2

Notation

2.1 Introduction

In this chapter we introduce some general expressions that have facilitated the development of many formulas found in this book. All notations are listed in the index of notations at the end of the book.

Demographic averages are introduced first, followed by an explanation of the reasoning behind the focus on averages. Next is a description of the operators used to measure changes over time and finally, the issue of data sources and an illustration of how the results from the estimations in the tables are presented.

2.2 Demographic Averages

The decompositions of change over time demonstrated here, mainly focus on averages, also known as expected values or expectations. Let the expectation operator at time t be denoted as $\bar{v}(t)$ the mean value of the function $v(x,t)$ over variable x, can be expressed as follows:

$$\begin{aligned} E(v) \equiv \bar{v}(t) &= \frac{\int_0^\infty v(x,t)w(x,t)dx}{\int_0^\infty w(x,t)dx},\ x\ continuous, \\ &= \frac{\sum_x v_x(t)w_x(t)}{\sum_x w_x(t)},\ x\ discrete, \end{aligned} \tag{2.1}$$

where $v(x,t)$ is some demographic function and $w(x,t)$ is some weighting function. For example, if $v(x,t)$ is substituted with the age-specific death rates and $w(x,t)$ with the age-specific population size we get a $\bar{v}(t)$ which is equal to the crude death rate. Equation (2.7) shows this substitution in detail.

The numerator and denominator of a ratio may or may not be subsets of the same variable. By using the average of ratios (2.3) the nominator and denominator are related. The sex ratio at birth, for example, is equal to the average of sex ratios by age of the mother.

Proportions are a special type of ratio where the denominator includes the numerator. A well-known demographic variable is the proportion of the total population belonging to an age group, for example the proportion below age a. It is possible to express proportions as averages by using an indicator function $I_a(x)$. This function has values of 1 when a condition is satisfied, x below a for example, or 0 otherwise. By using this function it is possible to include only the desired elements in the numerator,

$$\frac{v.(t)}{w.(t)} = \frac{\sum_x v_x(t)}{\sum_x w_x(t)} = \frac{\sum_x I_a(x) w_x(t)}{\sum_x w_x(t)} = \bar{I}_a(t). \tag{2.4}$$

The right hand side of equation (2.4) shows that the proportion of the variable $v.(t)$ over $w.(t)$ is equal to the average of the indicator weighted by $w_x(t)$. The proportion of the total population below age 15, for example, is equal to the average indicator $\bar{I}_a$ with $a = 15$ or less, and weighted by the age-specific population size at age a and time t, $N_a(t)$,

$$\frac{N_{0-15}(t)}{N.(t)} = \frac{\sum_{a=0}^{15} N_a(t)}{\sum_{a=0}^{\omega} N_a(t)} = \frac{\sum_{a=0}^{\omega} I_{0-15}(a) N_a(t)}{\sum_{a=0}^{\omega} N_a(t)} = \bar{I}_{0-15}(t). \tag{2.5}$$

Percentages are a special type of proportion in which the ratio is multiplied by a constant, normally 100, so that the result is expressed per hundred.

A *rate* refers to the occurrence of events over a given time interval, where the denominator is the duration of persons exposed to risk during this time interval. The duration of persons "exposed to risk" is specified by using the concept of "person-years lived". The crude death rate (CDR) for year t, for example, is sometimes denoted by $d(t)$ and calculated as the total number of deaths during the year t, $D.(t)$, over the number of person-years lived during the period, $N.(t)$. The number of deaths at age a, $D_a(t)$, is equal to the product of the age-specific death rates $m_a(t)$ at age a and time t, and the population size at that age, $N_a(t)$; $D_a(t) = m_a(t)N_a(t)$,

$$d(t) = \frac{D.(t)}{N.(t)} = \frac{\sum_{a=0}^{\omega} D_a(t)}{\sum_{a=0}^{\omega} N_a(t)} = \frac{\sum_{a=0}^{\omega} m_a(t) N_a(t)}{\sum_{a=0}^{\omega} N_a(t)} = \bar{m}(t), \tag{2.6}$$

where ω is the highest age attained. The population size functions as an estimate of the person-years lived.

In the continuous case when the age interval gets shorter the death rates can be changed to the "force of mortality", denoted $\mu(a,t) = \lim_{\Delta a \to 0} m_{a,a+\Delta a}(t)$. The continuous expression of the CDR is the average of the force of mortality weighted by the age-specific population size, $N(a,t)$,

$$d(t) = \frac{D(t)}{N(t)} = \frac{\int_0^{\omega} D(a,t)da}{\int_0^{\omega} N(a,t)da} = \frac{\int_0^{\omega} \mu(a,t) N(a,t)da}{\int_0^{\omega} N(a,t)da} = \bar{\mu}(t). \tag{2.7}$$

Due to the fact that demographers come from many academic disciplines and for various historical reasons, the actual rates used by demographers are not always consistent with the rates described here. The crude birth rate (CBR), for example, is defined as the total number of births over the total population size. Demographers working in the field of fertility mainly use female age-specific fertility rates. In our research we follow this convention and refer to female rates as age-specific fertility rates. Births occurring during the year t, denoted as $B(t)$, are the product of age-specific fertility rates, $b(a,t)$, and the female population size denoted at age a and time t as $N_f(a,t)$; $B(a,t) = b(a,t)N_f(a,t)$. The crude birth rate contains the female population size in the numerator as weights, while the total population size is the denominator,

$$CBR(t) = \frac{B(t)}{N(t)} = \frac{\int_0^\omega B(a,t)\,da}{\int_0^\omega N(a,t)\,da} = \frac{\int_0^\omega b(a,t)\,N_f(a,t)\,da}{\int_0^\omega N(a,t)\,da}. \tag{2.8}$$

It is possible to transform this rate into an average by carrying out an arithmetic manipulation as in equations (2.3) and (2.4). The CBR is modified to an average of the product of the age-specific fertility rates and the proportion of females,

$$CBR(t) = \frac{\int_0^\omega b(a,t)\,c_f(a,t)N(a,t)\,da}{\int_0^\omega N(a,t)\,da} = \overline{bc_f}(t), \tag{2.9}$$

where $c_f(a,t) = \frac{N_f(a,t)}{N(a,t)}$ is the proportion of females in the population at age a and time t.

Normally, ratios, proportions and percentages are used for analyzing the composition of a set of certain events or a population. Rates, in contrast, are used to study the dynamics of change. A *probability* is similar to a rate except for one important difference: a probability's denominator consists of all persons in a given population at the beginning of the observation period. The probability of dying in the year 2005 in a population closed to migration given that one survives to age 50 in January 2005 is equal to the deaths that occur during the year in this age group divided by the population of age 50 and above present on January 1, 2005. Arithmetic manipulations, similar to those for rates, can also be carried out to study a probability as an average.

As can be seen, averages have a central position within mathematical expressions of demographic measures. Demographic rates and other measures are influenced by the population composition. The study of changes over time, as described above in terms of averages, leads us to the field of decomposition methods. The mathematical formula resultants of derivation over time for demographic variables have different elements which represent the components responsible for the change. The following section presents the notation for the operators needed to study changes over time for demographic variables.

2.4 Operators

Let a dot over a variable denote the derivative with respect to time, t,

$$\dot{v}(t) \equiv \dot{v}(x,t) = \frac{\partial}{\partial t} v(x,t). \tag{2.10}$$

The change over time in the crude death rate, as seen in equation (2.7), is expressed following (2.10) as $\dot{d}(t) = \frac{\partial}{\partial t} d(t)$.

An acute accent is used to denote the relative derivative or intensity with respect to time, t

$$\acute{v}(t) \equiv \acute{v}(x,t) = \frac{\frac{\partial}{\partial t} v(x,t)}{v(x,t)}. \tag{2.11}$$

In this way, the relative change in the crude birth rate over time, as seen in equation (2.8), is then denoted as $C\acute{B}R(t) = \frac{\frac{\partial}{\partial t} CBR(t)}{CBR(t)}$. The use of the acute accent, which reduces the clutter in many demographic formulas, originated from Vaupel (1992) and is used in Vaupel and Canudas Romo (2000). Note that for simplicity we often omit the arguments x and t. Derivation is how change over time is measured and the notation presented here has proven to be very useful in developing direct vs. compositional decomposition formulas.

The covariance operator is widely used in statistical analysis but is rare in formal demography. The results and applications presented by Vaupel (1992), and Vaupel and Canudas Romo (2002), also found in Part III, suggest that this and various other covariances are of general significance to the understanding of population dynamics. In this context, the covariance measures the extent to which a demographic variable rises and falls with another demographic measure. For example, the covariance between age and the population growth rate is used by Preston, Himes and Eggers (1989) to analyze population aging. The covariance operator is defined as the average of the product of the deviations of two variables from their respective means,

$$C(v,u) = E\left[(v-\bar{v})(u-\bar{u})\right]$$

$$= \frac{\int_0^\infty \left[v(x,t)-\bar{v}(t)\right]\left[u(x,t)-\bar{u}(t)\right] w(x,t)dx}{\int_0^\infty w(x,t)dx}. \tag{2.12}$$

The term $[v(x,t)-\bar{v}(t)]$ corresponds to the distance between the variable $v(x,t)$ and its mean $\bar{v}(t)$, and it is similar to the interpretation for $[u(x,t)-\bar{u}(t)]$. Another expression for the covariance which is used in this book is the difference between the expectation of a product and the product of expectations

$$C(v,u) = E(vu) - E(v)E(u) = \overline{vu} - \bar{v}\bar{u}$$

$$= \frac{\int_0^\infty v(x,y)u(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} - \frac{\int_0^\infty v(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx}\frac{\int_0^\infty u(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx}. \tag{2.13}$$

For example, $C(\mu,r)$ is the covariance between the force of mortality $\mu(a,t)$ and the age-specific growth rates $r(a,t)$, weighted by the population size $N(a,t)$. The basic idea is that the covariance will be positive if the force of mortality tends to be higher (or lower) than average at those ages when the age-specific growth rate tends to be higher (or lower) than average.

Note that equation (2.13) implies that the expectation of a product can be decomposed as

$$\overline{uv} = \bar{u}\bar{v} + C(u,v), \tag{2.14}$$

which is a useful result in many contexts.

We also need to define a special notation for the operators of difference and relative difference. Parallel to the dot notation for the derivation operator let the difference operator between $t+h$ and t be denoted as

$$\Delta v(t) \equiv v(t+h) - v(t), \tag{2.15}$$

and the relative difference operator as

$$\grave{v}(t) \equiv \frac{\Delta v(t)}{v(t)} \equiv \frac{v(t+h) - v(t)}{v(t)}. \tag{2.16}$$

2.5 Data and Applications

The applications presented throughout this book come from freely available data. This section describes some of these data sources. The applications of decomposition formulas are organized in tables or figures which show the results. Following the presentation of the data sources is a subsection explaining how to read these tables.

2.5.1 Data Sources

The data sources used in the applications are listed below in the order of their appearance in the text. There is also a brief description of the databases and the tables which we used for the illustrations.

The United Nations Demographic Yearbook (2001): is a comprehensive collection of international demographic statistics, prepared by the Statistics Division of the United Nations. It contains historical demographic statistics from 1948 to 1997 and presents time series of population size by age, sex and urban/rural residence, natality, mortality and nuptiality as well as selected derived measures concerning these components of population change for a 50-year period. Applications in the book using this data source are employed in Tables 3.1, 6.1, 6.3, 6.4, 7.1, 9.1 and 10.1. Figure 7.1 is also based on this data source.

The Human Mortality Database (2002) was created to provide detailed mortality and population data for researchers interested in the history of human longevity. The main goal of the database is to document the longevity revolution of the modern era and to facilitate research into its causes and consequences. There is open international access to these data. At present the database contains detailed data for a collection of 17 countries. However, this database is limited to countries where death registration and census data are virtually complete, since this type of information is required by the uniform method followed in this database. As a result, the countries included here are relatively wealthy and for the most part highly industrialized. The book employs this data source in Tables 3.2, 3.3, 3.4, 3.5, 4.1, 4.7, 4.8, 4.10, 4.11, 5.5, 6.2, 6.7, 8.2, 8.3, 8.5, 10.2 and 10.3. Figures 6.1, 6.2, 6.3, 8.1, 8.2 and 9.1 are among the figures that are based on this data source.

The Berkeley Mortality Database (2001) has been replaced by a bigger and better project, i.e., the Human Mortality Database. All except one item are available from the Human Mortality Database. This item is the Japanese cause-of-death data and is used in Tables 4.2, 7.4, 7.5 and 8.4. Figure 7.5 is based on this data source.

The Eurostat Database (2000) is accessed through the New Cronos interface which allows one to navigate in the database to find and download the data of interest. The database collects information on an annual basis from 36 countries comprising EU countries, EFTA countries, and other east and central European countries. Data are collected by the respective National Statistical Institutes and depend on the registration systems used in each country. Eurostat collects raw numbers, not indicators directly from the countries. Time series for the EU and EFTA countries begin from 1950 and continue through 1999, but some demographic data by age for the period 1960-1990 has not been fully processed. This data source is used in Tables 4.3, 4.4, 6.5, 6.6, 8.6 and 10.4.

The International Data Base (IDB) is a computerized data bank containing statistical tables of demographic and socio-economic data for 227 countries and areas of the world. The IDB provides quick access to specialized information, with emphasis on demographic measures, for individual countries or selected groups of countries. The IDB combines data from country sources (especially censuses and surveys) with estimates and projections to provide information dating back as far as 1950 and as far ahead as 2050. Because the IDB is maintained as a research tool in response to sponsor requirements, the amount of information available for each country may vary. The database covers the national population, and selected data by urban/rural residence, and by age and sex. Tables 4.9, 7.2, 7.3, 8.1 and 10.5 employ this data from the U.S. Census Bureau (2001).

The National Retrospective Demographic Survey (1998) or EDER in Spanish, was carried out in Mexico in 1998. The data pertains to Mexican residents in 1998. This retrospective survey with duration data is representative of the entire population of the country. A total of 2,496 persons were interviewed in the EDER survey. Three cohorts of Mexicans are interviewed: those born in 1936-1938, 1951-1953, and 1966-1968. For each person, individual events are related to calendar time, from the year of birth to the moment of the survey in 1998. The event-history data of each individual can be divided into five data sets: migration history, labor history, education history, family history, and fertility history. Chapter 5 includes four tables which used this database: Tables 5.1, 5.2, 5.3 and 5.4.

The data used in Figures 7.2, 7.3 and 7.4 were supplied by Professor Virgilio Partida of the National Population Board from Mexico, Conapo (2002). A data source which comes from a table in the article by Islam et al. (1998) is used in Tables 4.5 and 4.6 in this book. Historical data for Japan in Table 4.11 is based on Japan Statistical Association (2002).

2.5.2 Tables of Applications

In general the configuration of the tables follows the same presentation as illustrated by Table 2.1. has five parts. There are five features, the first of which is the table number and caption. top heading has the number and name of the Table. At the beginning of the book is compilation of all tables used in this study.

In the actual table, the first row is for the heading of the columns, separated from the results by a line. For example, sometimes there is need to know information about different years, different regions, or the contribution of different components, in which case each column contains the respective information.

The next section of the table is the observed values of the variable under study. For example, the information of the demographic variable in the years under study is provided

Variable x in (2.1) can be continuous or discrete. This is denoted respectively as $v(x,t)$, or in the discrete case as a subindex, $v_x(t)$. In the applications presented in this book, x can denote age, region of residence or marital status, although other applications are also of interest. The variable t, which here denotes time exclusively, is always continuous.

In the text, the variables x, $v(x,t)$ and $w(x,t)$ are used for general formulas and are substituted with known demographic variables depending on the application. For example, in several applications the weighting function $w(x,t)$ equals $N(a,t)$ the age-specific population size at age a and time t. Here it is implicit that variable x is changed by a, representing age. (See Encyclopedia of the Social and Behavioral Sciences (2001) for a discussion on this basic but subtle quantity.)

If the demographic function $w_x(t)$ is aggregated over all the possible values of the discrete variable x then this is denoted with a dot in place of the variable x, $w.(t) = \sum_x w_x(t)$. For example, the total population at time t denoted $N.(t)$ is equal to the addition over all ages of the single age population sizes, denoted $N_a(t)$ for age a to $a+1$ at time t. This is expressed as $N.(t) = \sum_a N_a(t)$. In the case of two discrete variables, x and z, adding all values of one of the variables, we denote this with a dot instead of with the variable. When adding over z we use $w_{x\cdot}(t) = \sum_z w_{xz}(t)$, and adding over x we get $w_{\cdot z}(t) = \sum_x w_{xz}(t)$. So when adding over both x and z we obtain $w_{\cdot\cdot}(t) = \sum_x \sum_z w_{xz}(t)$.

For the continuous case the variables are simply omitted, i.e., integrating $w(x,t)$ over all values of x we obtain $w(t) = \int_0^\infty w(x,t)dx$ and $w(x,t) = \int_0^\infty w(x,z,t)dz$. The new variable z is useful when two compositional components, for example, age and country of residence, are included in the same decomposition.

The reason we focus on averages is that many demographic measures can be expressed as averages. Demographic formulas that can be generalized by a mathematical expectation, or average, are presented in the next section.

2.3 Demographic Measures: Ratios, Proportions, Rates, and Probabilities

As pointed out by Palmore and Gardner (1983), the most commonly used measures in demography are ratios, proportions, rates, percentages and probabilities. This section shows that all these measures can be considered as averages.

A *ratio* is a single number that expresses the relative size of two numbers. A commonly used ratio in demography is the ratio of male to female births, the sex ratio at birth. The result of dividing a number $v.(t) = \sum_x v_x(t)$ by another number $w.(t) = \sum_x w_x(t)$ is the ratio of v to w,

$$\frac{v.(t)}{w.(t)} = \frac{\sum_x v_x(t)}{\sum_x w_x(t)}. \tag{2.2}$$

This ratio can be seen as an average of ratios by carrying out an arithmetic manipulation of dividing and multiplying the numerator by the same term,

$$\frac{v.(t)}{w.(t)} = \frac{\sum_x \left(\frac{v_x(t)}{w_x(t)}\right) w_x(t)}{\sum_x w_x(t)} = \overline{\left(\frac{v}{w}\right)}(t). \tag{2.3}$$

Table 2.1: Example of how the results are presented in the tables.

Heading for the columns of the table
Observed demographic changes
Estimated components of the change

Source of the data and notes.

here. Furthermore, the observed change in the period under study is stated here. Again there is another line separating this information from the next part of the table.

The next section provides estimations of the change in the demographic variable and of the components of this change. This section is crucial for our work since it contains the results of the methods explained here. Sources and special notes accompany the table.