000 \text{ population.}$$

The corresponding rates for Egypt and Japan were 1.8 and 2.0 per 1,000 population respectively (United Nations 2011).

It is useful to consider not only the age at which divorces take place but also the duration of marriage. A percentage distribution of divorces by duration would be a simple way of presenting divorce data. Figure 9.2 shows the percentage distributions of divorces by duration of marriage for Spain, Egypt and Japan. The median marriage durations were 13.3 years for Spain, 4.0 years for Egypt and 8.2 years for Japan. In calculating these median ages, divorces where duration was not known were distributed on a *pro-rata* basis (see Box 5.1). This affected 21 % of divorces in Egypt, 6 % in Japan but none in Spain. Like other demographic statistics, the quality of data on marriages and divorces varies from country to country.

The most appropriate denominators for marriage and divorce rates are the persons who are eligible to get married or divorced. In the case of marriages, they

Table 9.2 Some indicators of registered marital status: Australia, 2006 census

Marital status (age 15 years or older)	Number (000's)	%	Median age	Sex ratio
Married	7,901	49.6	46.1	99.7
Separated	495	3.1	45.7	83.7
Divorced	1,307	8.2	50.5	76.6
Widowed	937	5.9	75.8	24.8
Never married	5,279	33.2	25.1	117.8
All	15,919	100.0	40.7	95.5
De facto/cohabitation	*1,243*	*7.8*	*31.5*	*98.3*

Source: Table B05 and Table B06 from the Basic Community Profile of Australia for 2006 Census (see Box 5.3)
Note: *De facto/cohabitation* category includes same sex relationships

are those persons who are currently not married (*i.e.*, those who are single, widowed or divorced), while divorce is only possible for those who are currently married. Generally, marital status data are available only on census years.

9.2.3 Marital Status

Population censuses and surveys collect information on marital status and living arrangements. In various stages of people's lives, they enter into relationships that may be formal like marriage or may be informal as cohabitation or *de facto* relationships. While these may involve people of different sexes (heterosexual), in some countries homosexual (involving people of the same sex) relationships are also being recorded, particularly in those countries where such relationships are legal. Formal relationships may change through separation, divorce or the death of a partner.

Table 9.2 presents some data on registered marital status from the 2006 census of Australia. The table shows that the majority of the population consisted of married people, one-third had never married, divorced and widowed constituted 8 % and 6 % of the population respectively. In terms of their median ages, never married were the youngest and widowed were the oldest, other groups were in the range of 45–49 years of age. There was a slight predominance of females in the population 15 years of age and over. This was reflected in most marital status groups, except among the never married where there were many more males. The rather low sex ratio among widowed people is because females tend to survive their partners due to their lower mortality.

In the 2006 Australian census, information was collected on social marital status. It included data on *de facto* relationships and same-sex couples. This is shown in the last row of Table 9.2.

9.3 Families and Households

9.3.1 Concepts and Definitions

Family and household are two related concepts. Nevertheless, definitions of family and household may vary from country to country (United Nations 1973). Generally, there are two main types of families: the *nuclear family* that consists of mother and or father and their children, and the *extended family* where other persons related by blood and/or marriage (or civil partnership) are also included. With the increasing incidence of divorce and remarriage in some countries a third family type – the *blended family* – is also emerging in some countries. This includes the offspring of a couples union and from their previous partnerships. In demographic analysis, definitions of family and household need to be consistent across geographical areas and over time. A household may be defined as one person living alone in a single dwelling or two or more persons living together in the same dwelling who may share meals and other household resources. A family may be defined as two or more people living in the same dwelling who are related by blood (such as children, but including adopted children) or by marriage, civil partnership or cohabitation. Accordingly, it is possible to have two or more families in a given household.

Information on household size and composition is available from most population censuses. However, data on families are not available as frequently from the same source.

It is useful to have one person as the head of the household/family, so that relationships are recorded with reference to this person. In some countries, headship is assumed by the senior male member of the household; however, female headships are also common, particularly in single parent families and when a female is the main income earner. In some instances, such as in the case of Australia, the term *head* of the household has been replaced by that of *reference person* of the household.

9.3.2 Household Size, Headship and Growth

A number of measures are useful in the analysis of households and their characteristics such as size, headship and growth:

$$\overline{h} = \frac{P_t}{H_t} \tag{9.3}$$

$\overline{h}$ is the *average household size*, P_t is the total population at time t and H_t is the total number of households at time t.

$$h_{i,j,t} = \frac{H_{i,j,t}}{P_{i,j,t}} \tag{9.4}$$

Table 9.3 Number of household reference persons and population by age, headship rates, Australia: HES 2009–2010

Age (years)	Reference persons (000's)	Population (000's)	Headship rate (%)
(1)	(2)	(3)	(4) = (2)*100/(3)
18–24	324.0	2,237.7	14.5
25–34	1,417.1	3,112.4	45.5
35–44	1,720.3	3,147.5	54.7
45–54	1,707.6	3,012.7	56.7
55–64	1,471.7	2,477.0	59.4
65+	1,757.9	2,914.3	60.3

Sources: Australia (2011) column (2), and Australia (2010b) column (3)
Note: In each household one person was identified as a reference person. This may be considered a proxy for the household head

$h_{i,j,t}$ is the *headship rate* for age i, sex j, at time t and $H_{i,j,t}$ and $P_{i,j,t}$ are, respectively, the number of household heads and population of age i, sex j, at time t (United Nations 1973).

Just like the rate of growth of population discussed in Sect. 4.3.2, the *rate of growth of households* can also be calculated as follows:

$$r = \frac{LN\left(\frac{H_t}{H_0}\right)}{n} \tag{9.5}$$

r is the average rate of growth of households between time 0 and t, H_0 is the number of households at time 0, H_t is the number at time t, and n is the number of years between 0 and t (Sect. 4.3.2).

In the 2006 census the total population of Australia was enumerated as 20.698 million, of which 1.77 % were living in non-private dwellings. The remaining 20,329,274 people were living in 7,780,193 million households (Australia 2010a). The application of Eq. (9.3) gave the average household size as:

$$\bar{h} = \frac{20,329,274}{7,780,193} = 2.61 \text{ persons per household.}$$

However, the household composition was such that the majority (87.53 %) of the 20.329 million people were living in family households ($\bar{h} = 3.16$), while 3.32 % were living in group households ($\bar{h} = 2.32$), and the remainder were living as lone person households (Australia 2010a).

To calculate headship rates, data were taken from the Household Expenditure Survey (HES) conducted by the Australian Bureau of Statistics from mid-2009 to mid-2010 (Australia 2011). As noted earlier, the term *reference person* in Table 9.3 may be considered equivalent to *head of household*. Population figures were the average of the midyear populations in 2009 and 2010. Headship rates were calculated by using Eq. (9.4) without regard to the sex (j) in both the numerator and denominator.

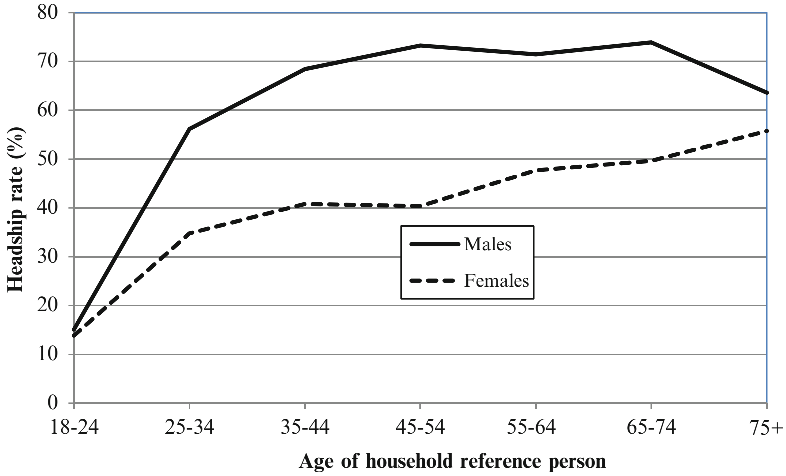


Fig. 9.3 Headship rates by age and sex: Australia, HES 2009–2010 (*Source*: Unpublished Household Expenditure Survey data)

Figure 9.3 gives the headship rates for males and females calculated from unpublished data on age and sex of reference persons from the HES. The figure shows that the sex-specific headship rates were converging in older ages, probably because of higher male mortality.

Age-sex specific headship rates can be calculated by household type and any other characteristics of the households. For this purpose the numerator and denominator in Eq. (9.4) will need to be adjusted accordingly.

In Great Britain there were 16.3 million households in 1961 that increased to 25.5 million by 2011 (United Kingdom 2012). According to Eq. (9.5), the average annual rate of household growth was:

$$r = \frac{LN\left(\frac{25.5}{16.3}\right)}{50} = \frac{0.447513}{50} = 0.895\ \%\ \text{per annum.}$$

During the same period the population of Great Britain increased from just over 50 million to 61 million resulting in an average annual growth rate of 0.38 %. The faster growth in the number of households was due to a number of factors. These include late family formation, propensity to stay single, marriage dissolution and the relative longevity of females.

Following Eq. (4.11) and assuming that the household growth rate remains constant at 0.895 % p.a. in future, the number of households in Great Britain in 2011 would double in just over 77 years. On the other hand, the population of 2011 would take much longer to double: about 180 years.

9.4 Education and Training

Education in a restricted sense is usually associated with learning in schools and other structured educational settings. It enhances means of communication and social interaction as well as economic activity. Together with training it can be seen as an investment in human capital and greater capacity for future social and economic activity.

Literacy is a basic step in the education ladder. There are many definitions of literacy. It is often expressed in terms of the "3 Rs" of reading, writing and arithmetic. Literacy is often defined as "*the ability to read and write with understanding a simple statement related to one's daily life. It involves a continuum of reading and writing skills, and often includes also basic arithmetic skills (numeracy)*" (UNESCO 2004).

Measures used in education and training may relate to the whole population or to a particular age and/or sex category or some other population characteristic.

9.4.1 Literacy

The *literacy rate* is defined as:

$$LR_t^x = \frac{LP_t^x}{P_t^x} \tag{9.6}$$

LR_t^x is the literacy rate for persons with characteristic x at time t, and LP_t^x is the number of literate persons of characteristic x at time t while P_t^x is the population with characteristic x at time t. In the above equation x may represent one or more characteristics of the population. Generally age and/or sex are the commonly used characteristics. In some countries, this rate is specifically calculated for adults and is therefore called the *adult literacy rate*.

According to the Human Development Report (UNDP 2007), India had an estimated population of 1,134.4 million people, of which 67 % were 15 years or older. The adult literacy rate was 61 %. Thus, the number of literate adults was estimated as 760.048 million. This provided the data for Eq. (9.6) as follows:

$LP_t^x =$	Literate persons age 15 years or older in 2005	= 463.629 million
$P_t^x =$	All persons age 15 years or older in 2005	= 760.048 million

$$LR_t^x = \frac{463.629}{760.048} = 0.61 \text{ or } 61\ \%.$$

Table 9.4 Enrolments, estimated populations and enrolment ratios: Mexico, 2009

Education level (h)	Enrolments (E_h^{2009})	Estimated ($P_{h,x}^{2009}$)	GER_h^{2009}
(1)	(2)	(3)	(4) = (2)/(3)
Primary	14,861,000	12,701,700	1.17
Secondary	11,475,000	12,750,000	0.90
Tertiary	2,705,000	9,660,700	0.28

Source: United Nations (2011)
Note: Column (3) was estimated using the procedure outlined in Box 4.1

Table 9.5 Gross enrolment ratios (percentage) for selected countries, 2009

Country	Primary level	Secondary level	Tertiary level
Kenya	113	59	4
Indonesia	121	79	24
Slovenia	98	97	88
United Sates	98	94	86

Source: United Nations (2011), Tables 3, 6 and 10

9.4.2 Gross Enrolment Ratio

The *gross enrolment ratio* reflects the extent to which a country caters for the educational needs of its population, and is calculated as follows:

$$GER_t^h = \frac{E_t^h}{P_t^{h,x}} \tag{9.7}$$

GER_t^h is the *gross enrolment ratio* for education level h, at time t, E_t^h is the number of students enrolled at that level (irrespective of their age) at time t, and $P_t^{h,x}$ is the population in the age group x for that level at time t that corresponds to the official age for studying at the particular education level h (UNESCO 2011). For example, if the entry to primary school is at age 5 and primary school education is 6 years, x will be taken as age 5–10 years (both ages included). Gross enrolment ratios can be calculated for any education level. Usually, they are estimated separately for primary, secondary and tertiary education levels. This index varies between 0 and 1, or 0 and 100 if expressed in percentage terms. However, it may take a value beyond its upper limit if the reported enrolments include pupils who are outside the eligible age group for that level.

Data related to various educational levels for Mexico in 2009 (United Nations 2011) are stated in Table 9.4 to estimate the gross enrolment ratios. The ratio for primary education was greater than 1 indicating that pupils outside the official age for primary education (6–12 years) were also included.

Information on gross enrolment ratios at different education levels for selected countries is given in Table 9.5. In international comparisons, it is important to consider the education policies and practices in different countries, particularly with regard to the age groups relevant for the particular education level.

Table 9.6 Net enrolment rates (percentages) in primary, secondary and tertiary levels for selected countries, 2009

Country	Primary level	Secondary level	Tertiary level
Kenya	83	50	Not available
Indonesia	98	69	Not available
Slovenia	98	92	Not available
United States	92	88	Not available

Source: United Nations (2011), Tables 3, 6 and 10

9.4.3 Net Enrolment Rate

The *net enrolment rate* differs from the *GER* in that it counts only those pupils who were in the official age of the given level of education and relates them to the population of that age:

$$NER_t^h = \frac{\hat{E}_t^h}{P_t^{h,x}} \tag{9.8}$$

NER_t^h is the net enrolment rate for education level h, at time t, and $\hat{E}_t^h$ and $P_t^{h,x}$ are enrolments and population, at time t in age group that corresponds to the official age for studying at educational level h. The denominator for both the gross enrolment ratio and the net enrolment rate is the same; the numerator of the latter is a sub-set of the numerator of the former.

An illustration of the estimation of the net enrolment rate for secondary education in Mexico in 2009 (United Nations 2011) is:

$\hat{E}_{2009}^{secondary}$	=	Pupils in secondary education aged 12–17 completed years being the official age for secondary school education in Mexico, in 2009	= 9,307,500
$P_{2009}^{secondary,12-17years}$	=	Estimated number of people of 12–17 years of age in Mexico in 2009 (estimated using the procedure 1 outlined in Box 4.1)	= 12,750,000

Therefore:

$$NER_{2009}^{secondary} = \frac{9,307,500}{12,750,000} = 0.73 \text{ or } 73\ \%.$$

the net enrolment rate in secondary education in Mexico was 73 % in 2009.

Information on net enrolment rates at different education levels for selected countries is given in Table 9.6. As mentioned before, in making international comparisons it is important to consider the education policies and practices in different countries, particularly with regards to the age groups relevant for the particular education level. It is difficult to calculate the net enrolment rates for tertiary students as there is no fixed age group for tertiary education.

9.4.4 Retention Rate

Retention involves the concept of continuity in education. The retention rate is used to measure the proportion of students that completed a given course, such as high school or a particular subject within a course, as a proportion of those that started in that course. This involves tracking a cohort of students and is, therefore, a longitudinal measure, and is estimated as follows:

$$RR_f^c = \frac{P_f^c}{P^c} \tag{9.9}$$

RR_f^c is the *retention rate* for a course c finishing in year f, P^c is the number of students who started the course and P_f^c is the number of that cohort of students that completed the course in year f.

Australian school data give an example (Australia 2007). There were 196,431 students in the final year of secondary schools in 2006 ($P_{2006}^{secondary}$), the number of that cohort of students that started first year of secondary school was 262,960 ($P^{secondary}$). Accordingly:

$$RR_{2006}^{secondary} = \frac{196,431 * 100}{262,960} = 74.7\ \%.$$

the retention rate was 74.7 %. The remaining 25.3 % represents those who dropped out of secondary school as well as those that repeated one or more years.

9.4.5 Educational Attainment

The highest level of education attained is an important characteristic of a population. Education systems and the labelling of levels of education vary from country to country. Some countries have their own standard classification of education and qualifications.

Table 9.7 presents data from the 2006 Australian Census Basic Community Profile (Box 5.3). These statistics give an overview of people 15 years of age and over who were deemed to have completed their education. The table shows that about 30 % of the population had a degree or a higher qualification. The level of education could not be determined for about a quarter of the population. These people were, on the average, somewhat older than those for whom information was available. Apart from postgraduate degree and certificate qualifications where there was a preponderance of males, in all other categories females out-numbered males.

Table 9.7 Level of education (non-school qualification): Australia, 2006 Census

Level of education (persons aged 15 years and over)	Number (000's)	%	Sex ratio
Postgraduate degree	413	4.9	133.4
Graduate diploma/graduate certificate	229	2.7	58.9
Bachelor degree	1,841	22.0	81.4
Advanced diploma/diploma	1,130	13.6	74.0
Certificate	2,663	31.9	207.8
Inadequately described	235	2.8	76.7
Not stated	1,851	22.1	91.3
All levels	8,362	100.0	111.7

Source: Table B39 from the Basic Community Profile of Australia for 2006 Census (see Box 5.3)
Note: This table excludes those still at school and persons with a qualification out of the scope of the Australian Standard Classification of Education

9.5 Labour Force

In most countries labour force statistics are collected through censuses as well as periodic labour force sample surveys. The International Labour Organization has a set of resolutions, and recommendations and guidelines (ILO 2013) to help member countries to collect internationally comparable labour statistics.

Generally, labour force statistics are collected for population aged 15 years and over. However, certain criteria are used to determine whether or not a person is a member of the labour force and on what basis. The criteria include whether a person is employed or unemployed but looking for employment. Employment may be on a full-time or part-time basis depending on the number of hours worked.

9.5.1 Labour Force Participation Rate

The *labour force participation rate* relates people considered to be in the labour force at a point in time to the population at that time. The labour force is made up of people employed or unemployed but actively seeking work. Those solely engaged in domestic and other work in their own households are generally not considered as members of the labour force. This rate is usually calculated for a period or a given date and on a proportional basis:

$$LFPR^x_{t\to t+1} = \frac{LF^x_{t\to 1}}{\overline{P}_t} \tag{9.10}$$

$LFPR^x_{t\to t+1}$ is the labour force participation rate of persons with characteristic x during the time t, $LF^x_{t\to t+1}$ is the number of persons of characteristic x in the labour force during the same time t and $\overline{P}_t$ is the average population of characteristic x in

Table 9.8 Labour force, population and labour force participation rates by age and sex: Japan, 2000 census

	Labour force (000's)		Population (000's)		Labour force participation rate (%)	
Age (years)	Males	Females	Males	Females	Males	Females
(1)	(2)	(3)	(4)	(5)	(6) = (2)/(4)	(7) = (3)/(5)
15–19	665.9	561.0	3,834.0	3,654.2	17.4	15.4
20–24	3,024.0	2,901.7	4,307.2	4,114.2	70.2	70.5
25–34	8,765.5	5,828.3	9,402.1	9,164.8	93.2	63.6
35–44	7,674.4	5,054.1	8,020.5	7,894.6	95.7	64.0
45–54	9,216.5	6,589.4	9,677.8	9,680.2	95.2	68.1
55–64	6,659.9	4,075.2	8,039.8	8,430.2	82.8	48.3
65+	3,244.0	1,837.8	9,222.2	12,783.1	35.2	14.4
15–64	36,006.2	25,009.7	43,281.4	42,938.2	83.2	58.2
15+	39,250.2	26,847.5	52,503.6	55,721.3	74.8	48.2

Source: Japan (2011)

the middle of t. Both the numerator and denominator in Eq. (9.10) refer to the population aged 15 years or older, or the population 15–64 years of age group, assuming that retirement is at 65 years of age. These rates may be calculated by age and separately for males and females. However, other characteristics such as marital status or number of children (particularly in the case of female labour force) may also be specified; for example, age-specific labour force participation rates for married women with children.

Table 9.8 shows the estimation of age-sex specific labour force participation rates using the data from the 2000 census of Japan. The rates in columns (6) and (7) were estimated using Eq. (9.10). For example, the rate for males in the 15–64 years of age group was calculated as $\frac{36{,}006.2}{43{,}281.3} = 0.832$ or 83.2 %.

Except for the 20–24 years age group, male participation rates were higher than those of females. Even in the 20–24 years age group the difference was only marginal. As expected, the participation rates for persons aged 15–64 years were higher than those for all persons 15 years and over (Table 9.8). The reason for this is that the majority of those aged 65 and over are retired. This applied to both males and females.

Labour force participation rates for males in Japan compared to those in Australia and the United States are drawn in Fig. 9.4. Rates for Japanese males were somewhat lower up to mid-20s. However, they were consistently higher than the Australian and United States rates after that age. Even among persons aged 75 years and over, the Japanese rate was nearly double that for the other two countries.

Female labour force participation rates for Japan in Fig. 9.5 show a bi-modal distribution that peaks at ages 20–24 and 45–54 years. This could be due to some Japanese women temporarily leaving the labour force to have children and returning once the children are of school age. This trend has also been observed in other countries such as Australia. However, the bi-modality of the female rates was not evident from the United States data.

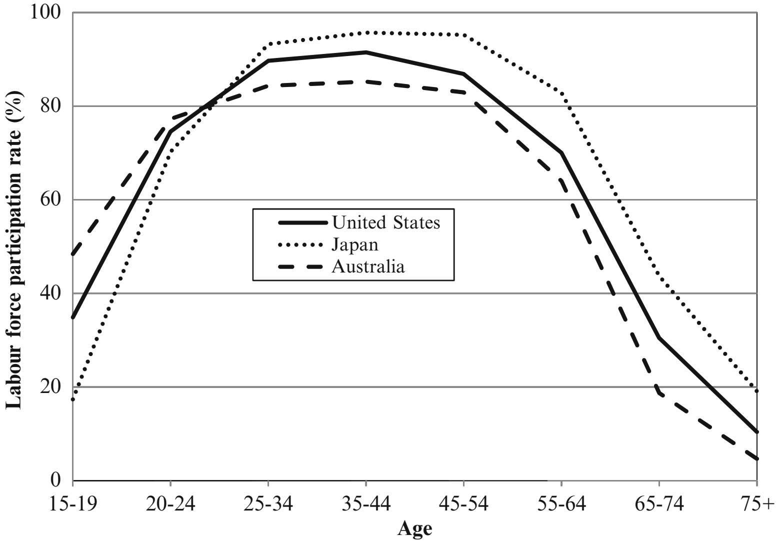


Fig. 9.4 Labour force participation rates of males (percentage) by age: United States, 2010; Japan, 2000 and Australia, 2006 (*Notes*: The first age group in the United States data was16–19 rather than 15–19. *Sources*: Toossi (2012); Japan (2011); Table B36 from the Basic Community Profile of Australia for 2006 Census (see Box 5.3))

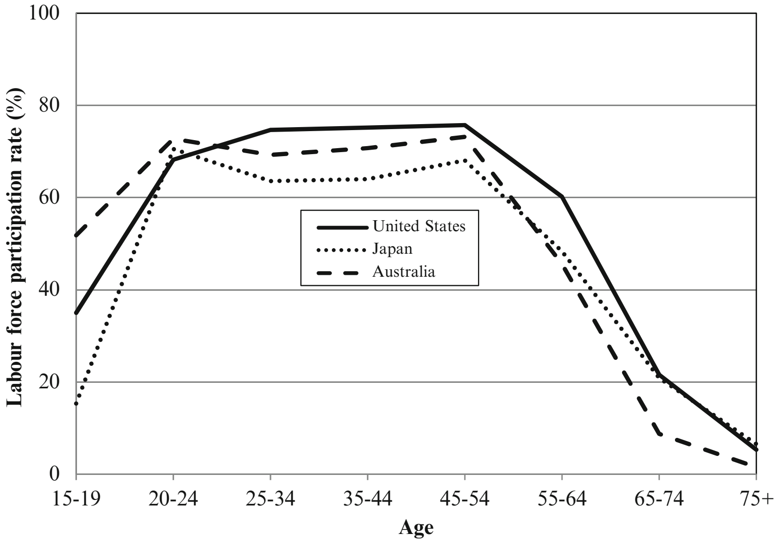


Fig. 9.5 Labour force participation rates of females by age (percentage): United States, 2010; Japan, 2000 and Australia, 2006 (*Notes*: The first age group for US data was 16–19 rather than 15–19. *Sources*: Toossi (2012); Japan (2011); Table B36 from the Basic Community Profile of Australia for 2006 Census (see Box 5.3))

Table 9.9 Employment and unemployment rates: Japan, 2000 census

	Labour force (000's)		Rate (%)	
Age (years)	Employed	Unemployed	Employment	Unemployment
(1)	(2)	(3)	(4) = (2)/(2) + (3)	(5) = (3)/(2) + (3)
15–19	1,066.1	160.9	86.9	13.1
20–24	5,429.6	496.1	91.6	8.4
25–34	13,788.5	805.4	94.5	5.5
35–44	12,315.5	413.1	96.8	3.2
45–54	15,324.7	481.1	97.0	3.0
55–64	10,162.1	573.0	94.7	5.3
65+	4,891.4	190.3	96.3	3.7
15–64	58,086.5	2,929.6	95.2	4.8
15+	62,977.9	3,119.9	95.3	4.7

Source: Japan (2011)

9.5.2 Employment and Unemployment Rates

Measurement of employment and unemployment rates is useful to assess the engagement of the labour force stock. These rates are expressed as a proportion (or percentage) of those in the labour force. Accordingly, the *employment rate* is specified as:

$$ER^x_{t\to t+1} = \frac{E^x_{t\to t+1}}{LF^x_{t\to t+1}} \tag{9.11}$$

$ER^x_{t\to t+1}$ is the employment rate for persons of characteristic x during year $t, E^x_{t\to t+1}$ is the number of employed people of characteristic x during the same t and $LF^x_{t\to t+1}$ is the number of people in the labour force of the same characteristic during the same time.

Similarly the *unemployment rate* is defined as:

$$UR^x_{t\to t+1} = \frac{U^x_{t\to t+1}}{LF^x_{t\to t+1}} \tag{9.12}$$

$UR^x_{t\to t+1}$ is the unemployment rate for persons with characteristic x during the time t and $U^x_{t\to t+1}$ is the number of unemployed persons of the same characteristic during the same time, and $LF^x_{t\to t+1}$ has the same definition as in Eq. 9.11. In some countries, unemployment rates are calculated on a quarterly or even monthly basis.

Since the labour force by definition consists of employed and unemployed people, it is obvious that $ER^x_t + UR^x_t = 100\ \%$.

Table 9.9 gives data for employed and unemployed people in Japan, at the time of the 2000 census. Equation 9.11 was applied to estimate the employment rate for the population aged 15 years and over, while Eq. (9.12) was used to calculate the unemployment rate for the same age group, as shown below:

$$ER_{2000}^{15+} = \frac{62,978}{(62,978 + 3,119.9)} = 0.953 \text{ or } 95.3\ \%$$

$$UR_{2000}^{15+} = \frac{3,119.9}{(62,978 + 3,119.9)} = 0.047 \text{ or } 4.7\ \%.$$

9.5.3 Job Creation Rate

The *job creation rate* relates the growth in the number of people employed over a period of time, usually 1 year, to the number of people employed at the start of that period. The rate may be calculated for any specific characteristic such as age or sex. The rate is defined as:

$$JCR_{t\to t+1}^{x} = \frac{E_{t+1}^{x} - E_{t}^{x}}{E_{t}^{x}} \tag{9.13}$$

$JCR_{t\to t+1}^{x}$ is the job creation rate between time t and $t + 1$ for persons of characteristic x, and E_t^x and E_{t+1}^x is the number of persons of characteristic x employed at time t and $t + 1$ respectively. When the measurement involves more than 1 year, average annual job creation rates can be estimated using a similar calculation as that used for the rate of population growth (Sect. 4.3.2).

The number of people employed in Australia in November 2009 and 1 year later (Australia 2012) provide figures for the estimation of the job creation rate for people aged 15 years and over:

Number of employed people aged 15 years and over in November 2009 (E_t^x)	= 10,664,900
Number of employed people age 15 years and over in November 2010 (E_{t+1}^x)	= 11,323,200

$$JCR_{t\to t+1}^{15+} = \frac{(11,323,200 - 10,664,900)}{10,664,900} = 0.062 \text{ or } 6.2\ \%$$

the job creation rate was 6.2 % in Australia over that 1-year period.

9.5.4 Labour Force Flows and Growth

From a demographic perspective, the labour force could be said to have its own balancing equation (Sect. 4.2): the number of people employed at date t plus hirings from t to $t + 1$ and separations during the same period equal the number employed at time $t + 1$:

$$E_{t+1} = E_t + H_{t\to t+1} - S_{t\to t+1} \tag{9.14}$$

E_{t+1} is the number of people employed at time $t + 1$, E_t is the number employed at t, $H_{t\to t+1}$ is the number of people hired from t to $t + 1$ and $S_{t\to t+1}$ is the number of people that were separated from their jobs due to (a) voluntary separation initiated by the employee, (b) involuntary separation initiated by the employer including layoffs and discharges, and (c) other separations due to death, disability, transfers to other locations and retirement.

These absolute numbers can be translated into hirings and separation rates and the difference between these two rates gives the labour force growth rate.

Usually, these rates are expressed as a percentage:

$$AHR_{t\to t+1} = \frac{H_{t\to t+1}}{\overline{E}_{t\to t+1}} \tag{9.15}$$

$AHR_{t\to t+1}$ is the average hiring rate in period $t \to t + 1$, $H_{t\to t+1}$ is hirings as defined in Eq. (9.14), and $\overline{E}_{t\to t+1}$ is the average of people employed during the period.

$$ASR_{t\to t+1} = \frac{S_{t\to t+1}}{\overline{E}_{t\to t+1}} \tag{9.16}$$

$ASR_{t\to t+1}$ is the average separation rate in period $t \to t + 1$, $S_{t\to t+1}$ is separations as defined in Eq. (9.14), and $\overline{E}_{t\to t+1}$ is as defined in Eq. (9.15).

$$AEGR_{t\to t+1} = AHR_{t\to t+1} - ASR_{t\to t+1} \tag{9.17}$$

$AEGR_{t\to t+1}$ is the average employment growth rate in period $t \to t + 1$, and the other two terms are as previously defined. It is apparent that the average employment growth rate can be positive or negative depending whether the hiring rate is greater or smaller than the separation rate.

United States employment statistics for 2011 (United States 2012) can be used to give examples of the estimation of these rates. The number of people hired during 2011 was estimated as 50,006,000, the number of people separated was estimated as 48,138,000 and the average number of people employed during the year was 131,561,800 (the estimation of the average number of people employed followed procedure 1 in Box 4.1 because the data were only available for the number and rate of hires and separations in each month).

Accordingly, the average hiring rate in the United States for 2011 was:

$$AHR_{2011} = \frac{50,006,000}{131,561,800} = 0.380 = 38.0\ \%$$

and the average separation rate:

$$ASR_{2011} = \frac{48,138,000}{131,561,800} = 0.366 = 36.6\ \%.$$

These two rates indicate that:

$$AEGR_{2011} = 0.380 - 0.366 = 0.014 = 1.4\ \%$$

employment in the United States grew by 1.4 % during 2011.

9.6 Occupation and Industry

9.6.1 Occupation

Information on occupation of members of the labour force and the industry where they work tends to be recorded in censuses and surveys. The International Labour Organization has a classification of occupations and industries but individual countries adapt them to cater for their particular circumstances (ILO 2010). The occupational classification has various versions some more detailed than others.

As an illustration, Table 9.10 presents some data on occupations from the 2006 census of Australia (Basic Community Profile Table B44). The categories listed are an abridged version of the detailed classification. The table shows that professionals were the largest occupational group and it was one of the four occupational groups where women outnumbered men. The second largest category was clerical and administrative workers, where there was a preponderance of females. Men out-numbered women particularly among managers, technicians and trade workers and machinery operators and drivers.

9.6.2 Industry

Another important characteristic is the industry in which people are employed. Again, there is an international standard industrial classification of economic activities which has been used by many countries subject to changes in line with local circumstances.

Data on employment by industry from the 2006 census of Australia (Basic Community Profile Table B42) is shown in Table 9.11. The table indicates that the three largest industries of employment were retail trade, manufacturing and health care and social assistance. These accounted for nearly one-third of employment in Australia (Australia 2013).

Table 9.10 Occupation classification: Australia, 2006 census

Occupation (persons aged 15 years and over)	Number (000's)	%	Median age	Sex ratio
Managers	1,202	13.2	39.7	190.7
Professionals	1,806	19.8	38.0	88.9
Technicians and trade workers	1,309	14.4	36.3	578.5
Community and personal service workers	802	8.8	36.1	45.3
Clerical and administrative workers	1,366	15.0	38.0	30.2
Sales workers	896	9.8	28.1	62.1
Machinery operators and drivers	605	6.6	38.8	860.0
Labourers	953	10.5	37.0	174.1
Inadequately described/not stated	166	1.8	37.9	148.7
All occupations	9,105	100.0	37.5	117.1

Source: Table B44 from the Basic Community Profile of Australia for 2006 Census (see Box 5.3)

Table 9.11 Industry of employment: Australia, 2006 census

Industry (persons aged 15 years and over)	Number (000's)	%
Agriculture, forestry and fishing	281	3.1
Mining	107	1.2
Manufacturing	952	10.5
Electricity, gas, water and waste services	89	1.0
Construction	710	7.8
Wholesale trade	396	4.3
Retail trade	1,033	11.3
Accommodation and food services	575	6.3
Transport, postal and warehousing	428	4.7
Information media and telecommunications	177	1.9
Financial and insurance services	349	3.8
Rental, hiring and real estate services	154	1.7
Professional, scientific and technical services	602	6.6
Administrative and support services	287	3.2
Public administration and safety	609	6.7
Education and training	698	7.7
Health care and social assistance	956	10.5
Arts and recreation services	127	1.4
Other services	338	3.7
Inadequately described/Not stated	236	2.6
Total	9,104	100.0

Source: Table B42 from the Basic Community Profile of Australia for 2006 Census (see Box 5.3)

References

Australia. (2007). *2006 Schools Australia* (Catalogue No. 4221.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/9DDA83611950C66-FCA25728B000CFC92/$File/42210_2006.pdf. Accessed Feb 2013.

Australia. (2010a). *Household and family projections*, *Australia*, *2006 to 2031* (Catalogue No. 3236.0). Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3236.02006%20to%202031?OpenDocument. Accessed Feb 2013.

Australia. (2010b). *Population by age and sex*, *Australian States and Territories*, *Jun 2010* (Catalogue No. 3201.0). Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3201.0Jun%202010?OpenDocument. Accessed Feb 2013.

Australia. (2011). *2009–10 Household expenditure survey*, *Australia*: *Summary of results* (Catalogue No. 6530.0). Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/CB07CC895DCE2829CA2579020015D8FD/$File/65300_2009-10.pdf. Accessed Feb 2013.

Australia. (2012). *Australian labour market statistics*, *Jan 2012* (Catalogue No. 6105.0). Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6105.0Jan%202012?OpenDocument. Accessed Feb 2013.

Australia. (2013). Census home: Data & analysis. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/websitedbs/censushome.nsf/home/data?opendocument#from-banner=LN. Accessed 8 Feb 2013.

ILO. (2010). *International standard classification of occupations*. Geneva: International Labour Organization. http://www.ilo.org/public/english/bureau/stat/isco/index.htm. Accessed Feb 2013.

ILO. (2013). *Standards and guidelines on labour statistics*. Geneva: International Labour Organization. http://www.ilo.org/global/statistics-and-databases/standards-and-guidelines/lang–en/index.htm. Accessed Feb 2013.

Japan. (2011) e-Stat. Portal site of official statistics of Japan. 2000 Population census (table 1). Tokyo: Statistics Bureau. http://www.e-stat.go.jp/SG1/estat/ListE.do?bid=000000030097&cycode=0. Accessed Feb 2013.

Toossi, M. (2012). Labor force projections to 2020: A more slowly growing workforce. *Monthly Labor Review*, January issue, 43–64. http://www.bls.gov/opub/mlr/2012/01/art3full.pdf. Accessed Feb 2013. .

UNDP. (2007). *Human development report 2007/2008*. New York: United Nations Development Programme. http://hdr.undp.org/en/media/HDR_20072008_EN_Indicator_tables.pdf. Accessed Feb 2013.

UNESCO. (2004). *The plurality of literacy and its implications for policies and programmes* (UNESCO Education Sector Paper). Paris: United Nations Educational, Scientific and Cultural Organization. http://unesdoc.unesco.org/images/0013/001362/136246e.pdf. Accessed Feb 2013.

UNESCO. (2011). *Global education digest 2011*: *Comparing education statistics across the World*. Paris: United Nations Educational, Scientific and Cultural Organization. http://www.uis.unesco.org/Library/Documents/global_education_digest_2011_en.pdf. Accessed Feb 2013.

United Kingdom. (2012) *Measuring national well-being*: *Households and families*, *2012*. Newport: Office of National Statistics. http://www.ons.gov.uk/ons/rel/wellbeing/measuring-national-well-being/households-and-families/art—households-and-families.html. Accessed 17 Feb 2013.

United Nations. (1973). *Manual VII*: *Methods of projecting households and families*. New York: Department of Social and Economic Affairs. http://www.un.org/esa/population/techcoop/SocInd/manual7/manual7.html. Accessed Feb 2013.

United Nations. (2011). *Demographic yearbook*, *2009*. New York: Department of Social and Economic Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dybsets/2009-2010.pdf. Accessed Feb 2013.

United States. (2012). *News release*: *Job openings and labor turnover – January 2012*. Washington: Bureau of Labor Statistics. http://www.bls.gov/news.release/archives/jolts_03132012.pdf. Accessed Feb 2013.

Chapter 10
Multiple Decrement Life Tables

10.1 Purpose

This chapter extends the concept of a life table where a cohort of persons is diminished as a result of death. In demographic analysis consideration is given to some situations where cohorts may be subjected to multiple decrements, such as different causes of death. In addition, the chapter widens the concept to the building of multistate life tables concerned not only with decrements from life to death but also with possible movements among various active states such as entry or exit from the labour force and different marital status categories.

10.2 The Multiple Decrement Perspective

In the life table concept, a cohort of persons is diminished as a result of death or one decrement. Multiple decrement tables deal with circumstances where a person may be subjected to two or more decrements. Multiple decrement life tables use similar principles to those in the building of a life table but allow the possibility of other decrements.

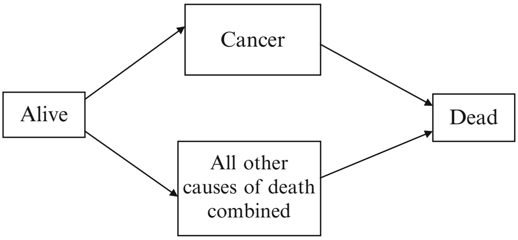
In a cohort of single persons, as they go from one age to another, some of them will get married and others will die. This is an example of a cohort experiencing two decrements: marriage and death or two different states. In demographic analysis situations arise where cohorts may be subjected to multiple decrements, such as different causes of death. Multiple decrement life tables are built using the method described in Chap. 7, except that there are more than one q_x (probability of dying or change in state) and related columns depending upon the number of decrements considered.

Another type of multiple decrement tables are *multistate* life tables. Such life tables take into account not only the movements between life (an *active state*) and death (an *absorbing state*) but also possible movements among various types of

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_10,

Fig. 10.1 Perspective of a single decrement life table

Fig. 10.2 Perspective of a multiple decrement life table



active states. Unlike conventional life tables or multiple decrement tables, multistate models do not have a standard table format. While the measures found in multistate life tables usually have counterparts in conventional or multiple decrement life tables, they may be presented in ways not found in those tables. Output measures could include the proportion of the population dying while in each state, expected duration in a given state (*e.g*., never married), and number of transitions to each state per person. These measures could, in turn, be assembled into a set of transition probabilities that cover long periods of time, including life times.

As an introduction to this perspective, movements from one state to another are illustrated in Figs. 10.1 and 10.2. In these diagrams, the origin of a given arrow represents a decrement from a *state of origin* and its destination represents an increment to a *receiving state*. Figure 10.1 represents the conventional life table discussed in Chap. 7 in which a person moves from the *state of alive* to the *state of dead*. In this case, *alive* is the *origin state* and *dead* is the *receiving state*. Death is not only a receiving state, but also what is known as an *absorbing state* because there is no exit from death. It is a one way street – people can move from life to death, but not back.

Figure 10.2 indicates that one can move from the state of alive to the state of dead *via* one of the many causes of death rather than simply all causes combined. The figure shows a specific cause of death, *cancer*, and *all other causes of death combined*, but this figure can be generalized to show a range of causes.

Smallpox was a major killer that has been virtually eliminated world-wide. Deaths due to smallpox can be examined in the presence of other (all other) causes by the analysis of the historical records *via* a multiple decrement approach, to assess the effect of eliminating smallpox on life expectancy. With the elimination of smallpox, people who may have otherwise died from it now die from other causes. This implies that deaths from *other causes* have increased. Given this trade-off, it is appropriate to ask what effect the elimination of smallpox has had on age-specific probabilities of dying. The multiple decrement approach can be employed to answer this type of questions.

The effect of eliminating smallpox using a multiple decrement approach was developed by the French scientist and mathematician, Daniel Bernoulli in 1776.

In so doing, he is credited with producing the first *double decrement* life table. Bernoulli estimated that in the absence of smallpox the expectation of life at birth was 29.9 years and in its presence it was 26.7 years, a difference of 3.2 years (Daw 1979). While the work of Bernoulli and his contemporaries was useful, their data were manufactured, to some degree, and at least some of their results were subjected to question (Daw 1979).

The complete answer to the effect of the elimination of smallpox has had on age-specific probabilities of dying is problematic. This can be illustrated by taking an example from Feeney (1974). Suppose there are N persons who have attained age x and who experience D deaths before reaching age $x + n$ of which D_s are deaths due to smallpox. It seems reasonable, at first glance, to estimate the probability of dying by causes other than smallpox as $P_{os} = \frac{D - D_s}{N}$. However, upon further consideration, it becomes clear that it underestimates the probability because the D_s persons who do not die of smallpox are still subject to death by other causes, and some of them will die. If $n = \infty$, for the N persons aged x, the probability of them dying from other causes before reaching age $x + \infty$ has to be 1.00 since nobody lives forever. However, this is not what is suggested by $P_{os} = \frac{D - D_s}{N}$, where P_{os} is smaller than 1.00 since $\frac{D - D_s}{N}$ is smaller than 1.00 for $D_s > 0$. Alternatively, if $D_s = 0$, then smallpox is not a cause of death and $\frac{D - D_s}{N} = \frac{D - 0}{N} = 1.00$ since $D = N$ when $n = \infty$. Daw (1979) makes a similar observation about the absurdities that can arise in dealing with multiple decrements, if care is not taken.

10.3 Multiple Decrement Life Tables

It is apparent that there are issues underlying an attempt to deal with multiple causes of death. This is especially the case when looking at the effect on life expectancy if one or more causes are eliminated. Multiple decrement life tables provide a solution to this problem. Put simply, a multiple decrement life table is one in which deaths are considered by cause. Since there are many causes of death, it is common practice to look at one cause (or a combined set of specific causes) at a time so that the life table is manageable.

10.3.1 Example Using Cause of Death Data for the United States

The total number of deaths in 2007 for the United States population is given in Table 10.1 by age (both sexes combined) as well as deaths due to four causes: (1) diseases of heart; (2) malignant neoplasms; (3) unintentional accidents; and (4) deaths from all causes other than (1) to (3). Obviously, the sum of deaths by these four types of causes equals total deaths.

Table 10.1 Death rates, estimated population and deaths by cause of death and age, both sexes combined: United States, 2007

				Deaths from			
Age (years)	Death rate ${}_nm_x$	Estimated population	Deaths from all causes	Diseases of heart	Malignant neoplasms	Unintentional accidents	All other causes
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	0.006845	4,256,830	29,138	424	72	1,285	27,357
1–4	0.000286	16,444,056	4,703	173	364	1,588	2,578
5–14	0.000153	40,176,471	6,147	241	959	2,194	2,753
15–24	0.000794	42,798,489	33,982	1,084	1,653	15,897	15,348
25–34	0.001049	40,583,413	42,572	3,223	3,463	14,977	20,909
35–44	0.001844	43,170,282	79,606	11,839	13,288	16,931	37,548
45–54	0.004209	43,878,831	184,686	37,434	50,167	20,315	76,770
55–64	0.008777	32,711,633	287,110	65,527	103,171	12,193	106,219
65–74	0.020113	19,352,558	389,238	89,589	138,466	8,753	152,430
75–84	0.050116	13,023,426	652,682	171,257	163,608	13,736	304,081
85+	0.129465	5,512,277	713,647	235,249	87,656	15,803	374,939
Total	0.008036	301,908,266	2,423,511	616,040	562,867	123,672	1,120,932

Source: Xu et al. (2010) Tables 3–4 for columns (2) and (4) respectively, and Table 10 for columns (4) to (7)

Note: There were 201 deaths where the age at death was not known. These have been excluded from further analysis. Column (3) was estimated by dividing column (2) into column (4), as discussed in Box 4.1. Column (8) was the residual of column (4) minus sum of columns (5) to 7

Regardless of how multiple causes of death are looked at, a *regular* life table is needed in which all deaths are taken into account. Table 10.2 provides this in the form of an abridged life table for the United States population in 2007 (both sexes combined). As can be seen in this table, life expectancy at birth was 78.5 years. This provides a benchmark of overall mortality due to all causes of death combined.

The cause-specific deaths in Table 10.1 were used to calculate life tables assuming that a particular cause of death was eliminated (Tables 10.3, 10.4, 10.5 and 10.6). The number of deaths for each of these tables was calculated by ignoring the deaths due to diseases of heart in Table 10.3, malignant neoplasms in Table 10.4, unintentional accidents in Table 10.5, and all other causes in Table 10.6. For each of these tables, the ${}_nm_x$ values were calculated using the estimated total population by age from column (3) of Table 10.1. Again the methodology used for the calculation of the remaining columns was the same as described in Sect. 7.6.

Column (9) in each of the Tables 10.3, 10.4, 10.5 and 10.6 represents the following:

$e_x^{\,1}$ – life expectancy at age x if only deaths due to diseases of heart were eliminated;
$e_x^{\,2}$ – life expectancy at age x if only deaths due to malignant neoplasms were eliminated;
$e_x^{\,3}$ – life expectancy at age x if only deaths due to unintentional accidents were eliminated;
$e_x^{\,4}$ – life expectancy at age x if only deaths due to other causes were eliminated.

Table 10.2 Abridged life table for both sexes and all causes of death combined: United States, 2007

Age (years)	${}_nm_x$	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	e_x
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	0.006845	0.006845	100,000	685	99,658	7,854,035	78.5
1–4	0.000286	0.001143	99,315	114	397,032	7,754,377	78.1
5–14	0.000153	0.001529	99,201	152	991,250	7,357,345	74.2
15–24	0.000794	0.007909	99,049	783	986,575	6,366,095	64.3
25–34	0.001049	0.010435	98,266	1,025	977,535	5,379,520	54.7
35–44	0.001844	0.018272	97,241	1,777	963,525	4,401,985	45.3
45–54	0.004209	0.041222	95,464	3,935	934,965	3,438,460	36.0
55–64	0.008777	0.084080	91,529	7,696	876,810	2,503,495	27.4
65–74	0.020113	0.182752	83,833	15,321	761,725	1,626,685	19.4
75–84	0.050116	0.400742	68,512	27,456	547,840	864,960	12.6
85+	0.129465	1.000000	41,056	41,056	317,120	317,120	7.7

Note: Column (2) was copied from Table 10.1, and the remaining columns were calculated using the methodology discussed in Sect. 7.6. For age 0, m_0 was taken as an approximation of q_0. Eq. (7.16) was utilized for estimating L_{85+}

Table 10.3 Abridged life table for both sexes combined if deaths due to diseases of heart eliminated: United States, 2007

Age (years)	*Deaths*	${}_nm_x$	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	$e_x{}^1$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	28,714	0.006745	0.006745	100,000	675	99,663	8,276,349	82.8
1–4	4,530	0.000275	0.001101	99,325	109	397,082	8,176,686	82.3
5–14	5,906	0.000147	0.001469	99,216	146	991,430	7,779,604	78.4
15–24	32,898	0.000769	0.007657	99,070	759	986,905	6,788,174	68.5
25–34	39,349	0.000970	0.009649	98,311	949	978,365	5,801,269	59.0
35–44	67,767	0.001570	0.015575	97,362	1,516	966,040	4,822,904	49.5
45–54	147,252	0.003356	0.033005	95,846	3,163	942,645	3,856,864	40.2
55–64	221,583	0.006774	0.065519	92,683	6,072	896,470	2,914,219	31.4
65–74	299,649	0.015484	0.143711	86,611	12,447	803,875	2,017,749	23.3
75–84	481,425	0.036966	0.311995	74,164	23,139	625,945	1,213,874	16.4
85+	478,398	0.086788	1.000000	51,025	51,025	587,929	587,929	11.5

Note: Column (2) in this table was estimated by subtracting in Table 10.1 column (5) from column (4). Life table methodology discussed in Sect. 7.6 was used. For age 0, m_0 was taken as an approximation of q_0. Eq. (7.16) was utilized for estimating L_{85+}

Table 10.7 indicates the impact of the eradication of various diseases on overall life expectancy. Obviously, the impact was largest if all causes other than diseases of heart, malignant neoplasms and unintentional accidents were eliminated. The second highest influence was that of the elimination of the diseases of heart, followed by malignant neoplasms and unintentional accidents.

With decreasing levels of mortality experienced by most countries, life tables dealing only with mortality have lost some of their usefulness as indicators of the

Table 10.4 Abridged life table for both sexes combined and deaths due to malignant neoplasms eliminated: United States, 2007

Age (years)	*Deaths*	${}_nm_x$	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	$e_x{}^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	29,066	0.006828	0.006828	100,000	683	99,659	8,210,805	82.1
1–4	4,339	0.000264	0.001055	99,317	105	397,058	8,111,146	81.7
5–14	5,188	0.000129	0.001290	99,212	128	991,480	7,714,088	77.8
15–24	32,329	0.000755	0.007525	99,084	746	987,110	6,722,608	67.8
25–34	39,109	0.000964	0.009590	98,338	943	978,665	5,735,498	58.3
35–44	66,318	0.001536	0.015245	97,395	1,485	966,525	4,756,833	48.8
45–54	134,519	0.003066	0.030194	95,910	2,896	944,620	3,790,308	39.5
55–64	183,939	0.005623	0.054693	93,014	5,087	904,705	2,845,688	30.6
65–74	250,772	0.012958	0.121696	87,927	10,700	825,770	1,940,983	22.1
75–84	489,074	0.037553	0.316168	77,227	24,417	650,185	1,115,213	14.4
85+	625,991	0.113563	1.000000	52,810	52,810	465,028	465,028	8.8

Note: Column (2) in this table was estimated by subtracting in Table 10.1 column (6) from column (4). Life table methodology discussed in Sect. 7.6 was used. For age 0, m_0 was taken as an approximation of q_0. Eq. (7.16) was utilized for estimating L_{85+}

Table 10.5 Abridged life table for both sexes combined and deaths due to unintentional accidents eliminated: United States, 2007

Age (years)	*Deaths*	${}_nm_x$	${}_nq_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	$e_x{}^3$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	27,853	0.006543	0.006543	100,000	654	99,673	7,959,590	79.6
1–4	3,115	0.000189	0.000757	99,346	75	397,234	7,859,917	79.1
5–14	3,953	0.000098	0.000983	99,271	98	992,220	7,462,683	75.2
15–24	18,085	0.000423	0.004217	99,173	418	989,640	6,470,463	65.2
25–34	27,595	0.000680	0.006777	98,755	669	984,205	5,480,823	55.5
35–44	62,675	0.001452	0.014413	98,086	1,414	973,790	4,496,618	45.8
45–54	164,371	0.003746	0.036771	96,672	3,555	948,945	3,522,828	36.4
55–64	274,917	0.008404	0.080653	93,117	7,510	893,620	2,573,883	27.6
65–74	380,485	0.019661	0.179010	85,607	15,325	779,445	1,680,263	19.6
75–84	638,946	0.049061	0.393970	70,282	27,689	564,375	900,818	12.8
85+	697,844	0.126598	1.000000	42,593	42,593	336,443	336,443	7.9

Note: Column (2) in this table was estimated by subtracting in Table 10.1 column (7) from column (4). Life table methodology discussed in Sect. 7.6 was used. For age 0, m_0 was taken as an approximation of q_0. Eq. (7.16) was utilized for estimating L_{85+}

general health. This led to the interest in *healthy life expectancy* and the incorporation of such indicators into life tables (Mathers et al. 2000). An example of this concept is a life that is free of major disabilities with a single decrement in which there is a move into a life with a major disability. A life table for a given population could be built based on the information on the proportion of people by age who are free of major disabilities.

Table 10.6 Abridged life table for both sexes combined and deaths due to all causes other than diseases of heart, malignant neoplasms and unintentional accidents eliminated: United States, 2007

Age (years)	*Deaths*	$_nm_x$	$_nq_x$	l_x	$_nd_x$	$_nL_x$	T_x	$e_x^{\,4}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	1,781	0.000418	0.000418	100,000	42	99,979	8,917,083	89.2
1–4	2,125	0.000129	0.000517	99,958	52	399,728	8,817,104	88.2
5–14	3,394	0.000084	0.000844	99,906	84	998,640	8,417,376	84.3
15–24	18,634	0.000435	0.004344	99,822	434	996,050	7,418,736	74.3
25–34	21,663	0.000534	0.005324	99,388	529	991,235	6,422,686	64.6
35–44	42,058	0.000974	0.009695	98,859	958	983,800	5,431,451	54.9
45–54	107,916	0.002459	0.024295	97,901	2,379	967,115	4,447,651	45.4
55–64	180,891	0.005530	0.053811	95,522	5,140	929,520	3,480,536	36.4
65–74	236,808	0.012237	0.115310	90,382	10,422	851,710	2,551,016	28.2
75–84	348,601	0.026767	0.236077	79,960	18,877	705,215	1,699,306	21.3
85+	338,708	0.061446	1.000000	61,083	61,083	994,091	994,091	16.3

Note: Column (2) in this table was estimated by subtracting in Table 10.1 column (8) from column (4). Life table methodology discussed in Sect. 7.6 was used. For age 0, m_0 was taken as an approximation of q_0. Eq. (7.16) was utilized for estimating L_{85+}

Table 10.7 Impact of the elimination of certain causes of death: United States, 2007

	Increase in life expectancy due to the elimination of			
Age (years)	Diseases of heart $e_x^{\,1} - e_x$	Malignant neoplasms $e_x^{\,2} - e_x$	Unintentional accidents $e_x^{\,3} - e_x$	All other causes of death $e_x^{\,4} - e_x$
0	4.3	3.6	1.1	10.7
1–4	4.2	3.6	1.0	10.1
5–14	4.2	3.6	1.0	10.1
15–24	4.2	3.5	0.9	10.0
25–34	4.3	3.6	0.8	9.9
35–44	4.2	3.5	0.5	9.6
45–54	4.2	3.5	0.4	9.4
55–64	4.0	3.2	0.2	9.0
65–74	3.9	2.7	0.2	8.8
75–84	3.8	1.8	0.2	8.7
85+	3.8	1.1	0.2	8.6

Note: $e_x^{\,1}$ to $e_x^{\,4}$ values were taken from column (9) of Tables 10.3, 10.4, 10.5 and 10.6 and e_x values from column (8) of Table 10.2

10.4 Other Decrements

Conventional life tables, including the *cause of death* ones shown in this chapter, represent the effects of death only. This can be extended to look at the effects of factors other than death, to include socio-economic factors such as labour force participation, marriage, giving birth, dropping out of school, and retirement. Mortality is still included as a factor, but decrements of a cohort of people can be examined in terms of additional factors as well.

It has been common to develop *double decrement tables* when socio-economic factors are involved (Kintner 2004). This is a type of multiple decrement table in which one decrement is death while the other decrement is a change in socio-economic status. An example of such double decrement life tables is the working life table.

A working life table is based on two decrements: (1) entry in the labour force, and (2) death. The basic data required are an appropriate life table and a set of age-specific labour force participation rates. These tables are usually calculated separately for males and females because of the different pattern of their participation in the labour force as shown in Figs. 9.4 and 9.5.

10.4.1 Example of a Working Life Table for Japan

An example of a working life table for Japanese males in 2000 is given in Table 10.8. Column (1) indicates the starting age in each 5-years age group beginning at 15 years of age. Column (2) was taken from the complete life table for Japanese males (Japan undated), and refers to the number of males surviving to exact age x. On the other hand, data given in the remaining eight columns refer to the 5-year age groups except the last row which refers to males aged 85 years and over.

Column (3) was estimated using the labour force and population data from the 2000 Japanese census (Japan 2011). Here the rates are calculated as per person and not in a % format.

Columns (4) and (7) to (10) were calculated using the life table methodology described in Chap. 7. Thus the ${}_5L_x$ values for each age group (where $x = 15 \ldots 80$ and $n = 5$) were calculated as $2.5 * (l_x + l_{x+5})$, as per Eq. (7.20). Similarly using Eq. (7.16), the value of L_{85+} was estimated as $l_{85} * log(l_{85})$.

The ${}_5L_x$ values in column (4) were disaggregated into those for males who were in the labour force shown in column (5), and those for males who were not in the labour force given in column (6). The former were simply the product of columns (3) and (4), while the later were the residual, column (4) minus column (5).

As per Eq. (7.17), the T_x values given in columns (7) and (8) were obtained by cumulating from the oldest to the youngest age the ${}_5L_x$ values in columns (5) and (6) respectively.

Finally, using Eq. (7.18), the e_x values given in columns (9) and (10) were estimated by dividing each of the columns (7) and (8) by column (2) being the number of males alive at a given age x. The value of e' represents the expected number of years (future lifetime) in the labour force for those who were alive at age x, while e''_x shows the expected number of years (future lifetime) not in the labour force for those who were alive at age x. Obviously, the life expectancy at age x, $e_x = e'_x + e''_x$.

Another double decrement table is the net nuptiality table in which a cohort of single persons is subjected to marriage and mortality rates. If column (3) of

Table 10.8 Working life table of males: Japan 2000

Age	Survivors at age x	Labour force participation rate	Person-years lived between age x and $x + 5$			Person-years lived beyond age x		Expected future lifetime at age x	
			Total	In labour force	Not in labour force	In labour force	Not in labour force	In labour force	Not in labour force
x	l_x	${}_5P_x$	${}_5L_x$	${}_5L'_x$	${}_5L''_x$	T'_x	T''_x	e'_x	e''_x
(1)	(2)	(3)	(4)	(5) = (3)*(4)	(6) = (4)−(5)	(7)	(8)	(9) = (7)/(2)	(10) = (8)/(2)
15	99,373	0.173686	496,308	86,202	410,106	4,377,777	1,847,181	44.1	18.6
20	99,150	0.702084	494,943	347,492	147,451	4,291,575	1,437,075	43.3	14.5
25	98,827	0.921327	493,318	454,507	38,811	3,944,083	1,289,624	39.9	13.0
30	98,500	0.944569	491,438	464,197	27,241	3,489,576	1,250,813	35.4	12.7
35	98,075	0.954708	488,950	466,804	22,146	3,025,379	1,223,572	30.8	12.5
40	97,505	0.959101	485,383	465,531	19,852	2,558,575	1,201,426	26.2	12.3
45	96,648	0.955314	479,768	458,329	21,439	2,093,044	1,181,574	21.7	12.2
50	95,259	0.949767	470,620	446,979	23,641	1,634,715	1,160,135	17.2	12.2
55	92,989	0.926460	456,490	422,920	33,570	1,187,736	1,136,494	12.8	12.2
60	89,607	0.716143	435,628	311,972	123,656	764,816	1,102,924	8.5	12.3
65	84,644	0.509287	404,520	206,017	198,503	452,844	979,268	5.3	11.6
70	77,164	0.346348	359,440	124,491	234,949	246,827	780,765	3.2	10.1
75	66,612	0.248597	297,085	73,854	223,231	122,336	545,816	1.8	8.2
80	52,222	0.160127	216,038	34,594	181,444	48,482	322,585	0.9	6.2
85+	34,193	0.089586	155,029	13,888	141,141	13,888	141,141	0.4	4.1

Notes: Column (1) gives the starting age in years for each 5-year age group
Column (2) was copied from the complete life table for Japanese females given in Japan (undated)
Column (3) gives the labour force participation rates for Japanese females calculated from data given in Japan (2011)

Table 10.8 is replaced by age-specific marriage rates, it is possible to calculate the future lifetime of a person remaining as single or as married (Pollard et al. 1995). Other examples of double decrement tables could include a cohort of people who are subject to death and, for example, one of the following decrements: joining the armed forces, enrolling in a higher education institution, having a first child, moving out of the parental household.

10.5 Multistate Models

Multistate models can also incorporate both increments and decrements to the initial cohort. These increments and decrements include labour force entries and exits, school enrolment and withdrawal, marriages and divorces (Kintner 2004). Multistate life tables are very flexible in that they allow not only for movements between life (an active state) and death (an absorbing state) but also for all possible movements among various types of active states. For example, a multistate nuptiality model allows individuals to move from being unmarried to married, from married to divorced, from divorced to remarried, and from remarried to divorced. Because of this flexibility, multistate life tables have a broad range of potential applications (Kintner 2004).

As stated earlier, multistate models do not have a standard table format. While the measures found in multistate life tables usually have counterparts in conventional or multiple decrement life tables, they may be presented in ways not found in these tables. Output measures, for example, could include lifetime transition probabilities, proportion of the population dying while in each state, expected duration in a given state (*e.g.*, never married), and number of transitions to each state per person.

The key measures in multistate life tables are the transition probabilities, which govern the movements among states. These probabilities can be calculated from census population counts and vital statistics. They may also be obtained from surveys such as a cross-sectional survey with a retrospective question on the previous year's status or a longitudinal (panel) survey.

Multistate models can become complicated, as can be illustrated with a series of diagrams that proceed from simple to higher levels of complexity (Figs. 10.3, 10.4, 10.5 and 10.6). As was the case in previous diagrams, the origin of a given arrow represents a decrement from a *state of origin* and its destination represents an increment to a *receiving state*.

The simplest example is in Fig. 10.1 that shows the relationship of life and death, namely that everybody living eventually dies and there is no return back to life. Death represents a *decrement* from alive (*origin state*) and at the same time is an increment to dead (*receiving state*). Dead is both a *receiving* and *absorbing* state. Eventually it receives all who are alive. Very few other states, if any, have these properties. For instance, not everybody alive will eventually marry and marriage is not an absorbing state from which there is no exit.

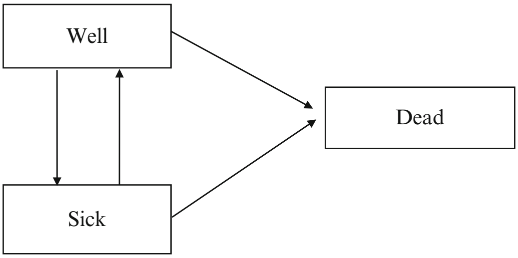


Fig. 10.3 Relationships between wellness, sickness, and death

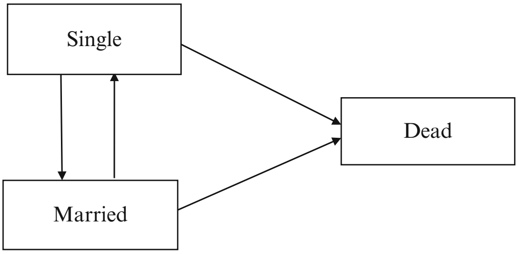


Fig. 10.4 Relationships of being single, married, and dead

The relationships for *wellness*, *sickness* and *death* are shown in Fig. 10.3. Those who are well can remain well and ultimately die or become sick, while those who are sick can remain sick and ultimately die or become well: those who are well can decrement into sickness or death, thus incrementing these two states respectively; and those who are sick can decrement into wellness or death, thus incrementing these two states.

The relationships among the states of being *single*, being *married*, and *dead* are portrayed in Fig. 10.4. Those who are single can either remain single until they eventually die or become married, while those who are married can either remain married and eventually die or become single again.

Movements concerned with marital states and death are expanded further in Fig. 10.5 from *single* and *married* by introducing the restrictive terms of *never married* and *no longer married* to go along with being married. Here, those who never marry can either remain never married and ultimately die or they can marry; those who are married can either remain married until they die or become no longer married; and those who are no longer married can either remain so until they die or they can remarry.

Additional states are added in Fig. 10.6 by replacing the state of *no longer married* with the states of being *divorced* or being *widowed*. As before, those who are never married can either remain so until they die or they can marry; those who are married can either remain so until they die or they can become divorced or widowed; those who are divorced can either remain so until they die or remarry while those who are widowed can either remain so until they die or they can remarry.

Although all states in a multistate model are mutually exclusive and discrete, Figs. 10.3 through 10.6 suggest that how a given state is conceptualized has a major

Fig. 10.5 Relationships of being never married, married, no longer married and dead

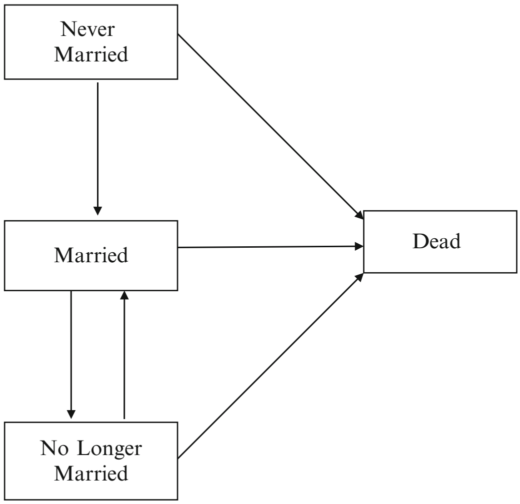
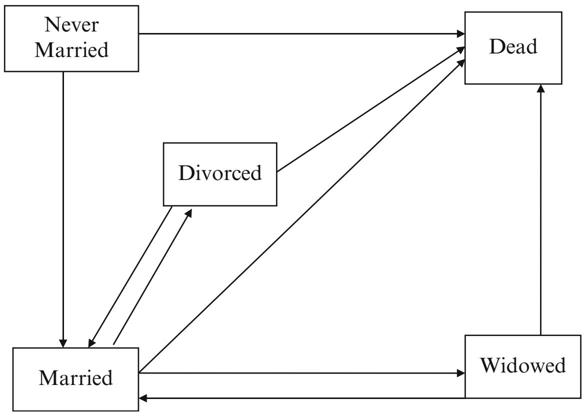


Fig. 10.6 Relationships of being never married, married, divorced, widowed, and dead



bearing on how a multistate model is set up. There are *absorbing states* such as death that only permit entries, while *transient states* such as married allow both entries and exits. Still other states permit only exits including a *state of origin* such as never married. Taken altogether, the states in a given multistate model form what is known as a *state space*. An example of the conceptualizations of what forms a state space could be a view of being not married as being (a) single or never married, or (b) either widowed or divorced. Although they deal basically with the same subject, these two conceptualizations would lead to different models, with the latter being more complicated. Data required for the calculation of multistate life tables are not easy to obtain from the usual sources as they require more probabilities including conditional (transitional) probabilities.

10.5.1 Example of a Multistate Life Table for Japan

An example of a multistate life table is given in Table 10.9, with some caveats. It relates to Japanese females 15 years of age and over who were either in the labour force or not participating in the labour force. Five states were considered – two for those in the labour force and three for those not in the labour force. These were: *employed* or *unemployed* for those in the labour force, and engaged only in *housework* or being a *student* or *other* reasons for not being in the labour force.

Japanese females may go from any of these five states (*origin states*) to death (*absorbing state*) and may go from any of these five states (*origin state*) to another (*receiving state*), or move to and from each of the five states making them *transient states*. Some simplifying assumptions are made in this example similar to those mentioned in Sect. 7.3 regarding the building of life tables. It is assumed that the age-specific rates in each labour force state will remain constant over a lifetime, in addition to the assumption of unchanged age-specific death rates. It is also assumed that there are no moves from one state to another during the 5-year periods.

Column (1) of Table 10.9 indicates the starting age in each 5-years age group beginning at 15 years of age. Column (2) was taken from the complete life table for Japanese females (Japan undated), and refers to the number of females surviving to exact age x. On the other hand, data given in the remaining eight columns refer to the 5-year age groups except the last row which refers to males aged 85 years and over.

Columns (3) to (7) contain the probabilities of being in one of the five labour force states at the given ages. These were estimated separately for each age group and represent the proportion of females in each age group who were in labour force and *employed* or *unemployed*, and those not in the labour force who were engaged in *house-work*, *student* or engaged in *other* activities. Obviously, the sum of these proportions was 1 for each age group.

Column (8) was calculated using the life table methodology described in Chap. 7. Thus, the ${}_5L_x$ values for each age group (where $x = 15 \ldots 80$ and $n = 5$) were calculated as $2.5 * (l_x + l_{x+5})$, as per Eq. (7.20). Similarly using Eq. (7.16), the value of L_{85+} was estimated as $l_{85} * log(l_{85})$.

Columns (9) through (13) were estimated by multiplying the ${}_5L_x$ values in column (8) by the probabilities for each of the five labour force states. Thus, for example, column (9) was equal to column (8) multiplied by column (3), column (10) was equal to column (8) multiplied by column (4) and so on.

As per Eq. (7.17), the T_x values given in columns (14) through (18) were obtained by cumulating from the oldest to the youngest age the ${}_5L_x$ values in columns (9) through (13) respectively.

Finally, using Eq. (7.18), the e_x values given in columns (19) through (23) were estimated by dividing each of the columns (14) through (18) by column (2), being the number of females alive at a given age x. These represent the average number of years in each labour force state a Japanese female is expected to live. Thus, for example, a 15-years old Japanese female is expected to spend 31.0 years in the

Table 10.9 Multistate working life table of females: Japan 2000

Age		Probability of being					Person-years lived between age x and $x+5$					
		In labour force		Not in labour force				In labour force		Not in labour force		
		Emp-loyed	Unemp-loyed	House-work	Student	Others	All females	Emplo-yed	Unemp-loyed	House-work	Student	Others
x	l_x	${}_5P'_x$	${}_5P''_x$	${}_5\tilde{P}_x$	${}_5\hat{P}_x$	${}_5\bar{P}_x$	${}_5L_x$	${}_5L'_x$	${}_5L''_x$	${}_5\tilde{L}_x$	${}_5\hat{L}_x$	${}_5\bar{L}_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
15	99,502	0.1369	0.0176	0.0134	0.8195	0.0126	497,283	68,064	8,776	6,662	407,502	6,279
20	99,411	0.6651	0.0549	0.0880	0.1696	0.0225	496,710	330,350	27,262	43,721	84,222	11,156
25	99,273	0.6612	0.0455	0.2613	0.0103	0.0218	495,958	327,927	22,548	129,579	5,093	10,810
30	99,110	0.5434	0.0324	0.4010	0.0039	0.0193	495,010	268,968	16,051	198,491	1,928	9,572
35	98,894	0.5815	0.0231	0.3765	0.0017	0.0171	493,700	287,092	11,424	185,885	848	8,450
40	98,586	0.6673	0.0187	0.2973	0.0008	0.0159	491,763	328,158	9,188	146,200	385	7,833
45	98,119	0.6888	0.0177	0.2750	0.0005	0.0180	488,818	336,708	8,639	134,427	234	8,809
50	97,408	0.6493	0.0166	0.3111	0.0004	0.0227	484,230	314,393	8,019	150,653	185	10,979
55	96,284	0.5594	0.0150	0.3835	0.0003	0.0418	477,643	267,193	7,147	183,195	142	19,967
60	94,773	0.3750	0.0136	0.5130	0.0004	0.0981	468,370	175,651	6,348	240,261	171	45,939
65	92,575	0.2449	0.0044	0.5812	0.0002	0.1694	454,255	111,232	1,992	264,011	70	76,950
70	89,127	0.1649	0.0020	0.5805	0.0002	0.2525	432,053	71,232	843	250,822	79	109,076
75	83,694	0.1003	0.0013	0.5283	0.0002	0.3699	395,355	39,655	498	208,874	82	146,247
80	74,448	0.0532	0.0011	0.4129	0.0004	0.5324	334,473	17,809	360	138,096	119	178,088
85+	59,341	0.0224	0.0014	0.2159	0.0004	0.7599	283,256	6,336	402	61,159	123	215,235

Age	Person-years lived beyond age x					Expected future lifetime at age x				
	In labour force		Not in labour force			In labour force		Not in labour force		
	Emp-loyed	Unemp-loyed	House-work	Student	Others	Emp-loyed	Unemp-loyed	House-work	Student	Others
x	T'_x	T''_x	$\tilde{T}_x$	$\hat{T}_x$	$\bar{T}_x$	e'_x	e''_x	$\tilde{e}_x$	$\hat{e}_x$	$\bar{e}_x$
(1)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
15	2,950,769	129,496	2,342,036	501,182	865,390	29.7	1.3	23.5	5.0	8.7
20	2,882,705	120,720	2,335,374	93,680	859,112	29.0	1.2	23.5	0.9	8.6
25	2,552,356	93,458	2,291,653	9,459	847,956	25.7	0.9	23.1	0.1	8.5
30	2,224,428	70,909	2,162,074	4,366	837,146	22.4	0.7	21.8	0.0	8.4
35	1,955,460	54,859	1,963,584	2,438	827,573	19.8	0.6	19.9	0.0	8.4
40	1,668,368	43,435	1,777,698	1,590	819,123	16.9	0.4	18.0	0.0	8.3
45	1,340,210	34,247	1,631,499	1,205	811,290	13.7	0.3	16.6	0.0	8.3
50	1,003,502	25,608	1,497,071	971	802,481	10.3	0.3	15.4	0.0	8.2
55	689,109	17,589	1,346,418	786	791,502	7.2	0.2	14.0	0.0	8.2
60	421,916	10,442	1,163,223	644	771,535	4.5	0.1	12.3	0.0	8.1
65	246,264	4,094	922,962	473	725,596	2.7	0.0	10.0	0.0	7.8
70	135,032	2,103	658,951	403	648,647	1.5	0.0	7.4	0.0	7.3
75	63,800	1,260	408,129	323	539,571	0.8	0.0	4.9	0.0	6.4
80	24,145	762	199,256	242	393,324	0.3	0.0	2.7	0.0	5.3
85+	6,336	402	61,159	123	215,235	0.1	0.0	1.0	0.0	3.6

Notes: Column (1) gives the starting age in years for each 5-year age group

Column (2) was copied from the complete life table for Japanese females given in Japan (undated)

Columns (3–(7) are the probabilities of being in one of the five states (employed, unemployed, housework, student or others) for each age group estimated from data given in Japan (2011). Obviously, the sum of these probabilities for each age group is 1 labour force (29.7 years *employed* and 1.3 years *unemployed*) and 37.2 years outside the labour force (23.5 doing *house-work*, 5 years as a *student* and 8.7 years engaged in *other* activities). It is apparent that by 65 years of age they will spend only 2.7 years in the labour force and continue to do house work and other activities for another 17.8 years (10.0 plus 7.8 respectively) Obviously, the life expectancy at age x, $e_x = e'_x + e''_x + \tilde{e}_x + \hat{e}_x + \bar{e}_x$.

References

Daw, R. H. (1979). Smallpox and the double decrement table: A piece of actuarial pre-history. *Journal of the Institute of Actuaries, 106*, 299–318.

Feeney, G. (1974). *Multiple decrement theory*. Honolulu: East–West Center. http://gfeeney.com/notes/multdecr/multdecr.pdf. Accessed Feb 2013.

Japan. (2011). *e-Stat. Portal site of official statistics of Japan. 2000 Population census, labour force status of population (table 1)*. Tokyo: Statistics Bureau. www.e-stat.go.jp/SG1/estat/ListE.do?bid=000000030097&cycode=0. Accessed Feb 2013.

Japan. (undated). *Abridged life tables for Japan 2000*. Tokyo: Ministry of Health, Labour and Welfare. http://www.mhlw.go.jp/english/database/db-hw/lifetb00/index.html. Accessed Feb 2013.

Kintner, H. (2004). The life table. In J. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 301–340). San Diego: Elsevier Academic Press.

Mathers, C. D., Sadana, R., Salomon, J. A., Murray, C. J. L. & Lopez, A. D. (2000). *Estimates of DALE for 191 countries*, *methods and results* (Global Programme on Evidence for Health Policy Working Paper No. 16). Geneva: World Health Organization. www.who.int/health-systems-performance/docs/articles/paper16a.pdf. Accessed Feb 2013.

Pollard, A. H., Yusuf, F., & Pollard, G. N. (1995). *Demographic techniques*. Sydney: A.S. Wilson.

Xu, J., Kochanek, K. D., Murphy, S. L. & Tejada-Vera, B. (2010). Deaths: Final data for 2007. *National Vital Statistics Reports*, *58*(19), 1–136. http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58_19.pdf. Accessed Feb 2013.

Chapter 11
Population Projections

11.1 Purpose

This chapter provides an introduction to the projection of populations. It offers some basic concepts and definitions. It deals with approaches to the projection of populations as aggregates. The chapter covers in detail the cohort-component method that is the most used technique of projecting populations by age and sex. It also discusses and illustrates the cohort-change method (often referred to as the Hamilton-Perry method) that requires less data. Further, it looks at projections of particular segments of the population such as people in the labour force and of school age.

11.2 Concepts: Estimates, Projections and Forecasts

In demographic analysis, it is customary to differentiate the terms *estimates*, *projections* and *forecasts*. Usually, a population *estimate* refers to a *present* and *past* population for which no census counts are available. The terms *projection* and *forecas*t are sometimes used without distinction about future populations. In this chapter the term *projection* is the numerical outcome of a particular set of assumptions regarding future population. A population projection is a conditional calculation showing what a future population would be like if a particular set of assumptions were to hold true. A *forecast* is a projection that has been selected as being the most likely to provide an accurate prediction of a future population (George et al. 2004). Yet, an option often used is to build projections based on different set of assumptions that represent alternative scenarios. The methods described in this chapter are concerned with population projections.

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_11,

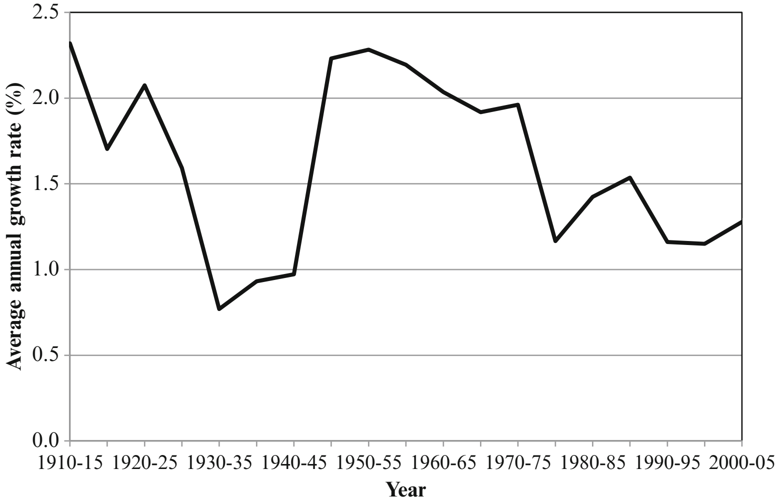


Fig. 11.1 Estimated annual population growth rates (% per annum): Australia, 1910–2005 (*Source*: Australia 2008)

11.3 Projections of the Total Population

11.3.1 Constant Growth Rate

A simple method of population projection was described in Sect. 4.3.3 where an average annual population growth rate was used to extrapolate the population in future or in the past. The main assumption in that method was that a certain growth rate would remain constant, and a population subjected to such a rate would continue to grow (or decline if the rate was negative) steadily.

Over time, human populations experience changing rates of population growth. For example, the population of Australia experienced an average annual growth rate of 1.6 % during most of the past century, but it varied markedly within this period ranging from below 1 % during the Economic Depression of 1930s, to the highest (about 2.3 %) immediately after the Second World War (Fig. 11.1).

11.3.2 Mathematical Models

Certain mathematical models can be fitted to the population data in order to project the total population (but not disaggregated by age and sex) in future, or in the past, or even interpolate the population at any time within a given period. The linear model: $Y = a + b * X$ was discussed in Sect. 3.9.2. Another model is the exponential model:

$$Y = a * b^X \tag{11.1}$$

Following the logarithm rules in Box 3.1, Eq. 11.1 can be converted into a linear model by taking its logarithm. Thus: $\log(Y) = \log(a) + X * \log(b)$. This is a linear model between X and $\log(Y)$ values, and can be fitted using the method outlined in Sect. 3.9.2.

Both the linear and exponential models assume that the population would go on increasing (or decreasing) over time depending upon the value of b, which is an indicator of the rate of population growth. Since this is unlikely to happen in human populations, mathematical models that involve an upper limit (asymptote) of population size are considered as more appropriate.

There are two commonly used asymptotic models. One of these, the Gompertz model, was developed in an attempt to discover a law of mortality (Gompertz 1825). Another commonly used asymptotic model is the Logistic model. (Verhulst 1845; Verhulst 1847; Pearl & Reed 1920). Both of these models are quite commonly used in demography as well as other sciences.

Equations for these two models are:

$$Y = k * a^{b^X} \quad \text{(Gompertz)} \tag{11.2}$$

$$Y = \frac{1}{k + a * b^X} \quad \text{(Logistic)} \tag{11.3}$$

Taking the logarithms, Eq. (11.2) becomes:

$$\log(Y) = \log(k) + \log(a) * b^X \tag{11.4}$$

while Eq. (11.3) can be re-written as:

$$\frac{1}{Y} = k + a * b^X \tag{11.5}$$

In the above equations Y is the population, k, a and b are the parameters of the model, and X is a measure of time. In the Gompertz model k is the asymptote while in the Logistic model $\frac{1}{k}$ is the asymptote. The asymptote is a mathematical constant for a given set of population data provided the past trends continue to prevail in future. Under these conditions it indicates the maximum possible size of a population in the future if it continues to follow the model.

11.3.3 Fitting the Gompertz and Logistic Models

There are many procedures to fit these models. One such procedure is the method of partial totals. This involves the following five steps (Croxton et al. 1968):

1. Time variable (X) is converted into (x) by taking deviations from its first value and dividing by the common factor: the distance between each value of X, which is supposed to be constant. This procedure will result in the first value of (x) as 0, second as 1 and so on.
2. Population variable (Y) is transformed into log(Y) for the Gompertz model and into $\frac{1}{Y}$ for the Logistic model.
3. The transformed values are then divided into three equal parts and the sum of each part is S_1, S_2 and S_3. Each of these sums contains n observations. If the number of observations is not a multiple of 3 then one or two observations have to be dropped. If the residual is 1 then either an observation at the beginning or at the end may be dropped. When the residual is 2 then one observation at the beginning and one at the end, or both in the beginning or both at the end, may be dropped.
4. Parameters in Eqs. (11.4) and (11.5) are calculated as follows:

Gompertz	Logistic	
$b^n = \dfrac{S_3 - S_2}{S_2 - S_1}$	$b^n = \dfrac{S_3 - S_2}{S_2 - S_1}$	(11.6)
$\log(a) = \dfrac{(S_2 - S_1) * (b-1)}{(b^n - 1)^2}$	$a = \dfrac{(S_2 - S_1) * (b-1)}{(b^n - 1)^2}$	(11.7)
$\log(k) = \dfrac{1}{n} * [S_1 - \dfrac{b^n - 1}{b - 1} * \log(a)]$	$k = \dfrac{1}{n} * [S_1 - \dfrac{b^n - 1}{b - 1} * a]$	(11.8)

5. Substituting in Eqs. (11.4) and (11.5) the values of the three parameters and the scaled variable (x) as 0, 1, 2, and so on, the expected values, $\log(\hat{Y})$ and $\frac{1}{\hat{Y}}$ are estimated. These are then converted into the corresponding expected (trend) values denoted by ($\hat{Y}$).

11.3.4 An Example of Fitting the Gompertz and Logistic Models

Table 11.1 illustrates the fitting of the Gompertz and Logistic models using population data for Australia during the period 1950–2000 at 10-year intervals (Australia 2008).

Column (1) refers to the year or any other measure of time (X), and column (2) contains the midyear population (Y). Column (3) gives the scaled value of X as described in the *first step*. In this case the difference from 1950, the starting point, is divided by 10 because the observations are 10 years apart.

Table 11.1 Fitting the Gompertz and Logistic models to the population data for Australia, 1950–2000

Year (X)	Population in millions (Y)	$(x) = \frac{(X)-1950}{10}$	$\log(Y)$	$\log(\hat{Y})$	Gompertz ($\hat{Y}$)	$(\frac{1}{Y})$	$(\frac{1}{\hat{Y}})$	Logistic ($\hat{Y}$)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1950	8.31	0	0.919601	0.919217	8.30	0.120337	0.120128	8.32
1960	10.39	1	1.016616	1.017000	10.40	0.096246	0.096456	10.37
1970	12.66	2	1.102434	1.100685	12.61	0.078989	0.079400	12.59
1980	14.81	3	1.170555	1.172304	14.87	0.067522	0.067111	14.90
1990	17.17	4	1.234770	1.233598	17.12	0.058241	0.058257	17.17
2000	19.27	5	1.284882	1.286054	19.32	0.051894	0.051878	19.28
					Gompertz model		Logistic model	
		Following step (3)			$S_1=$	1.936217	$S_1=$	0.216583
		Following step (3)			$S_2=$	2.272989	$S_2=$	0.146511
		Following step (3)			$S_3=$	2.519652	$S_3=$	0.110135
		Using Eq. (11.6)			$b^2=$	0.732433	$b^2=$	0.519115
		Square root of b^2			$b=$	0.855823	$b=$	0.720496
		Using Eq. (11.7)			$\log(a)=$	−0.678214	$a=$	0.084694
		Using Eq. (11.8)			$\log(k)=$	1.597431	$k=$	0.035434
					Asymptote=	39.58 million	Asymptote=	28.22 million

Source: Australia (2008)

According to the *second step*, the log(Y) and $(\frac{1}{Y})$ values were calculated as shown in columns (4) and (7) respectively.

Following the *third step*, the values of S_1, S_2, S_3 were computed. Since the total number of observations was 6, each one-third consisted of 2 observations.

In the *fourth step*, using the appropriate version (*i.e*., Gompertz or Logistic) of Eqs. (11.6), (11.7) and (11.8) the three parameters of the two models were calculated as shown in the bottom half of Table 11.1.

The two models were:

$$\text{Gompertz}: \ \log(\hat{Y}) = 1.597431 + 0.855823^x * (-0.678214) \tag{11.9}$$

$$\text{Logistic}: \ \frac{1}{\hat{Y}} = 0.035434 + 0.084694 * 0.720496^x \tag{11.10}$$

Following the *fifth step*, each value of (x) given in column (3) was substituted in Eqs. (11.9) and (11.10) to obtain the corresponding $\log(\hat{Y})$ or $\frac{1}{\hat{Y}}$ values respectively and hence the expected (trend) values $(\hat{Y})$ under each model.

The asymptote for the Gompertz model, k, was 39.58 million while the asymptote for the Logistic model, $\frac{1}{k}$, was 28.22 million. These are mathematically determined parameters and it is difficult to say if either of these figures is realistic. Attempts to mathematically predict the long term limits to the growth in human populations have not been very successful so far.

Projecting the population at a point in time into the future (or into the past) would require the estimation of (x) and substituting it in Eq. (11.9) for the Gompertz and Eq. (11.10) for the Logistic model. For illustrative purposes, the population of Australia was projected forward to mid-2011 and backward to the beginning to 1945. The relevant values of (x) were:

- for forward projection to mid-2011, $(x) = \frac{2011 - 1950}{10} = \frac{61}{10} = 6.1$
- for backward projection to the end of 1944 or the beginning of 1945, $(x) = \frac{1944.5 - 1950}{10} = -0.55$ because the period from the end of 1944 or the beginning of 1945 was -5.5 years from the middle of 1950.

Substituting the value of (x) as 6.1 in Eqs. (11.9) and (11.10) gave the projected population of Australia in mid-2011 as 21.63 and 21.32 million respectively. These projections were quite close to the 2011 census population of 21.51 million (Australia 2012).

11.4 Projections by Age: Cohort-Component Method

11.4.1 Data Requirements and Steps

The cohort-component method is based on the fundamental demographic equation examined in Sect. 4.2. In the case of a projection of population into the future, by age and sex, the data requirements are the population at the *launch year* by sex and age, age-specific birth rates, sex ratio at birth, sex and age-specific survival ratios (based on life tables) and sex and age-specific migration rates. The launch population is usually from the census count or an estimate. Age cohorts are usually defined in 5-year age groups but single or 10-year age groups may also be used. A feature of the process is that birth, survival and migration rates must be consistent with the age-groups selected for the projection and also with the projection time cycle. Accordingly, if the base population is in 5-year age groups then the projection is usually also for a 5-year period, and the survival and migration rates must also be in 5-year age groups. However, the projection may be extended to more than one projection cycle. This can be done by taking the projection for the first cycle as the base for the next cycle but either keeping constant or varying the fertility, mortality and migration rates for some or all subsequent projection cycles.

The *first step* in the process is to establish the *launch year* (time = t) and the population at that time by sex in, say, 5-year age groups and assemble the data on fertility, mortality and migration. The *second step* is the estimation of the number of persons in each sex-age group that survive to the next cycle (time = $t + 5$). This is done by estimating age-sex specific survival ratios for each group in the launch population at time t. Estimates from relevant life tables of males and females usually form the basis for these calculations. This process results in population aged 5 years and over at time $t + 5$. The *third step* involves the adjustment of the number of survivors aged 5 and over for the effect of net migration. In the *fourth step*, the 0–4 population at time $t + 5$ is calculated and adjusted for net migration. The age-specific birth rates are used to estimate the number of births during the period t to $t + 5$. These are then disaggregated by sex and adjusted for child mortality and net migration to estimate 0–4 population at time $t + 5$. As mentioned earlier, the projected population at the end of the first projection cycle will, in turn, become the starting point for the next cycle. This process is repeated until the target date is reached.

11.4.2 Assumptions

It is apparent that a number of assumptions are implicit in this process and that a degree of judgment is involved in the set of assumptions made in the selection of birth, survival and migration rates. Unless there is contrary evidence, one of the simplest assumptions is that the sex and age-specific birth, survival and migration rates prevalent at time t would remain constant during the period of at least the first projection cycle of 5 or so years. For the subsequent cycles, different set of fertility, mortality and migration assumptions may be made so as to prepare more than one projection scenarios. Generally, at least three scenarios – high, medium and low growth – are used in preparing population projections in most countries.

11.4.3 Illustrative Projections of the Population of Estonia

All of the input data required for this section were taken from the website of Statistics Estonia (see Box 8.1). This included the age-sex distribution from the 2000 population census, and the data on births by age of mother, deaths by age and sex and net migration by age and sex, all for the calendar year 2000 (See Table A11.1 for the input data and Table A11.2 for some preliminary calculations such as the adjustment for unknown ages, age-sex specific rates and selected abridged life table functions).

The preliminary analysis shown in Tables A11.1 and A11.2 indicates that in 2000 the total fertility rate (*TFR*) in Estonia was 1.38, the country lost 1,749 persons through migration and the life expectancy at birth (e_0) was 65.1 years for males and 75.9 for females. Thus, starting with the 2000 census (though conducted in April 2000 it was assumed to represent the midyear population), Estonia's population would be projected in two cycles, one ending in mid-2005 and the other in mid-2010 under the assumption that the *TFR*, net migration rates and the mortality rates underlying the e_0 values observed in 2000 would prevail through both the projection cycles. Table 11.2 shows the calculation and results of projections during the two projections cycles and Table 11.3 presents the estimation of the 0–4 population in mid-2005 and mid-2010.

In Table 11.2 columns (2) and (3) containing the age distributions by sex from the 2000 Estonian census were copied from columns (4) and (5) of Table A11.1. Columns (4) and (5) of Table 11.2 show the life table populations (${}_5L_x$) by sex and age. These were taken from columns (9) and (12) of Appendix 11.1, Table A11.2. Columns (6) and (7) of Table 11.2 present the age-sex specific net migration rates and column (8) contains the age specific fertility rates. These were copied from columns (5), (6) and (4) respectively of Table A11.2. This concludes the first step.

The *second step* involved the estimation of survival ratios for each age cohort over a 5-year period. In a life table population (see Sect. 7.8.1), the survival ratios of people, say, aged 5–9 in 2000 to age 10–14 in 2005 is $\frac{{}_5L_{10}}{{}_5L_5}$, or more generally $\frac{{}_5L_{x+5}}{{}_5L_x}$, where *x* takes values 0, 5, 10 and so on. This is the formula used to estimate the 2nd to the 17th values in columns (10) and (11). However, the first values in these columns were estimated using the formula $\frac{{}_5L_0}{500,000}$ because in a life table population 100,000 births occur every year of which ${}_5L_0$ reach the age 0–4. The last values in columns (10) and (11) were calculated as $\frac{L_{85+}}{{}_5L_{80}+L_{85+}}$, because the life table population aged 80+ (denominator) would survive to age 85+ (numerator) 5 years' later.

Table 11.2 Projecting the population of Estonia, 2000–2010

	Population, 2000		$_5L_x$		Migration rate				Survival ratio			Projection cycle I: 2000–2005				Projection cycle II: 2005–2010			
												Survivors in 2005		Migration adjusted		Survivors in 2010		Migration adjusted	
Age	Males	Females	Males	Females	Males	Females	Fertility rate	Survival from	Males	Females	Age group	Males	Females	Males	Females	Males	Females	Males	Females
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(1)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
0–4	31,574	29,801	496,698	497,620	−0.000728	−0.000906	…	Birth to 0–4	0.993396[a]	0.995240[a]	0–4	33,482	31,947	33,360	31,802	33,698	32,152	33,575	32,006
5–9	40,872	38,804	492,883	494,858	−0.001125	−0.000954	…	0–4 to 5–9	0.992319	0.994450	5–9	31,331	29,636	31,155	29,495	33,104	31,625	32,918	31,474
10–14	55,079	52,421	491,945	494,263	−0.001416	−0.001202	…	5–9 to 10–14	0.998097	0.998798	10–14	40,794	38,757	40,505	38,524	31,096	29,460	30,876	29,283
15–19	52,893	50,913	490,385	493,543	−0.002042	−0.001395	0.025671	10–14 to 15–19	0.996829	0.998543	15–19	54,904	52,345	54,343	51,980	40,377	38,468	39,965	38,200
20–24	48,265	46,414	486,328	492,425	−0.001471	−0.002241	0.087021	15–19 to 20–24	0.991727	0.997735	20–24	52,455	50,798	52,069	50,229	53,893	51,862	53,497	51,281
25–29	47,223	47,241	480,328	491,010	−0.001546	−0.001905	0.084884	20–24 to 25–29	0.987663	0.997126	25–29	47,670	46,281	47,302	45,840	51,427	50,085	51,029	49,608
30–34	43,846	45,519	473,203	488,890	−0.001140	−0.001714	0.054263	25–29 to 30–34	0.985166	0.995682	30–34	46,523	47,037	46,258	46,634	46,600	45,642	46,334	45,251
35–39	46,824	49,831	463,513	485,578	−0.001516	−0.001465	0.019707	30–34 to 35–39	0.979523	0.993225	35–39	42,948	45,211	42,622	44,880	45,311	46,318	44,968	45,979
40–44	47,437	52,029	447,423	480,693	−0.001560	−0.001807	0.004805	35–39 to 40–44	0.965287	0.989940	40–44	45,199	49,330	44,846	48,884	41,142	44,428	40,821	44,027
45–49	44,390	50,710	423,083	472,680	−0.001284	−0.001775	0.000177	40–44 to 45–49	0.945600	0.983330	45–49	44,856	51,162	44,568	50,708	42,406	48,069	42,134	47,642
50–54	39,005	46,782	392,365	461,025	−0.001282	−0.001026	…	45–49 to 50–54	0.927395	0.975343	50–54	41,167	49,460	40,903	49,206	41,332	49,458	41,067	49,204
55–59	32,745	41,442	355,228	444,545	−0.000458	−0.000748	…	50–54 to 55–59	0.905351	0.964254	55–59	35,313	45,110	35,232	44,941	37,032	47,447	36,947	47,270
60–64	34,638	48,134	309,020	423,403	−0.000635	−0.000706	…	55–59 to 60–64	0.869920	0.952441	60–64	28,486	39,471	28,396	39,332	30,649	42,804	30,552	42,653
65–69	27,270	42,193	254,560	394,603	−0.000660	−0.000901	…	60–64 to 65–69	0.823765	0.931980	65–69	28,534	44,860	28,440	44,658	23,392	36,657	23,315	36,492
70–74	21,027	39,856	194,300	348,835	−0.000999	−0.000828	…	65–69 to 70–74	0.763278	0.884015	70–74	20,815	37,299	20,711	37,145	21,708	39,478	21,600	39,315
75–79	10,548	28,605	134,095	282,525	−0.001043	−0.000909	…	70–74 to 75–79	0.690144	0.809910	75–79	14,512	32,280	14,436	32,133	14,294	30,084	14,219	29,947
80–84	4,552	13,640	79,448	199,508	−0.001098	−0.000587	…	75–79 to 80–84	0.592475	0.706161	80–84	6,249	20,200	6,215	20,141	8,553	22,691	8,506	22,624
85+	3,553	13,863	42,945	138,878	−0.001092	−0.000505	…	80+ to 85+	0.350878[a]	0.410413[a]	85+	2,882	11,288	2,866	11,259	3,186	12,887	3,169	12,854
Total	631,851	738,200	…	…	…	…	…	…	…	…	Total	618,120	722,472	614,227	717,791	599,200	699,615	595,491	695,110
											5+	584,638	690,525	…	…	565,502	667,463	…	…

Notes: Small discrepancies between the totals for actual and adjusted columns are due to rounding

[a]The formulae used to the first and last values were $\frac{_5L_0}{500,000}$ and $\frac{L_{85+}}{_5L_{80}+L_{85+}}$ respectively

[b]Copied from Table 11.3

Table 11.3 Estimating the 0–4 population of Estonia, 2005 and 2010

Age	Fertility rate	Number of females 2000	2005	2010	Expected births 2000	2005	2010
(1)	(2)	(3)	(4)	(5)	(6) = (2)*(3)	(7) = (2)*(4)	(8) = (2)*(5)
15–19	0.025671	50,913	51,980	38,200	1,307	1,334	981
20–24	0.087021	46,414	50,229	51,281	4,039	4,371	4,463
25–29	0.084884	47,241	45,840	49,608	4,010	3,891	4,211
30–34	0.054263	45,519	46,634	45,251	2,470	2,531	2,455
35–39	0.019707	49,831	44,880	45,979	982	884	906
40–44	0.004805	52,029	48,884	44,027	250	235	212
45–49	0.000177	50,710	50,708	47,642	9	9	8
Expected births per year					13,067	13,255	13,236

	Projection cycle I: 2000–2005	Projection cycle II: 2005–2010
Average number of expected births in 5 years	65,805	66,228
Sex ratio at birth (males per 100 females)	105	105
Proportion male	0.512195	0.512195
Estimated male births in 5 years	33,705	33,922
Estimated female births in 5 years	32,100	32,306
Survival rate from birth to age 0–4:		
Males (from column (10) of Table 11.2)	0.993396	0.993396
Females (from column (11) of Table 11.2)	0.995240	0.995240
Estimated 0–4 survivors at the end of cycle		……
Males (copied to Table 11.2)	33,482	33,698
Females (copied to Table 11.2)	31,947	32,152
Estimated deaths aged 0–4: (births – survivors)		
Males	223	224
Females	153	154

Having calculated the survival ratios, male survivors aged 5 years and over in 2005 shown in column (12) of Table 11.2 were estimated by multiplying the 1st value in column (2) by the 2nd value in column (10) to estimate population aged 5–9 in 2005, and repeating this process up to and including the age group 75–79 in column (2). The last value in column (12) was obtained by multiplying males aged 80+ (*i.e.*, aged 80–84 plus 85+) in column (2) by the last value in column (10). Female survivors shown in column (13) were similarly calculated using multiplying column (3) with appropriate values in column (11).

In the *third step* adjustments were made for the impact of net migration. This was achieved by multiplying the number of survivors in each age-sex category estimated in columns (12) and (13) by 5 times the corresponding net migration rates shown in columns (6) and (7). Adding this product to the number of survivors gave the migration adjusted projection. For example, the number of survivors among males aged 10–14 in 2005 was 40,794 and the net-migration rate was

−0.001416. The number of migration adjusted males aged 10–14 in 2005 were: [40,794 + (−0.001416*5*40,794)] = 40,505. This procedure was repeated for all age groups and the result is given in columns (14) and (15) for cycle I.

Table 11.3 shows the *final step* of estimating the 0–4 population in 2005. This requires first the estimation of births during the projection cycle I (2000–2005). The fertility rates in column (2) of Table 11.3 were copied from column (8) of Table 11.2, and the number of women in 2000 and the projected number in 2005 shown in columns (3) and (4) of Table 11.3 were copied from column (3) and (15) of Table 11.2. Expected numbers of births in 2000 and 2005 were estimated by multiplying the fertility rate with the numbers of women in 2000 and 2005, respectively. Births during 2000 and 2005 were averaged and multiplied by 5 to obtain the number of births during the projection cycle. They were disaggregated by sex using the sex ratio at birth of 105, and survived to age 0–4 by multiplying with the survival ratio from birth to age 0–4 shown as the first value in columns (10) and (11) of Table 11.2. These were adjusted for net migration using the procedure shown in the previous paragraph.

For the next cycle, 2005–2010, columns (14) and (15) served as the starting points and the *second* to the *fourth* steps discussed above were repeated. Again the underlying assumptions were that the fertility, net migration and survivorship rates used in projection cycle I would continue to prevail during projection cycle II.

Table 11.4 presents some of the summary results of the population projection of Estonia to 2005 and 2010. Census populations by sex were copied from the last row in columns (2) and (3) of Table 11.2. Estimated male and female births during each projection cycle were taken from Table 11.3. Estimated deaths by sex consisted of two components: (a) the difference between those alive at the beginning of the projection cycle and those aged 5+ at the end of the cycle, and (b) deaths among babies born during the projection cycle. Thus, for example, component (a) for male deaths during projection cycle I was 631,851–584,638 = 47,213, and component (b) was 223 deaths among male children born during 2000–2005. The sum, 47,213 + 223 = 47,436, is the number of male deaths during projection cycle I. The estimated net migration was equal to the difference between migration adjusted total and the total number of survivors at the end of a projection cycle. Thus, for instance, estimated net migration of males during the projection cycle I equals 614,227 − 618,120 = −3,893.

The crude annual rates were estimated by dividing the average annual number of vital events by the average population for the projection cycle. The population growth rates were calculated using Eq. (4.9).

According to Table PO0211 from the Statistics Estonia database (see Box 8.1) the midyear (average) population of the country was 1,346,097 in 2005 and 1,340,160 in 2010. This means that the illustrative projection for 2005 was an underestimate by 14,079 persons and the 2010 projection was off by 49,559 persons. These deviations have occurred because of the fact that the demographic rates in Estonia did not remain constant at 2000 levels.

Table 11.4 Summary results of the population projection of Estonia

Population/vital events	Males	Females	Persons
Census population in 2000	631,851	738,198	1,370,049
Estimated births 2000–2005	33,705	32,100	65,805
Estimated deaths 2000–2005	47,436	47,826	95,262
Estimated net migration 2000–2005	−3,893	−4,681	−8,574
Projected population in 2005	614,227	717,791	1,332,018
Estimated births 2005–2010	33,922	32,306	66,228
Estimated deaths 2005–2010	48,949	50,351	99,300
Estimated net migration 2005–2010	−3,709	−4,505	−8,214
Projected population in 2010	595,491	695,110	1,290,601
Indicator (average per year)	2000–2005		2005–2010
Crude death rate (per 1,000 population)	9.74		10.10
Crude death rate (per 1,000 population)	14.10		15.15
Crude net migration rate (per 1,000 population)	−1.27		−1.25
Population growth rate (%).	−0.56		−0.63

Source: Tables 11.2 and 11.3

11.4.4 Some General Considerations

As mentioned earlier, projections are generally made under different scenarios using varying fertility, mortality and migration rates. Thus, the *TFR*, life expectancies at birth and net migration rates (or numbers) may be assumed to increase or decrease from those in the launch year. While sophisticated models could be used to forecast the fertility and mortality rates in future the migration levels, both internal and international, are much more difficult to forecast. It may also be noted that projections are periodically revised as new data become available through census and/or other sources.

11.4.5 Availability of Population Projections

The statistical bureaus of most countries prepare population projections at national and, in some cases, even at sub-national levels also. For example, projections for the population of Estonia and its capital city, Tallinn, are available under various scenarios (Table PO09 and PO091 in *Population indicators and composition* section) on their website (see Box 8.1). The United Nations produces population projections for most countries of the world based on various scenarios (United Nations 2012a). These are revised from time to time. More recently, a set of projections has been released by them based on probabilistic modelling of total fertility and life expectancy at birth for countries not affected with high prevalence of HIV/AIDS (United Nations 2012b).

11.5 Projections by Age: Cohort-Change Method

11.5.1 Data Requirements and Methodology

An alternative and somewhat simpler method for population projections is described in this section. It has fewer requirements of input data compared to the cohort-component method discussed earlier. This method is often ascribed to Hamilton and Perry (1962), though more recently many improvements have been incorporated by others (for example: Swanson et al. 2010; Swanson & Tedrow 2012).

The basic input data required consist of the age distribution of a population at two points in time, say, t and $t + k$. These are usually taken from the two most recent censuses that are k years apart. As noted in Chap. 2, censuses in most countries are 10 or 5 years apart. Censuses at other intervals can be used but they would require some adjustments to the formulae presented in this section.

The method involves two steps:

- calculation by age (and sex if required) of *cohort-change ratios* from data in the two most recent censuses conducted at time t and $t + k$, and
- applying these ratios to the population at time $t + k$ to project the population to time $t + 2k$, $t + 3k$ and so on.

The general formula for the estimation of the cohort-change ratio (${}_nR_{x\to x+k}^{t\to t+k}$) for a cohort aged x to $x + n$, at time t that moves to age $x + k$ to $x + k + n$ at time $t + k$ is:

$$ {}_nR_{x\to x+k}^{t\to t+k} = \frac{{}_nP_{x+k}^{t+k}}{{}_nP_x^t} \tag{11.11} $$

${}_nP_x^t$ is the population aged x to $x + n$, at time t, ${}_nP_{x+k}^{t+k}$ is the population aged $x + k$ to $x + k + n$, at time $t + k$, k is the interval between the two censuses, and n is the width of the age interval. It is essential that k is a multiple of n.

To estimate the projected population (${}_nP_{x+k}^{t+2k}$) aged $x + k$ to $x + k + n$ in year $t + 2k$, the following equation is used:

$$ {}_nP_{x+k}^{t+2k} = {}_nR_{x\to x+k}^{t\to t+k} * {}_nP_x^{t+k} \tag{11.12a} $$

For the open-ended age group such as ≥ 85 (also referred to as 85+), Eq. (11.11) will take the form:

$$ R_{\geq x-k\to\geq x}^{t\to t+k} = \frac{P_{\geq x}^{t+k}}{P_{\geq x-k}^t} \tag{11.13} $$

$R^{t\to t+k}_{\geq x-k\to\geq x}$ is the cohort-change ratio for a cohort aged $\geq x-k$ at time t to age $\geq x$ at time $t+k$, $P^t_{\geq x-k}$ is the population aged $\geq x-k$ at time t and $P^{t+k}_{\geq x}$ is that aged $\geq x$ at time $t+k$. Now Eq. (11.12a) will take the form:

$$P^{t+2k}_{\geq x} = R^{t\to t+k}_{\geq x-k\to\geq x} * P^{t+k}_{\geq x-k} \tag{11.12b}$$

$P^{t+2k}_{\geq x}$ is the population aged $\geq x$ at time $t+2k$ and $P^{t+k}_{\geq x-k}$ is the population aged $\geq x-k$ at time $t+k$.

Depending upon the values of n and k, Eq. (11.12a) would only provide the projected population, at time $t+2k$, for ages $x+k$ and older. Thus, for example, if $k=5$, Eq. (11.12a) would only give the projected population for age 5 and beyond. However, if $k=10$, the projected population can only be calculated for ages 10 and older. In such circumstances, a different method may be used. Hamilton and Perry suggested the use of birth data but that may or may not be easily available. Therefore, an alternative method could be used. It would involve the calculation of *child-adult ratios* defined similarly to the child-woman ratios, described in Sect. 4.6.4, except in this case the denominator would consist of the age groups of adults that are closely associated with the age of children (Swanson et al. 2010; Swanson & Tedrow 2012).

The average child-adult ratio for children x to $x+n$, over the period t to $t+k$ $\left({}_nC_x^{t\to t+k}\right)$ is:

$${}_nC_x^{t\to t+k} = \frac{{}_nP^t_x + {}_nP^{t+k}_x}{{}_{30}P^t_{15} + {}_{30}P^{t+k}_{15}} \tag{11.14}$$

the numerator is the sum of the number of children in the age group x to $x+n$, and the denominator is the sum of the population aged 15–44, both at times t and $t+k$ respectively. The projected population for the x to $x+n$ age group at time $t+2k$, $\left({}_nP_x^{t+2k}\right)$, would be:

$${}_nP_x^{t+2k} = {}_nC_x^{t\to t+k} *_{30} P^{t+2k}_{15} \tag{11.15}$$

${}_{30}P^{t+2k}_{15}$ is the projected population aged 15–44 at time $t+2k$.

As noted earlier, if the two censuses are 10 years apart and the age distributions are in 5-year age groups, then Eqs. (11.14) and (11.15) would be used twice: once by substituting in these equations $x=0$ and $n=5$ (for the 0-4 age group), and once by substituting $x=5$ and $n=5$ (for the 5-9 age group). However, if the two censuses are 5 years apart then only the estimate for 0-4 age group would be required.

This method assumes that the demographic changes between the two censuses would continue to prevail in future. While this may not be a problem for short-term projections it may result in unlikely scenarios for long-term projections. The method can also be used to develop projections for other characteristics such as ethnicity, religion or social class as well as developing projections at small-area levels. In using this method for this purpose it is useful, though not necessary, to ensure that the geographic boundaries of the area(s) concerned have not changed over time.

Table 11.5 Projection of the population of Ryde Municipality, 2006–2011, using the cohort-change method

Age group	Population		Cohort change ratio (1996–2001)	Projected pop. 2006	Cohort change ratio (1996–2001)	Projected pop. 2011
	1996	2001				
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0–4	5,814	5,604	0.130822	5,779	0.129181	5,715
5–9	5,522	5,436	0.934985	5,240	0.935046	6,263
10–14	5,093	5,138	0.930460	5,058	0.930464	4,876
15–19	5,331	5,737	1.126448	5,788	1.126508	5,698
20–24	7,073	7,180	1.346839	7,727	1.346871	7,796
25–29	7,701	7,534	1.065177	7,648	1.065181	8,231
30–34	8,524	8,018	1.041163	7,844	1.041147	7,963
35–39	7,927	8,116	0.952135	7,634	0.952108	7,468
40–44	6,780	7,358	0.928220	7,533	0.928167	7,086
45–49	6,243	6,472	0.954572	7,024	0.954607	7,191
50–54	4,700	5,943	0.951946	6,161	0.951947	6,686
55–59	4,015	4,428	0.942128	5,599	0.942117	5,804
60–64	3,595	3,703	0.922291	4,084	0.922313	5,164
65–69	3,992	3,342	0.929624	3,442	0.929517	3,796
70–74	3,607	3,466	0.868236	2,902	0.868342	2,989
75–79	2,652	3,001	0.831993	2,884	0.832083	2,415
80–84	1,862	2,011	0.758296	2,276	0.758414	2,187
85+	1,353	1,757	0.546501	2,059	0.546501	2,369
Total	91,784	94,244	...	96,682	...	98,838

Note: (1) Overseas visitors were not included in columns (2) and (3)
(2) The first values in columns (4) and (6) were calculated using Eq. (11.14). The last values were estimated using Eq. 11.13, and the middle vales using Eq. (11.11)
(3) In columns (5) and (7) Eq. (11.15) was used for the first values, Eq. (11.12b) for the last values and Eq. (11.12a) for the middle values
(4) Actual population was 96,764 in 2006 and 103,097 in 2011

11.5.2 Illustrative Projections of the Population of Ryde Local Government Area

Age data from the 1996 and 2001 Australian censuses were used to project the population of the Ryde Local Government Area which is within the Sydney Metropolitan Area. The 2006 census *Time Series Profile* for this area provided the data shown in columns (2) and (3) of Table 11.5 (See Box 5.3 for further details).

The cohort-change ratios given in column (4) for age groups 5–9 to 80–84 were estimated using Eq. (11.11). In the equations used in this section, both n and k were 5 and t was 1996. The first and last values in the column were computed using Eqs. (11.14) and (11.13) respectively.

In the first projection cycle (2001–2006), 2001 was the *launch year*. The projected populations by age in 2006, shown in column (5) for age groups 5–9 to 80–84, were estimated using Eq. (11.12a). The first and last values in the column were computed using Eqs. (11.15) and (11.12b) respectively.

The above process was repeated for the second projection cycle (2006–2011) by taking 2006 as the *launch year* and using equations mentioned previously.

The above projections indicate that the population of Ryde was expected to increase by 2,438 persons between 2001 and 2006 but a slightly lower increase (2,156 persons) during the next 5 years. The actual population enumerated in the 2006 and 2011 censuses was 96,764 and 103,097 respectively. This indicates that while in the short-run the method gave quite robust results, the projected population for 2006 was only 82 persons less than the 2006 census population, the projection for 2011 was about 4 % lower than the population enumerated in the 2011 census.

11.5.3 Some General Considerations

A disadvantage of the cohort-change method, is that it can lead to unreasonably high estimates in rapidly growing places and unreasonably low projections in places experiencing population losses (Smith et al. 2001; Swanson et al. 2010). Geographic boundary changes are an issue, particularly for small geographic areas such as census tracts or collection districts. Since the cohort-change and other extrapolation methods are based on population changes within a given area, it is useful to have geographic boundaries that remain constant over time. For some sub-national areas, this presents a major challenge. A disadvantage of this method is that it does not allow the creation of different scenarios involving a range of assumptions about future levels of fertility, mortality, and migration, as is the case with the cohort-component method.

11.6 Socio-Economic Projections

11.6.1 Concepts and Methods

The projection of socio-economic characteristics has an important feature that distinguishes it from strictly demographic projections, namely, that these are *achieved characteristics* and not *ascribed characteristics*. The *ascribed characteristics* are essentially set at birth such as sex and date of birth (which determines the age at a particular point in time), while the *achieved characteristics* are not determined at birth such as educational attainment, labour force status and marital status. While race, and ethnicity are ascribed characteristics because they are largely set at birth, but they can change depending on social context and circumstances. The implication is that projections of socio-economic characteristics involve assumptions in addition to those for projections of strictly demographic characteristics, and that they can be more directly tied to policy decisions. For example, the age at which compulsory school attendance ends and the age at which people can enter the labour force. Achieved characteristics are often related to those ascribed characteristics that are associated with the life cycle, for example, age at first marriage, age at first birth, age at first entry into the labour force. In turn, this means that these characteristics are linked to the age structure of the population, and it is these linkages that provide a basis for projecting socio-economic characteristics.

Two fundamental approaches are frequently used to prepare socio-economic projections: the *participation rate or ratio method* and the *cohort-progression method* (George et al. 2004).

In the first method, socio-economic characteristics are related to demographic characteristics through the use of participation rates or ratios. For example, to prepare labour force projections, it is essential to have population projections as well as a set of labour force participation rates. The assumption would be that the population would experience in future the given labour force participation rates. The population projections may have been prepared under different scenarios using the cohort-component method (Sect. 11.4) or using the cohort-change method (Sect. 11.5) which would result only in one scenario. The participation rates or ratios may, however, be assumed to remain constant at current levels or may be extrapolated to follow a different pattern in future.

The second method uses only the cohort-change method (Sect. 11.5) to directly project the population of a group of people belonging to a certain socio-economic class. For instance, using age data from two consecutive censuses for a particular ethnic group, it would be possible to project their population in future following the procedure illustrated in Table 11.5.

11.6.2 Example of Projections of the Employed Persons Using the Participation Rate Method

To use this method two sets of data are needed: (a) the employment rates, and (b) population projections. Both datasets are by age and, if required, separately for males and females. Columns (2) and (3) of Table 11.6 show the number of employed persons and the population, both classified by age, in Estonia in 2000. Employment rates calculated based on these data are presented in column (4). Columns (5) and (6) contain the projected population of Estonia in 2005 and 2010. These were derived from Table 11.2. Assuming that the employment rates during 2000 would continue to prevail in the next 10 years, the projected number of employed persons, columns (7) and (8), were calculated by multiplying the projected populations by the employment rates.

Table 11.6 Projection of employed persons by age using participation rate method: Estonia, 2005 and 2010

	In 2000			Projections for			
				Population (M + F)		Employed persons	
Age group	Employed persons	Population (M + F)	Employment rate per person	2005	2010	2005	2010
(1)	(2)	(3)	(4) = (2)/(3)	(5)	(6)	(7) = (4) *(5)	(8) = (4) *(6)
15–24	62.3	198.5	0.313854	208.6	182.9	65.5	57.4
25–49	362.2	475.1	0.762366	462.5	457.8	352.6	349.0
50–74	148.0	373.1	0.396676	369.0	368.4	146.4	146.1

Sources: Column (2) was copied from Table ML202 in the "*Social life → Labour market → Employed persons → Annual statistics*" section of Statistics Estonia website (see Box 8.1 for the link). Column (3), (5) and (6) were based on Table 11.2
Note: Data in columns (2), (3) and (5)–(8) are given in thousands
M males, *F* females

Table 11.7 Projection of enrolments by grade using the cohort-progression method: Los Angeles County School District, 2012–13 and 2013–2014

Grade	Actual enrolments 2010–2011	2011–2012	Grade to grade	Grade progression ratio	Projected enrolments 2012–2013	2013–2014
(1)	(2)	(3)	(4)	(5)	**(6)**	(7)
K	115,760	122,647	0–K	0.081490	117,855	117,749
1	118,768	128,874	K–1	1.113286	143,474	159,728
2	116,734	117,347	1–2	0.988035	115,943	114,556
3	115,873	115,261	2–3	0.987382	113,807	112,371
4	117,380	115,491	3–4	0.996703	115,110	114,731
5	119,928	105,021	4–5	0.894709	93,963	84,070
6	117,957	117,123	5–6	0.976611	114,384	111,709
7	119,963	118,017	6–7	1.000509	118,077	118,137
8	123,576	119,423	7–8	0.995499	118,885	118,350
9	142,854	138,212	8–9	1.118437	154,581	172,889
10	134,503	131,921	9–10	0.923467	121,825	112,501
11	124,365	124,835	10–11	0.928121	115,862	107,534
12	119,828	122,336	11–12	0.983685	120,340	118,377

Source: Box 11.1

11.6.3 Example of Projecting School Enrolments Using the Cohort-Progression Method

Enrolment data for the Los Angeles County School District in California are presented in Table 11.7 (See Box 11.1 for further details). The data consist of enrolment by grade level as recorded during the school year 2010–2011, column (2) and the 2011–2012 school year, column (3). The grade levels consist of kindergarten (K), and grade levels 1, 2, 3, . . ., 12. The Cohort-Progression Method was used to project the enrolments for the next 2 years, 2012–2013 and 2013–2014. In this context, this method is usually referred to as the *Grade-Progression Method*.

All except the first value in column (5) were calculated using Eq. (11.11). As an example, the 4th value in this column was estimated by dividing the number of students in grade 3 in 2011–2012 (=115,261) by the number in grade 2 in 2010–2011 (=116,734). The grade-progression ratio from grade 2 to 3 was estimated as 0.987382.

The first value in column (5) was estimated using Eq. (11.14) as:

$$\frac{115,760 + 122,647}{1,471,729 + 1,453,861} = 0.081490$$

The denominator was the sum of students from grade 1 to 12 in both years.

Following Eq. (11.12a), projected enrolments in 2012–2013 shown in column (6) were estimated by multiplying all, except the 1st values, in column (3), being the launch year, with those in column (5). Similarly, using Eq. (11.15) the enrolments in K in 2012–2013 were estimated as:

$$0.081490 \ * \ 1,446,251 = 117,855$$

where 1,446,251 was the total number of projected students in grades 1–12 in 2012–2013.

Using 2012–2013 as the launch year, projected values for 2013–2014 were estimated by using Eqs. (11.12a) and (11.15) as explained in the previous paragraph.

Box 11.1 Obtaining Data from the California Department of Education

The link to the website is: http://dq.cde.ca.gov/dataquest/dataquest.asp. It will open the DataQuest screen. There are three rectangles, the first two have drop down menus. Select Level as *County* in the first, and Subject as *Enrollment* in the second (found under **Student Demographics**). Click on *Submit*.

On the next screen, select year of data and county from drop down menus. First select year as *2010–2011*, county as *Los Angeles* and *Submit*.

On the next screen, select the third option: *County Enrollment by Grade*, and *Submit*. The next screen will display the data for 2010–2011 for the Los Angeles County. This is the series shown in column (2) of Table 11.7.

Change the year to 2011–2012 to see the data shown in column (3) of Table 11.7.

11.6.4 Other Socio-Economic Projections

Two other types of projections that are often produced are the projections of households and families, and the projections of urban–rural segments of the population. The participation rate method can be used for both types of projections. For example, the average size of households/families may be calculated for different household/family types and assumed to either remain constant or change over time. These rates can then be applied to the independently prepared population projections in order to estimate the future number of households/families. Similar methodology may be used to estimate the future size of urban and rural populations.

11.7 Other Methods and Some Free United Nations Publications

Apart from the techniques described in this chapter, there are other techniques that can be used for population projections. These include methods such as ratio extrapolation and sophisticated approaches including ARIMA and stochastic models (see for example: Smith et al. 2001). The techniques discussed in this chapter are basic demographic techniques and the ones that can be used without specialised knowledge of statistics or mathematics.

Finally, the United Nations has published manuals dealing with various types of projections. These can be downloaded without charge as shown below:

Manual III. Methods for population projections by sex and age
http://www.un.org/esa/population/techcoop/PopProj/manual3/manual3.html
Manual V. Methods of projecting the economically active population
http://www.un.org/esa/population/techcoop/SocInd/manual5/manual5.html
Manual VII. Methods of projecting households and families
http://www.un.org/esa/population/techcoop/SocInd/manual7/manual7.html
Manual VIII. Methods for projections of urban and rural population
http://www.un.org/esa/population/techcoop/PopProj/manual8/manual8.html

Appendix 11.1: Input Data for Population Projections of Estonia

Table A11.1 Data on census population, deaths, net migration and births, and adjusted data for ages not known: Estonia, 2000

	Population		Adjusted[a]		Deaths		Adjusted[a]		Net migration				
Age group	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Age group	Births	Adjusted[a] births
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
0–4	31,561	29,794	31,574	29,801	83	57	84	57	−23	−27			
5–9	40,855	38,795	40,872	38,804	17	12	17	12	−46	−37			
10–14	55,056	52,409	55,079	52,421	19	9	19	9	−78	−63	<16	27	
15–19	52,871	50,901	52,893	50,913	48	21	49	21	−108	−71	16–19	1,280	1,307
20–24	48,245	46,403	48,265	46,414	114	23	116	23	−71	−104	20–24	4,038	4,039
25–29	47,203	47,230	47,223	47,241	119	31	121	31	−73	−90	25–29	4,009	4,010
30–34	43,828	45,508	43,846	45,519	147	49	150	49	−50	−78	30–34	2,470	2,470
35–39	46,804	49,819	46,824	49,831	224	82	228	82	−71	−73	35–39	982	982
40–44	47,417	52,017	47,437	52,029	435	125	443	125	−74	−94	40–44	250	250
45–49	44,371	50,698	44,390	50,710	573	219	583	220	−57	−90	45+	9	9
50–54	38,989	46,771	39,005	46,782	657	264	669	265	−50	−48			
55–59	32,731	41,432	32,745	41,442	736	370	749	371	−15	−31			
60–64	34,623	48,123	34,638	48,134	1,141	507	1,161	509	−22	−34			
65–69	27,259	42,183	27,270	42,193	1,203	750	1,224	753	−18	−38			
70–74	21,018	39,847	21,027	39,856	1,348	1,282	1,372	1,287	−21	−33			
75–79	10,544	28,598	10,548	28,605	881	1,539	897	1,545	−11	−26			
80–84	4,550	13,637	4,552	13,640	589	1,222	599	1,227	−5	−8			
85+	3,661	13,860	3,663	13,863	769	2,541	783	2,551	−4	−7			
NK	265	175	...	...	162	35	...	...	...	...	NK	2	...
Total	631,851	738,200	631,851	738,198	9,265	9,138	9,264	9,137	−797	−952		13,067	13,067

Notes: The reported age of mother in column (12) was aligned with column (1) by combining <16 and 16–19 age groups and designating them as 15–19 and 45+ age group was assumed to be 45–49. Small discrepancies between the totals for actual and adjusted columns are due to rounding. Data in columns (2) and (3) were taken from Table PC201, for columns (6) and (7) from Table PO045, for columns (10) and (11) from Table POR03, and for column (13) from Table PO114 available from the Statistics Estonia website (see Box 8.1)

NK age not known

[a]In columns (4), (5), (8), (9) and (14), cases of age NK were distributed on *pro-rata* basis as explained in Box 5.1

Table A11.2 Death, fertility and net migration rates and selected columns of the life tables for males and females: Estonia, 2000

	Death rate		Fertility rate	Net migration rate		Male life table columns			Female life table columns		
Age group	Males	Females		Males	Females	${}_5q_x$	l_x	${}_5L_x$	${}_5q_x$	l_x	${}_5L_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0–4	0.002660	0.001913		−0.000728	−0.000906	0.013212	100,000	496,698	0.009519	100,000	497,620
5–9	0.000416	0.000309		−0.001125	−0.000954	0.002078	98,679	492,883	0.001544	99,048	494,858
10–14	0.000345	0.000172		−0.001416	−0.001202	0.001724	98,474	491,945	0.000860	98,895	494,263
15–19	0.000926	0.000412	0.025671	−0.002042	−0.001395	0.004619	98,304	490,385	0.002058	98,810	493,543
20–24	0.002403	0.000496	0.087021	−0.001471	−0.002241	0.011943	97,850	486,328	0.002477	98,607	492,425
25–29	0.002562	0.000656	0.084884	−0.001546	−0.001905	0.012728	96,681	480,328	0.003275	98,363	491,010
30–34	0.003421	0.001076	0.054263	−0.001140	−0.001714	0.016960	95,450	473,203	0.005366	98,041	488,890
35–39	0.004869	0.001646	0.019707	−0.001516	−0.001465	0.024052	93,831	463,513	0.008196	97,515	485,578
40–44	0.009339	0.002402	0.004805	−0.001560	−0.001807	0.045630	91,574	447,423	0.011938	96,716	480,693
45–49	0.013134	0.004338	0.000177	−0.001284	−0.001775	0.063582	87,395	423,083	0.021457	95,561	472,680
50–54	0.017152	0.005665		−0.001282	−0.001026	0.082234	81,838	392,365	0.027929	93,511	461,025
55–59	0.022874	0.008952		−0.000458	−0.000748	0.108184	75,108	355,228	0.043780	90,899	444,545
60–64	0.033518	0.010575		−0.000635	−0.000706	0.154633	66,983	309,020	0.051513	86,919	423,403
65–69	0.044884	0.017847		−0.000660	−0.000901	0.201778	56,625	254,560	0.085424	82,442	394,603
70–74	0.065249	0.032291		−0.000999	−0.000828	0.280491	45,199	194,300	0.149395	75,399	348,835
75–79	0.085040	0.054012		−0.001043	−0.000909	0.350651	32,521	134,095	0.237932	64,135	282,525
80–84	0.131591	0.089955		−0.001098	−0.000587	0.495084	21,117	79,448	0.367197	48,875	199,508
85+	0.213759	0.184011		−0.001092	−0.000505	1.000000	10,662	42,945	1.000000	30,928	138,878

Notes: The death, fertility and net migration rates in columns (2) to (6) were calculated using data from Table A11.1. The total fertility rate of 1.38 was obtained from column (4) using Eq. (5.5a). Columns (7) to (12) were calculated using the abridged life table methodology (Chap. 7). The death rates in columns (2) and (3) were converted into ${}_5q_x$ by using Eq. (7.19). Equation (7.8) was used for l_x, and Eq. (7.20) for ${}_5L_x$ except for the last age group for which Eq. (7.16) was used. Using $e_0 = \frac{\sum {}_5L_x}{100{,}000}$ the life expectancy at birth was estimated as 65.1 years for males and 75.9 years for females

References

Australia. (2008). *Australian historical population statistics 2008*. Catalogue No. 3105.0.85.001. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3105.0.65.0012008?. OpenDocument Accessed Feb 2013.

Croxton, F. E., Cowden, D. J., & Klein, S. (1968). *Applied general statistics*. Chapter 13. London: Sir Isaac Pitman and Sons Ltd.

George, M. V., Smith, D. A., Swanson, D. A., & Tayman, J. (2004). Population projections. In J. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 561–601). San Diego: Elsevier/Academic.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality; and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society, 115*(27), 513–585.

Hamilton, C. H., & Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces, 41*, 163–170.

Pearl, R., & Reed, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of National Academy of Science, 6*(1), 275–288.

Smith, S., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis*. New York: Kluwer Academic Press.

Swanson, D. A., Schlottmann, A., & Schmidt, R. (2010). Forecasting the population of census tracts by age and sex: An example of the Hamilton-Perry method in action. *Population Research and Policy Review, 29*(1), 47–63.

Swanson, D. A., & Tedrow, L. (2012). Using cohort change ratios to estimate life expectancy in populations with negligible migration: A new approach. *Canadian Studies in Population, 39*, 83–90.

United Nations. (2012a). *World population prospects 2010 revision*. New York: Department of Economic and Social Affairs.http://esa.un.org/unpd/wpp/index.htm. Accessed Feb 2013.

United Nations. (2012b). *Probabilistic population projections based on the 2010 revision of the World population prospects*. New York: Department of Economic and Social Affairs. http://esa.un.org/unpd/ppp/index.htm. Accessed Feb 2013.

Verhulst, P. F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouv. mém. de l'Academie Royale des Sci. et Belles-Lettres de Bruxelles, 18*, 1–41.

Verhulst, P. F. (1847). Deuxième mémoire sur la loi d'accroissement de la population. *Mém. de l'Academie Royale des Sci., des Lettres et des Beaux-Arts de Belgique, 20*, 1–32.

Chapter 12
Testing the Quality and Smoothing of Demographic Data

12.1 Purpose

This chapter describes some of the commonly used methods to test the quality of demographic data, and procedures to adjust and smoothing data to improve their quality. The procedures described in this chapter can be used to test demographic data whether obtained through population censuses, surveys or other sources.

12.2 Types of Error and Their Sources

There are three main types of errors that can occur in any demographic dataset, whether it represents an entire population of interest (a census or a population register) or a sample of this same population. A fourth type of error only affects sample surveys.

Coverage errors result from a certain segment of the population being missed from the data collection. For example, people living in remote areas, nomadic populations or those travelling during the data collection period may be missed. Under some circumstances, particularly in sample surveys, the scope of a data collection may exclude certain population groups which may be difficult or too costly to canvass. Such deliberate exclusions are not considered as coverage errors.

Response errors may result because the respondent may not understand the question asked because of poor wording or vagueness in the questionnaire. For example, if a survey asked a voter if he/she liked a candidate it may be that some voters in fact liked the candidate, but did not vote for this particular person. A more appropriate question to determine the voting intentions would have been: "Would you vote for (name of candidate)?" Some questions are sensitive and the respondents may deliberately give a vague or incorrect response or even refuse to answer.

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_12,

Processing errors occur at various stages of the data processing, for example, when coding data into specific categories or when transcribing original answers from one medium (usually paper) to an electronic medium. Sometimes these errors are caused by humans while at other times they may result due to failure of the soft- and hard-ware used.

The above three types of errors are collectively referred to as the *non-sampling errors*. On the other hand, *sampling errors* are the result of obtaining answers from some but not all members of the population of interest. This means that by the luck of the draw one may have gotten higher income households in a sample than the average household income for the entire population. This may happen even though there was no coverage error, response error or processing error.

12.3 Some General Principles

Depending upon the source of data, a variety of methods can be used to test the accuracy and adequacy of data. Most methods involve comparing two or more statistics that have a known, or assumed, relationship. Methods of testing the accuracy of demographic data usually involve checking the internal consistency of data and comparison with some other data sources. While there is no specific checklist, a few examples of possible checks are given in the following paragraphs.

Any irregularities (peaks and troughs) in an age distribution may be examined in the light of the contribution of past events such as migration, wars, baby booms and famines. As far as possible, an attempt should be made to distinguish the real variations due to such events from the random fluctuations due to response errors.

Age-specific sex ratios generally decline by age because of the relatively higher male mortality. Sex-selective under- or over-enumeration as well as the type of events mentioned in the previous paragraph may cause the sex ratios to depart from the norm, and would thus require further investigation.

The number of currently married males should be approximately equal to the number of currently married females. If not, then reasons for this discrepancy, such as sex-specific in- or out-migration and cultural practices, should be examined.

Given that the age distribution of a population is based on the interaction of the recent and past fertility, mortality and migration levels, the consistency of births, deaths and migration data with the age pyramid may be considered. For example, as noted in Chap. 4, populations with higher birth rates in past 5 years are likely to have broad based population pyramids.

For small populations, annual rates may fluctuate considerably from year to year. In such populations, for example, the deaths of three teenagers in the 15–19 age group in a single automobile accident could cause a significant spike in the age-specific death rate for age group 15–19.

12.4 Quality of Age Data

Age data may be given in single years of age or in age groups. Sometimes these data exhibit large peaks and troughs that may be real or due to misreporting of ages. The single-year age data are particularly susceptible to such irregularities as they tend to be smoothed out when broad age groupings are used.

In this section two of the commonly used methods of testing the accuracy of age data are described – one for single-year age data and the other for data in age groups.

12.4.1 Digital Preference in Age Data

Figure 12.1 shows single-year age data from the 2000 censuses of Brazil and Indonesia (United Nations 2011) for both males and females combined in each country. A comparison of the two age distributions indicates that the Indonesian age distribution exhibited bigger peaks and troughs than that for Brazil. These irregularities may have resulted from age misreporting and digital preference, that is, people were more likely to report ages ending in certain digits than others.

The prevalence of digital preference in a population age distribution, say, from age 0 to 59, may be estimated by cumulating ages ending in each digit as shown below:

$$
\begin{array}{l}
\text{for terminal digit ``}\mathit{0}\text{'' : } T_0 = P_0 + \dots\dots\dots + P_{50} \\
\text{for terminal digit ``}\mathit{1}\text{'' : } T_1 = P_1 + \dots\dots\dots + P_{51} \\
\quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\text{for terminal digit ``}\mathit{8}\text{'' : } T_8 = P_8 + \dots\dots\dots + P_{58} \\
\text{for terminal digit ``}\mathit{9}\text{'' : } T_9 = P_9 + \dots\dots\dots + P_{59}
\end{array}
$$

where T_0 to T_9 are the totals for ages ending in *0, 1, . . ., 9* and the *P* values represent the numbers at each age. In the absence of migration, the relationship $P_x > P_{x+1}$ will be true for all values of x and thus $\mathrm{T}_0 > \mathrm{T}_1 > \mathrm{T}_2 \ldots > \mathrm{T}_9$. Myers (1940) proposed a method that involves calculating a *blended* population for each terminal digit, to give an equal weight to each terminal digit in the simulation process. In a population, where numbers decrease linearly by age, the expectation is that in the absence of any digital preference about 10 % of the total blended population will have ages ending in each of the ten terminal digits. Any deviations from this norm, when added irrespective of sign, would be an indication of the prevalence of digital preference. This is usually referred to as the Myers' index of digital preference.

Indonesian age data, for males and females combined, from the 2000 census shown in Table 12.1 have been used to illustrate the calculation of the above index. These are the same data that have been presented in Fig. 12.1.

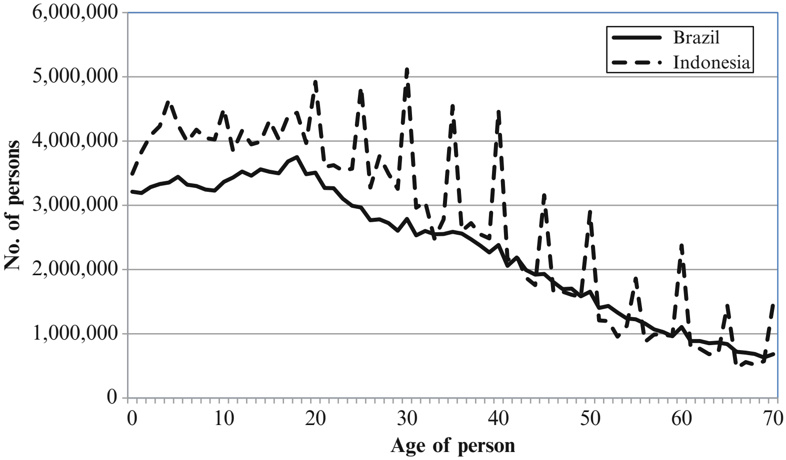


Fig. 12.1 Age distribution in single years: Brazil and Indonesia, 2000 censuses (*Source*: United Nations 2011)

Table 12.1 Single-year age data for Indonesia, males and females combined: 2000 census

Age	Persons	Age	Persons	Age	Persons	Age	Persons
0	3,492,259	18	4,441,813	36	2,600,870	54	1,123,261
1	3,834,175	19	3,971,307	37	2,722,768	55	1,860,625
2	4,089,996	20	4,921,618	38	2,548,602	56	874,332
3	4,232,178	21	3,597,374	39	2,483,767	57	989,520
4	4,653,768	22	3,626,812	40	4,474,171	58	987,528
5	4,252,803	23	3,541,645	41	2,149,854	59	966,659
6	3,989,618	24	3,570,652	42	2,211,118	60	2,375,464
7	4,178,930	25	4,848,603	43	1,873,558	61	792,149
8	4,047,752	26	3,278,048	44	1,759,147	62	767,710
9	4,024,988	27	3,765,187	45	3,156,156	63	681,797
10	4,506,698	28	3,492,531	46	1,659,968	64	703,899
11	3,847,492	29	3,256,568	47	1,657,431	65	1,443,103
12	4,161,790	30	5,117,317	48	1,607,406	66	465,351
13	3,948,136	31	2,962,140	49	1,575,044	67	561,834
14	3,989,616	32	3,056,358	50	2,896,670	68	520,094
15	4,323,082	33	2,480,895	51	1,210,064	69	574,544
16	4,029,498	34	2,783,010	52	1,198,712	70+	5,554,022
17	4,383,817	35	4,548,219	53	956,261	Total[a]	201,241,999

Source: United Nations (2011)
Note: [a]Total includes 11,847 persons with unknown ages

Table 12.2 shows the various steps involved in the calculation of the index of digital preference. Because of the limitations of the available age range, only data from ages 0 to 69 have been used. Any age range can be used for the calculation of

Table 12.2 Estimation of the index of digital preference: Indonesia, 2000

	Persons aged 0–59			Persons aged 10–69					
Digit	Sum	Weight	Product	Sum	Weight	Product	Blended population	%	Absolute difference from 10 %
(1)	(2)	(3)	(4) = (2) * (3)	(5)	(6)	(7) = (5) * (6)	(8) = (4) + (7)	(9)	(10)
0	25,408,733	1	25,408,733	24,291,938	9	218,627,442	244,036,175	14.0	4.0
1	17,601,099	2	35,202,198	14,559,073	8	116,472,584	151,674,782	8.7	1.3
2	18,344,786	3	55,034,358	15,022,500	7	105,157,500	160,191,858	9.2	0.8
3	17,032,673	4	68,130,692	13,482,292	6	80,893,752	149,024,444	8.6	1.4
4	17,879,454	5	89,397,270	13,929,585	5	69,647,925	159,045,195	9.2	0.8
5	22,989,488	6	137,936,928	20,179,788	4	80,719,152	218,656,080	12.6	2.6
6	16,432,334	7	115,026,338	12,908,067	3	38,724,201	153,750,539	8.8	1.2
7	17,697,653	8	141,581,224	14,080,557	2	28,161,114	169,742,338	9.8	0.2
8	17,125,632	9	154,130,688	13,597,974	1	13,597,974	167,728,662	9.7	0.3
9	16,278,333	10	162,783,330	12,827,889	0	0	162,783,330	9.4	0.6
Total							1,736,633,403	100.0	13.2

Source: Table 12.1
Note: Persons aged 70 years and over and those not reporting their ages were excluded from calculations presented in this table

this index insofar as the range covers an equal number of ages ending in each of the ten digits 0–9 shown in column (1) of Table 12.2.

Columns (2) and (5) are based on age ranges 0–59 and 10–69 respectively. These columns give the sum of ages ending in each of the ten terminal digits within the specified age range. For example, the first number in column (2) is the sum of persons aged 0, 10, 20, 30, 40 and 50. Note that all of these ages end in digit 0. Similarly, the second figure in column (2) is the sum of persons aged 1, 11, 21, 31, 41 and 51. All ages in this case end in digit 1. This process is repeated for the remaining eight digits. Column (3) gives the first set of weights, and column (4) contains the product of columns (2) and (3).

Column (5) follows the procedure of column (2) except that in this case the age range is 10–69. Thus, the third figure in column (5) is the sum of persons aged 12, 22, 32, 42, 52 and 62. Column (6) gives the second set of weights, and column (7) shows the product of columns (5) and (6). These procedures are repeated for each of the 10 terminal digits. The sum of the weights in columns (3) and (6) for each digit is equal to 10.

Column (8) is the sum of columns (4) and (7). Meyers called it the *blended* population. Column (9) gives the percentage distribution of the blended population. Digits 0 and 5 were the most popular: 27 % of the Indonesian population reported ages ending in these two digits.

According to Myers (1940), the blended population for each digit is expected to be 10 % of the total blended population. The absolute values of deviations from this norm are shown in column (10) for each terminal digit. The sum of this column is the Myers' index of digital preference. This index was estimated as 13.2 for Indonesia.

Applying the same procedure using the Brazilian data resulted in an index of only 1.5, indicating the better quality of age data in the Brazilian census.

The Myers' index of digital preference may be calculated for different segments of a population and using different age ranges. For example, using the data from a large sample survey conducted in Pakistan, Yusuf (1967) estimated this index for males as 66 and 65 for females. His calculations were based on the age range 10–79 instead of 0–69 in Table 12.2.

The percentage blended population can be used to determine a 5-year age grouping that would minimise the effect of digital preference. Such a grouping is the one that gives a sum of the percentage blended populations closest to 50 %. For instance, using column (9) in Table 12.2, the sum of blended populations for groups of digits and their closeness to 50 % was as follows:

Digits	Sum of % blended population	Difference from 50 %
0–4	49.7	0.3
1–5	48.3	1.7
2–6	48.4	1.6
3–7	49.0	1.0
4–8	50.1	0.1
5–9	50.3	0.3

Age grouping 4–8 was the one closest to 50 %. It means that to minimise the impact of digital preference, the single-year age data from the 2000 census of Indonesia should be aggregated in 5-year age groups: 4–8, 9–13, 14–18 and so on to be preceded by the 0–3 age group.

12.4.2 Sex and Age Ratio Scores

The sex and age ratio scores were developed by the United Nations (1955) to facilitate the checking of the accuracy of data given in age groups. These groups may be based on 5-year, 10-year or other aggregations of ages as long as all the age groups have the same number of years.

The *sex ratio score* is defined as the average of the absolute differences between the sex ratios of the consecutive age groups.

An average population needs to be estimated for each age group prior to the calculation of the age ratios. This average is based only on the population of groups immediately preceding and succeeding the particular age group. Obviously, this average cannot be calculated for the first and the last age groups in the series. The age ratio is defined as the ratio of the actual to average population of a particular age group. The *age ratio score* is calculated by taking the average of the absolute value of the percentage by which the actual population exceeds the average population.

Table 12.3 illustrates the calculation of the sex ratio score and the age ratio scores for Indonesian males and females enumerated in the 2000 census.

Columns (2) and (3) are based on the age data for males and females in the 2000 census of Indonesia. Sex ratios in column (4) are calculated as described in Eq. (4.16). Column (5) gives the difference between the sex ratios for consecutive age groups. Column (5) is summed, irrespective of sign, and divided by 13 to give the average. This is called the *sex ratio score*.

Columns (6) and (7) give the average population for all except the first and the last age groups. These averages are based on the population of two groups – those immediately preceding and succeeding the particular age group. Thus, the second figure in column (6) is the average of 10,295,701 and 10,460,908, and similarly the third figure in column (7) is the average of 10,060,226 and 10,500,169.

Columns (8) and (9) show the age-specific ratios of actual to average populations for males and females respectively.

Columns (10) and (11) indicate the percentage by which the actual populations for each age group exceeded their average populations. This was estimated for males and females separately by taking the value in column (8) minus 1 and multiplied by 100 for males and the value in column (9) minus 1 and multiplied by 100 for females. Absolute values of columns (10) and (11) were summed and divided by 12 to get the *age ratio scores* for males and females respectively.

Table 12.3 Calculation of the sex ratio score, and the age ratio scores for males and females: Indonesia, 2000.

	Actual population				Average population		Actual/Average		% actual > average	
Age group	Males	Females	Sex ratio	Difference	Males	Females	Males	Females	Males	Females
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
0–4	10,295,701	10,006,675	102.89	0.82	...	...	...	...	...	...
5–9	10,433,865	10,060,226	103.71	0.97	10,378,305	9,999,750	1.0054	1.0060	0.54	0.60
10–14	10,460,908	9,992,824	104.68	−3.26	10,541,607	10,280,198	0.9923	0.9720	−0.77	−2.80
15–19	10,649,348	10,500,169	101.42	−9.24	9,849,186	10,006,731	1.0812	1.0493	8.12	4.93
20–24	9,237,464	10,020,637	92.18	3.83	9,889,926	10,005,301	0.9340	1.0015	−6.60	0.15
25–29	9,130,504	9,510,433	96.01	4.10	8,720,883	9,108,028	1.0470	1.0442	4.70	4.42
30–34	8,204,302	8,195,418	100.11	−0.63	8,281,672	8,490,910	0.9907	0.9652	−0.93	−3.48
35–39	7,432,840	7,471,386	99.48	7.13	7,318,870	7,114,914	1.0156	1.0501	1.56	5.01
40–44	6,433,438	6,034,410	106.61	4.74	6,260,046	6,020,070	1.0277	1.0024	2.77	0.24
45–49	5,087,252	4,568,753	111.35	−5.86	5,112,312	4,814,097	0.9951	0.9490	−0.49	−5.10
50–54	3,791,185	3,593,783	105.49	−2.35	3,985,239	3,682,096	0.9513	0.9760	−4.87	−2.40
55–59	2,883,226	2,795,438	103.14	−7.80	3,194,131	3,158,863	0.9027	0.8850	−9.73	−11.50
60–64	2,597,076	2,723,943	95.34	−7.59	2,274,709	2,347,087	1.1417	1.1606	14.17	16.06
65–69	1,666,191	1,898,735	87.75	...	...	...	...	...	...	...
Total	...	...	...	$\lvert 58.32 \rvert$	...	...	...	...	$\lvert 55.25 \rvert$	$\lvert 56.69 \rvert$
Average	...	...	...	**4.49**	...	...	...	...	**4.60**	**4.72**
				Sex ratio score					**Age ratio scores**	

Source: United Nations (2011)
Note: Persons aged 70 and over and those not reporting their ages were excluded from the above table

A combination of these three scores, called the *joint score*, is sometimes used. It is defined as: [3 * (Sex ratio score)] + (Age ratio score for males) + (Age ratio score for females).

Given below is a comparison of the various scores based on 2000 census age distributions of Brazil and Indonesia:

	Brazil	Indonesia
Sex ratio score	1.47	4.49
Age ratio score for males	2.68	4.60
Age ratio score for females	2.53	4.72
Joint score	9.62	22.79

12.5 Comparison with Other Data Sources

One method to validate the accuracy of data from a census, survey or other collections is to obtain corresponding data from some other source which may be considered as a *standard*. For example, the quality of census data in many countries is determined through post-enumeration surveys conducted in a sample of households. Data on specific characteristics from such surveys are then matched with the census data from the same households. The outcome of such a procedure may be as shown in Table 12.4.

In this table, cells *a* and *d* contain the cases where both data sources concurred, while cells *b* and *c* represent the cases where the two sources differed. Under such circumstances, the gross difference between the two sources is $(c + b)$ and the net difference is $(c - b)$. These values have to be divided by the total sample size, *n,* to calculate the corresponding rates as shown below:

$$\text{Gross difference rate} = \frac{(c + b) * 100}{n} \tag{12.1}$$

$$\text{Net difference rate} = \frac{(c - b) * 100}{n} \tag{12.2}$$

A positive value of the net difference rate indicates an over-count of the specific characteristic in the census, survey or other collection, while a negative value indicates an under-count.

As noted earlier, there are two consistent responses in Table 12.4: one where both sources concurred on the presence of a characteristic (cell *a*) and the other where both sources concurred on the absence of a particular characteristic (cell *d*). To calculate the two indexes of stability these two cells are expressed as a percentage of $(a + b)$ and $(c + d)$ respectively:

Table 12.4 Schematic presentation of the results of matching data from two sources with regard to a specific characteristic

Some dataset taken as the *standard*	Census, survey or other collection		
	No. with a specific characteristic	No. without a specific characteristic	Total
No. with a specific characteristic	*a*	*b*	*a* + *b*
No. without a specific characteristic	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *d*	*n* = *a* + *b* + *c* + *d*

Source: Spiegelman (1969)

$$\text{Index of stability for the presence of a characteristic} = \frac{a * 100}{(a+b)} \quad (12.3)$$

$$\text{Index of stability for the absence of a characteristic} = \frac{d * 100}{(c+d)} \quad (12.4)$$

Another measure of the stability of response is estimated by dividing the gross difference by the sum of variances of the particular characteristic in the two data sources.

Let p_1 be the proportion with a particular characteristic in the census, survey or other collection and p_2 be the proportion with the same characteristic in the standard. Using Table 12.4 the values of p_1 and p_2 are estimated as $p_1 = \frac{(a+c)}{n}$ and $p_2 = \frac{(a+b)}{n}$.

Statistically, the variance (*VAR*) of $p = p * (1 - p)$ (see Box 12.1). Therefore, the *VAR* of p_1 and p_2 can be calculated as follows:

$$VAR(p_1) = \frac{(a+c)}{n} * [1 - \frac{(a+c)}{n}] = \frac{(a+c)}{n} * [\frac{n-(a+c)}{n}]$$

$$VAR(p_1) = \frac{(a+c)}{n} * [\frac{a+b+c+d-a-c}{n}]$$

$$VAR(p_1) = \frac{(a+c)}{n} * \frac{(b+d)}{n}.$$

Similarly

$$VAR(p_2) = \frac{(a+b)}{n} * \frac{(c+d)}{n}.$$

Therefore, the index of inconsistency (I) is:

$$I = \frac{(b + c) * n * 100}{[(a + c) * (b + d)] + [(a + b) * (c + d)]} \tag{12.5}$$

This index varies between 0 and 100 (Spiegelman 1969). A large value of this index indicates a higher degree of response error.

Post-enumeration surveys also provide an indicator of the extent of errors of coverage. For example, according to the 2011 census of Australia, the overall under-enumeration in the census was of the order of 1.7 % (Australia 2012). The undercount varied according to the characteristic of the population. For example, it was 2.2 % for males and 1.2 % for females. Since these estimates are based on a sample, they are themselves subjected to sampling errors. For the three figures quoted here the standard error was ±0.2.

Box 12.1 Variance and Standard Deviation of a Proportion

Consider a population of size N in which P have a certain characteristic. Therefore, $N - P$ do not have that characteristic. Proportion of population with the characteristic, P is: $p = \frac{P}{N}$ and the proportion not having that characteristic is: $q = \frac{N-P}{P}$. Obviously, $p + q = 1$ and $q = 1 - p$. Statistically, the variance of $p = p * q = p * (1 - p)$, and standard deviation $= \sqrt{p * (1 - p)}$. For further details see a textbook on statistics.

12.6 Smoothing of Demographic Data: Some General Considerations

Before attempting any detailed analysis of data for a particular country, it is essential to know about the history of that country. This is vital before using any smoothing technique. It is expected that any major occurrence such as wars, epidemics, famines, and migration streams will have a noticeable effect on the demography of a country. The effects of such events should not be smoothed out as they are real. Thus, while smoothing a dataset one has to strike a balance between smoothing and maintaining the integrity of data.

If accurate data on births, deaths and migration are available on an annual basis it is possible, to construct the correct age and sex distribution of a population as of any given date. The true number of 20-year-olds in the year 2010, for example, should be equal to the number of births 20 years previously, minus the deaths experienced by this cohort in the intervening years (with appropriate allowances for immigration and emigration). This method of constructing the true age and sex distribution of a population may be considered as one form of smoothing. Practically speaking, very few countries have the required basic data of a quality good enough to warrant using this method in preference to other smoothing methods.

This section describes some of the techniques used for *smoothing* various demographic distributions to make them conform more closely to the expected patterns, on the assumption that the irregularities are due to errors. Unless the significant historical events are known – for example, those which have led to a shortage of births for a particular year or so – peculiarities of the data that have genuine foundation will be wrongly treated as errors. These predictable peculiarities in the demographic characteristics such as the age structure cause problems in the analysis, but must be allowed for, instead of forcing out their effect to make the data conform to a particular pattern.

Smoothing of data, both numbers and rates, may involve calculating averages using various measures of central tendency, deriving moving averages, aggregation of data or using certain mathematical models.

12.6.1 Using Various Measures of Central Tendency

Demographic data may be smoothed by averaging them over a specific period of time. For instance, as noted in Chap. 7, deaths data are averaged over a period of many years (usually 3 years) to smooth and iron out the random fluctuations. Similar averaging of births and other demographic events are also commonly used. These same methods can be used with birth and death rates. For averaging, any of the measures of central tendency may be used (Sect. 3.6).

12.6.2 Moving Averages

For a set of data there will be only one average, while there may be a number of moving averages based on various overlapping segments of the series. Such averages are called *moving averages.*

Table 12.5 shows a hypothetical series of 15 observations to illustrate the calculation of moving averages based on arithmetic means. The same concept can be used to compute other measures of central tendency.

Column (1) of the table shows the observation number, and column (2) the actual data that have to be smoothed.

Column (3) has the moving averages based on three observations. The first figure in this column is the average of the first three observations and is assigned to the second observation. The second figure is the average of observations 2, 3 and 4 and is ascribed to third observation and so on.

Similarly, column (4) exhibits the moving averages based on 5 observations each and columns (5) to (7) the moving averages using seven, nine and 11 observations respectively.

Table 12.5 Hypothetical numeric data and calculation of moving averages

Observations	Actual data	Moving average using consecutive observations				
		3	5	7	9	11
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	180	...	...	...	...	...
2	194	193	...	...	...	...
3	206	205	206	...	...	...
4	214	219	213	207	...	...
5	236	222	215	212	208	...
6	217	219	217	214	210	210
7	203	211	215	213	215	211
8	213	207	209	216	214	213
9	206	208	212	210	214	212
10	205	214	211	210	210	211
11	232	211	211	210	208	...
12	197	214	210	208	...	...
13	214	204	209	...	...	...
14	201	206	...	...	...	...
15	202	...	...	...	...	...

Moving averages may be based on any number of observations, though the number is usually an odd number. In this example, arithmetic means have been used; however, the geometric or harmonic means can also be used to calculate the moving averages.

One of the limitations of moving averages is that a certain number of observations in the beginning and end cannot be smoothed. For instance, using a three-observations moving average the first and last observations cannot be estimated and when using five observations, the first two and the last two cannot be derived.

Moving averages may be used for smoothing of age data as well as demographic rates such as fertility and mortality rates. For example, if the numbers provided in Table 13.5 were rates, the same *moving average* method could be used to smooth them.

12.6.3 Aggregation of Data

Aggregation of data would generally result in smoother distributions. For instance, the Brazilian and Indonesian data in single years of age when aggregated into 5-year age groups resulted in much smoother series (Fig. 12.2 compared to Fig. 12.1).

12.6.4 Smoothing Age Data

In practice, different methods are used to smooth data exhibiting different degrees of inaccuracy, so that fairly accurate statistics would be modified only slightly

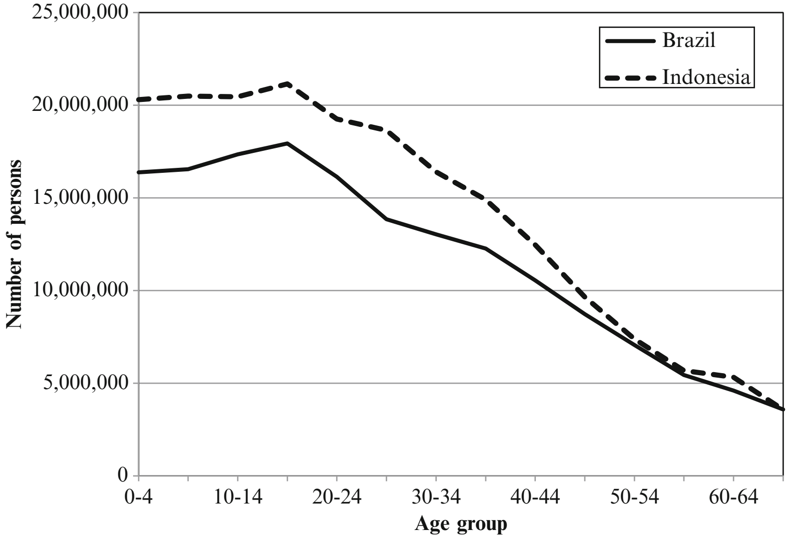


Fig. 12.2 Age distribution in 5-year age groups: Brazil and Indonesia, 2000 censuses (*Source*: United Nations 2011)

while less accurate data would be more radically transformed. A method to smooth 5-year age groups in the range from 10 to 74 years is given below. This procedure is appropriate for use where the data are markedly inaccurate. It is derived from a parabolic model (see Box 12.2) and has been commonly used (United Nations 1955). The procedure employs five terms; to adjust the figure for one 5-year age group, data for the two preceding and the two following age groups are inserted in the formula. If the age statistics are tabulated by 5-year groups up to age 85, smoothing can be effected by such a procedure for all groups between the ages of 10 and 75. The numbers at the youngest and oldest ages have to be dealt with separately. The formula may be stated as follows:

$$T_0 = 0.0625 * [-S_{-2} + 4 * S_{-1} + 10 * S_0 + 4 * S_1 - S_2] \qquad (12.6)$$

where T_0 is the smoothed number of persons in a particular 5-year age group, S_0 is the reported number of persons in that age group, S_{-1} and S_{-2} are the reported numbers in the two preceding age groups, and S_1 and S_2 are the reported numbers in the two subsequent age groups.

Box 12.2 Parabolic Model

The equation of a straight line or a linear model was defined as $Y = a + b * X$. In a parabolic model an extra term involving X^2 is included to allow a curvature in the model. Thus, the equation of a parabola is $Y = a + b * X + c * X^2$. The method of least squares can be used to fit this as well as the linear model (see, for example: Triola 2007).

When the data for the 0–4 age group are found to be seriously deficient, it should be replaced by an independent estimate. The deficiency is sometimes found to be important only in the case of infants under 1 year of age; if so, the independent estimate may be limited to these infants, and combined with the census figure for ages 1–4.

The method of deriving a corrected estimate of the population 0–4 utilises the number of births during the preceding 5-years time period, discounted by the probability of survival from birth to age 0–4 using an appropriate life table. As shown in Sect. 11.4.3 this probability is equal to ${}_5L_0$ divided by 500,000. This method depends upon the availability of accurate vital statistics.

The reported population aged 5-9 as reported in the census can generally be accepted, unless there are special reasons to doubt their accuracy. Examination of census data for many countries yields little evidence that numbers aged 5–9 years are subject to any special inaccuracy of reporting; in fact, even where the age data are very defective, the figures for this age group commonly appear to be more accurate than others. Experience has demonstrated that children of this age are not nearly as likely to be overlooked in the census enumeration as the other age groups. It is easier to obtain from their parents at least a nearly accurate estimate of age for children in this group than to get accurate reports on older groups.

To estimate the correct numbers for the advanced ages is generally difficult, whether by extrapolation or by reconstruction from birth and death statistics pertaining to a distant period in the past. For the purposes of demographic analysis, errors in numbers at the most advanced ages are only of limited significance, because the numbers in these cohorts are rapidly depleted by deaths.

12.6.5 Smoothing Using Interpolation Multipliers

One method often used, consists of applying a sets of constant multipliers to the reported 5-year age groups in order to obtain a distribution by single years of age. These multipliers are of special interest when the data by single years of age are not available. They are based on interpolation formulae derived by an actuary, T.B. Sprague, in 1880. His formulae were subsequently converted into a set of multiplier by J.W. Glover in 1921 (Jaffe 1951).

An important feature of these multipliers, usually referred to as the Sprague's multipliers, is that they redistribute the total of within a 5-year age group into single years of age in a manner that they add up to the original 5-year total. This means that if the 5-year age data are defective then the interpolated values using the multipliers would also be affected.

Consider an age distribution in 5-year age groups: 0–4, 5–9, … , 65–69 and 70+. All age groups except the last one (70+), with an open-ended age interval, can be converted into single years using Sprague's multipliers. The open-ended age intervals cannot be disaggregated using these multipliers.

Table 12.6 Sprague's multipliers to disaggregate 0–4 age group

	Population in age groups			
	N_0	N_1	N_2	N_3
n_0	+.3616	−.2768	+.1488	−.0336
n_1	+.2640	−.0960	+.0400	−.0080
n_2	+.1840	+.0400	−.0320	+.0080
n_3	+.1200	+.1360	−.0720	+.0160
n_4	+.0704	+.1968	−.0848	+.0176

Source: Jaffe (1951)

Table 12.7 Sprague's multipliers to disaggregate 5–9 age group

	Population in age groups			
	N_{-1}	N_0	N_1	N_2
n_0	+.0336	+.2272	−.0752	+.0144
n_1	+.0080	+.2320	−.0480	+.0080
n_2	−.0080	+.2160	−.0080	+.0000
n_3	−.0160	+.1840	+.0400	−.0080
n_4	−.0176	+.1408	+.0912	−.0144

Source: Jaffe (1951)

Table 12.8 Sprague's multipliers to disaggregate 10–14, 15–19 to the third last 5-year age group

	Population in age groups				
	N_{-2}	N_{-1}	N_0	N_1	N_2
n_0	−.0128	+.0848	+.1504	−.0240	+.0016
n_1	−.0016	+.0144	+.2224	−.0416	+.0064
n_2	+.0064	−.0336	+.2544	−.0336	+.0064
n_3	+.0064	−.0416	+.2224	+.0144	−.0016
n_4	+.0016	−.0240	+.1504	+.0848	−.0128

Source: Jaffe (1951)

Table 12.9 Sprague's multipliers to disaggregate the second last 5-year age group

	Population in age groups			
	N_{-2}	N_{-1}	N_0	N_1
n_0	−.0144	+.0912	+.1408	−.0176
n_1	−.0080	+.0400	+.1840	−.0160
n_2	+.0000	−.0080	+.2160	−.0080
n_3	+.0080	−.0480	+.2320	+.0080
n_4	+.0144	−.0752	+.2272	+.0336

Source: Jaffe (1951)

Sprague's multipliers are shown in Tables 12.6 to Tables 12.10. Multipliers given in Tables 12.6 and Table 12.7 have to be used for the 0–4 and 5–9 year age groups respectively. For the last two 5-year age groups (*e.g.*, 60–64 and 65–69), multipliers shown in Tables 12.9 and 12.10 have to be utilized in that order. For all other age groups, except the one with an open-ended interval, multipliers given in Table 12.8 are used.

Table 12.10 Sprague's multipliers to disaggregate the last 5-year age group

	Population in age groups			
	N_{-3}	N_{-2}	N_{-1}	N_0
n_0	+.0176	−.0848	+.1968	+.0704
n_1	+.0160	−.0720	+.1360	+.1200
n_2	+.0080	−.0320	+.0400	+.1840
n_3	−.0080	+.0400	−.0960	+.2640
n_4	−.0336	+.1488	−.2768	+.3616

Source: Jaffe (1951)

In Sprague's multipliers, N_0 refers to the population of the 5-year age group that is to be converted into single years of age. N_{-1} and N_{-2} are the populations of the two preceding groups, and N_1 and N_2 are the populations of the two following age groups.

The procedure to use the Sprague's multipliers involves multiplying each column of the relevant table with the appropriate value of $N_.$ The sum of each of the five rows gives the values of n_0 to $n_{4.}$

For the purposes of illustration data in Table 12.3 have been used in each of the Tables 12.11 to 12.16. For example, the second column in Table 12.11 was arrived at by multiplying 10,295,701 with each value in the N_0 column of Table 12.6. Similarly, the third column in Table 12.11 was calculated by multiplying 10,433,865 with each value in the N_1 column of Table 12.6. This procedure was repeated in the next two columns. Finally, each row was summed to arrive at the estimated population aged 0, 1, 2, 3 and 4. As expected, the sum of the interpolated values was the actual value of N_0: 10,295,701.

The same procedure was used in Tables 12.12 to 12.16 to interpolate the 5–9, 10–14, 15–19, 60–64 and 65–69 age groups. Table 12.8 was used for ages 20–24 to 55–59 with appropriate values of N_{-2}, N_{-1}, N_0, N_1 and $N_{2,}$ such that N_0 was always the age group to be disaggregated and the other four consisted of the two preceding and the two following age groups.

Results of applying Sprague's multipliers to the 5-year age distribution of Indonesian males are shown in Fig. 12.3 along with the original data. It appears that this method not only disaggregated the 5-year age distribution into single-years of age but was also instrumental in smoothing the resulting age distribution.

Sprague's multipliers can be used for a great variety of data besides age distributions, provided that the data are continuous and that quantitative variables are being studied. For instance, age-specific mortality and labour force rates in 5-year age groups can be disaggregated using these multipliers. However, discrete data such as age-specific hours worked per week are not suitable as these have peaks at 35 hours and/or 40 hours that must not be smoothed out if the results are to have any meaning. These multipliers are purely mathematical in their approach, and do not contain within themselves any means of taking into consideration, irregular but true fluctuations in the basic data.

Table 12.11 Disaggregating the 0–4 age group to single years of age: Indonesian males, 2000 (using multipliers in Table 12.6)

Notation	N_0	N_1	N_2	N_3			
Age group	0–4	5–9	10–14	15–19			
Population	10,295,701	10,433,865	10,460,908	10,649,348	Row sum	Rounded row sum	Age in single years
$n_0 =$	3,722,925.5	−2,888,093.8	1,556,583.1	−357,818.1	2,033,596.7	2,033,597	0
$n_1 =$	2,718,065.1	−1,001,651.0	418,436.3	−85,194.8	2,049,655.6	2,049,656	1
$n_2 =$	1,894,409.0	417,354.6	−334,749.1	85,194.8	2,062,209.3	2,062,209	2
$n_3 =$	1,235,484.1	1,419,005.6	−753,185.4	170,389.6	2,071,693.9	2,071,694	3
$n_4 =$	724,817.4	2,053,384.6	−887,085.0	187,428.5	2,078,545.5	2,078,546	4
Total	. . .	. . .	. . .	. . .	. . .	10,295,701	. . .

Note: Total of the "Rounded row sum" column may be slightly different from the actual value of the age group being interpolated because of rounding

Table 12.12 Disaggregating the 5–9 age group to single years of age: Indonesian males, 2000 (using multipliers in Table 12.7)

Notation	N_{-1}	N_0	N_1	N_2			
Age group	0–4	5–9	10–14	15–19			
Population	10,295,701	10,433,865	10,460,908	10,649,348	Row sum	Rounded row sum	Age in single years
$n_0=$	345,935.6	2,370,574.1	−786,660.3	153,350.6	2,083,200.0	2,083,200	5
$n_1=$	82,365.6	2,420,656.7	−502,123.6	85,194.8	2,086,093.5	2,086,094	6
$n_2=$	−82,365.6	2,253,714.8	−83,687.3	0.0	2,087,661.9	2,087,662	7
$n_3=$	−164,731.2	1,919,831.2	418,436.3	−85,194.8	2,088,341.5	2,088,342	8
$n_4=$	−181,204.3	1,469,088.2	954,034.8	−153,350.6	2,088,568.1	2,088,568	9
Total	...	...	...	...	...	10,433,865	...

Note: Total of the "Rounded row sum" column may be slightly different from the actual value of the age group being interpolated because of rounding

Table 12.13 Disaggregating the 10–14 age group to single years of age: Indonesian males, 2000 (using multipliers in Table 12.8)

Notation	N_{-2}	N_{-1}	N_0	N_1	N_2			
Age group	0–4	5–9	10–14	15–19	20–24			
Population	10,295,701	10,433,865	10,460,908	10,649,348	9,237,464	Row sum	Rounded row sum	Age in single years
$n_0=$	−131,785.0	884,791.8	1,573,320.6	−255,584.4	14,779.9	2,085,522.9	2,085,523	10
$n_1=$	−16,473.1	150,247.7	2,326,505.9	−443,012.9	59,119.8	2,076,387.4	2,076,387	11
$n_2=$	65,892.5	−350,577.9	2,661,255.0	−357,818.1	59,119.8	2,077,871.3	2,077,871	12
$n_3=$	65,892.5	−434,048.8	2,326,505.9	153,350.6	−14,779.9	2,096,920.3	2,096,920	13
$n_4=$	16,473.1	−250,412.8	1,573,320.6	903,064.7	−118,239.5	2,124,206.1	2,124,206	14
Total	…	…	…	…	…	…	10,460,908	…

Note: Total of the "Rounded row sum" column may be slightly different from the actual value of the age group being interpolated because of rounding

Table 12.14 Disaggregating the 15–19 age group to single years of age: Indonesian males, 2000 (using multipliers in Table 12.8)

Notation	N_{-2}	N_{-1}	N_0	N_1	N_2			
Age group	5–9	10–14	15–19	20–24	25–29			
Population	10,433,865	10,460,908	10,649,348	9,237,464	10,649,348	Row sum	Rounded row sum	Age in single years
$n_0=$	−133,553.5	887,085.0	1,601,661.9	−221,699.1	17,039.0	2,150,533.3	2,150,533	15
$n_1=$	−16,694.2	150,637.1	2,368,415.0	−384,278.5	68,155.8	2,186,235.2	2,186,235	16
$n_2=$	66,776.7	−351,486.5	2,709,194.1	−310,378.8	68,155.8	2,182,261.3	2,182,261	17
$n_3=$	66,776.7	−435,173.8	2,368,415.0	133,019.5	−17,039.0	2,115,998.4	2,115,998	18
$n_4=$	16,694.2	−251,061.8	1,601,661.9	783,336.9	−136,311.7	2,014,319.5	2,014,320	19
Total	...	...	...	...	...	...	10,649,348	...

Note: Total of the "Rounded row sum" column may be slightly different from the actual value of the age group being interpolated because of rounding

Table 12.15 Disaggregating the 60–64 age group to single years of age: Indonesian males, 2000 (using multipliers in Table 12.9)

Notation	N_{-2}	N_{-1}	N_0	N_1			
Age group	50–54	55–59	60–64	65–69			
Population	3,791,185	2,883,226	2,597,076	1,666,191	Row sum	Rounded row sum	Age in single years
n_0=	−54,593.1	262,950.2	365,668.3	−29,325.0	544,700.4	544,700	60
n_1=	−30,329.5	115,329.0	477,862.0	−26,659.1	536,202.4	536,202	61
n_2=	0.0	−23,065.8	560,968.4	−13,329.5	524,573.1	524,573	62
n_3=	30,329.5	−138,394.8	602,521.6	13,329.5	507,785.8	507,786	63
n_4=	54,593.1	−216,818.6	590,055.7	55,984.0	483,814.2	483,814	64
Total	...	...	...	...	...	2,597,076	...

Note: Total of the "Rounded row sum" column may be slightly different from the actual value of the age group being interpolated because of rounding

Table 12.16 Disaggregating the 65–69 age group to single years of age: Indonesian males, 2000 (using multipliers in Table 12.10)

Notation	N_{-3}	N_{-2}	N_{-1}	N_0			
Age group	50–54	55–59	60–64	65–69			
Population	3,791,185	2,883,226	2,597,076	1,666,191	Row sum	Rounded row sum	Age in single years
n_0=	66,724.9	−244,497.6	511,104.6	117,299.8	450,631.7	450,632	65
n_1=	60,659.0	−207,592.3	353,202.3	199,942.9	406,211.9	406,212	66
n_2=	30,329.5	−92,263.2	103,883.0	306,579.1	348,528.4	348,528	67
n_3=	−30,329.5	115,329.0	−249,319.3	439,874.4	275,554.6	275,555	68
n_4=	−127,383.8	429,024.0	−718,870.6	602,494.7	185,264.3	185,264	69
Total	...	...	...	...	...	1,666,191	...

Note: Total of the "Rounded row sum" column may be slightly different from the actual value of the age group being interpolated because of rounding

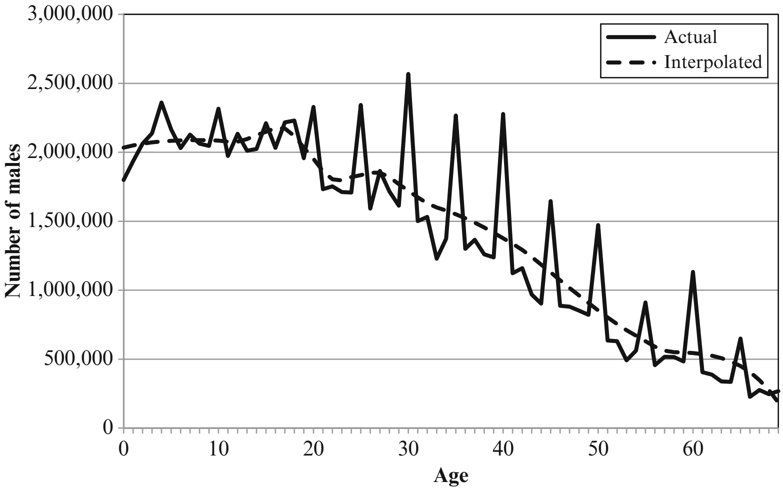


Fig. 12.3 Actual number of males by age and interpolated numbers using Sprague's multipliers: Indonesia, 2000 census (*Sources*: United Nations (2011) for the actual data. Tables 12.11 to 12.14 for interpolated values for individual ages 0–19, Tables 12.15 and 12.16 for individual ages 60–69. Estimates for ages 20–59 were calculated by authors using the multipliers given in Table 12.8), and following the procedure illustrated in Table 12.14

In disaggregating rates in 5-year age groups, they need to be multiplied by 5 before using the rates as N_0 to N_4.

Some of the limitations of Sprague's multipliers as enunciated by Jaffe (1951) are given in the following paragraphs.

If any particular 5-year age group is greatly in error due to under- or over-enumeration, this method will not correct such deficiencies; they must be corrected by graphic interpolation or by other methods such as calculating the expected number of survivors from the preceding census.

If the original data curves very rapidly from one single year to the next, as is the case with the proportion of single marital status during the teen ages, the multipliers are somewhat inadequate. In such cases graphic smoothing may be the best. Thus, for example, in smoothing the proportion single for each year of age from age 15 to age 100, it may be best to use graphic smoothing for the age group 15–19 and for the older age groups. The specific older age groups to which these multipliers should be applied will depend on how *regular* or *irregular* the data appear to be. These *irregularities* may represent respondents' biases which should be smoothed out by actually altering the reported numbers or percentages for the given 5-year age groups. For all intermediate 5-year age groups the multipliers should be adequate.

The Sprague's multipliers may also be inadequate for the very youngest age groups such as the under five, and 5–9 year olds. If birth and death statistics are available, it is preferable to calculate these ages by single years on the basis of the vital statistics.

12.7 Other Methods

As the examples in this chapter suggest, there are many ways to smooth and adjust demographic data, whether in the form of numbers or rates. The ones presented here are relatively straightforward and, as such, suitable for this book. Others exist that are important tools but they are beyond the scope of this book. One important tool is called *spline* interpolation (see for example: Currie et al. 2004; McNeil et al. 1977). Advanced methods, including splines, for smoothing rates are discussed by many authors such as De Beer (2011), and Oullette and Bourbeau (2011).

References

Australia. (2012). *Census of population and housing: Details of undercount, 2011*. Catalogue No. 2940.0. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/ausstats/abs@.nsf/mf/2940.0. Accessed Feb 2013.

Currie, I. D., Durban, M., & Eilers, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling, 4*(4), 279–298.

De Beer, J. (2011). A new relational method for smoothing and projecting age specific fertility rates: TOPALS. *Demographic Research, 24*(18), 409–454. http://www.demographic-research.org/volumes/vol27/20/27-20.pdf. Accessed Feb 2013.

Jaffe, A. J. (1951). *Handbook of statistical methods for demographers*. Washington, DC: Government Printing Office.

McNeil, D. R., Trussell, T. J., & Turner, J. C. (1977). Spline interpolation of demographic data. *Demography, 14*(2), 245–252.

Myers, R. J. (1940). Errors and bias in the reporting of ages in census data. *Transactions of the Actuarial Society of America, 41*(104), 395–415.

Oullette, N. & R. Bourbeau. (2011). Changes in the age-at-death distribution in four low mortality countries: A Non parametric approach. *Demographic Research*, *25*(19), 595–628. http://www.demographic-research.org/volumes/vol25/19/25-19.pdf. Accessed Feb 2013.

Spiegelman, M. (1969). *Introduction to demography*. Cambridge: Harvard University Press.

Triola, M. F. (2007). *Elementary statistics using excel* (3rd ed.). Boston: Pearson Addison Wesley.

United Nations. (1955). *Manual II. Methods of appraisal of quality of basic data for population estimates*. New York: Department of Economic and Social Affairs. http://www.un.org/esa/population/techcoop/DemEst/manual2/manual2.html. Accessed Feb 2013.

United Nations. (2011). *Demographic yearbook special census topics*. New York: Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/products/dyb/dybcens.htm. Accessed Feb 2013.

Yusuf, F. (1967). On the extent of digital preference in reporting of ages in Pakistan. *Pakistan Development Review, 4*, 519–532.

Chapter 13
The Stable Population Model

13.1 Purpose

The purpose of this chapter is to extend the concept of stationary population introduced in Chap. 7 to a more flexible and versatile model generally known as the *stable population model*. Some important properties and characteristics of this model are examined and its use in demographic analysis is discussed.

13.2 The Model

The stationary population model assumes that the following conditions prevail in a population over a long period of time:

- no migration in or out of the population
- constant age-specific mortality rates, and
- same number of births occurs every year.

Under these conditions, the age distribution of the population would be represented by the L_x column of its complete life table or the ${}_nL_x$ column of its abridged life table. Such a population would have a growth rate of zero.

A stable population is a population with an invariable relative age structure and a constant rate of growth (Lotka 1907; Sharpe & Lotka 1911; Dublin & Lotka 1925). That is, the proportion of people in each age group remains constant over time. When the absolute number of people in each group is also constant over time, the population becomes a stationary population, which is a special case of a stable population in which the growth rate is zero.

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_13,

13.3 Age Distribution of a Stable Population

It is assumed that a female population experienced over a long period of time, in the past, the following conditions:

- population was closed to migration
- experienced constant mortality levels, thus an unchanging life table
- births increased annually at a constant rate (r).

In such a population, b_{2010-x} is the number of births x years prior to a particular year, say 2010, and the survivors of these births at the end of 2010 (*i.e.*, females aged x, $F_x^{beginning2010}$) would be:

$$F_x^{beginning2010} = b_{2010-x} * \frac{l_x}{l_0} \tag{13.1}$$

Since births were increasing annually at a constant rate (r), knowing the number of births in any year, again say 2010, would enable the estimation of births in a previous year (say, 2000) using Eq. (4.7). Accordingly,

$$\begin{aligned} b_{2010} &= b_{2000} * e^{10r} \qquad \text{or} \\ b_{2000} &= \frac{b_{2010}}{e^{10r}} = b_{2010} * e^{-10r} \end{aligned} \tag{13.2}$$

Equation (13.2) can be generalised as

$$b_{2010-x} = b_{2010} * e^{-rx} \tag{13.3}$$

Substituting in Eq. (13.1)

$$F_x^{beginning2010} = b_{2010} * e^{-rx} * \frac{l_x}{l_0} \tag{13.4}$$

Two points are worth noting:

- the size of each birth cohort increases at a constant rate (r) and since the mortality levels do not change, each cohort of the population and consequently the whole population also increases at a constant rate (r)
- the proportion of the population at a given age x ($f_x^{beginning2010}$), where x varies from zero to the highest attainable age (ω) is:

$$f_x^{beginning2010} = \frac{b_{2010} * e^{-rx} * \frac{l_x}{l_0}}{\sum_{x=0}^{x=\omega} b_{2010} * e^{-rx} * \frac{l_x}{l_0}} = \frac{b_{2010} * e^{-rx} * \frac{l_x}{l_0}}{b_{2010} * \sum_{x=0}^{x=\omega} e^{-rx} * \frac{l_x}{l_0}} \quad \text{or}$$

$$f_x^{beginning2010} = \frac{e^{-rx} * \frac{l_x}{l_0}}{\sum_{x=0}^{x=\omega} e^{-rx} * \frac{l_x}{l_0}} \tag{13.5}$$

Thus, the proportion of population at any age remains constant and is a function of the value of (r) and the constant mortality rates. It does not depend on the number of births in any particular year.

Further, to calculate the midyear population, which is usually the denominator for most demographic rates, Eq. (13.4) is re-written as:

$$F_x^{mid2010} = b_{2010} * e^{-r(x+0.5)} * \frac{L_x}{l_0} \tag{13.6}$$

where $F_x^{midyear}$ is the midyear population at age x. Similarly, the proportion of midyear population at age x ($f_x^{midyear}$) is:

$$f_x^{mid2010} = \frac{e^{-r(x+0.5)} * \frac{L_x}{l_0}}{\sum_{x=0}^{x=\omega} e^{-r(x+0.5)} * \frac{L_x}{l_0}} \tag{13.7}$$

The x year olds would be $x + 0.5$ years of age in midyear and the l_x is replaced by L_x being the average life table population aged $x + 0.5$.

When the data are given in 5-year age groups, Eqs. (13.6) and (13.7) are re-written as:

$${}_5F_x^{mid2010} = b_{2010} * e^{-r(x+2.5)} * \frac{{}_5L_x}{l_0} \tag{13.6a}$$

$${}_5f_x^{mid2010} = \frac{e^{-r(x+2.5)} * \frac{{}_5L_x}{l_0}}{\sum_{x=0}^{x=\omega} e^{-r(x+2.5)} * \frac{{}_5L_x}{l_0}} \tag{13.7a}$$

13.4 Estimation of (*r*) in a Stable Population

Because a stable population is closed to migration, its rate of growth (r) is its rate of natural increase and is one of the most important determinants of a stable population. It is sometimes also referred to as the *intrinsic rate of natural increase* (*IRNI*). One of the simplest ways to estimate *IRNI* is by using the net reproduction rate *(NRR)* and the mean length of generation (*MLG*) of a female population (see Table 5.4 for the method of their estimation).

Since the *NRR* represents the ratio between the size of the daughters' generation (G_d) and the mothers' generation (G_m):

$$NRR \cong \frac{G_d}{G_m} \quad (\cong \text{ means approximately equal to}) \tag{13.8}$$

As discussed in Sect. 5.5.2, *MLG* is the average age difference between the daughters and their mothers. Following Eq. (4.8):

$$NRR \cong e^{r*MLG} \tag{13.9}$$

Taking the natural logarithms of both sides

$$LN(NRR) \cong r * MLG \; (\text{since the } LN(e) = 1).$$

$$r \cong \frac{LN(NRR)}{MLG} \tag{13.10}$$

The above equation can be re-written as:

$$r \cong \frac{LN(R_0)}{\frac{R_1}{R_0}} \tag{13.11}$$

The terms R_0 and R_1 were described in Eqs. (5.7) and (5.8) and their computational details were given in Table 5.4.

A more precise estimate involves, first, the calculation of an additional term, R_2, as defined below (using the symbols used in Sect. 5.5):

$$R_2 = \sum_{i=15}^{i=44} \{[f^i_{t\to t+1} * S * PS^{0\to i+0.5}] * (i + 0.5)^2\} \tag{13.12}$$

if data are given in single years of age, or alternately

$$R_2 = n * \sum_{i=15}^{i=44} \{[f_{t \to t+1}^{i \to i+n} * S * PS^{0 \to i+0.5n}] * (i + 0.5n)^2\} \tag{13.12a}$$

when only the age group data are available.

Having estimated R_0, R_1 and R_2, and solving the following quadratic equation for r would result in a more precise estimate of r (Dublin & Lotka 1925):

$$\frac{r^2}{2} * [(\frac{R_1}{R_0})^2 - (\frac{R_2}{R_0})] - r * \frac{R_1}{R_0} - LN(R_0) = 0 \tag{13.13}$$

The solution of the quadratic Eq. (13.13) is:

$$r = \frac{\frac{R_1}{R_0} - \sqrt{\{(\frac{R_1}{R_0})^2 + 2 * [(\frac{R_1}{R_0})^2 - (\frac{R_2}{R_0})] * LN(R_0)\}}}{[(\frac{R_1}{R_0})^2 - (\frac{R_2}{R_0})]} \tag{13.14}$$

The preceding equations are taken from a "Taylor series," that has more terms than are shown in the equations given here. Because the additional terms are very small, they are omitted, which is the usual procedure for dealing with them (Dublin & Lotka 1925; McCann 1973).

Coale (1957) suggested another method that does not involve the calculation of R_2 but still gives quite robust estimates of r:

$$r \cong \frac{LN(R_0)}{\frac{R_1}{R_0} - 0.7 * LN(R_0)} \tag{13.15}$$

Australian data given in Table 5.4 were used to apply the three methods of calculating r described above. Values of R_0 and R_1 respectively 0.913870 and 27.748615 were taken from Table 5.4, and R_2 was estimated as 874.244145 being five times the sum of the product of columns (10) and (7) from the same table. The three estimates of r are:

Simple method	Eq. (13.11)	−0.002966
Dublin & Lotka's method	Eq. (13.14)	−0.002960
Coale's method	Eq. (13.15)	−0.002959

Since all three methods give very similar results, the authors have adopted the simple method, Eq. (13.11), in the rest of this chapter.

13.5 Estimation of the Birth and Death Rates in a Stable Population

Following Eq. (5.2), the crude birth rate, referred here as the *stable birth rate* (*SBR*), for a given year, say 2010, is the ratio of births in 2010 divided by the midyear population:

$$SBR = \frac{b_{2010}}{\sum_{x=0}^{x=\omega} \left(b_{2010} * e^{-r(x+0.5)} * \frac{L_x}{l_0}\right)} \tag{13.16}$$

The term within brackets in the denominator of Eq. (13.16) represents the midyear population in 2010 aged x, for all ages from zero to the highest attainable age, ω. Since b_{2010} is a constant, it can be taken out of the summation sign, which will result in it being cancelled with the b_{2010} in the numerator. Equation (13.16) then becomes:

$$SBR = \frac{1}{\sum_{x=0}^{x=\omega} e^{-r(x+0.5)} * \frac{L_x}{l_0}} \tag{13.17}$$

Obviously, the stable death rate (*SDR*) is calculated as: *SBR* – r, because by definition r is the difference between the birth and death rates (natural increase rate), when the population is closed to migration.

13.6 Fitting the Stable Population Model to Australia and Indonesia

There are three steps involved in fitting the stable population model. They are: (1) estimating the value of r, (2) calculating the age distribution of the stable population, and (3) determining the birth and death rates for the stable population.

In *step 1*, the value of r has already been estimated for Australia as −0.002965 in section (13.4). Table 13.1 illustrates the estimation of r for Indonesia. To convert the age-specific fertility rates in column (2) from all births to those for daughters only column (3), the sex ratio at birth is assumed to be 105 male births per 100 female births. This gave a multiplier of 0.487805 as the proportion of total births that were female. Column (6) is the estimated probability of survival from birth to the mean age of the mother and column (7) is the adjusted fertility rates in column (3) for female survival. In column (8), the adjusted mean age of women in each age group is computed multiplying column (4) by column (7). Values of R_0 and R_1 are estimated along with the mean length of generation of 28.022757, being the ratio R_1 divided by R_0. Using Eq. (13.11) the value of r for Indonesia is estimated to be 0.004617.

Step 2 involves the calculation of the proportionate stable populations for Australia and Indonesia (Table 13.2). Column (1) and (2) give the age groups and

Table 13.1 Calculation of the r for Indonesia

Age group	Fertility rates per woman: All births	Fertility rates per woman: Female births	Mean age of woman	${}_5L_x$ from female life table	Probability of survival	Mortality adjusted fertility rate	Adjusted mean age of women
(1)	(2)	(3)	(4)	(5)	(6) = (5)/ 500,000	(7) = (3) * (6)	(8) = (4) * (7)
15–19	0.052300	0.025512	17.5	469,922	0.939844	0.023977	0.419598
20–24	0.133100	0.064927	22.5	465,494	0.930988	0.060446	1.360035
25–29	0.138200	0.067415	27.5	459,428	0.918856	0.061945	1.703488
30–34	0.098500	0.048049	32.5	452,224	0.904448	0.043458	1.412385
35–39	0.061800	0.030146	37.5	444,167	0.888334	0.026780	1.004250
40–44	0.021000	0.010244	42.5	433,596	0.867192	0.008884	0.377570
45–49	0.005200	0.002537	47.5	420,240	0.840480	0.002132	0.101270
15–49	0.510100	0.248830	...	...	...	0.227622	6.378596
*5	2.550500	1.244150	...	...	...	1.138110	31.892980
	= *TFR*	= *GRR*	...	...	...	$= R_o$	$= R_1$

Following Eq. (13.11), the value of $r = \dfrac{LN(1.138110)}{\dfrac{31.892980}{1.138110}} = \dfrac{0.129369}{28.022757} = 0.004617$

Sources: United Nations (2011) for column (2). While accessing the UN database, select Indonesia and 1995–2000 from the drop down menu and click on Apply Filters. Indonesia (2002) for column (5)
Note: Fertility rates are the average for 1995–2000 and the life table is for 2002. Estimation method of the probability of survival in this table is an alternative to that used in Table 5.4 when complete life table is not available. Both methods give approximately the same results

mid-points for each age group. Columns (3) and (4) give the ${}_5L_x$ values copied from the female life tables of the two countries. Taking the values of $r = -0.002966$ for Australia and 0.004617 for Indonesia, $e^{-r*(x+2.5)}$ values are calculated in columns (5) and (6) for each age group in Australia and Indonesia. Multiplying column (3) by column (5) and column (4) by column (6) results in the stable age distributions for Australian and Indonesian females. These are then converted into percentage distributions of the two stable populations. The procedure followed in this table is based on Eq. (13.6a). The only difference is that since l_0 is a constant, being 100,000 births for both life tables, it is ignored in estimating columns (7) and (8) and it has no effect on the proportionate age distributions.

In *step 3* the number of female births in Australia and Indonesia is assumed to be 100,000 each, being the radix of the life tables on which the stable populations for these countries are based. Dividing 100,000 by the total of columns (7) and (8) results in 0.010528 and 0.018084 as the annual female stable birth rates, or 10.53 and 18.08 per 1,000, for Australia and Indonesia respectively. Since the values of r of these two countries are −0.002966 and 0.004617, their annual stable death rates for females are estimated as 0.013494 and 0.013467, or 13.49 and 13.47 per 1,000. While there is not much difference in the stable death rates of the two countries, the life expectancy at birth for Australian females was 18.8 years higher than their Indonesian counterparts.

13.7 Some Important Characteristics of the Stable Population Model

So far it has been shown that a stable population, under the conditions of no migration and constant age-specific fertility and mortality rates, may increase or decrease in total size at a constant rate. This rate applies not only to the total population but also to each age group, and consequently its proportionate age distribution becomes constant.

There are two other properties of this model that are worth noting. These properties apply to any number of stable populations with different fertility and mortality schedules.

13.7.1 Relative Impact of Fertility and Mortality in Determining the Shape of a Stable Age Distribution

Two female stable populations such as those for Australia and Indonesia shown in Table 13.2, columns (9) and (10) have different age distributions. What would happen if their fertility and mortality patterns are exchanged as indicated in Table 13.3?

Figure 13.1 shows the female stable populations based on the four scenarios in Table 13.3. The stable populations labelled as A and C are taken from Table 13.2, columns (9) and (10). Stable populations B and D are derived by using the same procedure as outlined in Table 13.2 but reversing the life tables and using the appropriate values of r. Thus, in stable population B the Indonesian life table is used with the value of r as -0.007169 and for stable population D the r is taken as 0.007391 with the Australian life table.

It is apparent from Fig. 13.1 that for a given fertility level in a stable population, its proportionate age distribution is not much affected whether it is subjected to higher or lower mortality. On the other hand, for a given mortality level in a stable population, its proportionate age distribution is substantially affected whether it is subjected to higher or lower fertility. This leads to the *quasi-stable* population models that need not assume constancy of the age specific mortality rates. This property has been useful in developing many indirect methods of estimation for countries where mortality may be declining without substantial changes in fertility levels.

13.7.2 Age Distributions of Populations Subjected to Constant Fertility and Mortality Rates

In the case of two populations with different proportionate age distributions, if it is assumed that for the next 100 years both populations are subjected to a given set of

Table 13.2 Calculation of the proportionate stable age distributions for Australia and Indonesia

Age group	Mid-point	${}_5L_x$ from female life tables		$e^{-r*(x+2.5)}$		$e^{-r*(x+2.5)} *{}_5L_x$		% Stable population	
$(x- < x + 5)$	$(x + 2.5)$	Australia	Indonesia	Australia[a]	Indonesia[b]	Australia	Indonesia	Australia	Indonesia
(1)	(2)	(3)	(4)	(5)	(6)	(7) = (3) * (5)	(8) = (4) * (6)	(9)	(10)
0–4	2.5	497,860	512,657	1.007443	0.988524	501,565	506,774	5.28	9.16
5–9	7.5	497,393	476,969	1.022494	0.965965	508,581	460,735	5.35	8.33
10–14	12.5	497,168	473,349	1.037771	0.943921	515,946	446,804	5.43	8.08
15–19	17.5	496,745	469,922	1.053276	0.922380	523,209	433,447	5.51	7.84
20–24	22.5	496,063	465,494	1.069012	0.901331	530,297	419,564	5.58	7.59
25–29	27.5	495,268	459,428	1.084984	0.880762	537,358	404,647	5.66	7.32
30–34	32.5	494,313	452,224	1.101194	0.860663	544,334	389,212	5.73	7.04
35–39	37.5	493,003	444,167	1.117646	0.841022	551,003	373,554	5.80	6.76
40–44	42.5	491,045	433,596	1.134345	0.821829	557,014	356,342	5.86	6.44
45–49	47.5	488,048	420,240	1.151292	0.803075	561,886	337,484	5.92	6.10
50–54	52.5	483,558	400,745	1.168493	0.784748	565,034	314,484	5.95	5.69
55–59	57.5	477,020	372,541	1.185951	0.766840	565,722	285,679	5.96	5.17
60–64	62.5	466,805	332,343	1.203670	0.749340	561,879	249,038	5.92	4.50
65–69	67.5	450,995	282,745	1.221653	0.732240	550,959	207,037	5.80	3.74
70–74	72.5	426,253	223,847	1.239905	0.715529	528,513	160,169	5.56	2.90
75+	77.5	1,108,571	264,261	1.258430	0.699201	1,395,059	184,772	14.69	3.34
Total		8,360,108	6,484,528	…	…	9,498,359	5,529,742	100.00	100.00

Sources: Table A7.5 for column (3) and Indonesia (2002) for column (4)

Notes: The values of e_0 for females in Australia and Indonesia were calculated as 83.6 and 64.8 years respectively

These were arrived at by dividing the total of columns (3) and (4) by 100,000

[a] Section 13.6 provided the r for Australia (−0.002966)

[b] The source for Indonesian r (0.0046717) was Table 13.1

Table 13.3 Four scenarios for the stable populations of Australia and Indonesia

Scenario	Fertility of	Mortality of	r
A	Australia	Australia	−0.002966
B	Australia	Indonesia	−0.005759
C	Indonesia	Indonesia	0.004617
D	Indonesia	Australia	0.007400

Note: The values of r for scenarios A and C are as shown in the Notes for Table 13.2
The r for scenario B is estimated by substituting the values in column (8) of Table 5.4 by those in column (6) of Table 13.1 and recalculating R_0 and R_1 and hence the r. Similarly, the r for scenario D is estimated by replacing column (6) of Table 13.1 by column (8) of Table 5.4. The probability of survival for Australian females from birth to 45–49 is estimated as 0.976096

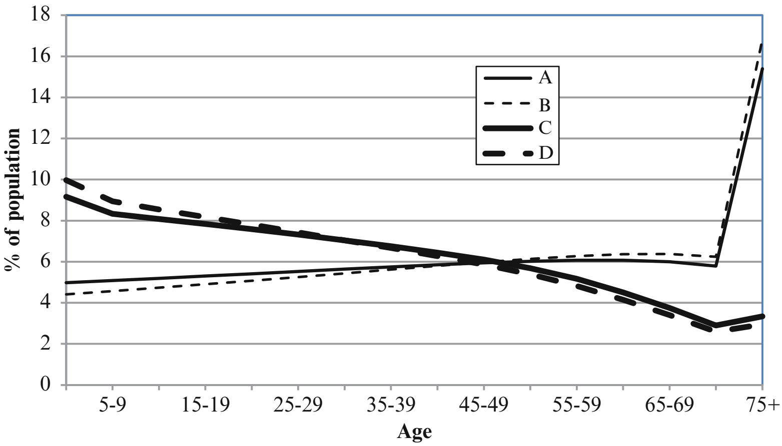


Fig. 13.1 Female stable populations (%) under the four scenarios

fertility and mortality rates, of either country or some other, would the differences in the proportionate age distribution persist over the long run? In this section an empirical example is presented to give an answer to this question.

Table 13.4 presents the age data for Australia and Indonesia in 2000, the $_5L_x$ column from the 2006 Australian female life and the 2006 age-specific fertility rates for Australia. The cohort component method, Sect. 11.4, was used to project the female populations of both countries from 2000 to 2100, under the assumption that both populations would be closed to migration during the projection period and the fertility and mortality levels of Australia, shown in Table 13.4 columns (5) and (4), would also remain constant throughout the projection period and would apply to both countries.

Table 13.4 Input data for projections of the population of Australia and Indonesia, assuming no migration and constant fertility and mortality levels

	2000 female population		Australia 2006	
Age group	Australia	Indonesia	${}_5L_x$ for females	Fertility rates (female births)
(1)	(2)	(3)	(4)	(5)
0–4	623,100	8,607,111	497,860	. . .
5–9	657,321	9,941,780	497,393	. . .
10–14	652,475	10,103,622	497,168	. . .
15–19	649,402	10,160,283	496,745	0.007512
20–24	635,881	8,871,160	496,063	0.025171
25–29	727,009	8,955,603	495,268	0.049171
30–34	718,323	7,792,422	494,313	0.058585
35–39	756,421	7,715,167	493,003	0.030878
40–44	728,900	5,973,181	491,045	0.005512
45–49	674,128	4,830,852	488,048	0.000293
50–54	623,134	4,026,984	483,558	. . .
55–59	473,483	3,237,430	477,020	. . .
60–64	396,853	2,895,687	466,805	. . .
65–69	345,081	2,009,169	450,995	. . .
70–74	333,643	1,278,851	426,253	. . .
75+	652,895	1,206,891	1,108,571	. . .
Total	9,648,049	97,606,193	. . .	. . .

Sources: Australia (2010) for column (2), Indonesia (2002) for column (3), Table A7.5 for column (4) and Australia (2007) for column (5). The sex ratio at birth is assumed to be 105 males per 100 females in both countries

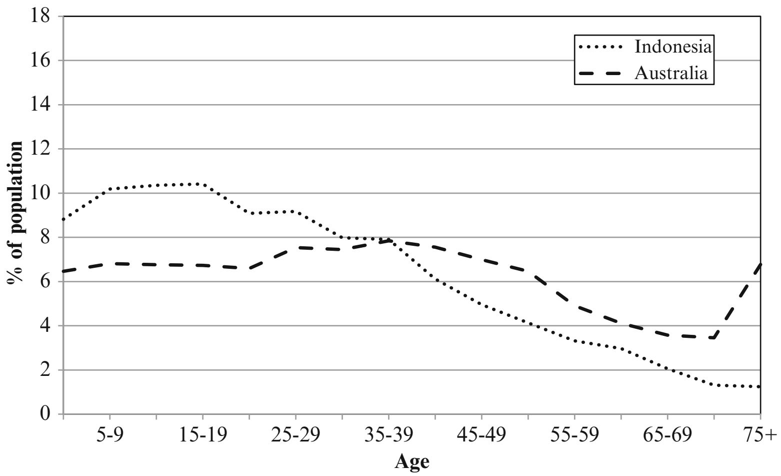


Fig. 13.2 Actual female populations (%) of Australia and Indonesia: 2000

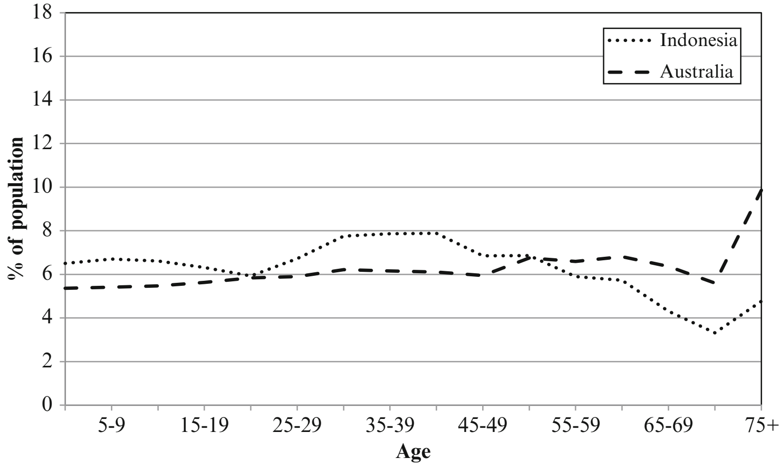


Fig. 13.3 Projected female populations (%) of Australia and Indonesia: 2025

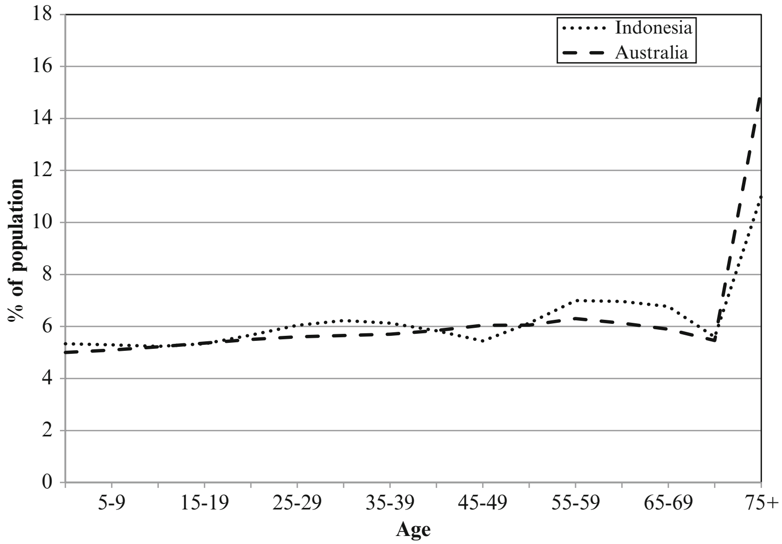


Fig. 13.4 Projected female populations (%) of Australia and Indonesia: 2050

Figures 13.2, 13.3, 13.4, 13.5 and 13.6 show the female populations (percentages) for both countries in the year 2000, and projections for 2025, 2050, 2075 and 2100. It is obvious from this projection that by 2100 the proportionate age distributions of Australia and Indonesia would be identical. This does not mean that

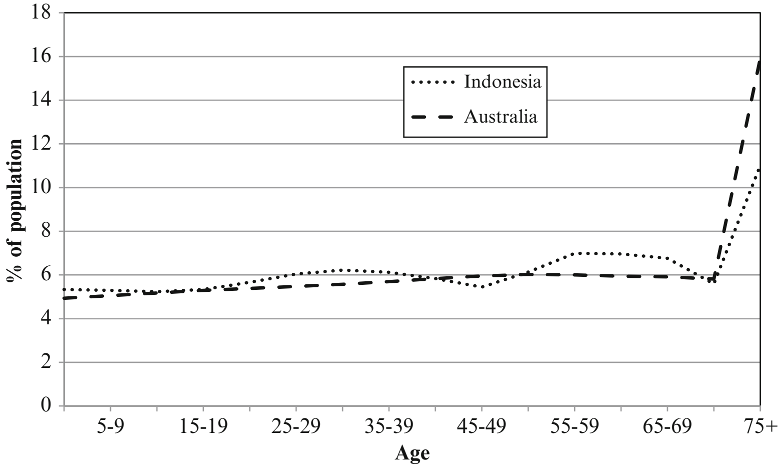


Fig. 13.5 Projected female populations (%) of Australia and Indonesia: 2075

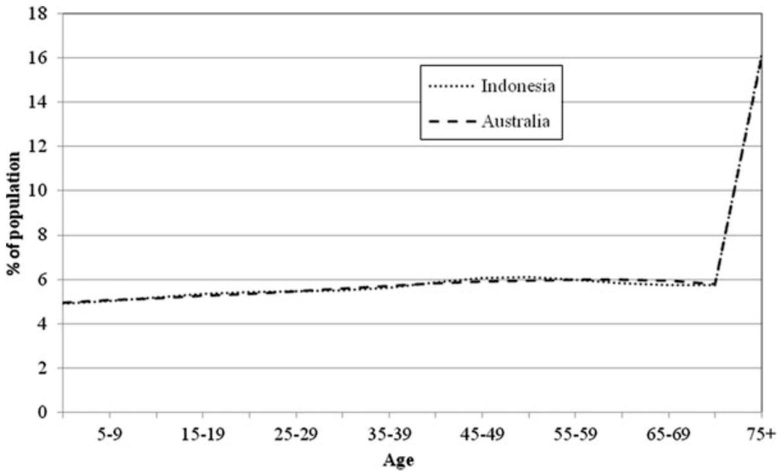


Fig. 13.6 Projected female populations (%) of Australia and Indonesia: 2100

both countries would be identical in actual size. The same conclusion could be drawn if this exercise was repeated by projecting the population of both countries using the Indonesian fertility and mortality rates.

This property of a human population to *forget* its past age distribution, when it is subjected to constant fertility and mortality rates, is known as *ergodicity* (Coale 1972; Preston et al. 2001).

13.8 Stable Population Models for Males

So far the discussion of stable populations has been limited to females. The methods described for female populations are equally applicable to male populations. However, one of the main problems is in the estimation of the intrinsic rate of natural increase for the male populations. Analysis of male fertility becomes difficult for two main reasons: (1) their reproductive period is much longer than and not as well-defined as that of females, and (2) information about the father may not always be available for a certain proportion of births. Fertility analysis of males and females generally results in different results – the former generally yield higher results for various fertility indictors compared to the latter. The idea of males in a population reproducing at a faster rate than the females is untenable unless the population practices polygamy. While several attempts have been made unsuccessfully to resolve this dilemma, it is an agreed convention to apply the female intrinsic rates of natural increase to males too (Pollard et al. 1995).

Table 13.5 illustrates the estimation of the stable population and the stable birth and death rates for males using the Indonesian life table for males and the value of r for females. The stable birth rate for Indonesian males is calculated as $\frac{100,000}{5,254,653}$ = 0.019031 or 19.03 per 1,000 and since the value of r was 0.004617, the stable death rate is 0.014414 or 14.41 per 1,000. The corresponding figures for females, estimated in Sect. 13.6, were 18.08 and 13.45 respectively.

Table 13.5 Calculation of the proportionate stable age distribution for Indonesian males

Age group $x- < (x+5)$	$x + 2.5$	${}_5L_x$	$e^{-r*(x+2.5)}$	$e^{-r*(x+2.5)} * {}_5L_x$	Stable population (%)
(1)	(2)	(3)	(4)	(5) = (3 * 4)	(6)
0–4	2.5	510,434	0.988524	504,576	9.60
5–9	7.5	474,127	0.965965	457,990	8.72
10–14	12.5	470,428	0.943921	444,047	8.45
15–19	17.5	465,675	0.922380	429,529	8.17
20–24	22.5	460,088	0.901331	414,692	7.89
25–29	27.5	453,375	0.880762	399,315	7.60
30–34	32.5	444,864	0.860663	382,878	7.29
35–39	37.5	436,106	0.841022	366,775	6.98
40–44	42.5	425,348	0.821829	349,563	6.65
45–49	47.5	407,190	0.803075	327,004	6.22
50–54	52.5	379,620	0.784748	297,906	5.67
55–59	57.5	346,151	0.766840	265,442	5.05
60–64	62.5	296,027	0.749340	221,825	4.22
65–69	67.5	230,980	0.732240	169,133	3.22
70–74	72.5	160,579	0.715529	114,899	2.19
75+	77.5	156,005	0.699201	109,079	2.08
Total	...	6,116,997	...	5,254,653	100.00
$e_0 = \frac{6,116,997}{100,000} = 61.2$					

Source: Indonesia (2002) for column (3)
Notes: r = 0.004617, see Table 13.1

13.9 Model Life Tables and Stable Populations

Empirical relationships between the q_x values for one age group and the next higher age group formed the basis for a set of model life tables that are useful particularly for those countries where accurate demographic data are not available.

The United Nation (1955) using life tables from 158 (mostly European) countries found empirical relationships between the mortality rates (q_x) for one age group and the next higher age group (q_{x+n}). Quadratic (parabolic) models of the type:

$$_5q_x = \alpha_i + \beta_i * ({}_5q_{x-5}) + \gamma_i * ({}_5q_{x-5})^2 \tag{13.18}$$

were developed for all age groups. It follows that knowing the value of one mortality indicator, say $_1q_0$, the remaining q_x values could be obtained by substituting the different values for α, β, and γ in Eq. (13.18). Accordingly, knowing one mortality index the whole life table can be determined.

Coale & Demeny (1966) analysed data from 192 life tables, including some from selected non-European countries, and found four distinct age patterns of mortality that were determined not only by geographic location but also by the pattern of their deviations from previously estimated regression equations. They were designated as North, South, East and West. Each regional pattern had distinguishing mortality characteristics. Twenty-four model life tables were calculated for females at mortality levels designated as (1) to (24) in a way that each increase in mortality level added 2.5 years to the life expectancy. At level (1) the female life expectancy was 20 years and at level 24 it was 77.5 years. Similarly, 24 model life tables were computed for males.

An excerpt from model life tables for females at mortality levels (9) and (11) is presented in Table 13.6. Each life table has the usual columns, as described in Chap. 7, as well as an extra column giving the proportion surviving from one age to the next age, shown as columns (6) for mortality level (9) and column (13) for mortality level (11).

Given that an infinite number of stable populations can be calculated associated with the same life table but different values of r, Coale and Demeny computed a set of 13 stable populations at each mortality level using the values of r ranging from −0.010 to +0.050 at 0.005 intervals. Table 13.7 gives an excerpt from the stable populations based on mortality levels (9) and (11) associated with six selected values of r. The first panel in the table presents the proportionate distribution by age group, the middle panel gives the proportion under age 1, 5, 10, . . . , 65, and the last panel gives selected parameters for each stable population.

Subsequently, Coale and Demeny revised and extended their analysis by including more recently available mortality data and increasing the number of mortality levels to 25, and extending the maximum value for females life expectancy to 80 years and males to 76.65 years (Coale & Demeny 1983).

Coale and Demeny's life tables and stable populations have been used for demographic estimation in many developing countries, particularly where reasonably accurate data from one or two censuses are available and the population

Table 13.6 Excerpt from Coale and Demeny's West region model life tables for females at mortality levels (9) and (11)

	Mortality level (9)							Mortality level (11)						
Age x	l_x	${}_nm_x$	${}_nq_x$	${}_nL_x$	$\frac{{}_5L_{x+5}}{{}_5L_x}$	T_x	e_x	l_x	${}_nm_x$	${}_nq_x$	${}_nL_x$	$\frac{{}_5L_{x+5}}{{}_5L_x}$	T_x	e_x
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
0	100,000	0.2010	0.1777	88,447	0.7835[a]	4,000,000	40.00	100,000	0.1615	0.1461	99,502	0.8219[a]	4,500,000	45.00
1	82,226	0.0320	0.1179	303,316	0.9100[b]	3,911,553	47.57	85,388	0.0250	0.0937	320,442	0.9288[b]	4,409,498	51.64
5	72,530	0.0069	0.0338	356,520	0.9698	3,608.237	49.75	77,389	0.0055	0.0272	381,683	0.9758	4,089,056	52.84
10	70,078	0.0054	0.0264	345,762	0.9694	3,251,718	46.40	75,285	0.0043	0.0212	372,430	0.9752	3,707,373	49.25
15	68,227	0.0071	0.0350	335,172	0.9606	2,905,956	42.59	73,687	0.0058	0.0284	363,207	0.9679	3,334,942	45.26
20	65,842	0.0090	0.0440	321,964	0.9533	2,570,784	39.05	71,596	0.0073	0.0360	351,543	0.9618	2,971,735	41.51
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
75	14,505	0.1253	0.4772	55,221	0.3657[c]	87,064	6.00	19,311	0.1144	0.4448	75,083	0.3856[c]	122,207	6.33
80	7,584	0.2382	–	31,843	–	31,843	4.20	10,722	0.2275	–	47,125	–	47,124	4.40

Source: United Nations (1967)

Notes: [a]Proportion surviving from birth to 0–4

[b]${}_5L_{x+5}/{}_5L_0$

[c]T_{80}/T_{75}

Table 13.7 Excerpt from Coale and Demeny's West region model stable populations for females at mortality levels (9) and (11)

	Mortality level (9)							Mortality level (11)						
	Annual rate of increase							Annual rate of increase						
	−0.010	−0·005	0.000	0.005	0.010	…	0.050	−0.010	−0·005	0.000	0.005	0.010	…	0.050
Age interval	Proportion in age interval							Proportion in age interval						
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
<1	0.0158	0.0188	0.0221	0.0257	0.0295	…	0.0667	0.0142	0.0170	0.0201	0.0235	0.0272	…	0.0631
1–4	0.0557	0.0653	0.0758	0.0870	0.0988	…	0.2018	0.0516	0.0610	0.0712	0.0822	0.0939	…	0.1971
5–9	0.0684	0.0786	0.0891	0.1000	0.1111	…	0.1894	0.0643	0.0743	0.0848	0.0957	0.1069	…	0.1875
10–14	0.0698	0.0781	0.0864	0.0946	0.1024	…	0.1431	0.0660	0.0743	0.0828	0.0911	0.0992	…	0.1425
15–19	0.0711	0.0776	0.0838	0.0894	0.0945	…	0.1080	0.0676	0.0743	0.0807	0.0867	0.0920	…	0.1082
…	…	…	….	….	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
75–79	0.0213	0.0173	0.0138	0.0109	0.0085	…	0.0009	0.0255	0.0207	0.0167	0.0133	0.0104	…	0.0011
80+	0.0131	0.0103	0.0080	0.0061	0.0046	…	0.0004	0.0170	0.0134	0.0105	0.0081	0.0062	…	0.0005
Age	Proportion under given age							Proportion under given age						
1	0.0158	0.0188	0.0221	0.0257	0.0295	…	0.0667	0.0142	0.0170	0.0201	0.0235	0.0272	…	0.0631
5	0.0715	0.0842	0.0979	0.1127	0.1284	…	0.2685	0.0658	0.0780	0.0913	0.1057	0.1210	…	0.2602
10	0.1400	0.1627	0.1871	0.2127	0.2394	…	0.4579	0.1301	0.1523	0.1761	0.2014	0.2279	…	0.4477
15	0.2097	0.2408	0.2735	0.3073	0.3419	…	0.6010	0.1961	0.2266	0.2589	0.2926	0.3271	…	0.5902
20	0.2809	0.3185	0.3573	0.3968	0.4363	…	0.7090	0.2637	0.3009	0.3396	0.3792	0.4191	…	0.6984
…	…	…	….	….	…	…	….	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
65	0.8865	0.9059	0.9229	0.9374	0.9497	…	0.9934	0.8695	0.8913	0.9104	0.9270	0.9411	…	0.9921

(continued)

Table 13.7 (continued)

	Mortality level (9)							Mortality level (11)						
	Annual rate of increase							Annual rate of increase						
	−0.010	−0·005	0.000	0.005	0.010	...	0.050	−0.010	−0·005	0.000	0.005	0.010	...	0.050
Age interval	Proportion in age interval							Proportion in age interval						
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Birth rate	0.0175	0.0212	0.0250	0.0291	0.0336	...	0.0773	0.0156	0.0188	0.0222	0.0260	0.0302	...	0.0715
Death rate	0.0278	0.0262	0.0250	0.0241	0.0236	...	0.0273	0.0256	0.0238	0.0222	0.0210	0.0202	...	0.0215
GRR(27)	1.24	1.42	1.63	1.86	2.12	...	5.85	1.13	1.29	1.48	1.69	1.92	...	5.33
GRR(29)	1.25	1.44	1.66	1.91	2.20	...	6.54	1.13	1.30	1.50	1.73	2.00	...	5.94
GRR(31)	1.25	1.46	1.70	1.97	2.30	...	7.36	1.12	1.31	1.53	1.78	2.07	...	6.68
GRR(33)	1.25	1.47	1.73	2.04	2.40	...	8.39	1.12	1.32	1.56	1.83	2.15	...	7.56
Average age	36.2	33.9	31.6	29.5	27.4	...	15.6	37.6	35.2	32.9	30.6	28.5	...	16.1
B/P	0.042	0.048	0.056	0.065	0.075	...	0.221	0.038	0.044	0.051	0.059	0.068	...	.201

Source: United Nations (1967)

Notes: For each GRR the number in parentheses is the mean length of generation. B/P = births/population 15–44

Table 13.8 Population by age in 2011 census and the 2001 projected population using the model life tables based on various mortality levels: Hypothetical data

Age	Population in 2011	Projected population using life table for mortality level						Mortality level
		5	7	9	11	13	15	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20+	4,905	4,631	4,682	4,821	4,944	5,055	5,149	9–11
25+	4,212	3,838	3,981	4,109	4,223	4,325	4,411	9–11
30+	3,592	3,400	3,534	3,653	3,759	3,855	3,935	5–7
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
60+	720	603	650	693	731	767	795	9–11
65+	371	338	366	395	422	447	468	7–9

could be assumed to be stable at least in terms of its fertility and mortality rates and with negligible, if any, net migration. As noted earlier, the role of mortality in determining the age structure of a population is not as great as that of fertility and many indirect methods of demographic estimation based on the stable population theory can be used even in countries where mortality rates have been changing (United Nations 1967; United Nations 1982a).

13.9.1 A Hypothetical Example of the Use of Model Life Tables and Stable Populations

If data on fertility and mortality are deficient in a country, but there have been two recent censuses, preferably 10 years apart, and the conditions of stability or quasi-stability exist, then it is possible to identify an appropriate model life table that could be ascribed to such a country. Knowing the growth rate and using population totals from the two censuses, it would be possible to combine them with the appropriate model life table to estimate the stable population and its parameters as outlined in Sect. 13.5.

In a country where one census was conducted in 2001 and the other in 2011 and using the model life tables based at various mortality levels, it should be possible to project the population from 2001 to 2011. Following the cohort-component method (Sect. 11.4) this could be done by multiplying the 2001 population by survival ratios given in the model life tables at various mortality levels. This might involve preparing a number of population projections.

The projected populations are then summed into age groups such as 20+, 25+,. . ., 65+ and these are then compared with the actual population from the second census using the same age groups. Column (1) of Table 13.8 shows the cumulative ages, column (2) the population in 2011 census at those ages, and columns (3) to (8) the projected populations using the model life tables based on different mortality levels. Column (9) indicates the range of mortality levels that would result in the projected population being identical to the population in 2011 census. For instance, somewhere between mortality levels (9) and (11) is where the actual and projected population aged 20+ coincides.

Linear interpolation is used to pinpoint the mortality level that would result in projected populations of the first census to be the same as the actual populations in the second census.

Data from Table 13.8 are used to illustrate the linear interpolation technique:

Actual population aged 20+ in the second census	= 4,905
Projected population aged 20+ using model life table at level (9)	= 4,821
Projected population aged 20+ using model life table at level (11)	= 4,944

To calculate the level that would give projected population same as the actual population f_1 and f_2 are estimated:

$$f_1 = \frac{4,905 - 4,821}{4,944 - 4,821} = 0.6829 \text{ and } f_2 = 1 - f_1 = 0.3171.$$

The required mortality level that gives the projected population aged 20+ exactly the same as the actual number is:

$$(A * f_2) + (B * f_1) = 10.37$$

A stands for the mortality level that gives projected population less than the actual and B is the mortality level that gives the projected population more than the actual.

The above process is repeated for nine age groups like 20+ . . ..., 65+. The median of the resulting mortality level could be ascribed to the population of the country.

Having determined the mortality level, the next issue is to estimate the life table at this level. To illustrate this procedure, let the median mortality level found from the interpolations of data in Table 13.8 be 9.86. To estimate the ${}_nq_x$ values for mortality level (9.86) another set of multipliers needs to be calculated:

$$f'_1 = \frac{9.86 - 9}{11 - 9} = 0.43 \text{ and } f'_2 = 1 - f'_1 = 0.57.$$

Multiplying each value of the ${}_nq_x$ column of model life table level (9) with 0.57 and adding to it the corresponding value for level (11) multiplied by 0.43 gives an estimate of the ${}_nq_x$ column for level (9.86).

Having estimated the ${}_nq_x$ values a life table can be calculated using the procedures outlined in Chap. 7. This could be taken to represent the mortality situation in the country.

The value of r can be estimated from the population totals from the 2001 and 2011 censuses.

Combining the ${}_5L_x$ column of the estimated life table for mortality level (9.86) with r would result in a stable population and its parameters. These could be ascribed to the country.

13.10 Some Further Comments on Model Life Tables and their Role in Indirect Methods of Estimation

More recently other model life table systems have been developed (for example: Ledermann 1969; Carrier & Hobcraft 1971; United Nations 1982b). However, Coale and Demeny's tables are still commonly used.

In countries where vital registration systems are deficient or even non-existent, it is difficult to gain information about the components of population change. In these cases, the so-called *indirect* methods come into play. These methods depend on census counts and sample surveys. There are some methods that require two censuses while others require only one. Ideally, the census counts should be of good quality and provide age data at least in 5-year age groups.

Indirect estimation techniques are one of the fast growing areas in demography, partly because of the need for demographic parameters in countries where accurate data are still not available. Much research is going on and new methods are being found. The range of methods is quite vast and beyond the scope of this book. Nevertheless, some general comments are worth making.

Usually the stable population theory works well because fertility (as well as mortality) regimes have been found to be consistent across a wide range of populations, a finding that implies that these populations, at least as observed in the present tense, must be approximately stable (Popoff & Judson 2004). In turn, when the assumption of stability is reasonable for a given population, it offers the ability to estimate the components of change (and other characteristics) for this population. This assumption along with the concepts of the stable population theory can provide indirect measures in the absence of adequate direct ones.

As noted in Sect. 13.9.1, the *first step* is the generation of an underlying stable population model for the population in question; and, in the *second step* the stable population model's parameters are estimated and ascribed to the population concerned. The real power of this approach rests on two factors: the underlying stable population can be constructed with confidence, even on the basis of fragmentary data; and the estimated parameters yield a series of measures for which no accurate information exists otherwise. However, the stable model may not represent the actual situation if a country experiences large migratory movements or some unique outbreak of a fatal disease such as AIDS. Likewise, substantial, if temporary, deviations from past fertility and mortality rates (such as those created by epidemics, wars, or other unusual conditions) will have the same effect, even if both fertility and mortality levels have been constant. Even if the true situation is close to a stable state, the available data may be too fragmentary or biased to permit the derivation of an appropriate stable population. In this situation, there will be no reliable information to use in order to choose a particular stable model.

The preparation of stable estimates can be made mechanically by following the detailed rules. However, even when using pre-computed set of stable populations such as the Coale & Demeny (1966) regional life tables, it is important to have a full understanding of the logic underlying the method to apply the model in unusual situations.

Further, the idea that a stable population can only be constructed in the absence of migration started as a question of convenience, but recent research has shown that a stable population can be constructed by using not only a constant set of death and birth rates, but also migration rates (Preston et al. 2001).

The literature on methods of indirect estimation has grown substantially, as the need and demand for reliable demographic data increases. These methods are not covered in detail here, because of the scope of this book. However, there is much information available from a variety of sources such as the United Nations, national statistical offices and professional organizations like the International Union for the Scientific Study of Population.

Readers interested in the indirect methods of demographic estimation may wish to download the following United Nations manuals free of charge:

Manual IV. Methods for estimating basic demographic measures from incomplete data.
http://www.un.org/esa/population/techcoop/DemEst/manual4/manual4.html
Manual X. Indirect techniques for demographic estimation.
http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html

References

Australia. (2007). *2006 Births Australia*. Catalogue No. 3301.0. Canberra: Australian Bureau of Statistics. http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/0C4C341C51104DC4CA2573800015C2DC/$File/33010_2006.pdf. Accessed Feb 2013.

Australia. (2010). *Population by age and sex, Australian States and Territories, Jun 2010*. Catalogue No. 3201.0. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3201.0Jun%202010?OpenDocument. Accessed Feb 2013.

Carrier, N. H., & Hobcraft, J. (1971). *Demographic estimation for developing societies. Appendix 1, Brass model life table system*. London: Population Investigation Committee.

Coale, A. J. (1957). A new method for calculating Lotak's *r* – the intrinsic rate of growth in a stable population. *Population Studies, 11*(1), 92–94.

Coale, A. J. (1972). *The growth and structure of human populations: A mathematical investigation*. Princeton: Princeton University Press.

Coale, A. J., & Demeny, P. (1966). *Regional model life tables and stable populations*. Princeton: Princeton University Press.

Coale, A. J., & Demeny, P. (1983). *Regional model life tables and stable populations*. New York: Academic.

Dublin, L., & Lotka, A. J. (1925). On the true rate of natural increase. *Journal of the American Statistical Association, 20*, 305–339.

Indonesia. (2002). *The 2000 Indonesian life table*: *Final report*. Jakarta: National Institute of Health Research and Development. Ministry of Health. http://whoindonesia.healthrepository.org/handle/123456789/660. Accessed Feb 2013.

Ledermann, S. (1969). *Nouvelles tables-type de mortalité. Traveaux et document, Cahier No. 53*. Paris: Institut National d'études Démographiques.

Lotka, A. J. (1907). Relation between birth rates and death rates. *Science (New Series), 26*(653), 21–22.

McCann, J. (1973). A more accurate short cut method of approximating Lotka's R. *Demography, 10*(4), 567–570.

Pollard, A. H., Yusuf, F., & Pollard, G. N. (1995). *Demographic techniques*. Sydney: A.S. Wilson.

Popoff, C., & Judson, D. (2004). Some methods of estimation for statistically underdeveloped areas. In J. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 603–641). San Diego: Elsevier/Academic.

Preston, S., Heuveline, S. P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Malden: Blackwell Publishing.

Sharpe, F., & Lotka, A. J. (1911). A problem in age distribution. *Philosophical Magazine, 21*, 435.

United Nations. (1955). *Age and sex patterns of mortality: Model life tables for under-developed countries*. New York: Department of Economic and Social Affairs.

United Nations. (1967). *Manual IV*. *Methods for estimating basic demographic measures from incomplete data*. New York: Department of Economic and Social Affairs. http://www.un.org/esa/population/techcoop/DemEst/manual4/manual4.html. Accessed Feb 2013.

United Nations. (1982a). *Manual X*. *Indirect techniques for demographic estimation*. New York: Department of Economic and Social Affairs. http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html. Accessed Feb 2013.

United Nations. (1982b). *Model life tables for developing countries*. New York: Department of Economic and Social Affairs. http://www.un.org/esa/population/techcoop/DemMod/model_lifetabs/model_lifetabs.html. Accessed 10 Feb 2013.

United Nations. (2011). *UN data*: *Age-specific fertility rate*. New York: Department of Economic and Social Affairs. http://data.un.org/Data.aspx?d=GenderStat&f=inID%3A13. Accessed Feb 2013.

Chapter 14
Demographic Software

14.1 Purpose

The purpose of this chapter is to acquaint the readers with some of the software that are useful for demographic analysis. Given the rate at which the software (and hardware) technology is changing, readers would need to be aware of the latest developments. The list of software mentioned in this chapter is indicative only and should not be considered as complete. Also it is important to understand that some of the software packages mentioned here were developed some time ago and may or may not function in the latest computing environments. While the authors have used some of the software described here, the authors neither endorse any software nor guarantee that any of the downloads from the sites are free of viruses or any other problems.

14.2 Types of Software

It is difficult to think about doing substantial demographic work without a computer. It is safe to assume that most students, and other readers of this book, are likely to have access to a computer through their employers and/or their educational institutions, and that most computers have one or more software packages including those incorporating spreadsheet capabilities. For those readers who have no access to a computer, the good news is that every method of analysis presented in this book can be applied using a calculator with simple mathematical functions such as the logarithm and exponential functions.

There are three major types of software that demographers commonly use. They are:

- spreadsheets
- statistical analysis packages, and
- specialized demographic software.

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3_14,

14.2.1 Spreadsheets

Many computers are sold with some basic software including spreadsheets and word processing programs. There is a range of proprietary spreadsheet software available in the market. There are some that are proprietary and licenses need to be purchased while others are in public domain such as the Google Doc and Apache Open Office.

There are many advantages in the use of spreadsheets. For example they are fast, are very efficient in doing repetitive calculations and allow the preparation of templates, and, in general, are compatible across a wide range of computing environments, to name a few of their advantages.

Some spreadsheet software also enables the user to compile macros – which are sub-routines to do a particular procedure that can be done repeatedly. Most spreadsheets also have some statistical analysis capabilities, such as linear regression and analysis of variance. Though for larger datasets it may be better to use one of the statistical analysis packages mentioned below.

14.2.2 Statistical Packages

There are many statistical packages available in the market. These include, for example, *SPSS*, *SAS*, and *STATA*. They are particularly useful in the analysis of large datasets and for more complicated data analysis. Many readers may have access to these packages through their educational institutions and/or their workplace. These packages are generally quite expensive for individuals to purchase their licenses, although some give discounts for students and academics. One statistical package, *R*, is in the public domain and can be downloaded using the following link:

http://www.r-project.org

Another package that is useful for tabulations and other data processing applications is *CSPro*, though it has very limited analytic capabilities. The following link can be used to download this software:

http://www.census.gov/population/international/software/cspro/csprodownload.html

This package works in some Windows operating systems, though not necessarily in the most recent ones. The United States Bureau of Census states that it provides support for this package. Initial contact information is available at the preceding site.

Table 14.1 List of spreadsheets in *PAS* that are relevant to the methods discussed in this book

Spreadsheet	Brief description
ADJASFR:	Adjusting a set of age-specific fertility rates to get a desired total number of births.
ADJMX:	Adjusting m_x values from a life table to obtain a specified no. of deaths.
AGEINT:	Interpolation between two population age structures by linear or exponential models.
AGESEX:	Calculation of the United Nations' accuracy indexes by age and sex.
AGESMTH:	Smoothing of a population age distribution using several methods.
BASEPOP:	Preparing the age and sex distribution for making a projection.
CSRMIG:	Estimating intercensal net migration between two areas and several related indices.
FITLGSTC:	Fits a logistic function to 3 (or a multiple of 3) equidistant values of any index.
LTMXQXAD:	Constructing a life table from the m_x and q_x values.
LTPOPDTH:	Constructing a life table after smoothing the population and deaths data.
MOVEPOP:	Moving the age distribution on a specific date to another date.
PYRAMID:	Making a population pyramid using absolute numbers and %.
SINGAGE:	Calculating the Myers' and other indices of digital preference.
SP:	Fitting a stable population based on life tables by sex, ASFR and r.
URBINDEX:	Calculating several urbanization and population distribution indices.

14.2.3 Specialized Demographic Software

During the early 1970s and 1980s, many specialized demographic software packages were developed. Most of them pre-dated the micro-computer era, and practically all were based on the *disk operating system* (DOS). Many of them have not been updated and do not work on the more modern computers, particularly those using the Apple or the most recent Windows operating systems.

The *Population Analysis System (PAS)* consists of 45 Excel spreadsheets developed by the United States Bureau of Census. It covers a variety of indirect methods of estimation of fertility, mortality, migration and distribution of population particularly for countries where such data are not accurately recorded. Only 15 of the spreadsheets were found to be directly relevant to the material covered in this book. They are listed in Table 14.1 in alphabetic order with a brief description. The remainder would be of interest to anyone involved in using the indirect methods of estimation. As in February 2013, *PAS* did not work with the Windows 7, Windows 8 and Mac operating systems. For further information, readers may contact the United States Bureau of Census directly.

RUP is another package available from the United States Bureau of Census. Its main function is in the preparation of rural and urban population projections.

Both packages are in the public domain and can be downloaded using the link:

http://www.census.gov/population/international/software/pas/index.html

Another package that incorporates the most recent methods of indirect estimation along with comprehensive documentation on the rationale and theoretical background of each method is the *Tools for Demographic Estimation*. These tools mainly consist of template spreadsheets dealing with the indirect methods of demographic analysis, a topic outside the scope of this book. These tools are being developed under the auspices of the International Union for the Scientific Study of Population with support from the United Nations Population Fund. Although this is a continuing project, many of the tools are ready for use and can be downloaded from:

http://demographicestimation.iussp.org/

The tools are in the public domain; however, users would need to register using the above link.

MODGEN (Model Generator) is a set of programs for those interested in micro-simulation modeling. It was developed by Statistics Canada and has been used for population projections. It is a generic micro-simulation programming language supporting the creation, maintenance and documentation of dynamic micro-simulation models. Both the software and documentation on it are in the public domain and can be downloaded from the following link:

http://www.statcan.gc.ca/microsimulation/modgen/modgen-eng.htm

A package developed by the United Nations is the *MORTPAK*. This package consists of 17 programs that enable the computation of life tables, both empirical and model, estimation of fertility and mortality using incomplete data, and population projections. It has only three programs that are relevant to the material covered in this book. They deal with eth construction of life tables (LIFTB), calculation of a stable population (STABLE) and the graduation (smoothing) of a set of age-specific probabilities of dying (UNABR).

This package is also available in the public domain, and can be downloaded using the following link:

http://www.un.org/en/development/desa/population/publications/mortality/mortpak.shtml

14.3 Conclusion

New and more powerful demographic software is expected to become available subject to the demands from users and advancement in technology. As a concluding note, readers may keep in mind that most of the work presented in this book was done using spreadsheets with occasional help from a statistical analysis program.

Index

F. Yusuf et al., *Methods of Demographic Analysis*, DOI 10.1007/978-94-007-6784-3,