TRAINING MANUAL ON

DEMOGRAPHIC TECHNIQUES

भारत की जनगणना
CENSUS OF INDIA
2011
Our Census, Our Future

UNFPA
United Nations Population Fund – India

TRAINING MANUAL ON DEMOGRAPHIC TECHNIQUES

2011
Our Census, Our Future

United Nations Population Fund – India

Prepared by:

K. Srinivasan
Honorary Visiting Professor,
Institute for Social and Economic Change, Bangalore

E-14, TVH Park Rozalia,
Anand Nagar, Thuraipakkam,
Chennai - 600097

Foreword

The Office of Registrar General & Census Commissioner, India (ORGI) has taken the pioneering initiative of setting up a state-of-the-art resource and training centre of international standard to cater to in-house and international training on census and surveys. The centre is designed to offer routine/customized training programmes as well as provide specialized consultancy services. The UNFPA and other UN organizations have supported ORGI in this initiative. The training centre is in the process of developing course material on different topics. The present volume, *Training Manual on Demographic Techniques*, is the first in the series of training manuals developed by the Census Resource and Training Centre.

This manual is designed for a 10-day capacity-building programme covering introduction to demography, demographic data sources, population age-sex structure, quality of data and adjustments, basic measures and concepts of fertility, mortality, migration, urbanization and nuptiality. Further, it demonstrates standardization of rates and ratios, population projection, demographic models and selected software packages used in demographic analysis.

The manual has been prepared by Prof. K. Srinivasan after extensive consultation with stakeholders and after practical testing. We are sure this manual will serve an extremely useful purpose and will be widely used.

Dr. C. Chandramouli
Registrar General &
Census Commissioner, India
Government of India

Frederika Meijer
UNFPA Representative
India and Bhutan

Preface

This training manual has been prepared keeping in mind the recently arisen needs of the young professionals joining the Office of the Registrar General of India & Census Commissioner (ORGI). This number is really huge. Most of them had no formal training in demography, excepting the experience of their direct participation in the 2011 Census work and related short training.

It was felt necessary to give them, among other courses, a short orientation training in basic demographic techniques so that they will be useful in the coming years in the analysis of the large volume of census data that has been collected and in interpretation of the findings. This *Training Manual on Demographic Techniques* is expected to meet this immediate need. The manual was used in the first such teaching program of two weeks' duration from 3 to 13 June 2013 at the Office of RGI. There were 10 participants in the course and, based on the interaction we had with the participants, the first draft of the Manual has been suitably modified. There were also suggestions to make the training program on-line for the benefit of those from the Census Department who are not able to attend the training program as full time participants in established training centers.

I wish to acknowledge the assistance I had from Prof. Sulabha Parasuraman, former Professor at the International Institute for Population Sciences, Mumbai, who shared my teaching load, Dr. Preeti Dhillon, Consultant (UNFPA) who helped in the laboratory sessions and assisted in finalizing training materials and to Dr. K.M. Sathyanarayana and Dr. Sanjay Kumar of UNFPA who motivated me to write this manual and conduct the first training program. I gratefully acknowledge the Training Division of ORGI, TOT participants and Mr. C. Chakravorty, Consultant (UNFPA) for their unconditional support.

The errors of commission and omission are solely mine and I feel, like any training Manual, this has to be periodically modified and improved upon keeping in mind the training needs of the participants.

K. Srinivasan
Honorary Visiting Professor,
Institute for Social and Economic Change,
Bangalore

Contents

CHAPTER I

Introduction: Definition, Scope and Evolution of Demography

1.1 Definition and Scope of Demography

The term **demography** is derived from the Greek word ‘demos’ meaning a human being. The term seems to have been coined for the first time by a Belgian statistician, Achille Guillard, in 1855 in the article entitled *‘Elements de statistique humaine, ou demographique comparee’* (Elements of human statistics or comparative demography) (Shryock and Siegel, 1971). Like all other disciplines, demography admits to a narrow and a wider definition. In narrow terms demography is defined in the *Multilingual Demographic Dictionary* brought out by the International Union for the Scientific Study of Population (IUSSP) as ‘the scientific study of human populations, primarily with respect to their size, structure and development; it takes into account the quantitative aspects of their general characteristics’ (Van de Walle, IUSSP, 1982). Thus, it is a scientific study of human populations in their aggregate with regard to their size, composition or structure, spatial distributions and developments or changes in these over time. In its wider definition, where it is also called *Population Studies*, it not only deals with levels and changes in the size, composition and distribution of the population but also with the causes and consequences of the levels and changes. In this wider perspective, demography overlaps with a number of other disciplines such as economics, sociology, social psychology, law, political science and reproductive physiology (Bogue, 1969; UN, 1958). Thus demography in its wider definition as population studies is multi-disciplinary in nature and in recent years has attracted scholars from various other disciplines, especially economics and sociology, who have made valuable contributions to its development. These scholars are primarily associated with their parent discipline but interested in the population phenomena. They can be designated as sociologist-demographers, economist-demographers etc.

The practical central focus of demography, known as *formal demography*, is a set of techniques by which data collected in censuses, surveys and vital registration systems are described, summarized and manipulated. Again, the *Multilingual Demographic Dictionary* of the IUSSP defines formal demography as “the treatment of quantitative relations among demographic phenomena in abstraction from their association with other phenomena” (Van de Walle, 1982). In this sense, it can be considered as a toolkit containing the techniques commonly employed by demographers and others in their ‘scientific’ study of human populations. It is fundamentally descriptive or analytic rather

NOTES

than explanatory in nature. It answers questions of 'what' rather than 'how' regarding the population phenomena. In this manual, we are interested in this formal demography in describing and illustrating some of the basic demographic techniques and their utility in the planning and management of various developmental programmes.

In demographic parlance, *demographic processes* mean the phenomena of fertility or births, mortality or deaths and migration or movement of people from one location to another, which contribute to changes in the size and structure of a population. The variables of age, sex and marital status are considered the three basic demographic variables. Most of the techniques of formal demography are confined to an understanding of the interrelationships between demographic processes and size, growth and the age-sex-marital structure of a population. Thus, by a demographic analysis of a population is meant the analysis of its size, growth, spatial distribution, age-sex-marital status structure of the population and changes in these over time.

1.2 Organization of the Manual

NOTES

Many textbooks and manuals have already been written in the past to address the needs of scholars and students about basic demographic techniques. Notable among them are *Population Analysis with Microcomputers* by Arriaga et al. in 1994 for the US Census Bureau, *Basic Demographic Techniques and Applications* by Srinivasan (1998) mainly addressed to the public health personnel and *Demographic Methods and Concepts* by Rowland (2004). A very detailed volume on this topic used also as a standard text in the teaching of demography course is *Methods and Materials in Demography* (Vols. I and II) by Shryock and Siegel (US Census Bureau, 1971). The present manual should be regarded as no more than a basic toolkit containing some of the basic methods commonly employed by demographers in the analysis of population data, especially for India. It is intended as a first course in the analysis of population data, especially for the newly collected census data in the Indian context. The first chapter is introductory in nature, provides a brief discussion of population studies and formal demography and briefly traces the historical development of the subject. Chapter II deals with very basic concepts and definitions used in formal demography such as universe/population and samples, variables, rates, ratios, periods of exposure, period and cohort rates, and discusses the basic demographic equation and demographic transition theory. The recent extensions of the demographic transition theory, called the *Second Transition Theory*, are also discussed. Demography is considered to be an empirical science and its analysis generally depends on the materials available for use. Chapter III covers the sources of the materials or data from which a demographer usually draws materials for analysis, especially in the Indian context. The four different sources of data, namely census, vital statistics, sample surveys and service statistics, are described and the general characteristics of a population or events covered by such data are discussed in detail. The most popular publications on population data from the United Nations (UN) and the national organizations in India and their websites are also described, with the range of discrepancies of data collected from different sources.

Chapter IV describes the crucial role played by the variable of age in demographic analysis, importance of the population age-sex structure in explaining the demographic process and various indicators that can be derived such as the dependency ratios, employment rates, child-woman ratios etc. The various types of errors of coverage and accuracy, biases and random errors, sampling and non-sampling errors in any reported age distribution, indicators for identifying and measuring the same and methods of correcting and adjusting for such errors are also discussed. Chapter V describes the basic measures of fertility which are often utilized in the demographic analysis of structure, commonly used period and cohort measures, natural fertility, Coale's indices of fertility and the crucial role played by fertility in determining the age distribution of a population. Chapter VI discusses the various measures of mortality, the techniques used in the measurement and analysis of mortality in a population and specific measures of mortality such as the infant mortality rate, child mortality rate, maternal mortality rate etc. This chapter describes the basic functions and construction of a life table, which is regarded as giving a comprehensive picture of the mortality situation in a population. Chapter VII covers the third component of population change, namely migration, measurement of internal and international migration, calculation of indices and methods of measuring migration from census and survey data.

Chapter VIII deals with nuptiality analysis. Though the institution of marriage as the main social institution within which child bearing takes place is rapidly disappearing in the western societies, in India and in many other Asian countries the institution of marriage is still quite robust and more than 99% of the births in India take place among the married women. Hence, there is a need for analysis of proportion marrying and age at marriage for women in the Indian context.

NOTES

The technique of standardization was originally developed in demography but is finding wide applications in other fields too. Chapter IX deals with this and describes direct and indirect standardization techniques and decomposition of the differences in a given parameter between two populations due to different factors. Chapter X covers the population projections, their importance, some of the commonly used algebraic methods, the most commonly used method of component projection, period and cohort projections and recently developed probabilistic projections method. Chapter XI discusses the demographic models and model life tables including the field of stable, quasi-stable and general stable population theories. Particular attention is given to model life tables which are essential in population projections. Chapter XII, the final chapter, describes the oft-used software packages in demographic analysis in India – MORTPAK, PAS, and SPECTRUM – and some of the packages recently developed in India. Chapterization of the manual is to facilitate sequential organization of lectures by trainers so that all the topics are covered within the stipulated time.

1.3 Tracing the History of Formal Demography

Formal demography or technical demography, as mentioned earlier, is that branch of the subject which deals with the measures and methods or techniques used in demographic analysis. For an appraisal of the evolution and the present status of the subject, it is useful to consider how the subject has developed globally. Mortality, or the study of deaths in a population, has been the area of initial investigation for over two centuries.

John Graunt is generally acclaimed as the founding father of demography as a field of empirical research (quoted in Glass, 1950). In 1662, from the records of baptisms and burials in London, Graunt noted the excess of male over female births, variations in mortality by age, and temporal and spatial variations in the causes of death, and even attempted the construction of a crude mortality table. He brought out a series of handwritten statements on deaths in London, which came to be called *the diabolical bills*. It is interesting to note that Graunt was the son of a haberdasher with no formal school education. However, it was Edmund Halley (1693), famous for the discovery of the comet named after him, who constructed the first empirical life table showing a stationary population by single-years of age up to 84 years (UN, 1973). Johann Sussmilch covered a variety of demographic topics during 1741–65 including vital rates and life tables (UN, 1973).

The development of the subject prior to the late nineteenth century was essentially in the spheres of analysis and presentation of mortality data. The increased availability of vital registration data during the nineteenth century enabled William Farr, England's first Registrar General, to produce the decennial series of English life tables from 1841 (UN, 1973). The techniques of life table construction and related theory were further improved by treating mortality as a continuous function.

NOTES

The centrepiece of technical demography at present is the analytical theory of the relationship between fertility and mortality on the one hand, and population age structure on the other, and is called the *stable population theory*. This was first implied in the analysis that a Swiss mathematician, Leonard Euler, attempted as far as back in 1760, 150 years before Lotka, who is currently credited with this discovery (as quoted by Smith and Keyfitz in their 1977 work). Euler had mathematically demonstrated that constancy of the force of mortality matched by the constancy of the force of fertility in a closed population necessarily implied a constancy of age and sex distribution irrespective of the level of natural increase. However, the credit for a more complete presentation of the stable population model goes to Alfred J. Lotka (1922), who in the course of empirical investigations developed the mathematics of a stable population.

The most outstanding work in the field of fertility analysis is by Kuczynski (1935), who compiled and analyzed data on reproduction for a number of countries. Through this work, the concepts of total fertility and net reproduction rate (NRR) became widely recognized. The concept of cohort in the field of fertility analysis was first developed and applied by Whelpton in 1954 in his studies of American fertility trends. Coale and Trussell (1974) devised a fertility model with the intention of capturing the range of age patterns of fertility typically observed in human populations. This marital fertility model has been used extensively in studies of 'demographic transition' initially in the study of fertility levels and trends in different countries of Europe.

Another noteworthy development in this subject was the concept of singulate mean age at marriage developed by John Hajnal (1953). It is one of the most widely cited indices of nuptiality. The proportions of single men or women obtained for each age or age group from a census or survey are used as information from a single hypothetical cohort and used to estimate the average years of life lived in the single state by men and women in the population. Hence, it is called the singulate mean age at marriage (SMAM). Subsequently a more complex marriage model was devised by Coale (1971), who observed that the variations in age-specific patterns of fertility of ever-married persons could be represented by variations in just three parameters.

In recent years, the concept of different age patterns of mortality prevalent in human populations was first recognized and used by the UN (1955) in the development of model schedules of mortality or model life tables. In these model tables, hypothetical schedules describing variations in mortality by sex and age were provided, in terms of a limited number of constants. The major application of these schedules is in the estimation of life tables from limited data and their use in the projection of future population. The UN model life tables are an example of a one-parameter system of model life tables, because in these tables only the initial starting value of q_0 needs to be chosen and the values of the other ages are mere functions of this value. Subsequently, Coale and Demeney (1966) have produced model life tables from the actual experience of developed countries, developing four families of life tables: East, West, North and South. In wider use in recent years is the logit life table system constructed by Brass (1971), which is based on the assumption that once the probabilities of surviving to each age are transformed by a logit function a linear relation exists between such logits in different populations. Hence, from a standard life table, and knowing these constants of logit transformation, the life table of a specified population can be constructed. The work on population projections for different countries of the world was initiated by the United Nations Population Division in 1951, though projections were made by earlier scholars such as Frank Notestein for the United States of America and a few other countries and by Kingsley Davis for India and Pakistan. The component method of population projection was refined over the years and in recent years, population projections for different subgroups of the population and probabilistic projections have been pioneered by the International Systems Institute at Vienna (1978). The 1980s saw rapid developments in various indirect methods of estimation of fertility, mortality and migration. A theory of 'generalized stable population' in which different age groups grew at varying but constant rates over time, hence also called the variable 'r' method was developed by Preston and Hirouchi (1981) and used to estimate fertility and mortality rates in non-stable conditions under some assumptions.

NOTES

Because of paucity of reliable data from the vital registration systems of many developing countries on births and deaths occurring from year to year, a number of indirect methods to estimate fertility and mortality rates were developed beginning from the late 1950s. These techniques were first developed for the African countries, especially Kenya, for which Brass (1961) developed a series of methods to estimate fertility and mortality from data on children ever born and children living for women classified in different age groups. A series of other related indirect techniques were also developed by other scholars and a UN manual on 'indirect methods of estimation of demographic parameters' was brought out in 1983 (Manual X).

The development of 'population studies' or so-called substantive demography for which the 'formal demography' or 'demographic techniques' were innovated over time owes its origin to the essay written by Thomas Malthus in 1798 titled, *An Essay on the Principle of Population*, which focused attention on the competition between population growth and means of subsistence or food production. In essence, Malthus' principle of population was that human populations tended to increase at a more rapid rate, the geometric rate, while the food supply needed to sustain them could at best grow at an arithmetic rate. He pointed out that any series of numbers growing at the geometric rate will eventually outgrow the series growing at the arithmetic rate and when the population growth exceeds the level that can be sustained by the given level of food availability, further growth in population will be kept in check by positive checks, such as famines, wars and epidemics. He also pointed out that unless population growth is kept in check by prudent preventive checks such as celibacy and the use of natural methods of family planning within marriage, there would always be a conflict between population growth and the means of subsistence. Malthus's essay generated innumerable writings and discussions and he became one of the most frequently cited authors in history. A study of the literature on population problems from Malthus to Simon (1981), who argued that population growth contributes to innovations, entrepreneurship and development, reveals that our understanding of the population phenomenon in terms of its impact on development and environment remains vastly incomplete and is still in the stages of development. Over time, of course, the scope of explorations into the phenomena has progressively widened; several different approaches have been and are being tried, and many more are indicated for future tests. In general, the role of formal demography is to collect, compile and process population data to suit the requirements of other social sciences; it is a basic discipline in the service of different disciplines dealing with human life. However, newer methods of analysis are being constantly developed because of felt needs by scholars. Demography does not however wait upon these other disciplines to exercise their demands; it has on its own sought to develop knowledge about major ramifications of the population phenomena, which other disciplines have used to their advantage. Quantitative analysis of population has not indeed been the prerogative of demographers who have generally drawn from contributions from other social sciences; in fact, the development of demography is richer for the reason that scholars from several other disciplines have made valuable contributions to its development, both in the process of accumulating its store of knowledge and its technical knowledge. Before discussing various sources of data and techniques of demographic analysis, it is important to understand some of the basic concepts on the types of variables used and the nature of their underlying relationships. These are discussed in the next chapter.

NOTES

CHAPTER II

Basic Concepts and Measures

2.1 Universe and Variables

As demography is an empirical science, it needs appropriate data, statistical investigation and scientific inference. Any statistical investigation is concerned with one or more characteristics of a set of individuals or objects. This group of objects may be animate or inanimate, existent or non-existent, real or hypothetical, finite or infinite and is known as *population* or *universe* in a statistical sense. In statistics, any representative part of a population or universe is called a *sample*. Thus, in the statistical sense, a sample of oranges, animals or birds can be considered as a sample of a corresponding universe or population. In a demographic sense, however, the term usually refers only to human populations.

Often in the study of demography, it is assumed that some underlying, unobservable process is occurring and this underlying process can be better understood by studying the characteristics of the population. Generally, the value representing a characteristic of a population may vary from individual to individual and over time and space. This type of characteristic is called a *variable*. For example, age of individuals in a population varies from individual to individual and is normally considered to be a variable. Variables are usually classified into two types, discrete and continuous. The notion of continuous and discrete variables depends on mathematical concepts that are beyond the scope of this book. A continuous variable is a variable that takes on an infinite number of values; furthermore, it can assume a value between any two specified values, however small and close they may be. Discrete variables also assume a finite or infinite number of values, but may not assume a value between two given values. Integers are examples of discrete values. As far as the study of demography is concerned, it generally deals with discrete variables. The concept of continuous variables is used only in a few applications, such as in the study of stable or stationary populations. Variables may be grouped into four categories according to the nature of scale of values of the variables. These are nominal, ordinal, interval and ratio scales.

NOTES

In the nominal scales, the variable assumes only a limited set of values, which do not have any hierarchical relationship with each other. The variables of religion and sex of an individual come under this category. The binary scale is a particular subcategory of the nominal scale in which a variable takes only two possible values, zero or one. Variables that can be represented in an ascending or descending scale belong to the category of ordinal variables; for example, socio-economic status – normally represented by high, medium and low – falls into this category. Variables whose values can be represented as falling within specified class intervals comprise the third category of interval scales. Income and age are examples of this type of variables.

When a scale consists not only of equidistant points but also has a meaningful zero point, then we refer to it as a ratio scale. If we ask respondents their ages, the difference between any two years would always be the same, and 'zero' signifies the absence of age in completed years at birth. Hence, a 100-year-old person is indeed twice as old as a 50-year-old one. Ratio scales are used to gather quantitative information, and we see them perhaps most commonly when respondents are asked for their age, income, years of participation etc. In order to respect the notion of equal distance between adjacent points on the scale, each category must be of the same size. Therefore, if the first category is $0–19,999, the second category must be $20,000–39,999. Obviously, categories should never overlap and categories should follow a logical order, most often increasing in size. Ratio scales are the most sophisticated of scales, since they incorporate all the characteristics of nominal, ordinal and interval scales. As a result, a large number of descriptive calculations are applicable when this scale is used.

NOTES

2.2 Rates, Ratios and Proportions

The main aim of demographic analysis is generally to identify and quantify, as precisely as possible, the various demographic phenomena through a variety of measures and indicators with a view to using them to make comparisons between populations and in a given population over time. While a measure is a definitive quantitative value of the phenomenon being studied, an indicator is a proxy or an approximation to a measure. Thus, while income is a measure of the economic status of an individual, the type of house he/she lives in is an indicator of his/her income. These measures and indicators are also used to gain an understanding of the factors underlying the changes and differentials in them by relating them to various socio-economic conditions of the population. To deal with this problem the demographer often relates the number of events such as births, deaths, marriages, movements etc., with the size and characteristics of the base population from which such events arise and thus computes various ratios and rates. We cannot make comparisons of the demographic structure and processes between populations based only on the number of events occurring in them. There is a need for denominator data to which the events should be related before valid comparisons can be made. Information on the absolute numbers of events can convey only a limited meaning in demographic analysis, as the number of these components is closely related to the size and structure of the population. Such an analysis is called *numerator analysis*. However, most of the techniques of demographic analysis are dependent to a large extent on the nature of measures and indicators selected for the study of the phenomena and related to the population.

Several types of relative numbers are used in demographic analysis and are mostly derived from the simple concepts of ratios, rates and proportions. Hence, a general notion about these concepts is required for understanding the demographic measures.

Ratios and Proportions

A ratio is the result of dividing the size of one of the two non-overlapping groups possessing some common characteristics by that of the other. An example is the number of males in a population divided by the number of females, which is called the *sex ratio*. A ratio is an index helpful in comparing the relative strength of each group in populations at different times and territories. Such ratios can be computed for different age groups and in the Indian context, when computed for children under age 7, it is called the *child sex ratio*. Proportion is a relative number that expresses the size of one subgroup to the total of all subgroups, which is equated to 1. When the sizes of all subgroups are expressed as percentages, the result is called a *percentage distribution*. In other words, proportion is a special type of ratio in which the numerator is included in the denominator. If the characteristic under consideration is age, the distribution of persons at each age is called the *age distribution* or the age composition of the population.

Example of ratio

$$\text{Child sex ratio} = \frac{{}_6P_0^f}{{}_6P_0^m}$$

where

${}_6P_0^m$ = number of male children in a population at a specified time

${}_6P_0^f$ = number of female children in a population at a specified time

In 2011, child sex ratio in Kerala was $\left(\frac{17366387}{16021290}\right) \times 1000 = 1084$

Example of proportion

$$\text{Proportion of population aged } 0-6 = \left(\frac{\text{population of } 0-6}{\text{total population}}\right)$$

In Kerala, proportion of population aged $0-6$ in 2011 $= \left(\frac{3322247}{33387677}\right) = 0.100$

NOTES

Rates

The most commonly used demographic measures are rates. They express the number of events, say E, which occur in a population of size P in a given period of time, which is usually a year, as a fraction E/P. Rate is a measure of the speed of occurrence of events in the population. Thus the concept of rates is associated with dynamic phenomena such as growth, birth, death and movement. A rate refers to the occurrence of events over a given interval of time. In demographic applications, rates are normally considered as indicators of what is known in statistical parlance as occurrence/exposure, measures where the numerator is the number of events that have occurred in a population and the denominator is the duration of exposure of the population to such events. They contain a count of the number of events occurring within a defined time period in the numerator, and in the denominator, an estimate of the population during the middle of that time period. Simple illustrations of the rate are the crude birth rate (CBR) which is defined as CBR = 1000 * (B/P), where B denotes the number of live births in a specified population in one year and P denotes the size of the population at the middle of the year. It is common practice to express CBR per 1000 population.

The crude death rate (CDR) is defined as CDR = 1000 * (D/P), where D denotes the number of deaths in a specified population in one year and P denotes the size of the population at the middle of the year. Again, CDR is defined per 1000 population.

Example of rates

CBR for Kerala in the year 2001 was $\left(\frac{539427}{31841374}\right) * 1000 = 16.94$

Combining various indicators into a unitary index

NOTES

In modern cross national comparisons of the levels and pace of change in development, it has become necessary to combine many indicators such as infant mortality rates and literacy rates into a unitary index to facilitate comparisons of progress of the same country over time and across countries at a given point in time. This is a necessary consequence of globalization. In such a case, every indicator of a country is standardized into an index varying between 0 and 1 or 0 and 100 (if considered as percentiles).

An indicator X_i of a country is standardized as

$$\frac{X_i - X_{min}}{X_{max} - X_{min}}$$

if the variable is expected to increase in the desired direction towards the maximum (such as income or literacy levels), and

$$\frac{X_{max} - X_i}{X_{max} - X_{min}}$$

if it has to decrease to its lowest possible value (such as the infant mortality rate).

This standardization is done usually on a 0 to 1 scale. After standardizing the various indicators, they are combined into one index either as an arithmetic mean of different indices or a geometric mean.

For the construction of the human development index (HDI) in 2010 the following procedures were used:

The HDI combined three dimensions until its 2010 report:

- Life expectancy at birth, as an index of population health and longevity
- Knowledge and education, as measured by the adult literacy rate (with two-thirds weightage) and the combined primary, secondary and tertiary gross enrolment ratio (with one-third weightage).

- Standard of living, as indicated by the natural logarithm of gross domestic product per capita at purchasing power parity.

In Human Development Report (2010), the UNDP began using a new method of calculating the HDI. The following three indices were used:

1. Life expectancy index (LEI) = $\frac{LE - 20}{83.4 - 20}$

2. Education index (EI) = $\frac{\sqrt{MYSI \times EYSI}}{0.951}$

   2.1 Mean years of schooling index (MYSI) = $\frac{MYS}{13.2}$

   2.2 Expected years of schooling index (EYSI) = $\frac{EYS}{20.6}$

3. Income index (II) = $\frac{\ln(GNIpc) - \ln(100)}{\ln(107.721) - \ln(100)}$

Finally, the HDI is the geometric mean of the previous three normalized indices:

$$HDI = \sqrt[3]{LEI \,.\, EI \,.\, II}$$

where

LE : Life expectancy at birth

MYS : Mean years of schooling (years that a 25-year-old person or older has spent in schools)

EYS : Expected years of schooling (years that a 5-year-old child will spend with his/her education in his/her whole life)

GNIpc : Gross national income per capita at purchasing power parity

NOTES

2.3 Person-Years of Life or Exposure

The mid-year population is used in the denominator as an approximation of the person-years of life or exposure to the risk or occurrence of the event considered in the numerator computed over a year. This concept of person-years of life or exposure to risk is a fundamental one in the computation of various rates and demographic measures. It denotes the total number of years of life or exposure to risk of occurrence of an event experienced by a given population during a given time interval. For example, if the interval considered is one year, and if the size of the population is P_1 at the beginning of the year and P_2 at the end of the year, and if the change in the population over the year is due to deaths, immigration and out-migration and if they occur uniformly during the year, then it can be proved that the number of person-years lived by the initial population P_1 during the year will be $(P_1 + P_2)/2$. If however, the events that change the population size over time, births, deaths and movements do not occur uniformly over time, that is, unequal numbers occur in equal intervals of time, then the mid-year or mid-period population will no longer be equal to the person-years of life or exposure during the period. In most demographic studies, the mid-year population is used as an approximation to the population at risk during the year or period under consideration. It should be remembered that this may not always be the case, as in the case of deaths in the first year of life or at older ages when frequency of occurrence of events over time intervals within a period change drastically. While comparing the rates of occurrence of births, deaths and marriages based on the mid-year population as the denominator, it has to be recognized that such comparisons are strictly valid only if they are based on rates or ratios taking into account the population exposed to the risk of such events and controlling for the duration of occurrence of such events. Comparison of only numerator values, even when the population sizes are the same, may be invalid if the years of exposure to the risk are different between the populations. It is also possible that in some situations the periods of exposure to risk of an event may not be continuous but can occur with gaps, as in the case of a patient getting a new infection in a hospital which will depend on the number and duration of hospitalization.

NOTES

However, in some situations when the events are drawn at random from the population and when there is a strong association between selected characteristics of the events and the rate of occurrence of the event in the population, a comparative analysis based on the characteristics of the events may be valid. For example, there is a strong association between the distribution of births by birth order and fertility levels in a population. Populations that have higher fertility rates have a higher proportion of births in the higher birth order, that is, order 4 and above. Under such circumstances, an estimate of the fertility rates in a population may be obtained from just the proportion of births of order 4 and above in the population. Such an analysis is typical of numerator analysis.

In the domain of public health and epidemiology, there are two commonly used rates which are valuable in many types of analysis. These are the prevalence rate (P) and the incidence rate (I) of a disease and are defined as follows:

$$P = \frac{\text{Number of cases of a specified disease prevailing in a population at a particular point of time}}{\text{Population of the area exposed to the risk of the disease at that time}} \times 100$$

$$I = \frac{\text{Number of new cases of the disease occuring in a given period of time, usually a year}}{\text{Duration of the exposure of the population to the disease in the given period or the mid-year population of the area, as an approximation}} \times 100$$

It is obvious that in the case of diseases that have a short duration, such as the common cold or diarrhoea, the incidence (I) of the disease in a year, or the number of new cases occurring in the population is very high, but if we survey the population to identify the number of persons with such diseases suffering at a given point of time or the prevalence of the disease P, it may be quite low. A large number of cases crop up frequently, have a short duration and get cured on their own. On the other hand, if we consider diseases that have a long duration of sickness such as tuberculosis or leprosy, the number of new cases (I) occurring in a population may be relatively low, but the number of prevalent cases at a particular point of time P may be high. Even though there are only a few cases occurring each year, since no case gets cured quickly and lingers for a long time, there is a build-up of cases in the population and the prevalence rate of the disease rises in the population. If the incidence rate of a disease in a population remains constant and the duration of the disease also remains the same over time, then the prevalence rate of the disease remains constant. In such a situation, there is a simple but useful formula that links the incidence, prevalence and duration of the disease.

$$P = I \times D$$

where

P = prevalence rate of the disease

I = incidence rate of the disease

D = duration of the disease

NOTES

Though this is a simple formula, it has useful applications in demography and public health. It can also be used to calculate statistics for phenomena other than diseases. For example, if P is the number of pregnant women in a population at a given point of time and B is the number of births in the population (which in this case denotes the end of the episode of disease and not the beginning of the episode of disease), then

$P = B \times 0.75$ since in this case D is almost constant at nine months or 0.75 years.

Another example is the estimation of the duration of breastfeeding in a population. If at any given time we find in a population N mothers breastfeeding their children (on that day), which is the same as the prevalence, and there are B births in the population in the previous year, then the average duration of breastfeeding in the population is given by the equation

N (prevalence) = B (incidence) × D (average duration of breastfeeding)

or

$D = N/B$

2.4 Basic Demographic Equation

The most basic equation that expresses the change in population over time in a simple form is known as the *basic demographic equation* or the *balancing equation*.

It is the decomposition of the population change into its components. It is expressed as follows:

$$P_2 = P_1 + (B–D) + (I–E)$$

where

P_2 and P_1 are population at two different points of time

B = births

D = deaths

I = immigrants

E = emigrants

The events contributing to the equation fall into two categories: B–D gives the natural increase, while I–E stand for the net migration. Hence, the equation can also be expressed as

$$P_2 = P_1 + NI + NM$$

where NI is natural increase and NM is net migration.

NOTES

If we are interested in projecting the population to a future date, we can use this equation in principle by making assumptions about the future births, deaths and migration. Also, suppose we have two successive population counts and we also have vital statistics on births and deaths, then this equation can be used to estimate the net migration during the period.

Table 2.1 provides data on the rates of growth, natural increase and migration for the years 1901 to 2001. These rates have been converted to actual number of births, deaths, natural increase and migration for the decades 1951 to 2001 and these figures are given in Table 2.2. We find that in every decade the natural increase of population in Kerala (births

Table 2.1: Population Growth and Migration in Kerala, 1901–2001

Year	Census Population (000)	Growth Rate	Natural Increase Rate	Migration Rate
1901	6,396	–	–	–
1911	7,148	1.11	1.08	0.03
1921	7,802	0.88	0.84	0.04
1931	9,507	1.98	1.86	0.12
1941	11,032	1.49	1.49	–0.01
1951	13,549	2.06	2.16	–0.11
1961	16,904	2.21	2.41	–0.20
1971	21,347	2.31	2.47	–0.16
1981	25,454	1.75	1.96	–0.22
1991	29,011	1.32	1.63	–0.31
2001	31,839	0.91	1.20	–0.27

Sources: Mari Bhat and Irudaya Rajan (1990); Irudaya Rajan and Zachariah (1998)

Table 2.2: Components of Population Growth, Kerala, 1951–2001 (in ’000)

Decade	Population Growth	Births	Deaths	Natural Increase	Net Migration
1951-61	3,355	6,684	3,000	3,684	–330
1961-71	4,444	7,095	2,333	4,762	–318
1971-81	4,106	6,398	1,819	4,579	–473
1981-91	3,620	6,071	1,707	4,364	–744
1991-01	2,765	5,482	1,827	3,655	–890

minus deaths) was higher than the actual growth rate recorded between two consecutive censuses of the state between 1951 and 2001. This implied a net out-migration from Kerala which has been increasing steadily over the decades. For example, during the decade 1951–61 the natural increase was 3.68 million but the observed increase between the two censuses was only 3.35 million implying a net out-migration of 0.33 million (330,000). During the decade 1991–2001 the natural increase was 3.66 million while the observed increase was only 2.77 million, implying a net out-migration of 0.89 million (890,000). Part of this sharp increase was out-migration to other states within the country but rest part of it was emigration to other countries, especially to the Middle East since 1970s.

NOTES

The balancing equation in demography, though simple, can be used for more detailed analysis if we can consider the equation separately for different groups, such as gender or religion, and also by age groups when the analysis becomes a little more complex.

The raw materials of demographic analysis are data collected in censuses, surveys and registration systems. Information about population is collected in two main ways – by enumeration at a point of time, and by recording events as they occur over a period. Censuses and surveys are examples of the first type of data collection and provide ‘stock’ data, while birth or death registration records are examples of the second type and provide ‘flow’ data. The different sources and ways of collecting data are discussed in the next chapter.

2.5 Population Change

Population change is measured as the difference in population size between two points of time (i.e. two specific dates). It can be expressed in terms of absolute change, percentage, average annual absolute change, geometric or exponential growth rate. It can refer to changes in size, distribution or composition, or to any combination of the three.

2.5.1 Measures of population change

Let P_2 = population at later date, P_1 = population at the earlier date, and y = number of years between P_2 and P_1. Then

Absolute change = $P_2 - P_1$

(Negative sign indicates a population loss)

Percent change = $[(P_2 - P_1)/P1] * 100$

Average annual absolute change = $|(P_2 - P_1)|/y$

Linear growth rate r = $[(P_2/P_1) - 1]/y$

Geometric growth rate r = $(P_2/P_1)^{(1/y)} - 1$

Exponential growth rate r = $[\ln (P_2/P_1)]/y$,

where ln is the natural logarithm

Note: Measures of population change always refer to a specific population, a specific geographic area and a specific period of time

2.5.2 Components of population change

There are only three components of population change, namely, births, deaths and migration. Population grows through the addition of births and in-migrants. Population declines through the subtraction of deaths and out-migrants. Understanding these three demographic processes is essential to understanding the nature and causes of population change.

NOTES

Fertility

Fertility refers to the 'occurrence of a life birth (or births)' and is determined by a variety of biological, social, economic, psychological and cultural factors. Biological factors determine the physiological capacity to reproduce. Other factors determine perceptions regarding the costs and benefits of children. Contraceptive availability and effectiveness play a role in affecting the ability to control the number and timing of births. Fertility rates have declined over the last two centuries in Europe, North America and other high-income countries. Some of the developing countries have also experienced a drastic decline in their fertility rates in the recent past. A detailed discussion of the fertility component of population growth is given later.

Mortality

Mortality refers to the 'occurrence of deaths in a population'. Changes are determined primarily by changes in a population's standard of living and advances in medicine, public health and science. Low-income countries typically have higher mortality rates than high-income countries. Education also has a substantial impact on mortality rates. Mortality rates have declined tremendously during the last two centuries in Europe, North America and other high-income countries. Mortality rates have also declined dramatically in many low- and middle-income countries, especially during the last 50 to 60 years. Mortality rates have varied considerably more among low- and middle-income countries than among high-income countries.

Migration

Migration refers to the 'change in one's place of residence from one political or administrative area to another'. It refers solely to changes in place of usual residence, thereby excluding all short-term or temporary movements such as commuting to work, visiting friends or relatives and going away on vacation.

Migration is distinguished from local mobility. At the aggregate level, factors affecting migration are area-specific characteristics such as wage rates, unemployment rates, cost of living and amenities such as climate and recreational opportunities. At the individual level, factors affecting migration include factors mentioned for the aggregate level, as well as a host of personal characteristics such as age, sex, education, occupation and marital status.

Local mobility refers to 'a change of address within a particular political or administrative area'. Gross migration refers to 'the total number of migrants into or out of an area'. Net migration refers to 'the difference between the two (into an area minus out of same area)'. Internal (or domestic) migration refers to 'the changes in residence from one place to another within the same country' and is greatest at the sub-national level. International (or foreign) migration refers to 'the changes in residence from one country to another'. International migration is a minor component of population growth for most countries. Migration affects the total population of an area.

2.6 Demographic Transition Theory

There are three important concepts which carry the adjective 'demographic' before them and which are popular within and outside demography. These are 'demographic transition', 'demographic dividend or bonus' and 'demographic momentum'. These are briefly described in the following paragraphs and they are crucial in understanding the significance of some demographic parameters.

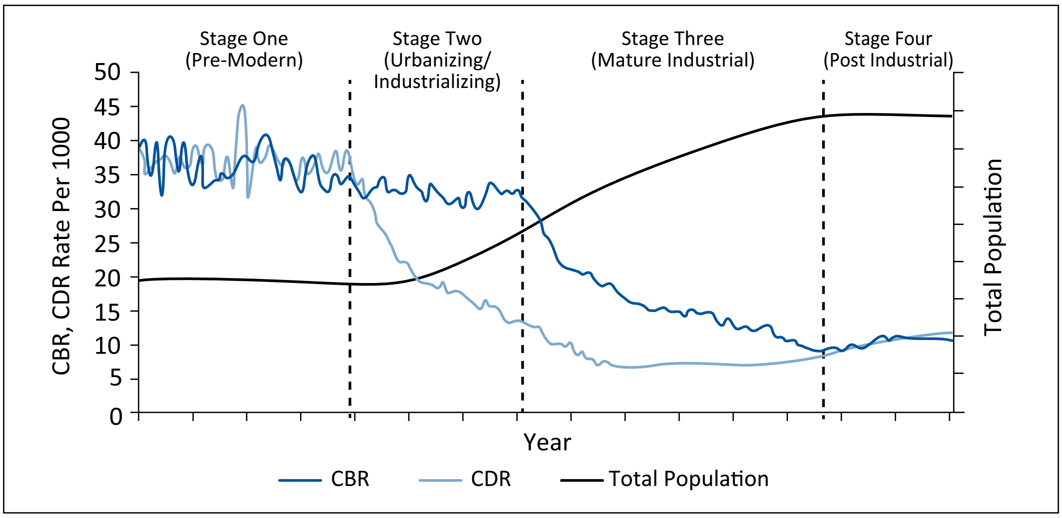
NOTES

2.6.1 Demographic transition

Studies on demographic transition probably occupy the largest segment of the literature on demographic research. The term 'demographic transition' connotes the transition of a human population from a high mortality–high fertility situation to a low mortality–low fertility situation over a period of time. In such a transition, generally the mortality rates start declining first on a secular basis and the fertility rates decline after a time lag. In the basic demographic transition model (DTM), four stages are observed: stage 1, the pre-transition stage when both fertility and mortality rates are high and population growth rates are very low, which has been the condition of human population for millennia; stage 2, when mortality rates decline with fertility remaining almost constant or even rising a little bit or declining more slowly than the fertility rates (during which period the growth rates of the population keep rising fast); stage 3, when after a time lag fertility rates start a faster downward trend than mortality and the growth rate slows down; and finally, stage 4 when both the fertility and mortality levels are low at or below replacement level. While in the first stage the CDRs fluctuate, in the fourth stage the CBRs fluctuate (Thompson, 1929) (See Figure 2.1).

There are many elaborations of the basic version of the DTM, some identifying five stages or even six. The differences between these models are essentially in the conceptual delineation of the last stage when the birth and death rates are low, below replacement levels of fertility. In the five-stage model, this stage is considered to have two segments: in the new stage 4 both the birth and death rates are low with the birth rates oscillating

Figure 2.1: Demographic Transition Model



NOTES

around the replacement level and the population growth rates are low but positive; in stage 5 the birth rates continue to fall below replacement level and death rates rise because of the ageing populations, and the growth rates are negative. The populations can even fear extinction unless padded up by fertility rise or large-scale immigrations. The Second Demographic Transition is an elaboration of this specific situation of stage 5, which is discussed in a subsequent section.

In the pre-industrial era in the developed countries and until the early decades of the twentieth century in the developing countries, the mortality rates tended to be very high because of famines, epidemics and wars and varied widely from time to time because of the unexpected impact of these calamities. Only when these were brought under reasonable control did secular declines in mortality occur. In some segments of the population the declines occurred at a faster pace and the differentials in mortality among socio-economic groups widened. The nobility and the wealthier groups experienced faster declines in their mortality rates than the poor and underprivileged. Later, with continuing declines in overall mortality rates such differentials tended to narrow down though they persist, in a smaller magnitude, to this day. Similar phenomena happened in the fertility transition that occurred at the second stage. Widening and narrowing of socio-economic differentials in fertility and mortality is a hallmark of demographic transition.

Before the onset of secular declines in mortality, the fertility rates were close to the natural fertility levels (where there is no deliberate control by the couples to limit or space their children). Louis Henry (1961) estimated the natural fertility levels of the French population at between 9 and 11 children per woman. Though such high levels were observed in special groups such as the Hutterites in Canada, for most of the populations, in the pre-transition stage the total fertility rates (TFRs) hovered around 6 to 7 children per woman.

There are still large uncertainties as to precisely when, where and how the demographic transition began. The trigger for declines in mortality seems to have been caused by the industrial revolution in Europe in the late seventeenth century. Studies based on available data and using indirect methods by the Princeton group of scholars concluded that the secular declines in mortality started around 1800 in France and spread to other countries of northern Europe. After a few decades fertility rates started to decline mainly because of use of condoms (called *French leather*) and *coitus interruptus*. This transition spread rapidly to all the European countries and to the other countries with immigrants of European origin and seems to have been completed by the mid-twentieth century, when most of these countries achieved replacement or below-replacement levels of fertility. Among the non-European populations, Japan was the first to start its demographic transition almost five decades later, around 1850, but by 1950 had completed the transition. The low mortality–low fertility situation has been prevailing in these countries for over four decades now, with the population growth rates in many of these countries being negative and the prospects of continuing loss of population becoming a matter of serious concern for policy makers.

Among the other developing countries the transition seems to have commenced almost a century later, possibly first in India by 1920 when the secular declines in mortality commenced, followed by other developing countries and also in China, Cuba, South and North Korea, Taiwan, Singapore, Malaysia and Sri Lanka where the transition occurred at an accelerated pace. Because of very rapid declines in mortality and time lags in the declines in fertility, the world has seen a historically high increase in human population growth during the past two centuries. The figures commonly quoted is that it increased from about 1 billion in 1800 to 5 billion in 1987 to 6 billion in 1999, 7 billion in 2011 and is expected to rise to 9.5 billion by 2100. While it took centuries, or even millennia, for the world human population to reach the first billion, in recent times a billion is added every 12 years. The population base has become so large that even with very small growth rates, large increases in populations will continue to occur in the future. The annual population growth rate increased from 0.50% around 1800 to 1.80% around 1950. It has been steadily declining thereafter to 1.22% in 2000 and is expected to decline to 0.33% by 2050 and to 0.04% by 2100. In the less developed countries the transitions occurred at a faster pace and at present 90% of the addition to the population are due to the developing countries.

NOTES

In the developing countries mortality declines have occurred at a much faster pace than in the developed world, because they had the benefit of the technological developments already made in the west and also because of national programmes of public health and family planning implemented by them with the support of international organizations such as the UN, World Health Organization (WHO) and the United Nations International Children's Education Fund (UNICEF). It took almost a century for Europe to reach a life expectancy of 45 years from around 25 years in 1800 or an increase of 0.2 year per year. On the other hand, in India life expectancy increased from 24 years in 1920 to 67 years in 2008 or an increase of 0.48 year per year, while in China the increase in life expectancy was higher – from 41 during 1950–55 to 73 during 2005–2010 or an increase of 0.58 year per year. Similar declines in the fertility rates were also observed because of availability of modern contraceptive methods and the implementation of national programmes of family planning by the governments in the developing world thus achieving replacement levels of fertility within three to four decades, compared to what was achieved over a century and a half in the developed counties (See Table 2.3).

Table 2.3: Global Population Trends over the Transition

Year	Pop. Size (Billion)	Life Expectancy	TFR	Growth Rate (Annual %)	Pop. <15 Years (%)	Pop. >65 Years (%)
1700	0.68	27	6.0	0.50	36	4
1800	0.98	27	6.0	0.51	36	4
1900	1.65	30	5.2	0.56	35	4
1950	2.52	47	5.0	1.80	34	5
2000	6.07	65	2.7	1.22	30	7
2050	8.92	74	2.0	0.33	20	16
2100	9.46	81	2.0	0.04	18	21

Source: Ronald Lee (2003)

NOTES

It is observed that during the 60-year period from 1950 to 2010, the population of the world as a whole increased from 2529 million to 6908 million or by 2.73 times. The increase during the same period in the 'more developed countries' was from 812 million to 1237 million or by 1.52 times compared to an increase from 717 million to 5671 million or by 7.90 times in the 'less developed countries'. The population growth during the second half of the twentieth century is essentially due to growth of population in the 'less developed countries'. During 1950–55 the TFR and life expectancy of females in the less developed regions of the world were 6.0 and 41.8 years respectively compared to 2.82 and 68.4 years in the developed countries, a surplus gap of 3.18 children and a deficit in female life expectancy of 26.6 years; by 2000–05 this gap came down to 1.31 children and 13.4 years, and by 2045–50, according to 'medium variant' projections by the UN, this gap will further reduce to 0.25 child per woman and 9.1 years in life expectancy. There is no doubt that the human race as a whole is rapidly converging to a uniform demographic condition of fertility and mortality and will be in the final stage of transition globally within the next two decades. There may be differences of opinion on the time span but the process is ongoing.

2.6.2 Factors underlying demographic transition

The first comprehensive statement of the possible factors underlying the demographic transition in the developed world was given by Notestein (1949). The underlying causal mechanisms, which triggered and sustained the declines in mortality and fertility, though not yet fully understood, are generally referred to by him as forces of 'industrialization', 'modernization', or 'development'. These terms include modern education especially of women, transformation of an agricultural society to an industrial society, rise in per capita income, personal hygiene, improved nutrition and provision of public health and medical services and a secular view of life. In the developed societies the declines in mortality occurred with the development of various vaccines and drugs and various innovations in the society as a part of industrialization including the use of soap for cleaning and washing and had very little to do with the governments of the time.

The declines in fertility followed after a time lag because of the perceived higher survival chances of children either by avoiding pregnancies or terminating them if unavoidable.

Governments of the time had no role in this transition; if at all, they only condemned the use of any form of contraception or abortion on religious grounds. The role of public policies, especially on the provision of organized public health and family planning services, were not among the earlier set of factors contributing to demographic transition.

Later however, in many developing countries, such as the countries in the Commonwealth after they achieved political independence from the colonial rule, the newly formed national governments assumed dominant roles in the development of their populations and launched various policies and programmes for the betterment of the economic and health conditions of their people including public health programmes and national family planning programmes. India was the first country to launch a family planning programme as early as 1951. High population growth rates were considered one of the key hurdles to economic development. Though China launched its development strategy on Marxian lines and though Marxian theory considered population an asset and was against state measures to reduce the population growth rates, China adopted an aggressive population policy route, first through the policy of *Wan, Ki, Shu* (later, longer and fewer) in 1972 and later with the one-child policy in 1979. As a result of these policies it succeeded in reducing its fertility levels by half, from over 5.8 children in 1970 to 2.7 in 1980 and to 1.6 in 2010.

While in most of the developed countries and in a number of developing countries the demographic transition was complete by the turn of the twentieth century, many less developed countries are still in the second stage and will enter the final stage in the next few decades. Public policies and government programmes for reduction of fertility and mortality assisted financially and technically by the bilateral donor agencies and international organizations have become major driving forces behind such an accelerated transition.

NOTES

2.7 Demographic Dividends

The demographic transition made significant changes not only in lengthening life, reducing fertility levels, increasing and later reducing population growth rates but also altered the age distribution, especially the ratio of economically productive population to dependent population. This enabled diversion of household and public expenditure from consumption-related items to savings and investment, leading to accelerated economic growth. The first major empirical study on the effects of high dependency ratios on savings and investments at the household and the governmental level was made by Coale and Hoover (1954) in their study on the effects of population growth on economic development in India and Mexico (Coale and Hoover, 1952). They empirically, but more intuitively, proved in their study that reduction of fertility through national programmes of family planning would contribute to overall economic development by increasing savings and investment at the household and national levels. Their study paved the way, by providing economic justification, for the launch of the national programmes of family planning in India and later in a number of other developing countries as a part of their developmental strategy. They did not use the term ‘demographic dividends’, which came into use in the 1990s.

The necessity for more systematic studies on the effects of changes in the dependency ratios on the economy arose in the 1990s because of what was considered an economic miracle in the East Asian countries, especially China, Korea, Singapore and Taiwan, during the late 1980s and 1990s after these countries experienced very rapid declines in their fertility and mortality levels. Their economic dependency ratios declined faster and the impact of these on their high economic growth had to be disentangled, leading to what is termed 'demographic bonus', 'demographic dividend' or 'window of opportunity' (Bloom and Williamson, 1998). While Coale and Hoover justified the launch of national family planning programmes in the 1950s in developing countries as programmes that would contribute to economic development, later authors measured the contribution of such programmes to economic development because of reductions in dependency ratios. These terms are also used to connote the period of time when the dependency ratio in a population started to decline because of earlier declines in their fertility levels, until such time when it started to rise again because of the rise in the proportion of older persons caused by continuing declines in fertility and increasing longevity. The second rise in the dependency ratios was due to the fact that the pace of increase in the old age dependency ratios was faster than to just compensate for the declines in the youth dependency ratios. Usually the period of demographic bonus is dependent on the pace of decline in the fertility levels of a population. If the switch to small families is fast, the demographic bonus can give considerable push to development as happened in China, beginning in the early 1980s with rapid declines in the fertility levels of the population because of its one-child policy.

NOTES

Many studies have empirically shown that the reduction in the child dependency ratios in a population, just by reducing the consumption expenditure at the household level and at the national levels, contributed to 10–20% of the increase in the per capita incomes of the populations of East Asia and South Asia, which have grown fast economically during the past three decades.

If the period of demographic bonus is considered a window of opportunity, and if public investments during this period are made wisely in health care and secular education with emphasis on skill development of the population as happened in the East Asian countries, the contribution of this reduction in dependency ratios could be higher. The Republic of Korea, for example, increased net secondary school enrolment from 38% to 84% between 1970 and 1990 while more than tripling expenditure per secondary pupil. Countries which failed to make such investments during the periods of demographic bonus did not record such high economic growth rates. It is estimated that up to 40% of the growth of per capita income can be derived as demographic dividend as in China and South Korea. During 1965–1990, GDP per capita in East Asia grew annually by an average of 6.1%; the changes in the age structure contributed an estimated 0.9 to 1.5 percentage points. Changing demographic structures now present similar opportunities and challenges in the South Asian countries and there is evidence that they are also poised for similar growth in the coming decades. Table 2.4 presents data on the reductions in the total dependency ratios (TDER) – ratio of number in age groups 0–14 and 60+ combined to number in the age group 15–59 – in different states in India between 1961 and 2001 and the reductions in TFR (average number of children born per woman) values between 1956

Table 2.4: Changes in Dependency Ratios and TFR by States in India: Changes from 1961 to 2008

State	Total Dependency Ratios (0–14) +(60+)/15–59 as %		% Decline in TDER		TFR	% Decline in TFR	
	1961	2001	1961–2001	1951–61	2003–05	2008	1956–2008
Andhra Pradesh	84.4	65.8	18.6	5.6	2.1	1.8	3.8
Bihar	92.1	92.3	–0.2	6.2	4.3	3.9	2.3
Gujarat	91.7	66.0	25.7	6.4	2.8	2.5	3.9
Haryana	NA	77.0		7.3	2.9	2.5	4.8
Karnataka	91.9	65.6	26.3	6.4	2.3	2.0	4.4
Kerala	94.1	57.6	36.5	5.5	1.7	1.7	3.8
Madhya Pradesh	85.1	83.0	2.1	6.6	3.7	3.3	3.3
Maharashtra	85.0	69.1	15.8	5.9	2.2	2.0	3.9
Odisha	81.1	70.9	10.1	5.5	2.6	2.4	3.1
Punjab	100.5	67.9	32.7	7.0	2.2	1.9	5.1
Rajasthan	91.6	88.3	3.3	6.4	3.7	3.3	3.1
Tamil Nadu	76.1	55.9	20.2	4.8	1.8	1.7	3.1
Uttar Pradesh	87.9	91.9	–4.0	6.6	4.3	3.8	2.8
West Bengal	85.0	67.8	17.2	6.9	2.2	1.9	5.0
India	87.4	75.1	12.3	6.03	2.9	2.6	3.43
Correlation Coefficient between decline in TFR, TFR and in TDER between 1961 and 2008							0.716

Source: Mari Bhat (1989) and Census reports of 1961 and 2001
TDER - Total dependency ratio; TFR - Total fertility ratio

and 2008. For the country as a whole the TDER declined from 87.5% in 1961 to 75.1% in 2001, a decline of 12.4 percentage points. On the other hand, in the state of Kerala the TDER declined from 94.1% to 57.6%, a decline of 36.5 percentage points. The TFR in the country as a whole declined from 6.03 during 1951–61 to 3.43 children in 2008, a decline of 2.6 children per woman while for Kerala the decline was from 5.5 to 1.73, a decline of 3.8 children per woman. This decline in TDER values has significantly contributed to greater development in Kerala and similar states compared to states where the fertility levels have not declined as fast. Based on an econometric analysis studying the effects of changes in the age structure of the population on the Indian economy in recent years James (2008) concludes, "The empirical analysis vividly exhibits the positive impact of the working age population boom on economic growth. This is despite the fact that the educational achievements and the health conditions of the people are far from adequate and employment creation is below the required level".

A simple way of measuring the demographic dividend is to consider it as the difference in the growth rate of the population of working age and the growth rate of the population as a whole. It can also be measured as the difference in the growth rate of workers and that of the population as a whole. Table 2.5 gives these data at the state level between 1991 and 2001. It can be seen that the maximum dividend during the decade 1991–2001 has been observed in the states of Andhra Pradesh, Karnataka, Haryana and Himachal Pradesh and the minimum in the states of Uttar Pradesh and Bihar, where it is negative and hence a demographic burden.

Table 2.5: Demographic Bonus Derived from the Growth of Working Age Group Population and Workers for the Period 1991–2001

State	Growth Rate of Population	Growth Rate of Working Population	Demographic Bonus	Growth Rate of Workers	Demographic Bonus using Workers
Andhra Pradesh	1.4	1.9	0.5	1.7	0.3
Karnataka	1.7	2.3	0.6	2.3	0.7
Kerala	0.9	1.3	0.3	1.2	0.3
Tamil Nadu	1.1	1.5	0.4	1.3	0.2
Assam	1.8	2.2	0.4	1.6	–0.1
Bihar	2.5	2.4	–0.1	3.1	0.6
Goa	1.4	1.8	0.4	2.3	0.9
Gujarat	2.1	2.5	0.4	2.5	0.4
Haryana	2.5	3.1	0.6	5.0	2.5
Himachal Pradesh	1.6	2.3	0.6	2.9	1.2
Madhya Pradesh	2.1	2.2	0.1	2.3	0.2
Maharashtra	2.1	2.4	0.3	1.9	–0.2
Odisha	1.5	1.8	0.3	1.8	0.3
Punjab	1.8	2.2	0.4	3.7	1.9
Rajasthan	2.5	2.6	0.1	3.1	0.6
Uttar Pradesh	2.3	2.2	–0.1	2.3	0.0
West Bengal	1.7	2.1	0.4	3.0	1.3
India	1.9	2.2	0.3	2.4	0.4

NOTES

2.8 Second Demographic Transition

The countries that have completed the demographic transition, that is, those that are at or below replacement levels of fertility are faced with the prospect of negative population growth rates in addition to rapid ageing of their populations. The originally stated theory of transition implied that human beings when modernized and industrialized would prefer to marry, have children within marriage (Malthusian desire) and have two children (one son and one daughter) as a biological desire to replace themselves in the next generation. In the fourth stage of demographic transition, it was implicitly assumed that the fertility levels would fluctuate but around the replacement level of fertility with population growth hovering around zero (stationary) or slight positive growth as in the pre-transitional phase but at much lower levels of mortality and fertility. The intellectual basis of this assumption was given by Aries in 1962 and Easterlin in 1973. The innate value of a child to the parents and the cyclical nature of the wages in labour markets based on supply and demand factors were supposed to underlie the fluctuations behind fertility below and above replacement levels in the fourth and final stage of transition. However what has happened during the post transitional stage in many of the developed countries is that fertility levels are continuing to decline well below replacement and the growth of white populations in many European countries is negative. Theories have been advanced to explain why their fertility levels continue to fall, and why there is a decreasing value of children in societies, preceded by declining rates of marriage and increasing proportions of children born outside wedlock. This phase was termed 'the second demographic transition' by Van de Kaa and Ron Lesthaeghe in a paper in 1986. They argued that the value of marriage and children declines because the adult population tends to pursue goals higher than materialistic pursuits, mainly self-actualization, as postulated in Maslow's theory of hierarchy of needs. In the first and second stages, parents beget children for lineage according to natural fertility (as God's gifts). In the third and fourth stages of transition, the parents desire to have better quality children in terms of education and health, and fertility regulation methods are adopted in the trade-off between quantity and quality. This desire to have better quality children is seen to be the main motivating factor behind contraceptive use and family limitation. With rising individualism and disenchantment with materialism in the western societies, children lose their importance in the lives of adults, with the need for self-actualization becoming more important for many individuals. Along with religiosity, marriage, as a scared institution within which procreation and child rearing is expected to take place, is losing its hold on the populations (Lesthague, 2010). In many of the western societies which have achieved low fertility and where this has happened, the percentage of births outside marriage is rising sharply, and it is predicted that this trend will envelope the whole human race in the coming decades. In the United Kingdom in 2005, more than 40% of the births were extramarital and in many European countries this percentage was higher and rising. On the other hand, in many Asian societies having low fertility the percentage of women remaining unmarried until 35 has dramatically increased, between 1970 and 2000, from 7.2 to 26.6 in Japan; 8.1 to 16.1 in Thailand; from 1.4 to 10.7 in Korea; from 11.1 to 21.6 for Singapore Chinese; and from 9.5 to 18.2 for Malaysian Chinese. While births outside marriage are still rare among the countries of Asia, an increasingly higher proportion of women chose to remain unmarried till the end of their reproductive periods. It is difficult to predict how these changes will evolve over time and hence what the future course of fertility will be and what shape the second demographic transition will take across the world.

NOTES

2.9 Demographic Momentum or Population Momentum

When fertility starts declining in a population, as stated earlier, the proportion of children below age 15 in the population decreases, simultaneously increasing the proportion of women in the reproductive ages 15–49. Hence, the potential of the population to have an increased number of births even when the age-specific fertility rates decline is set in motion. Thus, even when the average number of children born per woman decreases, the CBR and consequently the population growth rate can continue to increase for some time till the wave of higher proportion of women in the reproductive ages passes. Thus, we will have a peculiar situation where the CBRs continue to rise for some time even when there is an increase in the use of contraceptive methods by couples and the fertility rates of women are falling. This phase of possible increase in CBR when the fertility levels are falling and the increased number of births due to increasing proportion of women in the

NOTES

reproductive ages is called the *demographic momentum* or *population momentum*. This is similar to the additional time and distance which a car takes to come to a stop even after the brakes are applied.

The factors of demographic dividends and demographic momentum are solely due to changes in the age structure of a population; the first due to increase in the proportion of the population in the productive ages and the second due to increase in the proportion of women in the reproductive ages. Population momentum occurs when the growth rate of women in reproductive ages 15–49 is higher than the population growth rates in general and demographic dividend occurs when the growth rate of persons in the productive or working ages 20–59 is higher than the population growth rates. Since women in the reproductive ages are a subset of the population in productive ages 20–59, the population momentum factor can sometimes work against the demographic dividend factor.

Exercises

Exercise 2.1: Compile the following data from Census and RGI for selected state e.g. Rajasthan

a. District-wise population for 2001 and 2011

b. Single year age-sex distribution for 2001 (Source: c13 from reference tables)

c. Age-sex marital status (Source: Census reference table)

d. Compile five years age-sex population and then adjust age not stated to all age groups

   Note: Age not stated should be distributed to all age groups in proportion to age-specific population.

e. SRS-based age-specific fertility rates for year 2001 and 2011

f. SRS-based age-sex specific death rates for year 2001 and 2011

g. SRS-based abridged life table for 2001 (males and females)

Exercise 2.2: Calculate the linear and exponential growth rates for all districts of Rajasthan (Day 1 exercise).

a. Compare growth rates of Rajasthan with Kerala

b. List the top two and bottom two districts according to growth rate

Sl.No.	Districts of Rajasthan	2001 (P_0)	2011 (P_t)
01	Ganganagar	1,789,423	1,969,520
02	Hanumangarh	1,518,005	1,779,650
03	Bikaner	1,902,110	2,367,745
04	Churu	1,696,039	2,041,172
05	Jhunjhunu	1,913,689	2,139,658
06	Alwar	2,991,552	3,671,999
07	Bharatpur	2,100,020	2,549,121
08	Dholpur	983,258	1,207,293
09	Karauli	1,205,888	1,458,459
10	Sawai Madhopur	1,117,057	1,338,114
11	Dausa	1,323,002	1,637,226
12	Jaipur	5,251,071	6,663,971
13	Sikar	2,287,788	2,677,737
14	Nagaur	2,775,058	3,309,234
15	Jodhpur	2,886,505	3,685,681
16	Jaisalmer	508,247	672,008
17	Barmer	1,964,835	2,604,453
18	Jalor	1,448,940	1,830,151

Contd...

Sl.No.	Districts of Rajasthan	2001 (P_0)	2011 (P_t)
19	Sirohi	851,107	1,037,185
20	Pali	1,820,251	2,038,533
21	Ajmer	2,178,447	2,584,913
22	Tonk	1,211,671	1,421,711
23	Bundi	962,620	1,113,725
24	Bhilwara	2,020,969	2,410,459
25	Rajsamand	982,523	1,158,283
26	Udaipur	2,481,201	3,067,549
27	Dungarpur	1,107,643	1,388,906
28	Banswara	1,420,601	1,798,194
29	Chittorgarh	1,330,360	1,544,392
30	Kota	1,568,705	1,950,491
31	Baran	1,021,473	1,223,921
32	Jhalawar	1,180,323	1,411,327
33	Pratapgarh	706,807	868,231
State Total		56,507,188	68,621,012

Linear growth rate $P_t = P_0(1+rt)$

$R = (P_t - P_0)/P_0 * t, t = 10$

Exponential growth rate $P_t = P_0 * e^{(rt)}$

$r = (\ln (P_t/P_0))/t, t = 10$

CHAPTER III

Demographic Data Sources

3.1 Introduction

Any study of human populations depends primarily on the data available for analysis and research. This chapter identifies various types of data generally used in the demographic analysis of a population, the sources of the data and methods of their compilation and presentation. The types of data collected on populations and the methods of their collection vary considerably from country to country but there are also a number of common characteristics. To be truly meaningful and useful they have to be studied in the context of each country. However, the main aim of this chapter is to describe the major types of demographic data generally used by demographers in their analysis, the sources from which they can be compiled and some basic techniques of their presentation, with illustrations from India.

As a prerequisite for the analysis of demographic data, familiarity with the whole process of collection of such information is essential. Also demographic data can never be free from the limitations arising out of difficulties of data collection, biases and errors of the respondents and similar errors arising from the interviewers. In most demographic data, errors and biases are inevitable to some extent but their magnitude and type vary from country to country. The nature of analysis to be conducted on a body of demographic data and the validity of the conclusions based on such an analysis are circumscribed by the quality of data available. In this chapter only the issues related to the sources of data, compilation and presentation of the data are discussed. The problems connected with quality of data, assessment of errors and biases involved in various types of data and methods of adjustment for them are discussed under each chapter with respect to specific variables.

NOTES

The system of demographic data collection is the mechanism whereby information on some of the basic characteristics of the population such as its age–sex marital structure, and the various events that contribute to changes in this structure, such as births, deaths, marriages, migration and other related topics, are compiled and tabulated. All such demographic data can be categorized broadly into two types: stock data and flow data. The former denotes the information pertaining to the situation of a population at a given point of time, in terms of its age–sex distribution, marital structure, occupational structure etc. It is a snapshot of the population at that point of time. The second set of information pertains to events which occur over time that contribute to changes in the population picture. Deaths remove people, mostly at older ages; births add to the population at young ages; migration generally depletes people at productive ages from the areas of out-migration and adds to the same group at places of in-migration, etc. The data collection agencies and procedures for the two types of data are generally distinct in most countries.

Rarely is any set of data collected primarily for demographic purposes. They are on the other hand by-products of data compiled for administrative or political reasons. Censuses, the vital registration system, statistics compiled by the various ministries such as those of Internal Security, Health and Family Welfare, Population Control, the Department of Migration Control and Transportation are compiled for administrative reasons and are oriented towards planning and monitoring of their programmes. The population census and population registers serve as the basis for allocation of public resources to different areas and also serve as a basis for the political representation for different areas. From the point of view of demographic analysis and as the source of demographic data, there are four major sources: censuses, vital registration, sample surveys and service statistics.

NOTES

Among the various data sources currently available, the most important and widely used in the analysis of demographic stock data at the national and sub-national level is the population census. The vital registration system or the system of compulsory registration of births, deaths and marriages is the major source of flow data. The third and, in recent years, the most popularized source of demographic data which can provide information on both the stock and the flow variables are the sample surveys. Sample surveys are generally conducted when the analyst needs more detailed information on a specific topic of his/her own interest and also when the information available from the censuses and vital registration is not adequate in coverage and quality. The fourth source of demographic information are the various administrative records and service statistics, wherein the data are compiled as a part of the routine activities of the various wings of the government, such as by the airport authorities on immigration and emigration, schools, hospitals, public health centres etc. We discuss these four sources in some detail in the ensuing sections. More details on these sources can be obtained from the book by Shryock and Seigel (1971), the UN manuals and the large number of publications from the Demographic Health Surveys.

3.2 Censuses

The word 'census' is of Latin origin and was used during days of the Roman Republic to keep a list of all adult males fit for military service. Even as far back as the third millennium BC, the Egyptians during the early Pharaonic period during 3340 and 3056 BC kept a list of all adults useful for military service and for collection of taxes. In his book *Arthasastra*, Kautilya, an Indian scholar in the fourth century BC, writes about the need of kings to keep track of all adult members in the kingdom to collect taxes and raise revenues. Prior to the seventeenth century the focus of any census was to keep track of the adult population, especially the males, to be used in warfare or for collection of taxes. The modern census is essential for international comparisons of any kind of statistics and censuses collect data on many attributes of the population, not just how many people there are, although population estimates remain an important function of the census.

After the industrial revolution in the seventeenth century in Europe, a census came to be redefined as the procedure of systematically acquiring and recording information about the population of a country. Viewed in this light Sweden had the first population census carried out in 1748. All the European countries, the state of Quebec in Canada and the United States followed, carrying out their censuses in different years during the eighteenth and early nineteenth centuries and this has since become a regularly occurring and official count of the population of the countries. The term is used mostly in connection with national population and housing censuses; other common censuses include agriculture, business and traffic censuses. The United Nations Handbook of the Census 1958 still holding good states, 'The census is the total process of collecting, compiling and publishing of demographic, economic and social data pertaining to a specified time or times of all persons in a country or delimited territory'. It also states the essential features of population and housing censuses as "individual enumeration, universality within a defined territory, simultaneity and defined periodicity" and recommends that population censuses be taken at least every 10 years. United Nations recommendations also cover census topics to be collected, official definitions, classifications and other useful information to coordinate international practice. Viewed from this definitional perspective the British census conducted in 1841 can be considered the first modern census.

NOTES

In India the first synchronous population census was conducted in 1881, though some parts of the country were covered by a non-synchronous census in 1872. The arrangements for conducting a census start years before the actual count and the work usually continues for years after the count is over for the analysis of data and publication of results. Since 1881, there has been an unbroken chain of censuses at 10-year intervals in the country, the latest being in 2011. The 2011 Census enumerated 1,210.19 million people while in the 1991 Census it enumerated 843.42 million people. Though the census count is a costly affair, involving great administrative efforts with disruption of normal services, it is considered essential for any modern state interested in the security, welfare and development of its citizens and is also an activity of national pride.

The census refers to the population at a particular point of time in the year. In the Indian Census of 2011, this point of time called *the Census Referral Moment* (CRM) is 00.00 hrs of 1 March 2011. In general, census enumeration can adopt two types of counts. One method is to count the population according to where the persons are staying at the time of census enumeration or the night before, irrespective of their usual place of residence called the *de facto* method. Another method is to collect information about people who are the normal residents of a specified area, irrespective of whether they are present or not at the time of the census enumeration time in that particular household. This type is called the *de jure* method. Thus in a *de jure* census, information about all the persons who usually live in a household is collected, whether or not they are present at the time of enumeration. In practice, it is impossible to count the entire population at any instant of time. A close approximation to this is achieved by starting the enumeration sufficiently early and then adjusting for the deaths, births and movements that occur between the enumeration and the census moment through additional inquiries soon after the census moment. The population enumeration was carried out from 9 February 2011 to 28 February 2011 and the revisional round to identify those present in the households during the CRM was conducted from 1 March 2011 to 5 March 2011.

The primary unit of census enumeration is the individual. For the final classification and analysis of census data, information regarding each person is required. To study characteristics common to the whole household or family, the unit of enumeration could be the household or the family itself. The usual minimum questions asked in the census about each individual are name, age, sex, relationship to the head of the household, marital status, religion, educational attainment and school going status for children, employment status, occupation, whether seeking work if not working, place of birth and migration. The housing conditions and amenities in the house are usually covered in a Housing Census, which usually precedes the population census. In India the House Listing and Housing Census was conducted during April–December 2010. In the population census, in addition to the above listed questions, general questions about the number of children ever-born, how many children are still alive and the births during the one-year period during the census time are also asked to all ever-married women. The questions on the employment status of the individuals include whether he or she is working full time or part time (called main worker or marginal worker in India), the nature of work they are engaged in and whether he or she is seeking work. The questionnaire can be filled by the enumerator or by the respondent and handed over by the enumerator personally or sent by mail. The quality of the responses seems to be better and the extent of error lower where the questionnaires are kept short, the questions are simple and factual, and the questionnaires are administered by trained enumerators.

NOTES

It is important to distinguish the difference between censuses and sample surveys. As mentioned, the census is a complete recording of the entire population and does not involve sampling. Censuses relate to a single point of time, not a period, even though the enumeration itself may be spread over days or weeks. In a census, each individual should be enumerated separately. The success of a census depends upon the efficiency of its fieldwork. Many countries, including India, have a large census organization as a part of their permanent statistical system. This ensures the supply of trained technical persons, improved documentation and reduced expenditure.

Once the census enumeration has been done, all data collected by the enumerators are carefully checked and edited by supervisors. A post-enumeration check is usually carried out for assessing the coverage error. The processing of data collected involves entering the data into computers, validation and producing the final tabulations. A detailed tabulation plan may be prepared well before the enumeration. The final publication of the census will be produced from the tabulation plans initially formulated and funded for, by appropriate analysis of the data collected. Usually, the final publication of the results may take a long time, even many years, but provisional figures are generally published by the census organization as early as possible after the census enumeration. For a country with a large population like India, the census enumeration demands a large number of enumerators (approximately a million were used for the 1991 Census) and to maintain homogeneity in collecting data through them is really a difficult task. Population census in most countries is preceded by an enumeration of the households, a mapping of the villages and the urban blocks and in many situations by a housing census, wherein data on the conditions of the house in terms of number of rooms, construction materials used etc., is gathered. It has to be recognized that censuses provide the basic data needed on the structure of the population of the country in terms of size, age–sex marital status distributions in different areas and are essential for the planning of various developmental programmes, such as the number of children of school-going age in an area and the number of those to be brought under schooling, the number of people seeking work, the number of aged persons etc. The census also provides the population size of different areas, which forms the basis of division of political constituencies, the number of political leaders to be elected from different states or provinces to the state and central legislative bodies and is in a way crucial to setting up the power base of the country. Thus, the census is fundamental to the politics and planning for development of any country and the fact that it serves an important function in demography is only supplementary to its other administrative functions.

3.3 Vital Statistics

NOTES

The flow data on the demographic variables are largely derived from the vital registration system existing in most countries. Censuses only provide information about the population at one period of time. The gap between two censuses is usually very large (10 years) and often policy makers and planners require detailed information about the population in the intervening years for planning of public health programmes, implementing social welfare measures and drawing up schemes for distribution of public facilities. In such situations, the aggregates of vital registration along with the latest population census serve as basic data. Vital registration is a system for the registration of the demographic events occurring in a population – births, deaths and marriages – and is the basic source of information on population dynamics. The registration system of developed nations provides elaborate statistical data on many aspects of social change. In most countries, the registration of births and deaths is compulsory, as this information is also necessary for individuals for a variety of legal purposes. In organizational aspects the vital statistics system is different from the census operation. Vital statistics is the total process of registering, compiling and reporting of the aggregate of vital events, such as births, deaths, marriages, that occur during a specified duration of time among the members of the population residing in an area. Births can be reported by the parents or a relative, and deaths can be reported by a relative. A period of time is owed to report the event after its occurrence. In case of marriage, divorce etc., the informant could be the individual himself. A registrar is the official authorized to register the occurrence of a vital event and to record the required details with respect to the event. The quality of vital data generally depends upon the educational levels of the registrar and the informant, and the level of enforcement of the law on registration. A standard proforma is used to collect necessary information on a birth, death or marriage. Vital registration is mainly carried out for administrative and legal purposes, and only a few selected questions are asked. The questions on a birth include the date of the event, date of registration, place of birth, sex of the child, birth order and characteristics of the parents including their name, age, marital status (in some societies), religion and occupation. For a death, the minimal information includes the name of the deceased, age at death, sex, marital status, date and place of death, date of registration, cause of death and characteristics of the parents in the case of the death of a child.

The registrar transmits the information either in a duplicate form or as a statistical summary to the higher level of administration on a periodic basis. Generally, in India, in the urban areas the municipalities or the corporations are the places for registering the events, while in the rural areas the officials authorized to register the events vary from state to state: from the village headman, the village panchayat, the auxiliary nurse midwife (ANM) or the health inspector of the public health department. Usually it is the village headman from the revenue department who has to report births and deaths to the next higher authority, viz., the tehsildar. The Registration of Births and Deaths Act of 1969 gave the Registrar General of India the statutory authority to coordinate civil registration work throughout the country. According to this Act, registration of births and deaths has become compulsory, births to be registered within 14 days and deaths within a week; violators are punishable. Yet, the registration of events is usually far from complete. In India, and many other countries, the vital statistics system forms part of the general administration of statistics collection. Table 3.1 gives a classification of different states in India according to the level of registration of births and deaths in 2006 and 2007. In 2007 the percentage of births registered in the country as a whole was 74.5 and of deaths 69.3. There is a considerable degree of variation in the level of registration of births among the states with Kerala having almost complete registration of births and deaths for over a decade now, with the level of registration of births 100% since 1998 and 100% deaths in 2007, while in Bihar and Uttar Pradesh even in 2007, 13.7% of the births and 26.3% of the deaths remained unregistered.

NOTES

Table 3.1: Classification of Major States (Over 10 Million Persons) by Level of Registration, 2006 and 2007

Level of Registration	2006		2007	
	Births	Deaths	Births	Deaths
Above 90	7 states (Haryana, Karnataka, Kerala, Tamil Nadu, Punjab and West Bengal)	4 states (Karnataka, Kerala, Punjab and Tamil Nadu)	8 states (Gujarat, Haryana, Karnataka, Kerala, Tamil Nadu, Punjab and West Bengal	4 states (Karnataka, Kerala, Punjab and Tamil Nadu)
80–90	3 states (Maharashtra, Odisha and Rajasthan)	2 states (Haryana and Maharashtra)	2 states (Odisha and Rajasthan)	3 states (Chhattisgarh, Haryana and Maharashtra)

Contd...

Level of Registration	2006		2007	
	Births	Deaths	Births	Deaths
50–80	6 states (Andhra Pradesh, Assam, Chhattisgarh, Jammu & Kashmir, Madhya Pradesh and Uttarakhand)	7 states (Andhra Pradesh, Chhattisgarh, Gujarat, Madhya Pradesh, Odisha, Rajasthan and West Bengal)	7 states (Andhra Pradesh, Assam, Chhattisgarh, Jammu & Kashmir, Madhya Pradesh, Uttarakhand and Uttar Pradesh)	8 states (Andhra Pradesh, Gujarat, Jammu & Kashmir, Gujarat, Madhya Pradesh, Odisha, Rajasthan, Uttar Pradesh and West Bengal)
25–50	2 states (Jharkhand and Uttar Pradesh)	5 states (Assam, Jammu & Kashmir, Jharkhand, Uttarakhand and Uttar Pradesh)	2 states (Bihar and Jharkhand)	4 states (Assam, Bihar, Jharkhand and Uttarakhand)
Below 25	1 state (Bihar)	1 state (Bihar)	Nil	Nil
Total Major States	19	19	19	19

NOTES

3.4 Sample Surveys

The third type of data generally used by demographers is the information collected in sample surveys. In general, a sample survey involves taking a sample of the population to be enumerated with more trained investigators interviewing them more intensely in a few selected areas. The major advantage of this type of data collection is that it requires fewer resources – money and manpower – than required for censuses and registration. Data accuracy will also be enhanced since more time can be spent at each household or with the individual, and the interviews can be held in greater detail and by more skilled persons. However, this type of data involves a sampling error since only a sample is studied, but recent developments in sampling theory make it possible to get estimates of any characteristic in the population, called a *parameter*, with a fairly high degree of precision even with relatively small sample sizes. For example, it is possible to get estimates of fertility rates in most of the developing countries within 5% error from a sample survey of 2,000 women in reproductive ages. The advantage of the sample survey method is that the sample size needed for the estimation of most demographic parameters does not depend on the size of the population as a whole for which estimation is needed. A sample survey with 2,000 women in India with a population of over 950 million and from Sri Lanka with a population of 15 million will have almost the same degree of sampling error. The other advantages of sample surveys are that they can be organized and executed relatively quickly, and can collect information in much more detail than is possible in a census. The examples of large-scale surveys conducted almost simultaneously in many countries of the world on nationally representative samples and providing internationally comparable data are the World Fertility Survey (WFS) and DHSs. In India, a similar large-scale sample survey, the National Family Health Survey (NFHS), was carried out in most of the larger states and the National Capital Territory of Delhi during 1992–93. These surveys are designed not only to provide estimates of various demographic parameters but also with a view to test a number of research hypotheses. Sample surveys are generally designed with a view to assess the impact of existing policies, to aid the planning of those yet to be implemented, to influence governmental or other agencies or for academic purposes. They generally look into the levels and patterns of health, fertility, mortality, contraceptive use and the impact of economic, social, environmental and specific programme-related factors on them. Since the data are generally collected from a sample in order to produce fairly reliable results, the sample size cannot be too small. In a majority of sample surveys, the unit of observation has been the household and interviews within the household have been focused on married women in reproductive ages and their children.

Demographic sample surveys are also diverse in design as well as in their objectives but two main types of surveys are worth mentioning. First there is the single-round survey, which involves only one interview with the respondent. Such types are generally known as retrospective surveys. This design permits the use of larger samples and is especially suitable in countries where addresses are not well defined or the population is highly mobile. Another type is the multi-round survey, which encompasses a great variety of survey designs: surveys which re-interview the same respondent several times or surveys which take different samples in each round or a combination of both. Modifications on these are sometimes referred to as follow-up or panel or prospective surveys. These surveys primarily produce information on trends in the parameters over time and factors associated with such trends.

NOTES

As discussed, all sample surveys question only a fraction of the total population for which data are sought and the task of relating the survey findings to the larger population is a complex one. Sample design is therefore of major importance to any sample survey and has become a well-developed field of applied statistics. The strict rules of scientific probability sampling are many a time impossible to be followed exactly and some pragmatism is required to be introduced.

Survey questionnaires are also diverse and two main types are generally observed. In pre-coded forms with closed questions a respondent's answer must conform to some pre-set range of acceptable responses. Forms with open-ended questions which do not constrain the respondent in any way constitute the second type. The choice of questions, the sequence in which they are posed, the words in which they are expressed and even the orientation and attitude of the interviewer may have a powerful effect on the offered responses. In general, sample surveys often include census-type inquiries. Retrospective surveys collect information about the demographic events that happened in the past.

3.4.1 Sample Registration System

Another form of data collection system unique to India is the Sample Registration System (SRS) which is a dual record system, introduced in the early 1960s and became fully operational in 1969. In the absence of reliable data coming forth through the civil registration system on birth and deaths in most of the states, the SRS was introduced to provide reliable estimate of fertility and mortality at the state level for the rural and urban areas. Here, in sample areas – villages in the rural areas and enumeration blocks in the urban areas – duly selected on the basis of scientific sampling procedures, data are compiled on vital events, births and deaths, by two methods: first through a continuous registration system by a local registrar, who is a local teacher/Anganwadi worker/ASHA and second, through a half-yearly or annual survey conducted in the same area by independent full-time supervisors recruited by the office of the Registrar General of India. The events on which information has been compiled by the two systems are then matched event by event. They are categorized into three groups: those that have been identified by both systems, those identified only in the continuous registration and those identified only in the survey. The fourth category, viz., those omitted in both the systems is to be estimated by the formulae developed by Chandrasekhar and Deming (1949), but in the SRS the fourth category is not included since it is found to be exceedingly small. In case the number omitted in either of the two systems is small, there is no need to include the correction factor. In this case, the total numbers of events are those identified in either of the two systems. Sometimes this scheme involves running two statistically independent surveys covering the same population and matching the results. The difficulties of executing two surveys simultaneously and independently make this procedure less popular. Over the years the SRS has become the major and only reliable source of fertility and mortality estimates at the state level for every year for the rural and urban areas, since the civil registration system continues to be deficient in a number of states. One hypothesis that has been consistently made is that the reliability on SRS has reduced the emphasis that should have been placed in improving the Civil Registration System (CRS) over the decades since CRS is the only method that can provide continuous estimates of fertility and mortality at any level of aggregation.

NOTES

The universe from which the sample of villages and urban blocks are selected for SRS is the census. The sampling of villages and urban blocks to be followed over a 10-year period is selected after each census by an Expert Committee appointed by the Registrar General. Based on the 2001 Census, using a single-stage stratified random sample without replacement a total of 7597 units consisting of 4433 villages and 3164 urban blocks were selected from all the states and union territories in the country. A committee has recently been constituted to revise the sample based on the 2011 Census.

The half-yearly survey conducted as a part of the SRS is also used to collect data on causes of death reported in the rural sample areas in a project called *Survey on Causes of Death (Rural)* from 1 January 1999 using a refined method of Verbal Autopsy (VA) developed by a team of experts, and the supervisors who collect these data have been trained appropriately to collect such information focusing on infant and maternal deaths.

3.5 Service Statistics

In all the developed countries and in many developing countries, in recent years (e.g. Sri Lanka and China) the administrative records, called the *Service Statistics*, compiled by the various departments of government, vital registration, health, education, family welfare, women's empowerment and employment, as an integral part of their supplies and services, form the major database for the routine monitoring and evaluation of various social development programmes. Unfortunately in India such service statistics are woefully poor both in coverage and data quality. The birth and death registration can also be considered in this category.

Even 50 years after the passage of the act it is rather distressing to note that among the births that occurred between 2000 and 2005, only 41% were registered – 35% in the rural areas and 59% in the urban areas. Among the states the variation is very large, ranging from a high of 85% in Goa to a low of 6% in Bihar, 7% in Uttar Pradesh and 9% in Jharkhand. Since the official responsibility for the registration of births and deaths is vested with the peripheral state government officials such as the village munsif/patel, the ANM or male multipurpose worker, or the panchayat secretary or the thane official depending upon the department of the state government with which the responsibility for the registration of vital events is vested, and since the vital events occur all over the place at all times, the extent of registration of births can be considered as a good proxy of the effectiveness of governance of the state government. It can be considered as a simple but good index of effectiveness of state governance (IEG). Viewed in this context, the IEG is highest, more than 80% in Goa, Gujarat, Maharashtra, Himachal Pradesh, Kerala, Mizoram, Sikkim and Tamil Nadu and the lowest less than 20% in Rajasthan, Uttar Pradesh, Bihar and Jharkhand. Kerala is having close to 100% registration of vital events since the early 1990s. None of the states in north (excepting Himachal Pradesh), central and east India have a good IEG, above 80% registration.

NOTES

Similar deficiencies in data sets compiled by the various ministries of the government, central as well as state, can be pointed out. Significantly deficient in quality are the data with regard to enrolment of children in schools, health services offered and utilized, employment, identification of households below the poverty line and extent of utilization of various welfare services provided on subsidized basis to the population. It is depressing to note that even after six decades of independence the administrative records maintained by the various departments of health, education, employment and women's empowerment, and other social welfare departments cannot be used for routine monitoring and evaluation of programmes and there is increased dependence on sample surveys. These errors are estimated by comparing the results obtained from the service statistics with those estimated from sample surveys or estimated from other theoretical considerations.

Service statistics provide only numerator data in the estimation of any demographic parameter. However, it has to be recognized that errors in the numerators have a greater effect on the value of the indicators than similar magnitudes of errors in the denominator. This is of major significance, especially if the ratio or rate is of small magnitude. If the indicator of infant mortality rate in a population is indicated as 'k', it is defined as

$$k = 1000 * I/B$$

where I is the number of infant deaths in a population and B is the number of births in a given period in the same population. Below is the equation that relates the magnitude of the error in an indicator (k), indicated as d(k), in relation to error in 'I' and 'B':

$$d(k) = d(I)/B - I\,d(B)/B^2$$

The error term contributed by the denominator B is reduced by a division by the value B^2 (square of B) while it is increased by the error of I divided by B. The contribution of the numerator increases in proportion to the extent to which the denominator increases. When the events become rarer and the indicator values become lower and lower, the significance of the numerator increases proportionately. This is one of the primary reasons for monitoring the programmes on the basis of numerator data and analysis. This is especially significant for rates such as IMR, CMR, MMR, enrolment in higher education etc. In case the rates become larger than 50%, as in the case of 'literacy rate, enrolment rates in primary and secondary level education and employment rates, it is better to monitor the complement of these rates, number of persons illiterate, number of persons unemployed, infant deaths, maternal deaths etc.

NOTES

3.6 Noteworthy UN Publications

The Statistical Office and the Population Division of the United Nations have the major responsibility of collecting and compiling international statistics covering a variety of topics including population. They publish the *Demographic Yearbook* annually and a summary form in their Population and Vital Statistics Reports published quarterly. In recent years the Population Division of the United Nations publishes regularly the annual series entitled *World Population Prospects* online, which provides detailed data for all member countries on the population size as estimated for the latest year, age distribution, the vital rates as reported, migration and urbanization and population projections. They are essentially borrowed and updated from the Demographic Yearbooks. Sometimes they publish data on special topics including fertility, mortality, marriage, migration, divorce statistics and population census statistics. When fertility is the special topic, data on births by birth order, by age of mother, birth weight and gestational age are presented. Details of births by legitimacy status are also provided in these special issues. In addition to these, latest census data on female population by age, number of children born alive and number of children living are also shown. Under mortality, detailed data on foetal, perinatal and infant mortality rate are given. Detailed data on deaths by cause and marital status are also presented. The expectation of life at specified ages, the number of survivors at specified ages and life table mortality rates have also been furnished. These publications have also focused on the specialized topics of mortality, marriage rate by age of groom and bride, and marriages by age and previous marital status and first marriages. The United Nations Population Division also regularly publishes a series of alternative projections of population for the countries of the world and these projections are used extensively nationally and internationally. The latest list of publications listed on the UN website on population issues is as below:

Manuals I to X of the United Nations

These are important manuals that describe the various techniques of population analysis most of them developed during 1950s, 1960s and 1970s by eminent demographers of the world. Some of these may be outmoded now, but nevertheless are important in setting out the methodologies of population analysis with regard to quality of data, projections, estimation of fertility, mortality, migration and urbanization levels and various indirect methods of estimating demographic parameters. All the demographic techniques that are in use at present are described in one or the other of the manuals. Only headings of the manuals are presented here; the details are given in Appendix II.

- United Nations (1952). *Manual I: Methods for estimating total population for current dates* (United Nations Publications, Sales No. 52.XIII.5)
- United Nations (1955). *Manual II: Methods of appraisal of quality of basic data for population estimates* (United Nations Publications, Sales No. 56.XIII.2)
- United Nations (1956). *Manual III: Methods for population projections by sex and age* (United Nations publication, Sales No. 56.XIII.3)
- United Nations (1967). *Manual IV: Methods of estimating basic demographic measures from incomplete data* (United Nations Publications, Sales No. 67.XIII.2)
- United Nations (1971). *Manual V: Methods of projecting the economically active population* (United Nations publication, Sales No. E.70.XIII. 2)
- United Nations (1970). *Manual VI: Methods of measuring internal migration* (United Nations Publications, Sales No. E.70.XIII.3)
- United Nations (1973). *Manual VII: Methods of projecting households and families* (United Nations publication, Sales No. E.73.XIII.2)
- United Nations (1974). *Manual VIII: Methods for projections of urban and rural population* (United Nations publication, Sales No. E.74.XIII.3)
- United Nations (1979). *Manual IX: The methodology of measuring the impact of family planning programmes on fertility* (United Nations publication, Sales no. E.78.XIII.8)
- United Nations (1983). *Manual X: Indirect techniques for demographic estimation* (United Nations publication, Sales No. E.83.XIII.2)

In addition to the above technical manuals the UN also publishes periodical bulletins or articles specialized topics and wall charts on population. Many of them are available online. Some of them are listed below:

- *Completing the Fertility Transition,* Population Bulletin Special Issue, Nos. 48/49, 2002
- *Fertility, Contraception and Population Policies*
- *HIV/AIDS: Population Impact and Policies 2001* (Wall Chart)
- *International Migration 2002* (Wall Chart)
- *International Migration 2006* (Wall Chart)

NOTES

- *Long-range World Population Projections: Based on the 1998 Revision*
- *Methods for Estimating Adult Mortality*
- *Model Life Tables for Developing Countries*
- *MortPak: The United Nations Software Package for Mortality Measurement*
- *National Trends in Population, Resources, Environment and Development 2005: Country Profiles*
- *Population Ageing and Development 2009* (Wall Chart)
- *Population Challenges and Development Goals*
- *Population, Environment and Development 2001* (Wall Chart)
- *Review and Appraisal of the Progress Made in Achieving the Goals and Objectives of the Programme of Action of the International Conference on Population and Development, The 2004 Report* [Arabic] [Chinese] [English] [French] [Russian] [Spanish]
- *The World at Six Billion*
- *Urban Agglomerations 2007* (Wall Chart - PDF)
- *World Fertility Patterns 2009* (Wall Chart)
- *World Marriage Data 2008* (Data online)
- *World Marriage Patterns 2000* (Wall Chart)
- *World Mortality 2009* (Wall Chart)
- *World Population Ageing: 1950-2050*
- *World Population Ageing 2007*

Note: The executive summary of the above publications are available in all UN-approved international languages viz. Arabic, Chinese, English, French, Russian and Spanish Concise Report.

NOTES

The *Population and Vital Statistics Report* of the United Nations is issued quarterly and presents the total population from the latest population census, the latest official estimate of the total mid-year population and an estimate of population for a recent reference year. Vital statistics shown include the total number of births, deaths, infant deaths, CBRs and CDRs. Indications of quality of both the mid-year population estimates for the reference year and the registration of births, deaths and infant deaths are also presented. The other oft-referred-to international publications include FAO's *Production Yearbook*, which provides information on agricultural population. Detailed data on the economically active population are provided in the ILO's *Yearbook of Labor Statistics*; information on education, literacy and school attendance in the United Nations Educational, Scientific and Cultural Organization (UNESCO) *Statistical Yearbook* and 'detailed cause of death' data in WHO's *World Health Statistics Annual Numbers*. Other international organizations such as the World Bank also provide data on population and provide an independent series of population projections. The World Bank publishes the *World Development Report* annually, which covers basic demographic measures of major countries in the world. The

UNDP also publishes annually the social, economic and demographic indicators for all the countries in the world in their publication series begun in 1990, known as *Human Development Reports*.

3.7 National Publications

3.7.1 Publications by the Registrar General

The Office of the Registrar General of India and the Census Commissioner (RGI) publishes after every census a large number of tables based on the data collected in the census at the national, district level and as primary census abstract at lower levels. These form the backbone of any population analysis by planners, demographers, social scientists and others. These tables are published under different grouped categories and for the 2001 Census the list is as follows:

A – Series: General Population Tables

Number of villages, urban agglomerations and towns ~ Area ~ Houses ~ Total population ~ Rural population ~ Urban population ~ Scheduled Castes and Scheduled Tribes population ~ Primary Census Abstract ~ Slum population ~ Houseless population ~ Institutional population

NOTES

B – Series: Economic Tables

Workers (Main & marginal) ~ Industrial category of workers ~ Educational level ~ Main activity of marginal workers ~ Main activity of non-workers ~ Seeking/available for work

C – Series: Social and Cultural Tables

Population by religious communities ~ Age and marital status ~ Ever married and currently married population ~ Age at marriage ~ Literates and educational level ~ Attendance in educational institutions ~ Disabled population

D – Series: Migration Tables

Migrants by place of birth ~ Place of last residence ~ Duration of residence ~ Intra-district migration ~ Inter-district migration ~ Inter-state migration ~ International migration ~ Reason for migration ~ Age, marital status ~ Educational level of migrants ~ Economic activity of migrants

F – Series: Fertility Tables

Ever married women by present age, parity and total children ever born ~ Number of surviving children ~ Number of births last year ~ Birth order

HH – Series: Household Tables

Number of households ~ Household size ~ Male- and female-headed households ~ Normal households ~ Houseless households ~ Institutional households ~ Households with aged persons ~ By religion

Based on the SRS, the RGI publishes annually an SRS Bulletin giving data on the birth, death and infant mortality rates for the rural and urban areas for all the states and union territories (UTs) and the sampling errors associated with these estimates. These bulletins are eagerly looked forward to by researchers and programme personnel to ascertain the latest trends in fertility and mortality levels. In addition, the RGI also issues a detailed Statistical Report based on the SRS giving various fertility and mortality measures classified by age, sex, marital duration, birth order and other characteristics. Periodical reports based on Special Surveys conducted by the RGI's Office like the *Surveys on Cause of Deaths* and the latest one on Maternal Mortality Rates are also published.

3.7.2 Publications by other government agencies

The second major source of demographic data are the various rounds of National Sample Surveys conducted by the National Sample Survey Organization (NSSO), with its annual publications and special rounds devoted to population issues. The history of NSS surveys can be divided into three phases: the formative years (1st to 10th round), the period of growth and consolidation (11th to 27th round) and the period after formation of NSSO (since 28th round). The 28th round of NSS (October 1973 to June 1974) marks a significant watershed in the evolution of NSS as it introduced major changes in the design.

The basic design followed in these surveys is stratified two-stage (for smaller villages/urban blocks) or three-stage (for larger villages/blocks involving hamlet group/sub-block formation). The primary stage unit (PSU) is the village (in rural areas) or urban frame survey block (in urban areas), but the ultimate stage unit (USU) is invariably the household/enterprise for household/enterprise surveys. However most of the NSS surveys being multi-subject integrated ones, detail design varies from survey to survey. The most recent round is the 66th round.

The third major source of demographic and health data are the specialized surveys carried out under the title of 'National Family Health Surveys' conducted jointly by the International Institute for Population Sciences, Mumbai (IIPS) jointly with the Ministry of Health and Family Welfare and macro-international and international financial support providing scientifically valid and internationally comparable estimates of fertility, mortality, various health indicators and proximate determinants thereof. These surveys followed globalization of the Indian economy in the late 1980s and liberalization and making accessible to the Indian public/scholars and the rest of the world data on various developmental and demographic indicators of the Indian society. Three rounds of such surveys 1, 2 and 3 have been conducted during 1991–92, 1997–98 and 2005–06. These were followed by a series of District Level Household Surveys (DLHS-1, 2 and 3) conducted by IIPS on behalf of the Ministry of Health and Family Welfare. Recently similar surveys at the district level in Empowered Action Group (EAG) states are carried out by the RGI under the title 'Annual Health Surveys'. The list of publications from the Office of the Registrar General of India and Census Commissioner, The NSSO and NFHS 1, 2 and 3 and DLHS-1, 2 and 3 are available in their respective websites and the data sets are also available online. The Ministry of Health and Family Welfare publishes (used to publish) regularly an annual yearbook consolidating all the relevant population, family planning and health data for the most recent year and was a useful ready reckoner for the latest information on population trends, fertility and mortality levels, family planning acceptance by methods, couple years of effective protection (CEP) and comparison of achievement with targets. Unfortunately this yearbook has not been published regularly during the past 10 years.

NOTES

3.8 Websites

There are literally thousands of websites providing information on the population and its characteristics of any country in the world, all easily accessed through Google. However some of the commonly used websites for India and global demography are the following:

1. www.cenusindia.gov.in
2. www.mospi.nic.in
3. www.iipsindia.org
4. www.nfhsindia.org
5. www.indiasta.com
6. www.mohfw.nic.in
7. www.esa.un.org
8. www.unfpa.org
9. www.unstat.un.org
10. www.searo.who.int

NOTES

3.9 Graphic Representation of Demographic Data

The most common forms of graphical representation of demographic data are age pyramids, pie diagrams, frequency polygons and simple line graphs. Age pyramids are used to represent the age distributions of males and females, a population that is unique to demographic data. In the series of six pyramids given below (Tables 3.2 to 3.7 and Figures 3.1 to 3.6) we compare the age distributions of Kerala and Rajasthan for 1951, 1971, 1991 and 2001 and of Kerala and Japan for 1950–51 and 2000–01 to demonstrate the power of age pyramids in bringing out past demographic trends and the differences between Kerala and Rajasthan, and between Kerala and Japan. If we review the age pyramids of different societies, we observe that there is an empirical relationship between age structure and the rate of population growth. Usually a broad base indicates a high proportion of children, a rapid population growth and a low proportion of old age population. This is a common population characteristic of the developing countries, which gives the pyramid a triangular shape. A country which has long had low birth and death rates gives rise to a rectangular shape for the pyramid. A recent increase in births will also be reflected in a certain shape. This may give the pyramid a wider base. Similar observations can also be made by a comparison of the pyramids for Kerala and Rajasthan.

In the composition of the population in Japan after the Second World War there was a large difference between male and female populations aged 20–39 because of the large mortality rate of young men in the war. Population pyramids from successive censuses make it possible to follow the movement of a cohort through the population. This is considered an important means to learn the demographic history of a population. Pyramids can also be used to ascertain the quality of age data.

Another useful method of graphic representation of data on the proportions of a population in various subgroups, such as in different marital status categories – single, married, divorced and widowed – or in the broad age categories such as children, working population and old population, is through a pie diagram. The *areas* in a circle, called the *pie*, are proportional to the percentage of each subgroup to the total population. A pie diagram of the proportions of India's surface area in the global land area and its share of population in the global population are given as illustrations of this method of graphic presentation (Figure 3.7). Another method of graphic presentation is through frequency polygons. A line graph of trends in the population of India in the rural and urban areas over the years is also given in Figure 3.8. Another method is the conventional bar diagram and a trend in the sex ratios of population and among children aged 0–6 is given in Figure 3.9.

Table 3.2: Population of Kerala and Rajasthan, 1951

Age Group	Kerala				Rajasthan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
0-4	688900	670700	15.2	14.9	1096200	1062200	14.1	14.8
5-9	575500	567500	12.7	12.6	1049800	983400	13.5	13.7
10-14	541100	546900	11.9	12.1	927800	834100	11.9	11.6
15-19	504700	526300	11.1	11.7	797500	727300	10.2	10.1
20-24	440600	462700	9.7	10.2	680200	664300	8.7	9.2
25-29	362700	373500	8.0	8.3	609500	605700	7.8	8.4
30-34	303200	303500	6.7	6.7	555300	529500	7.1	7.4
35-39	261100	257400	5.8	5.7	494600	444300	6.3	6.2
40-44	225600	221600	5.0	4.9	424900	363400	5.5	5.1
45-49	191000	189500	4.2	4.2	350100	293400	4.5	4.1
50-54	158500	158800	3.5	3.5	279700	233900	3.6	3.3
55-59	123700	126500	2.7	2.8	220000	189800	2.8	2.6
60-64	90300	94600	2.0	2.1	169600	148400	2.2	2.1
65-69	61200	6500	1.3	0.1	117400	95900	1.5	1.3
70+	9100	9800	0.2	0.2	17900	13400	0.2	0.2
Total	4537200	4515800	100.0	100.0	7790500	7189000	100.0	100.0

Table 3.3: Population of Kerala and Rajasthan, 1971

Age Group	Kerala				Rajasthan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
0-4	1441758	1413646	13.6	13.1	2043849	1951590	15.2	15.9
5-9	1457604	1418369	13.8	13.2	2135421	1912612	15.8	15.6

Contd...

NOTES

Age Group	Kerala				Rajasthan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
10-14	1446053	1417832	13.7	13.2	1800468	1536979	13.4	12.5
15-19	1127098	1210207	10.6	11.2	1190177	996228	8.8	8.1
20-24	977869	1012032	9.2	9.4	981847	992596	7.3	8.1
25-29	664365	723141	6.3	6.7	947524	934513	7.0	7.6
30-34	587456	622125	5.5	5.8	884735	838473	6.6	6.8
35-39	617632	660645	5.8	6.1	738058	658614	5.5	5.4
40-44	497158	495924	4.7	4.6	682106	622617	5.1	5.1
45-49	498587	464704	4.7	4.3	500735	442696	3.7	3.6
50-54	351904	340445	3.3	3.2	548831	470083	4.1	3.8
55-59	288041	284182	2.7	2.6	286365	245618	2.1	2.0
60-64	234469	247426	2.2	2.3	378933	344525	2.8	2.8
65-69	165662	183850	1.6	1.7	148528	124340	1.1	1.0
70-74	108254	120822	1.0	1.1	125776	115522	0.9	0.9
75-79	66107	75882	0.6	0.7	35035	31379	0.3	0.3
80+	57310	67913	0.5	0.6	55086	61807	0.4	0.5
Total	10587327	10759145	100.0	100.0	13483474	12280192	100.0	100.0

NOTES

Table 3.4: Population of Kerala and Rajasthan, 1991

Age Group	Kerala				Rajasthan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
0-4	1365090	1298620	9.6	8.8	3076375	2879852	13.4	13.8
5-9	1459473	1425373	10.2	9.6	3381853	3049705	14.7	14.6
10-14	1564752	1534328	11.0	10.4	2963281	2597482	12.9	12.4
15-19	1467374	1551965	10.3	10.5	2252584	1856895	9.8	8.9
20-24	1475802	1636830	10.4	11.1	1871346	1836442	8.2	8.8
25-29	1262730	1398554	8.9	9.5	1769595	1703040	7.7	8.1
30-34	1046690	1060612	7.3	7.2	1537858	1433959	6.7	6.9
35-39	1013437	1036484	7.1	7.0	1329614	1170893	5.8	5.6
40-44	764030	724101	5.4	4.9	1122952	953870	4.9	4.6
45-49	657954	678628	4.6	4.6	874641	847813	3.8	4.1
50-54	523257	538602	3.7	3.6	876134	692389	3.8	3.3
55-59	465404	510493	3.3	3.5	498326	524742	2.2	2.5
60-64	417045	454527	2.9	3.1	623895	559932	2.7	2.7
65-69	324559	374175	2.3	2.5	291517	309199	1.3	1.5

Contd...

Age Group	Kerala				Rajasthan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
70-74	199223	232057	1.4	1.6	276179	263801	1.2	1.3
75-79	121978	154954	0.9	1.0	85440	91561	0.4	0.4
80+	128057	160790	0.9	1.1	124130	142216	0.5	0.7
Total	14256855	14771093	100.0	100.0	22955720	20913791	100.0	100.0

Table 3.5: Population of Kerala and Rajasthan, 2001

Age Group	Kerala				Rajasthan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
0-4	1409487	1355677	9.1	8.3	3781055	3452165	12.9	12.8
5-9	1295679	1248502	8.4	7.6	4242507	3825563	14.5	14.2
10-14	1523842	1463358	9.9	8.9	3847418	3394523	13.1	12.6
15-19	1484586	1499920	9.6	9.2	2987604	2502610	10.2	9.3
20-24	1440467	1543523	9.3	9.4	2470943	2278423	8.4	8.5
25-29	1296905	1489290	8.4	9.1	2127889	2069334	7.3	7.7
30-34	1185807	1330656	7.7	8.1	1917187	1903964	6.5	7.1
35-39	1154778	1311576	7.5	8.0	1788941	1664919	6.1	6.2
40-44	960397	990887	6.2	6.1	1476036	1284239	5.0	4.8
45-49	952021	974123	6.2	6.0	1183740	1085911	4.0	4.0
50-54	724701	712819	4.7	4.4	974037	801009	3.3	3.0
55-59	541668	588576	3.5	3.6	646505	702810	2.2	2.6
60-64	480345	551791	3.1	3.4	676223	685051	2.3	2.5
65-69	399671	502344	2.6	3.1	457287	525587	1.6	2.0
70-74	273293	340129	1.8	2.1	376076	381922	1.3	1.4
75-79	173558	225531	1.1	1.4	145849	170081	0.5	0.6
80+	157345	231668	1.0	1.4	173490	218706	0.6	0.8
Total	15454550	16360370	100.0	100.0	29272787	26946817	100.0	100.0

NOTES

Table 3.6: Population of Kerala and Japan, 1950–51

Age Group	Kerala				Japan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
0-4	688900	670700	15.2	14.9	4850000	4710000	13.8	12.8
5-9	575500	567500	12.7	12.6	4390000	4320000	12.5	11.7
10-14	541100	546900	11.9	12.1	4270000	4250000	12.2	11.5
15-19	504700	526300	11.1	11.7	3800000	3910000	10.8	10.6

Contd...

Age Group	Kerala				Japan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
20-24	440600	462700	9.7	10.2	2820000	3340000	8.1	9.0
25-29	362700	373500	8.0	8.3	2350000	2850000	6.7	7.7
30-34	303200	303500	6.7	6.7	2380000	2680000	6.8	7.3
35-39	261100	257400	5.8	5.7	2210000	2280000	6.3	6.2
40-44	225600	221600	5.0	4.9	2020000	1990000	5.8	5.4
45-49	191000	189500	4.2	4.2	1720000	1670000	4.9	4.5
50-54	158500	158800	3.5	3.5	1390000	1350000	4.0	3.7
55-59	123700	126500	2.7	2.8	1100000	1190000	3.1	3.2
60-64	90300	94600	2.0	2.1	790000	970000	2.3	2.6
65-69	61200	6500	1.3	0.1	540000	750000	1.5	2.0
70+	9100	9800	0.2	0.2	400000	650000	1.2	1.8
Total	4537200	4515800	100.0	100.0	35030000	36910000	100.0	100.0

NOTES

Table 3.7: Population of Kerala and Japan, 2000–01

Age Group	Kerala				Japan			
	Males	Females	Males (%)	Females (%)	Males	Females	Males (%)	Females (%)
0-4	1409487	1355677	9.1	8.3	3083431	2938358	5.2	4.8
5-9	1295679	1248502	8.4	7.6	3353150	3193462	5.7	5.2
10-14	1523842	1463358	9.9	8.9	3833984	3654181	6.5	5.9
15-19	1484586	1499920	9.6	9.2	4307242	4114218	7.3	6.7
20-24	1440467	1543523	9.3	9.4	4965277	4825032	8.4	7.8
25-29	1296905	1489290	8.4	9.1	4436818	4339792	7.5	7.0
30-34	1185807	1330656	7.7	8.1	4096286	4018579	6.9	6.5
35-39	1154778	1311576	7.5	8.0	3924171	3876048	6.7	6.3
40-44	960397	990887	6.2	6.1	4467772	4448236	7.6	7.2
45-49	952021	974123	6.2	6.0	5210038	5231952	8.8	8.5
50-54	724701	712819	4.7	4.4	4290239	4443933	7.3	7.2
55-59	541668	588576	3.5	3.6	3749528	3986305	6.4	6.4
60-64	480345	551791	3.1	3.4	3357281	3748658	5.7	6.1
65-69	399671	502344	2.6	3.1	2670270	3230306	4.5	5.2
70+	604196	797328	1.8	2.1	3194565	5804072	5.4	9.4
Total	15454550	16360370	100.0	100.0	58940052	61853132	100.0	100.0

Figure 3.1: Age-Sex Pyramid - 1951 (Kerala & Rajasthan)

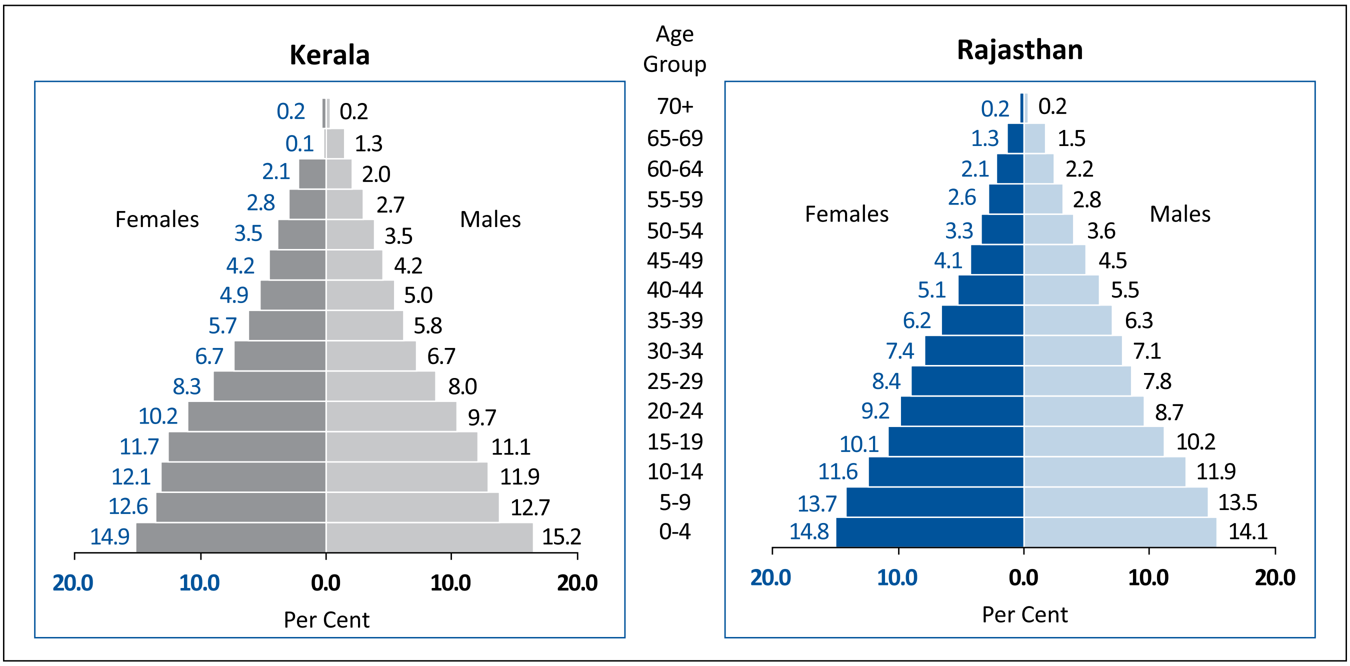


Figure 3.2: Age-Sex Pyramid - 1971 (Kerala & Rajasthan)

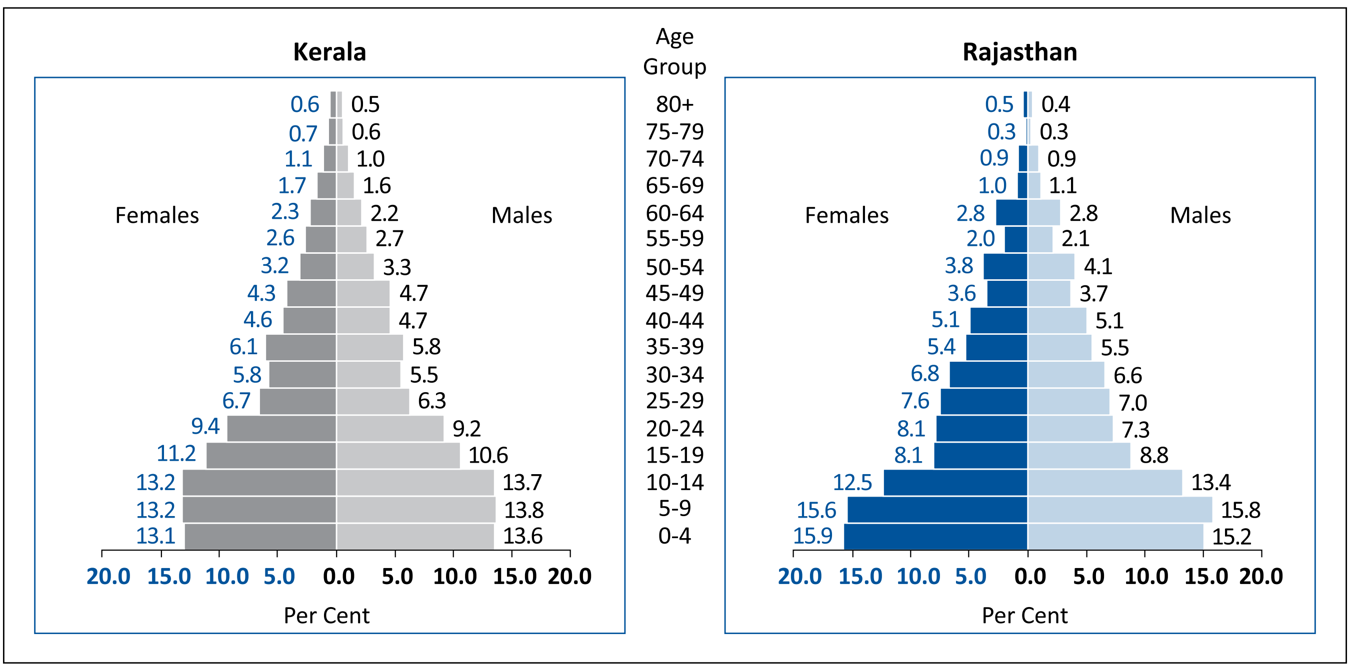


Figure 3.3: Age-Sex Pyramid - 1991 (Kerala & Rajasthan)

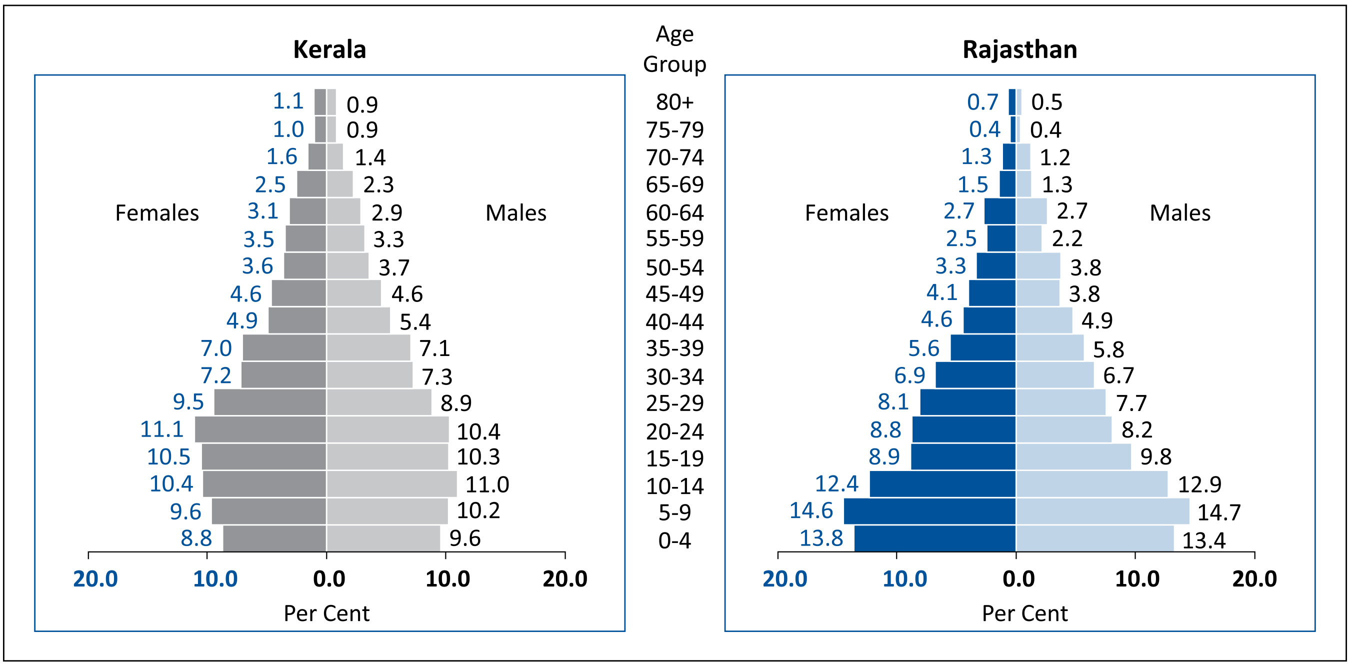


Figure 3.4: Age-Sex Pyramid - 2001 (Kerala & Rajasthan)

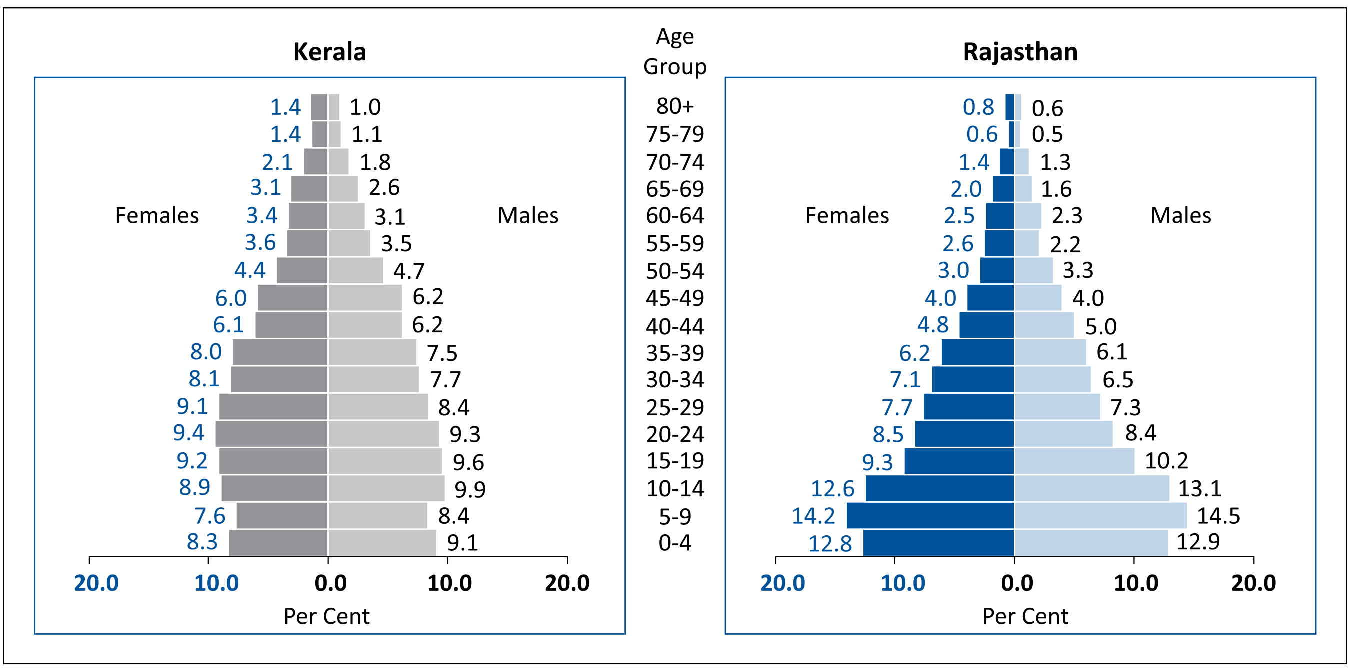


Figure 3.5: Age-Sex Pyramid - 1950–51 (Kerala & Japan)

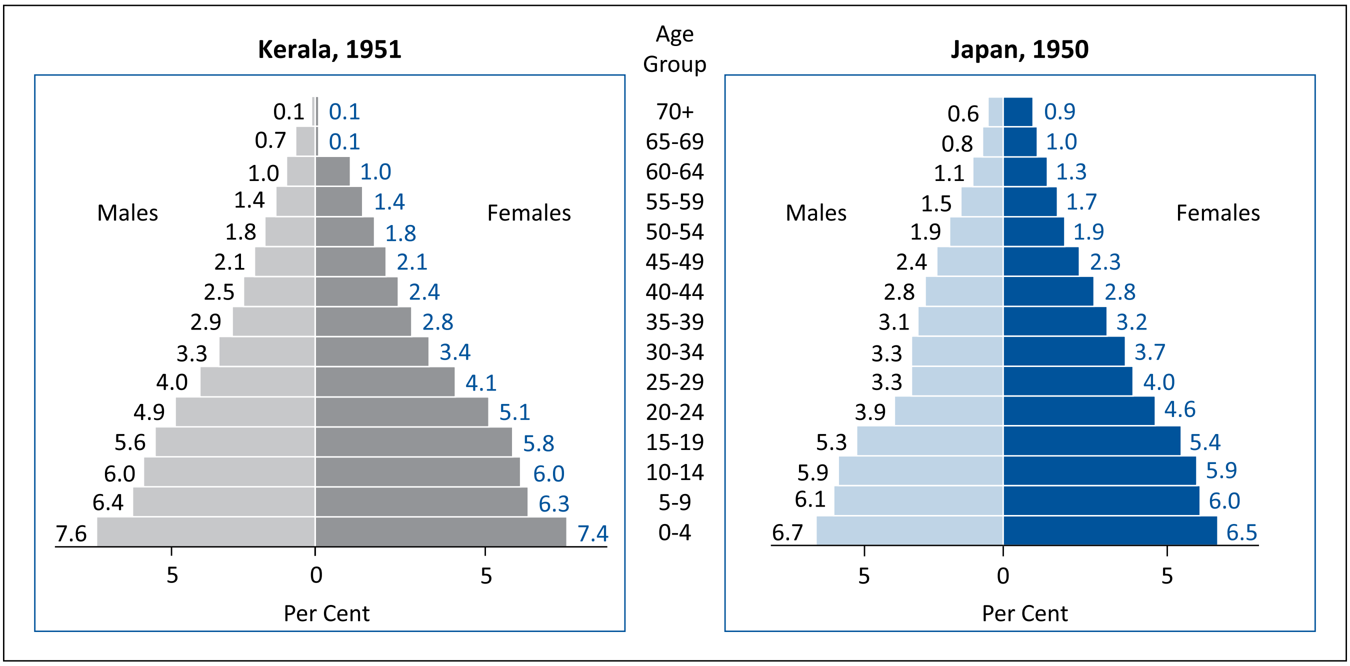


Figure 3.6: Age-Sex Pyramid - 2000–2001 (Kerala & Japan)

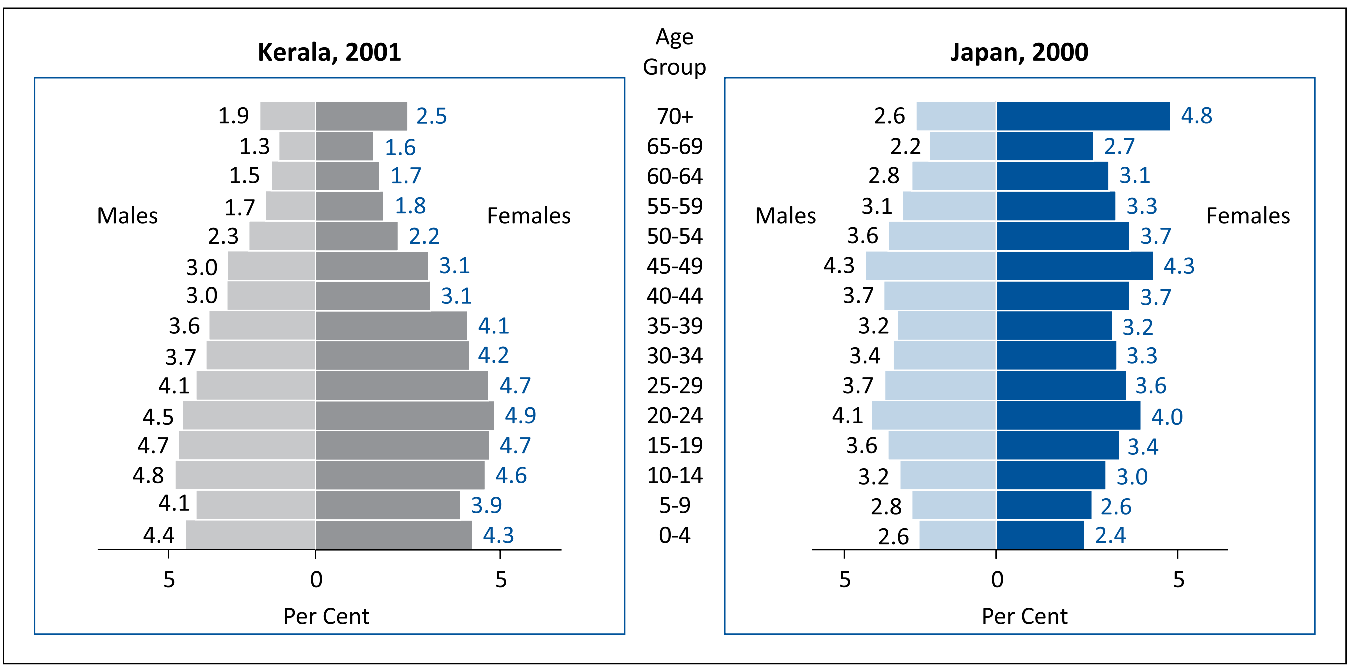
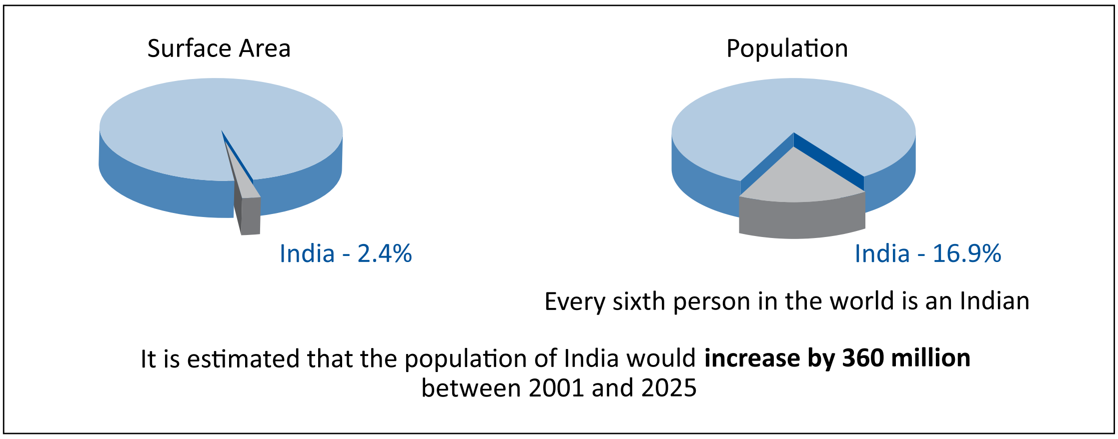
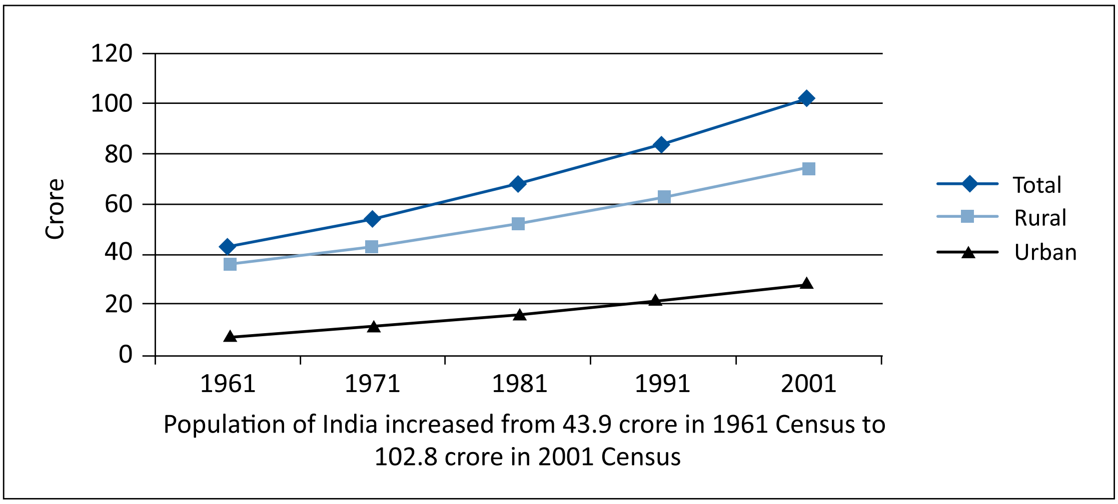


Figure 3.7: Data Highlights - 2001 Census



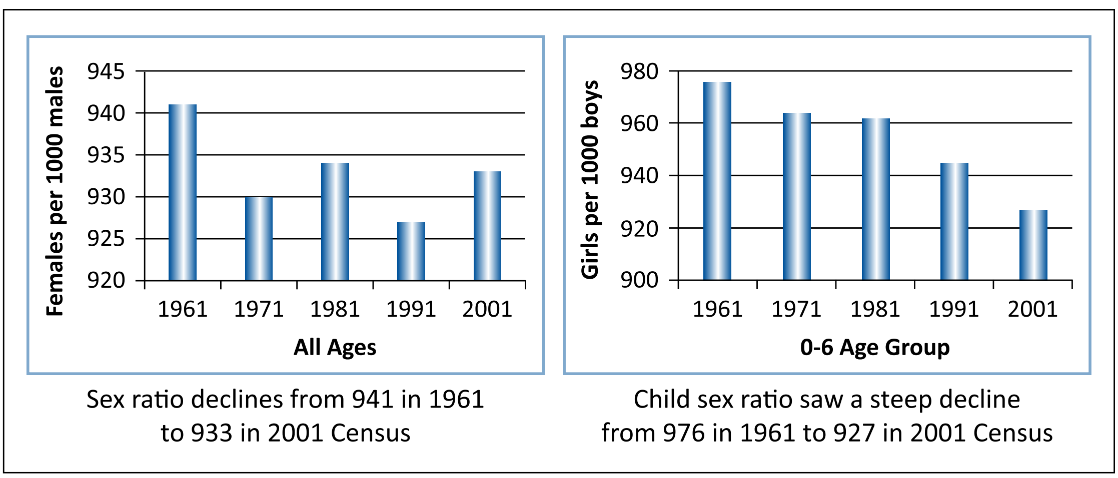
Source: Census of India

Figure 3.8: Population of India since 1961



Source: Census of India

Figure 3.9: Sex Ratio - India, 1961–2001



Source: Census of India

Exercises

Exercise 3.1: Prepare the appropriate graphs using the given data (Day 2 exercise).

Graph 1: Per cent distribution of live births by women age, 2010

Age group	India
15-19	7.4
20-24	39.6
25-29	31.3
30-34	13.2
35-39	5.9
40-44	1.8
45-49	0.8
All	100.0

Graph 2: Trends of Birth and Death rates (1971–2006), Rajasthan

Year	Birth Rates	Death Rates
1980	38.7	13.4
1981	37.1	14.3
1982	38.0	12.1
1983	40.1	13.6
1984	39.7	14.3
1985	39.7	13.2
1986	36.4	11.7
1987	35.1	11.6
1988	33.3	14.0
1989	34.2	10.7
1990	33.6	9.6
1991	35.0	10.1
1992	34.9	10.5
1993	34.0	9.1
1994	33.7	9.0
1995	33.3	9.1
1996	32.4	9.1
1997	32.1	8.9
1998	31.6	8.8
1999	31.1	8.4
2000	31.4	8.5
2001	31.1	8.0
2002	30.6	7.7
2003	30.3	7.6
2004	29.0	7.0
2005	28.6	7.0
2006	28.3	6.9

Graph 3: Levels of TFR

State	TFR
Rajasthan	3.1
Kerala	1.8
India	2.5

Exercise 3.2: Construct age-sex pyramids for Rajasthan (2001) and compare with Kerala.

Calculate age-dependency ratio and sex ratio

Age group	Population of Rajasthan, 2001		
	Persons	Males	Females
1	2	3	4
0-5	7270221	3800071	3470147
5-9	8109341	4263844	3845490
10-14	7278986	3866768	3412204
15-19	5518298	3002630	2515646
20-24	4773661	2483370	2290291
25-29	4218693	2138591	2080113
30-34	3840698	1926829	1913881
35-39	3471528	1797938	1673591
40-44	2774395	1483460	1290928
45-49	2281261	1189693	1091567
50-54	1784126	978936	805181
55-59	1356217	649757	706471
60-64	1368237	679624	688619
65-69	987902	459587	528325
70-74	761875	377967	383911
75-79	317546	146583	170967
80+	394202	174363	219845
Total	56507188	29420011	27087177

Step 1. Calculate proportion of age-sex specific population to total population.

Step 2. Multiply by (-1) to the male proportion.

Step 3. Select both columns and insert bar chart.
Edit chart by changing negative sign, colour, label fonts etc.

CHAPTER IV

Population Age-Sex Structures: Quality of Data and Adjustments

4.1 Concept and Significance of Age in Demographic Analysis

Age is probably the most important variable in demographic analysis. We need to distinguish the difference between the age of an individual and the age structure or distribution of a population. For demographic purposes, age of an individual at any time is defined as the number of completed years lived by the individual by that time. In other words, it is the 'age at last birthday' that is generally used. So, an age given as 30 years, means that the person is at present somewhere between 30 and 31 in terms of completed years of life. Every individual from the time of birth ages continuously and this is called his/her chronological age. In a population, the average age of its population distribution by age is called its *mean age*. The mean age of the population can increase or decrease with time, unlike that of an individual.

In developing countries with a high prevalence of illiteracy, age is one of the most difficult of personal characteristics to be ascertained accurately by direct enquiries. There are considerable errors and biases that have been observed in the reported ages.

There seems to be interdependence between the demographic processes of fertility, mortality and migration on the age structure of a population. Clearly the age structure of a population at a particular point of time is the result of past demographic processes of births, deaths and migration and all these processes are affected by age. Mortality rates fluctuate over a wide range at different ages. Child bearing is associated with only a limited part of the lifespan of women as is the event of marriage, and migration generally in adult ages seeking employment. The usefulness of age data is more visible when it is cross-classified with other demographic and socio-economic characteristics such as marital status, literacy, educational attainment, economic activity etc. While studying population statistics, it becomes evident that many of the above referred associated characteristics are unevenly distributed by age. Most of these characteristics are sufficiently significant only when they are observed according to sex and age and therefore the age data becomes an excellent instrument of separating and identifying the groups of people to be studied.

Apart from purely demographic concerns, the data on sex–age structure are widely required for a variety of administrative, scientific, technical and commercial purposes. It is of prime importance for formulating programmes to raise the standard of living.

NOTES

The dependency ratio, which is the ratio of economically active to economically inactive persons in a population, is dependent on the age composition. Many types of schemes, particularly planning of community institutions and services for the children, adolescents, youth and the elderly, require data on age composition. Age is an important variable in measuring mean age at marriage, potential school population, the potential voting population, potential manpower etc. Projections of the number of households and population, school enrolment, labour force, requirement of school teachers, health services, food and housing, problems of economic dependency etc. are some of the key areas in which the sex–age data are of prime importance. Several manufacturing and commercial ventures on consumer goods, leisure and entertainment industry, medical and health sector have immense utility for the age data to target their products and services. Hence it is important to know the efforts made in censuses and surveys to collect the information on age migrations. Also, the age structure of the population at present will affect the future trends of demographic events'.

NOTES

The change in the number of births first affects the child age group. After many years this effect can be seen in older age groups as well. In general, a decline in the number of births has the effect of narrowing the size of child population. If the decline is quite rapid, the effect on age structure will be quite dramatic. Decline in the level of infant deaths also produces the same results as caused by increase in the number of births, affecting the first years of life and then filtering towards higher ages. The effect of changes in deaths is generally to produce fewer variations than births in developed and developing nations. This is because deaths affect most age groups uniformly but births affect only the younger age groups. In developing countries, it is possible that a large number of deaths in a particular year is due to epidemics and natural catastrophes and these may have a big impact on some ages which is noticeable in the age structure for many years. Migration can also cause a big distortion in the age structure, as it is normally age selective. Migration is normally concentrated among young adults. Thus, the age structure of a population is primarily dependent on births, and on deaths and migration to a lesser extent. We tend to think of a population as growing older when people live longer, but it is when fertility declines, resulting in fewer births, that a population begins to age.

Age structure influences future demographic events as well. For example, a larger proportion of women in the childbearing ages will contribute to an increase in the number of births. High birth rate during one period will produce, a generation later, a comparatively larger number of women of childbearing ages and, consequently, a comparatively higher number of births could occur even if current birth rates have declined. In a number of developing countries, the birth rates are very high, and as a consequence the proportion of young children is high, because of high fertility in the past. Hence, even under rapid declines in fertility in the future, the number of births might keep on increasing because of rise in the number of women in the reproductive age in the future. This is simply because there is an increasing number and proportion of potential parents who will enter the childbearing ages, who are already born and thus the population will continue to grow in size. This aspect is generally known as population momentum discussed earlier. The population size in these countries will continue to grow rapidly for many years to come, irrespective of future levels and trends in births, deaths and migration. This is very much similar to the distance it takes for a speeding vehicle to come to a halt even after the breaks are applied. Age distribution of a population determines its momentum of growth.

Age is also the crucial variable in determining the size of the potential school-going population, manpower, voting population, women in the childbearing ages, older population requiring pension or other types of old age assistance etc. Tabulations by age are essential in the computation of basic measures relating to the factors of population change and in the study of economic dependency. Since the social and economic characteristics vary so much with age and also vary in time and place, populations cannot be meaningfully compared with respect to these characteristics unless age has been controlled. Accordingly, age is considered to be the variable of highest priority in demographic analysis. Tabulations by age are also very useful in the evaluation of the quality of data. Age composition of a population is of particular interest to the planner, as the planning of public institutions and services is largely dependent on it.

4.2 Sex Composition

NOTES

Sex of the individual holds an important place in demography. Data on sex is required for planning purposes, particularly health services. Sex is considered to be a biological characteristic that divides human beings into males and females. Though conventionally there are only two sexes, male and female, the third sex called the 'transgender' (neither male nor female) is becoming a vocal and assertive group in recent years and data on this group has been collected in the recent India census of 2011.

Demographic events are generally differentiated by sex. Fertility, or the number of children born to a couple, is measured in relation to the characteristics of the woman/wife, her age, duration of marriage and spacing between births. There are also substantial differences between the death rates of the sexes, especially at older ages, which are considered to be largely biological differences. Hence, it is important to consider the effect of variation in the sex composition of populations while comparing their fertility and mortality levels. The measure commonly used to represent the sex composition of a population is the sex ratio. It is simply the number of males per thousand/hundred females in a population. Though, internationally sex ratio is defined as males per 100 females, Indian censuses define sex ratio as females per 1,000 males.

$$\text{Sex ratio} = (P_m/P_f) * 1000$$

where

P_m = number of males in a population at a specified time

P_f = number of females in a population at a specified time

By definition, a ratio above 1,000 means an excess of males and a ratio below 1,000 means an excess of females. Hence, this simple measure can adequately describe the sex composition of the population. This ratio can be computed by individual age or age groups. Such age-specific ratios are useful in assessing the differential mortality, migration or under-enumeration in a population. The level of sex ratio in a population as a whole is affected by the sex ratio at birth, sex differentials in mortality and migration. Sex ratio for India has always been favourable to males and the decline in the proportion of females in the population has been a matter of great concern and debate. The sex ratio varies widely among the states in India, with Kerala being the only state with a consistently higher proportion of females in all recorded history. There is also some empirical evidence that there has been consistently an under-enumeration of females in the Indian censuses and that this factor has distorted the observed sex ratios in the population to some extent. Defined in this manner, the sex ratio for Kerala computed from the provisional census data for 2011 is

$$\text{Population sex ratio} = \left(\frac{\text{Number of females}}{\text{Number of males}}\right) \times 1000$$

$$= \left(\frac{17366387}{16021290}\right) \times 1000 = 1084$$

NOTES

(The population figures are the provisional figures given after the 2011 Census [Paper 1 of 2011]).

In 2001 based on the actual figures of the population by the two sexes in Kerala the population sex ratio was (16372760 / 15468614) ∗ 1000 = 1058.

Thus we can see that between 2001 and 2011, the female population in Kerala has grown faster than the male population, improving the sex ratio during the decade.

Another special type of sex ratio used in the Indian population is the child sex ratio (CSR) which is the ratio of female children below the age of 7 to the male children in the same age group. The Indian census collects data on those aged 7 and above – whether they are literate, attending a school or college and if completed education, the highest years of schooling. This is done primarily to study the trends in the literacy rates and educational levels of the population. This also makes it possible to separate the population of children below age 7 and study their characteristics even in the data published in the provisional totals. CSR has assumed enormous significance in the last two decades since a low female to male ratio is indicative of a higher incidence of sex-selective abortions which are now illegal.

$$\text{Child sex ratio in Kerala in 2011} = \frac{\text{Number of girls aged } 0-6}{\text{Number of boys aged } 0-6} \times 1000$$

$$= \frac{1626312}{1695935} \times 1000 = 959$$

Similarly CSR in 2001 based on the final totals of the census is

(1858119/1935027) ∗ 1000 = 960

Thus we see that the child population of Kerala, both boys and girls below the age of 7, in 2011 is significantly less than in 2001 both for boys and girls because of continuing declines in fertility during the decade 2001–2011 and the CSR has remained fairly stable.

4.2.1 Quality of age data

The data on the single year age–sex distribution of Kerala from the 2001 Census are given in Table 4.1. A line graph of the same is also given (see Figure 4.1). The graph shows that there are wild peaks at ages ending with digits 5 and 0 and troughs ending with digits 1 or 9. This indicates a strong digit preference in the reporting of ages by individuals even in the state of Kerala where the literacy rate in 2001 was high and close to 90%. There are definitely gross errors in the reporting of age data in Kerala and the data cannot be used as they are. The errors, their nature and magnitude and methods of correcting them are examined in the ensuing sections.

Table 4.1: Single-Year Age Distribution for Kerala, 2001

Age	Males	Females
0	268807	261599
1	269547	256679
2	271967	259994
3	299307	293449
4	299859	283956
5	279470	265751
6	246070	236691
7	254525	246946
8	265367	256307
9	250247	242807
10	302241	284532
11	262369	251243
12	329463	310105
13	315605	313011
14	314164	304467
15	292140	281682
16	284215	282637
17	285037	287513
18	347062	351575
19	276132	296513
20	334385	360833
21	252714	254044
22	294688	311645
23	275675	300394
24	283005	316607
25	288938	338713
26	271919	302061
27	247391	275145
28	287763	340418
29	200894	232953
30	351745	411899
31	172579	174839
32	261516	290248
33	205134	232658
34	194833	221012
35	319850	369310
36	215396	241872
37	181017	204843
38	267852	306417
39	170663	189134
40	334060	362040
41	130898	118918
42	217500	213831
43	142701	154642
44	135238	141456
45	292502	309553
46	147166	153662
47	144539	148095
48	228715	231476
49	139099	131337
50	248960	276261
51	100803	84330

Contd...

NOTES

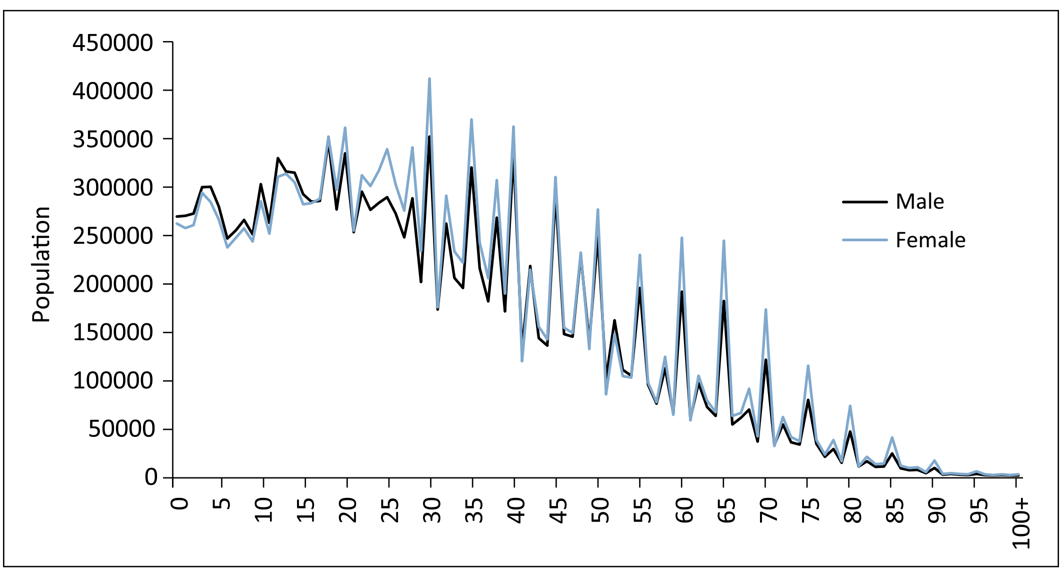
Age	Males	Females
52	161217	146984
53	109899	103405
54	103822	101839
55	195029	228911
56	94820	96822
57	74964	76487
58	111714	123228
59	65141	63128
60	190961	246873
61	59392	57312
62	96211	103616
63	71412	77915
64	62369	66075
65	181341	243843
66	53484	62058
67	60291	65194
68	68967	90229
69	35588	41020
70	120282	172372
71	32050	31000
72	53242	60922
73	34950	39959
74	32769	35876
75	78869	114231
76	33181	37292

Age	Males	Females
77	19949	21856
78	28000	36992
79	13559	15160
80	46199	72511
81	9735	10125
82	15343	19619
83	9324	11873
84	9818	12675
85	23416	39641
86	7798	10282
87	6107	7921
88	6243	8910
89	3013	3981
90	8165	15622
91	1511	1868
92	2016	2797
93	1383	2007
94	1164	1534
95	2320	4436
96	937	1371
97	700	917
98	874	1353
99	589	751
100+	690	1474
ANS*	14064	12390

* Age not stated

NOTES

Figure 4.1: Single-Year Age Returns in Kerala, 2001



4.3 Errors in Demographic Data

Demographic data pertaining to various characteristics of population such as age, marital status, duration of marriage and birth intervals are subject to a variety of errors, both in developed and developing countries. Of course, for various reasons, mainly because of the low levels of literacy of the people, the errors are much larger in developing countries than in the developed countries. These errors arise partly out of the illiteracy or lack of knowledge on the part of the respondent about the variable he is enquired about, but are also due to the errors and biases introduced by the investigators who interview the respondents for data collection and the methods adopted for data collection including the type of questionnaire used and the time spent on the interview. These errors, termed *Response Errors* and *Biases*, are different in nature and origin from the sampling errors which arise because of the procedures of sampling adopted in the data collection in which we do not cover the whole population. They are also referred to as *non-sampling errors* to distinguish them from *sampling errors*.

In developing countries, it has been found from a number of studies that non-sampling errors are much more important in determining the overall quality of data of the demographic variables than the sampling errors. The subject of assessment of and adjustment for such errors in demographic data is quite vast and has grown considerably during the past two decades. Errors that are found in many demographic variables can be broadly classified into two types:

1. Errors of coverage of the population supposed to be included in the study called *errors of omission*.
2. Errors on the accuracy of information on the various characteristics of the persons included in the study referred to as *errors of commission*.

The second type of errors is subdivided into two categories: 'errors' and 'biases'. Errors arise mainly because of wrong reporting of information on the characteristic by the respondent, unconsciously, without any motive to misreport, and are attributable to his/her ignorance or because he/she has not understood the question properly. There are also unintended errors in recording the data by the interviewers including errors in coding and tabulation. On the other hand, a bias arises when there is a purposive misreporting on a characteristic and consequently errors are introduced in the data. There is no randomness in such errors and hence they are termed as biases. For example, an unmarried person, male or female, when asked about his/her age may deliberately understate the age hoping thereby to remain in the marriage market; young men and women in many countries may understate or overstate their ages to avoid military conscription depending on the age limits fixed for military duty.

When information on age is not reported by a respondent it is customary to assume that for each sex, their age–marital status distributions would be the same as for those for whom the data is available. From Table 4.1 we find that the age data for 26,454 persons is not available but data on their sex and marital status are available. Hence these persons were distributed within each marital status category as per the age distributions of those for whom the information was available.

NOTES

There are also unconscious prejudices and biases for numbers ending with certain digits, particularly those ending with 0 and 5, in developing countries. Even educated people when asked about their age may state it close to the nearest digit ending with 0 or 5. This type of bias is called *heaping* or *digit preference* and is very common in the age data of developing countries. In developing countries, the order of preference of numbers in terms of digit preference has been found to be 0, 5, 2, 4, 6 and 8, and the least preferred are the other odd numbers. Though there is a universal bias of reporting of numbers ending with multiples of 5 or 0, the extent of preference varies between the developed and the developing countries. In the category of biases there is also another type of bias wherein there is systematic over-reporting or under-reporting of ages that is culturally oriented, such as widows and older persons in India over-reporting their ages. These biases tend to push the population totals up or down to higher or younger ages respectively.

In this chapter we deal with illustrative methods of studying the quality of age and birth interval data compiled in a survey or census. Data on the age distribution of a population as a whole or those suffering from specific diseases is very important for the planning and implementation of public health programmes, since the aetiology of many diseases varies with age and the planning and implementation of programmes for prevention, management and control of the diseases also varies with age. There are a number of techniques available to estimate the extent to which data on age and birth intervals (or any intervals) are misreported because of unconscious errors and due to conscious biases. A few simple techniques for assessing the extent of age displacement or digit preference and simple methods of adjustment for them are given in the following sections. Students interested in greater details on these methods may refer to the books by Shryock and Seigel (1971) and Arriaga (1994).

NOTES

4.4 Whipple's Index

Whipple's Index (WI) is a measure of preference for ages ending with the digits 0 and 5, observed in a given population. The preference for these two digits is studied only between the ages 23 and 62 because outside this range, shifting and other problems often tend to confuse the pattern of heaping. The index is calculated simply by expressing the population with reported ages ending with the digits 0 and 5 as a percentage of the total population aged 23 to 62, and then multiplying it by 500. The range of the index is then from a minimum of 100 indicating practically no preference to these two digits to a maximum of 500 when all report their ages with digits ending with 0 or 5. The index is studied separately for males and females since the extent of digit preference has been found to vary between the two sexes.

WI for Kerala using the 2001 Census data given in Table 4.2 is

$$\text{WI} = 500 * \frac{(\text{Sum of populations aged } 25,30,35,....60)}{(\text{Sum of populations aged 23 to 62})}$$

Table 4.2: Computation of Whipple's Index (WI) for Males and Females

Age	Males	Females	Age	Males	Females
1	2	3	4	5	6
23	275675	300394	25	288938	338713
24	283005	316607	30	351745	411899
25-29	1296905	1489290	35	319850	369310
30-34	1185807	1330656	40	334060	362040
35-39	1154778	1311576	45	292502	309553
40-44	960397	990887	50	248960	276261
45-49	952021	974123	55	195029	228911
50-54	724701	712819	60	190961	246873
55-59	541668	588576			
60	190961	246873			
61	59392	57312			
62	96211	103616			
Total	7721521	8422729		2222045	2543560
WI				144	151

Note: WI for males is obtained by dividing the total in column 5 by the total in column 2 and multiplying this value by 500. WI for females is obtained by dividing the total in column 6 by the total in column 3 and multiplying this value by 500

NOTES

Computed separately for men and women and for 2001 Census of Kerala given in Table 4.2 it works out to 144 for men and 151 for women. In 1991 these values were 165 for males and 173 for females. These values indicate the continuing prevalence of high digit preference in Kerala and there is a strong bias for digits ending with 5 and 0 in spite of high literacy levels in the state. The digit preferences appear to be more cultural than scientific.

The UN has given a possible rating on the quality of data on the basis of WI as follows:

Quality	Range of WI
Highly accurate	under 105
Fairly accurate	105-110
Approximate	110-125
Rough	125-175
Very rough	over 175

The Kerala age data compiled from the 2001 Census judged from the above criteria can be considered to be 'Rough'.

4.5 Myers' Blended Index

Whipple's Index measures the extent of preference for numbers ending with digits 0 and 5. It does not measure the extent of preference for other digits. It is also observed that in any population that has been growing or declining steadily due to the effects of changes in fertility and mortality over time, the number ending with digit 0 is found to be higher than the next age ending with digit 1, which will be higher than the next ending with digit 2 and so on, that is, population at age 10 will be more than population at age 11, which is higher than age 12 and so on. In some populations, because of rapid declines in fertility it can be the other way round also in younger ages. Myers has developed a method for assessing the extent of preference or dislike for all digits, 0 to 9, using a 'blended' method to avoid a possible bias in the index due to the fact that numbers ending in 0 would normally be larger than the following numbers ending in 1 to 9 because of the effect of mortality. The principle employed is as follows: first, to begin to count the number of people at ages ending with each of the 10 digits in the population beginning with age 10 and ending at 69 or 89, the maximum at which data are available; second, to do the same thing beginning with age 20 separately; third, to blend them (-weighted sum) in the ratio of x + 1 and 10 — x — 1 for the numbers ending with digit x; fourth, compute the percentages of the blended sum to the total; fifth, take the absolute deviation of this percentage from 10 (which is supposed to be the expected percentage of the blended number at any digit, in the absence of any digit preference); finally, compute the average of the results over the 10 digits. The abbreviated procedure of calculation calls for the following steps:

NOTES

Step 1: Add the populations ending in each digit over the whole range, starting with the lower limit of the range beginning with 10 (i.e. sum of population at ages 10, 20, 30, 80 placed at 0 digit; sum of populations at ages 11, 21, 31,...81 placed at digit 1 and so on).

Step 2: Ascertain the sum of populations ending with different digits starting from age 20 (i.e. sum of populations at ages 20, 30, 40, ... 90 placed at digit 0 and sum of populations at ages 21, 31, 41, ... 91 placed at digit 1 and so on). Commonly, the same number of populations with different digits is used. For example, in Step 1, if we use the sum of eight numbers at digit 0, that is, sum of populations at ages 10, 20, 30, and 80, in Step 2 we should also use eight numbers for digit 0, that is, sum of populations at ages 20, 30, 40 and at 90. But this is not necessary. The population above age 99 is not included.

Step 3: Weight the sums in Steps 1 and 2 and add the results to obtain a blended population (weights assigned are 1 and 9 for the 0 digits, weights 2 and 8 for 1 digit and so on).

Step 4: Convert the distribution in Step 3 into percentages.

Step 5: Take the absolute deviation of each percentage in Step 4 from 10.0, the expected value for each. The deviations are measures of digit preference or aversion. They should be close to 0 if there is no preference or aversion.

Step 6: The sum of these absolute deviations divided by 2 provides a 'Summary Index' of age preference across all digits. This index is expected to vary from 0 to 90.

The results in Step 6 indicate the extent of concentration on or avoidance of a particular digit. The higher the value of the absolute deviation for a digit, higher is the preference or avoidance of the *digit*. An illustrative application of Myers' Index for the age distribution of Kerala's male population is given in Table 4.3.

Table 4.3: Calculation of Preference Indexes for Terminal Digits by Myers' Blended Method for Males in Kerala, 2001

Terminal digit, a	Population with terminal digit 'a'		Weight for		Blended population		Absolute deviation of col.7 from 10
	Starting at age 10+a	Starting at age 20+a	col.2	col.3	(2*4)+(3*5)	Per cent	
1	2	3	4	5	6	7	8
0	1928833	1634757	1	9	16641646	14.60	4.60
1	1020540	759682	2	8	8118536	7.12	2.88
2	1429180	1101733	3	7	11999671	10.53	0.53
3	1164700	850478	4	6	9761668	8.57	1.43
4	1142095	829095	5	5	9855950	8.65	1.35
5	1672085	1382265	6	4	15561570	13.66	3.66
6	1107979	824701	7	3	10229956	8.98	1.02
7	1019295	734958	8	2	9624276	8.45	1.55
8	1346316	1000128	9	1	13116972	11.51	1.51
9	904089	628546	10	0	9040890	7.93	2.07
Total					113951135	100.00	20.60
Myers' Index = (Total of col.8/2)							10.30

Note: When the value of 'a' is 2, the corresponding population for 10+a and 20+a is obtained as follows:
Population for 10+2 is obtained by adding the population aged 12, 22, 32, 42, 52 and 62.
i.e. P(10+2) = P12+P22+P32+P42+P52+P62
Population for 20+2 is obtained by adding the population aged 22, 32, 42, 52, 62 and 72.
i.e. P(20+2) = P22+P32+P42+P52+P62+P72

NOTES

For Kerala, it is found that for the male population, preferences for different digits vary widely: the strong preferences being for numbers ending with digits 0 and 5, with mild preferences for digits 8 and 2 with strong aversions for numbers ending with digits 1, 9, 7, 4, 3 and 6 in that order. The summary index of preference or digit preference or aversion is found to be 12.32 and has declined marginally from 15.09 in 1991. Such indices can be computed for the population of the two sexes separately and for the total population.

4.6 Digit Preference Quotients for Birth Intervals

It is found that in developing countries in any retrospective survey answer for any question on the intervals between successive births, the date of onset of disease and current date of recovery or death is in years or in multiples of 6 months. In such cases if the reported intervals computed in months are divided by 12 or 6 and classified by their residues 0, 1, 2, 3, ... 11 in the first case and 0, 1, 2, ... 5 in the second case, there will be an undue heaping of frequencies at 0 and 6 in the first case and 0 in the second case. If the frequencies of birth intervals ending with a residue of 0, 1, 2 and 11 are denoted by f_1, f_2 and f_{11} and the total of all frequencies by f, then the digit preference quotient (Q_1) is defined as

$$Q_1 = \frac{100 * \text{Sum}\left(\left(12 * f_i\right) - f\right)}{\left(22 * f\right)}$$

where the sum of the numerators values over 'i' is only on the absolute values.

The value of Q_1 ranges from 0 to 100, where 0 is an indication of practically no bias in the birth intervals and 100 the case when the intervals are all clustered in intervals of 1 year or 6 months. Similarly, if we are considering the biases in terms of multiples of 6 months or six-monthly preferences, we have the digit preference quotient, Q_2, given as

$$Q_2 = \frac{100 * \text{Sum}\left(\left(6 * f_i\right) - f\right)}{10 * f}$$

From an analysis of the data collected in the Fiji Fertility Survey conducted in 1974 where the quality of data is good, the Q_1 index for the birth intervals of different birth orders 0 to 1 (i.e. marriage to the first child) and 1 to 2, that is, between the first and the second child, was found to be 9 and 5 respectively, indicating a fairly high quality of interval data. In order to compute the quotients we need data on birth intervals recorded in completed months. But there is a bias for digits ending with 0 and 5 in the reporting of age by individuals in India, there is also a bias in the reporting of birth intervals in multiples of 6 months or 1 year.

NOTES

4.7 UN Joint Score or Accuracy Index Based on Age Data in 5-year Age Groups

The Whipple's Index and the Myers' Index discussed earlier are based on the data compiled in single years of age. Even when the age data are grouped into five-year age groups, as is the convention in demographic analysis, there may still be errors of age misreporting. In many situations there is a shifting of data from one age group to the other, usually the next age group or the previous one. This is also called *age displacement* and it can vary between the two sexes. The UN Accuracy Index, also called the *UN Joint Score* (UNJS) is a combined or composite indicator of the overall age displacement in five-year age groups and the differential errors in age misreporting between males and females. This is computed in five steps as follows:

Step I. We define the age ratio for any age group 'a' to a+4: (${}_5P_a$) is defined as

$$\text{Age ratio} = \frac{{}_5P_a}{0.5 * ({}_5P_{a-5} + {}_5P_{a+5})} * 100$$

Step II. The age ratio scores are computed separately for men and women. The female age ratio score (FARS) over 'n' age groups is defined as

$$\text{FARS} = \frac{1}{n-2} * \sum \left[200 * \left(\frac{{}_5F_1}{{}_5F_{i-5} + {}_5F_{i+5}} \right) - 100 \right]$$

where

${}_5F_1$ = population of females in the age group i to i + 4; the sum is taken over all age groups with the absolute values ignoring the sign over all the age groups assumed to be 'n' in number.

Similarly the male age ratio score (MARS) is defined as

$$MARS = \frac{1}{n-2} * \sum \left[200 * \left(\frac{{}_5M_1}{{}_5M_{i-5} + {}_5M_{i+5}} \right) - 100 \right]$$

where

${}_5M_1$ = population of males in the age group i to i + 4 where the sum is taken over all the age groups assumed to be 'n' in number.

Step III. The sex ratio for any age group is computed as

$$100 * \frac{{}_5M_1}{{}_5F_1}$$

Step IV. The sex ratio score (SRS) is computed as

$$SRS = \left(\frac{1}{n} \right) * \sum \left(\frac{{}_5M_1}{{}_5F_1} - \frac{{}_5M_{i-5}}{{}_5F_{i-5}} \right) * 100$$

Step V. The UNJS or the UN Accuracy Index is defined as

UNJS = FARS + MARS + 3SRS

For the state of Kerala from the data on the age distribution by sex given in Table 4.4 it can be calculated that MARS = 7.39, FARS = 8.47 and SRS = 4.69 and hence the UNJS = 29.93.

NOTES

Table 4.4: Computation of UNJS for Males and Females in Kerala, 2001

Age Group	Males	Females	Sex Ratio	Absolute Difference between Successive Sex Ratio	Age Ratio (Males)	Absolute Difference of Age Ratio from 100	Age Ratio (Females)	Absolute Difference of Age Ratio from 100
1	2	3	4	5	6	7	8	9
0-4	1409487	1355677	103.97					
5-9	1295679	1248502	103.78	0.19	88.34	11.66	88.58	11.42
10-14	1523842	1463358	104.13	0.35	109.62	9.62	106.49	6.49
15-19	1484586	1499920	98.98	5.16	100.16	0.16	99.77	0.23
20-24	1440467	1543523	93.32	5.65	103.58	3.58	103.27	3.27
25-29	1296905	1489290	87.08	6.24	98.76	1.24	103.63	3.63
30-34	1185807	1330656	89.11	2.03	96.73	3.27	95.02	4.98
35-39	1154778	1311576	88.05	1.07	107.61	7.61	112.99	12.99

Contd...

Age Group	Males	Females	Sex Ratio	Absolute Difference between Successive Sex Ratio	Age Ratio (Males)	Absolute Difference of Age Ratio from 100	Age Ratio (Fe-males)	Absolute Difference of Age Ratio from 100
1	2	3	4	5	6	7	8	9
40-44	960397	990887	96.92	8.88	91.17	8.83	86.7	13.3
45-49	952021	974123	97.73	0.81	112.99	12.99	114.35	14.35
50-54	724701	712819	101.67	3.94	97.04	2.96	91.23	8.77
55-59	541668	588576	92.03	9.64	89.9	10.10	93.08	6.92
60-64	480345	551791	87.05	4.98	102.06	2.06	101.16	1.16
65-69	399671	502344	79.56	7.49				
70+	604196	797328	75.78					
Total (from age 5 to 65)				48.94		74.08		87.51
Average totals (for 12 values)				4.08 (S)		6.17 (M)		7.29 (F)
UNJS 3S+M+F					25.70			

Note: Age ratio score has been worked out excluding age group 5–9. Sex ratio score has been worked out, excluding age group 0–4. Sex ratio score (S) is computed as the average of the 12 values of the differences in the successive sex ratios irrespective of sign. Male age ratio score (M) is computed as the average of the deviation of male age ratio from 100 irrespective of sign. Similarly, the age ratio score for females (F) is computed

NOTES

In the above total, we start from age group 5–9, but omitting the last one, 65–69. The age ratios for 65–69 involve the 70+ age group, which normally have much larger number due to open age interval. Obviously, age ratios for 65–69 are low partly due to open interval in denominator and hence we exclude the same. The Population Analysis Spreadsheets (PAS) programme also computes in a similar fashion.

The UN has recommended the following criteria for adjudging the overall quality of data in terms of accuracy of age reporting on the basis of UNJS as follows:

Value of UNJS	Quality of Data
Below 20	Accurate
20 to 40	Inaccurate
Above 40	Highly inaccurate

According to the above classification the quality of age data in Kerala based on the UNJS of 26.29 can be judged as ‘inaccurate’ but not ‘highly inaccurate’. It is obvious that the lower the score, the higher is expected to be the accuracy of the data with regard to age displacement from one age group to the next and the better is the reporting with regard to the two sexes. Unlike the WI which measures the heaping of the age statements at ages with specified digits as 0 or 5, the UNJS measures the extent of displacement of the aggregate age distribution to adjacent age groups because of systematic over- or under-statement of age. It is really surprising that even with grouped data the age distribution of Kerala, which is the most literate state in the country, has such highly inaccurate age data.

4.8 Smoothening of Age Distributions

The distribution of population by age as obtained from the census data is, as mentioned above, subject to various types of errors of omission and commission. The extent of errors in age reporting can be gauged from the single-year-based line graph for Kerala given in Figure 4.1. The age graph has a number of spikes and dips which cannot be expected in any human population. Most serious among the observable errors is the digit preference that tends to cluster a large proportion of the population in ages with digits ending in 5 or 0. Even in the grouped data, in 5-year age groups, there are serious errors with figures shifted from one age group to adjacent groups. If the age data are to be useful for any planning purposes and meaningful demographic analysis including population projections, they have to be adjusted for these errors. These are called *smoothening procedures* and a number of such methods are available in the literature. There is no generalized method applicable to all populations. It depends on the nature and magnitude of errors in the age–sex distributions and as such the age structure has to be analyzed before deciding on the type of smoothening to be adapted, that is, whether it should be light or strong. As a first step we should look at the data carefully to take note of the troughs and bumps in the single-year and five-year age distributions.

NOTES

Generally, these methods assume that the total reported population in the census is correct but its distribution by age is wrong and needs correction. The jaggedness in the age distribution is smoothened out by these procedures. Notable among these methods are those developed by Carrier-Farrag, Karup-King Newton, Arriaga and the UN, and a strong moving-average method called the *strong method*. The first three methods assume that the population totals given in the census in 10-year age intervals are correct, need no adjustments, and hence are used in splitting the 10-year age intervals into component 5-year intervals. The latter two methods are used when even the 10-year totals are adjusted for possible spillover errors in successive intervals. [For details on each of these methods refer to Shryock and Seigel (1971) and for a computer-based application and illustration of the methods refer to *Population Analysis Spreadsheets (PAS)-Presentation of Techniques* by Arriaga (1994).] These methods are based on the assumption that the age distributions of any human population, throughout the world, have some common patterns which can be described by some well-known mathematical curves or functions. In the absence of any reporting errors, the age distribution of any population declines in its numbers rapidly from age 0 to age 1 because of high infant mortality, slows down in its pace of decline from ages 1 to 4, and slows down further from ages 5 to 15. The pace of decline rises with slow rise in mortality rates from age 15 until 50 and then rapidly declines in numbers with a fast rise in mortality rates in older ages, until it reaches very high levels after age 80. In the absence of a war or famine or other natural calamities, or sharp declines in fertility that tend to severely change the population age distributions, that is, in normal times, the populations by age follow well-known patterns, which can be captured by mathematical functions. This characteristic of existence of common patterns in the age structures of populations is used in the smoothening of the age distributions. Two approaches are used in such smoothening methods.

4.8.1 Method of moving averages: Strong method

In this method, the corrected population size at any age 'i' is considered to be equal to the average of the three observed values at adjacent ages i-1, i, i+1, that is,

$$P_i \text{ (Corrected)} = (P_{i-1} + P_i + P_{i+1}) / 3$$

This is termed as the 3-year moving average of the observed values P_{i-1}, P_i and P_{i+1} at ages i -1, i and i+1. The rationale behind this approach is that in the observed population while some members whose real age is i-1 would have reported it as i, some at i would have reported as i+1 and this shifting from age to age will be largely confined to adjacent ages and a moving average over 3 years will largely eliminate this type of error in reporting. In grouped data, with 5-year age groups the formula used is

$$P_i \text{ (Corrected)} = (P_{i-1} + 2P_i + P_{i+1}) / 4,$$

where P_i = Adjusted population in the age group i to i + 4

NOTES

In severe cases of errors in reporting of age, where it is felt that errors could spill over to adjacent ages, a 5-year moving average can be used. In this case for individual ages

$$P_i \text{ (Corrected)} = (P_{i-2} + P_{i-1} + P_i + P_{i+1} + P_{i+2}) / 5$$

In populations of very poor quality data, even a 11-point moving average has been used on ages. In such adjustments using moving averages, it has to be remembered that the total of adjusted populations over all ages or a given age range may differ slightly from the observed totals, and if we have to keep the observed and actual totals the same then we have to adjust the population sizes in each age pro rata after smoothening so that the totals will agree. Further, it has to be kept in mind that smoothening of any distribution by any method, however sophisticated it may be, will tend to mask or overwrite the real changes in the age distributions of a population. Hence, use should be made of the actual enumerated population to the extent possible excepting for population projections when the errors on age reporting of the base population get cumulated over time.

The method of moving averages can also be used to smoothen populations tabulated in 5-year age groups, usually as averages of three adjacent age groups. In the strong method of smoothening, when errors of age reporting are found to spill over to beyond the adjacent age groups, the following procedure has been adopted. First, the population totals are obtained by 10-year age groups 0–9, 10–19, 20–29 etc. Such a 10-year grouped population is first smoothened by using the moving average formula:

$$P_i \text{ (Corrected)} = (P_{i-10} + 2P_i + P_{i+10}) / 4,$$

where P_i = Population between ages i and i + 10

Table 4.5: Population by 10-year Age Group, Kerala

Age	Population
0–9	2707611
10–19	3011146
20–29	2739846

For example, as given in Table 4.5, for Kerala the adjusted population of males between ages 10 and 20 (called 10 to 19), is

$$P_{10} \text{ (Corrected)} = (2707611 + (2 \times 3011146) + 2739846) / 4 = 2867437$$

After getting the corrected population totals in successive 10-year age groups by this method, we adjust the total of the adjusted population to the actual on a pro rata basis. The adjusted totals are then split into 5-year age groups by using Arriaga's method:

$$_5P_{x+5} = (-\,_{10}P_{x-10} + 11 * {}_{10}P_x + 2 * {}_{10}P_{x+10}) / 24$$

$$\text{and } {}_5P_x = ({}_{10}P_x - {}_5P_{x+5})$$

where

$_5P_x$ = Population in the ages x and x + 5

$_{10}P_x$ = Population in age x and x + 10

The strong method of smoothening has been used routinely in the smoothening of the Indian census data for the purposes of population projection.

NOTES

4.9 Use of Polynomial and Other Curves

The standard procedure used in the developing countries for the smoothening of age distributions of populations is the fitting of mathematical curves to the observed age distributions and using these curves to obtain an estimate of the population for any age group. The mathematical curves used are the third- or fifth-degree polynomials and the logistic curve. Many other types of functions have also been used in the literature.

For the polynomial functions the population at any age x is written as

$P(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ for the third-degree and

$P(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5$ for the fifth-degree polynomial

It is assumed that the constant 'a' can be estimated from four successive observations in the case of a third-degree polynomial and from six successive values in the case of a fifth-degree polynomial. In the case of assumption of a logistic curve for describing the population age pattern the function used is

$$P(x) = \frac{K}{e^{(a+bx)}}$$

and the constants can be estimated from three adjacent values.

The UN (Carrier and Farrag, 1959 quoted in United States Bureau of Census, 1994) developed the following formula for the smoothening of populations in 5-year age groups from ages 10 to 74, assuming that 5-year age distributions of the population are available up to age 85. The formula is based on fitting second-degree curves through five successive 5-year age groups.

P_i (corrected) = $(-\,{}_5P_{i-10} + 4 * {}_5P_{i-5} + 10 * {}_5P_i + 4 * {}_5P_{i+5} - {}_5P_{i+10}) / 16$

P_i being the population in the age range i to i+4

NOTES

For the data on male population of Kerala smoothened by the strong method and given in Table 4.6 we have the following figures

Age	Population
0–4	1313082
5–9	1394528
10–14	1460746
15–19	1450827
20–24	1413372

Therefore, ${}_5P_{10}$ (Corrected) = [–1313082 + (4 × 1394528) + (10 × 1460746) + (4 × 1450827) – 1413372] / 16 = 1453902.

Table 4.6: Reported and Smoothened Population by Age and Sex, Kerala 2001

Sex and Age	Reported	Smoothened				
		Carrier-Farrag	Karup-King Newton	Arriaga	United Nations	Strong
Male Total, 0-79	15,311,027			15,311,027		15,311,027
Total, 10-69	12,156,162	12,156,162	12,156,162	12,156,162	12,113,515	12,156,162
0-4	1,410,761			1,291,912		1,313,082
5-9	1,296,850			1,415,699		1,394,528
10-14	1,525,219	1,503,346	1,503,558	1,515,534	1,470,673	1,460,746
15-19	1,485,927	1,507,800	1,507,588	1,495,612	1,508,269	1,450,827
20-24	1,441,769	1,412,893	1,411,701	1,414,322	1,427,600	1,413,372
25-29	1,298,077	1,326,953	1,328,145	1,325,523	1,303,350	1,336,700
30-34	1,186,878	1,223,825	1,222,956	1,223,610	1,205,084	1,241,408
35-39	1,155,821	1,118,875	1,119,744	1,119,089	1,118,739	1,129,378

Contd...

Sex and Age	Smoothened					
	Reported	Carrier-Farrag	Karup-King Newton	Arriaga	United Nations	Strong
40-44	961,265	1,030,414	1,024,272	1,028,815	1,008,451	1,008,834
45-49	952,881	883,732	889,874	885,331	911,082	879,416
50-54	725,356	695,053	698,340	692,925	736,979	738,016
55-59	542,157	572,460	569,173	574,588	555,825	614,990
60-64	480,779	497,429	491,672	492,648	473,603	497,111
65-69	400,032	383,382	389,139	388,163	393,858	385,364
70-74	273,540			279,774		276,559
75-79	173,715			167,481		170,696
80+	157,487					
	15,468,514					

In addition to the above there are three formulae which are used to split the data given in 10-year intervals into component 5-year groups assuming that the data in 10-year intervals are correct – (1) Carrier–Farrag; (2) Karup–King–Newton; and (3) Arriaga. The formulae used are briefly stated and illustrated below using Kerala data without explaining how the formulae are derived.

NOTES

4.9.1 Carrier–Farrag formula

This formula is used to split the 10-year age group data into component 5-year age groups and is used when the given 5-year age group data is considered not very reliable.

$_5P_{x+5}$ = 10 P_x / [1+ (10 P_{x-10} / 10 P_{x+10}) ¼] and

$_5P_x$ = 10 P_x – $_{-5}P_{x+5}$

where

$_5P_{x+5}$ is the population in the age-group x+5 to x+9;

$_{10}P_x$ is the population in the age-group x to x=9; and

$_5P_x$ is the population in the age-group x to x+4.

From actual data given in Table 4.6, the adjusted population of males in Kerala in the age group 20–24 and 25–29 is estimated using Carrier-Farrag formula:

Age Group	Males
10–19	3,011,146
20–29	2,739,846
30–39	2,342,699

From the above figures we estimate $_5P_{20}$ as follows

$_5P_{25}$ = 2,739,846 / [1 + (3,011,146 / 2,342,699) ¼] = 1,326,953 and

$_5P_{20}$ = 2,739,846 – 1,326,953 = 1,412,893

4.9.2 Karup–King–Newton formula

This is also another formula to split 10-year age group into component 5-year age groups.

$_5P_x = {}_{10}P_x / 2 + ({}_{10}P_{x-10} - {}_{10}P_{x+10}) / 16$ and

$_5P_{x+5} = {}_{10}P_x - {}_5P_x$

where the symbols are as defined above.

For example,

$_5P_{20}$ = 2,739,846 / 2 + (3,011,146 + 2,342,699) / 16 = 1,411,701 and

$_5P_{25}$ = 2,739,846 – 1,411,701 = 1,328,145

4.9.3 Arriaga formula

This is the third formula to split 10-year age groups into 5-year age groups given by Arriaga and widely used:

$_5P_{x+5} = (-{}_{10}P_{x-10} + 11{}_{10}P_x + 2{}_{10}P_{x+10}) / 24$ and

$_5P_x = {}_{10}P_x - {}_5P_{x+5}$

NOTES

where the symbols are as defined above.

For example,

$_5P_{25}$ = {–3,011,146 + (11 × 2,739,846) + (2 × 2,342,699)} / 24 = 1,325,523 and

$_5P_{20}$ = 2,739,846 – 1,325,523 = 1,414,323

Arriaga has also given the formula for smoothening population in the youngest and oldest age groups. When the 10-year age group to be separated is an extreme age group (the youngest or the oldest), the formulae are different. For the youngest age group, the formula is as follows:

$_5P_{x+5} = (8\,{}_{10}P_x + 5\,{}_{10}P_{x+10} - {}_{10}P_{x+20}) / 24$ and

$_5P_x = {}_{10}P_x - {}_5P_{x+5}$

For the oldest age group coefficients are reversed.

$_5P_x = (-{}_{10}P_{x-20} + 5\,{}_{10}P_{x-10} + 8\,{}_{10}P_x) / 24$ and

$_5P_{x+5} = {}_{10}P_x - {}_5P_x$

For example,

Age group	Males
0-9	2,707,611
10-19	3,011,146
20-29	2,739,846
50-59	1,267,513
60-69	880,811
70-79	447,255

${}_5P_5$ = {(8 × 2,707,611) + (5 × 3,011,146) – (2,739,846)} / 24 = 1,415,699 and

${}_5P_0$ = 2,707,611 – 1,415,699 = 1,291,912

Similarly,

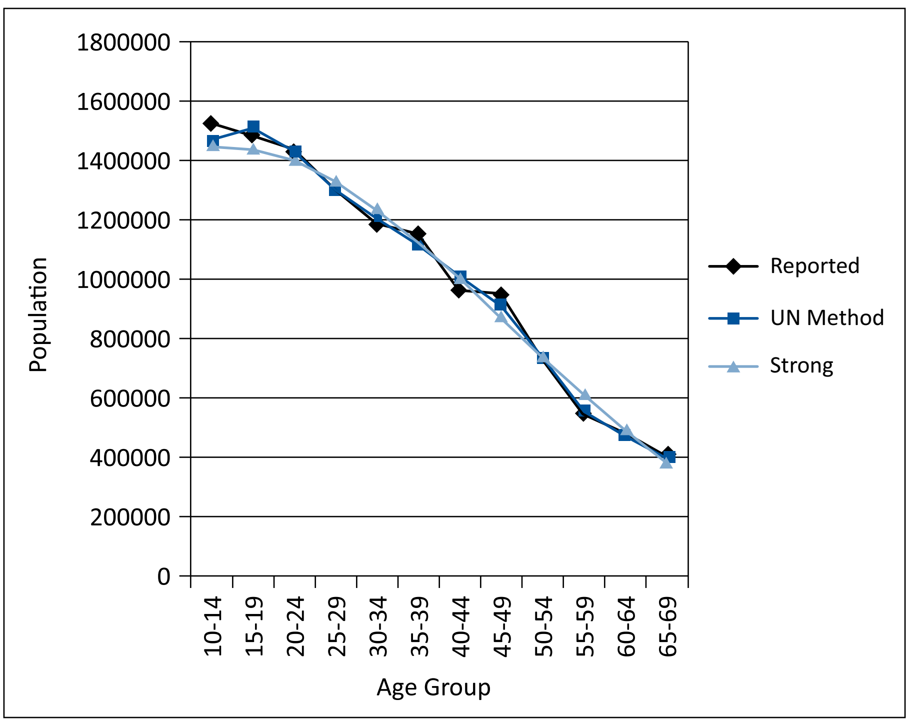
${}_5P_{70}$ = {(–1,267,513 + (5 × 880,811) + (8 × 447,255)} / 24 = 279,774 and

${}_5P_{75}$ = 447,255 – 279,744 = 167,481

The computations involved in the use of these methods are fairly complex. Fortunately there are computer software packages which provide smoothened distributions for any observed age distribution. The commonly used software package is PAS developed by the US Census Bureau. This is a user-friendly spreadsheet package like Excel in Windows. PAS contains a large set of packages, 51 in number, for demographic analysis of a given body of data, including estimation of demographic parameters through direct and indirect methods, population projections and construction of life tables and smoothening of age distributions. The package used for smoothening is called AGESMTH.WK1. It contains a worked-out example of the applications of five procedures of age smoothening, that is, by Carrier-Farrag, Karup-King Newton, Arriaga as well as the UN method, and a strong moving-average method called *Strong*. Among these, the most commonly used are the UN and the Strong methods. These methods are applied from the package PAS and the resulting smoothened distributions are given in Table 4.6. Figure 4.1 provides the age distributions of the reported population and Figure 4.2 provides the smoothened male populations in Kerala by the two methods. It can be seen that while the UNJS for Kerala is 26.24 for the raw population, it is 15.76 for the smoothened population according to the UN method and 7.05 according to the Strong method.

NOTES

Figure 4.2: Reported and Smoothened Age Distribution of Males in Kerala, 2001



4.10 Adjustments for Very Young and Old Ages

The smoothening procedures described above are applicable to populations within the age range 10 to 74 or 10 to 69, depending upon the availability of data by age. They are useful for population projections, since the births in future years will be determined by the number of women in the reproductive ages 15 to 49 and their fertility levels and for this purpose smoothened populations in the age range 10 to 69 or 74 are necessary. However, populations at younger ages (0 to 9) and at older ages (75 and above) are also subject to a variety of errors which are discussed in brief in the following section.

4.10.1 Under 5 age group

In most developing countries, children under 5 years of age are incompletely enumerated in the censuses as well as in surveys for various reasons. The major factor is the under-reporting of the births that took place in the last 5 years and the tendency to over-report the ages of young children. In populations which have a strong son preference there is a tendency to omit reporting of female births and girl children.

NOTES

Summary of Indices Measuring the Accuracy of Data

Index	Reported	United Nations	Strong
Sex ratio score	4.1	3.3	1.80
Male age ratio score	6.2	2.6	1.40
Female age ratio score	7.7	3.2	2.00
Accuracy index	26.2	15.7	8.87

Note: The accuracy index is the sum of the male and female age ratio scores plus three times the sex ratio score, all calculated using data for ages 10–14 through 65–69.

The Strong method of smoothening has given the lowest UN accuracy index indicating high quality of age data but this may be due to heavy smoothening procedures inflicted on the population.

One simple technique is to study the sex ratio at young ages and use the expected sex ratio at these ages to correct the number of female children, assuming that the number of male children is reported correctly. As discussed earlier, the sex ratio at birth is usually 106 male births to 100 female births and this ratio declines slowly and steadily over the years because of the increased mortality of males compared to females. This technique may prove defective in cases where male children are also reported incorrectly and the sex ratio of children changes *drastically* because of gross neglect of the girl child leading to higher death rates for the females contrary to biological expectations.

This deficiency is found even in the age group 0 to 4 in many situations, among infants below 1 year of age; if this is so, the independent estimate may be limited to these infants, based on the sex ratio or other method and combined with the census total for ages 1 to 4. The method of estimating the number of children in the age group 0 to 4 is generally based on first estimating the number of births that occurred in the population during the

5 years prior to the census and then estimating the survivors among them by subtracting the deaths. If B is the total number of births that occurred in the population during the 5 years prior to the survey and if the mortality experience of the population during the past 5 years has been captured in a life table, then the proportion surviving among births in the previous 5 years to the survey date is S (0-5) = ${}_5L_0/5 \times l_o$ and the children in the age group 0 to 4, P (0-4), can be estimated as B × S (0-5).

This method requires an independent estimate on the birth rate or the number of births in the population in the recent period and the survivorship probabilities. This will depend on the accuracy of the vital statistics data in the country. In India, fortunately, the system of sample registration provides a fairly reliable estimate of fertility and mortality at the state level on an annual basis. There are more complex methods of estimating the number of children in the age range 0 to 4, which can be studied in the book by Shryock and Seigel (1971).

4.10.2 Ages 5–9

The number of children in the age group 5–9 reported in a census is generally considered to be reliable. Studies of populations by age in different countries reveal that even in situations where there are gross errors in the reporting of ages the number reported in this age group is the most reliable of all ages. *Hence the census figures* reported in this age group are accepted unless there are valid reasons for not doing so.

NOTES

4.10.3 Older Ages

There is a universal tendency among people of all countries to overestimate their ages when they consider themselves old. The cut-off point when they consider themselves old may vary from population to population depending on the expectation of life, but once old age is perceived to be reached there is a tendency to over-report their ages. In populations where there are birth certificates this can be checked on a sample basis. To estimate the correct figures in old ages will be extremely difficult in populations which have no record of birth registration. A rough estimate can be obtained from the expected numbers in a model life table population with the current levels of fertility and mortality discussed in the previous chapter. However, for the purposes of population projection, errors in numbers at older ages are not of any significance, since only about 6% of the population of a developing country is currently above the age of 60 and any minor shifts in the distribution of this population in 5-year age groups beyond 60 may not have much significance in the future demographic trends in developing countries. This is not to underestimate the social, economic and political significance of this group in the society as a whole.

Exercises

Exercise 4.1: Compute the following indexes in the Excel file (Day 3 Exercise) and compare indexes with Kerala:

a. Whipple's index

b. Myers' index

c. UN Joint Score

Use data from Day 1 Exercise 2.1b and Exercise 2.1d and follow steps given in the Sections 4.4, 4.5 and 4.7 of Chapter IV.

Exercise 4.2: Adjust the age distribution using UN strong smoothening method.

CHAPTER V

Basic Measures of Fertility

5.1 Introduction

The three major demographic events which affect the population size of an area are births, deaths and migration. Births and deaths are technically referred to as fertility and mortality in demography. This chapter covers the basic measures used to study fertility, a major area of interest and research in the field of demography. These measures are of central importance to demography as births are crucial for biological replacement of populations and for the survival of the human society as a whole. Any increase or decrease in the population of a country is largely the result of a surplus or deficit of births over deaths. While the instinct for survival and the desire to live as long as possible are universal among all human beings and consequently programmes to reduce the incidence of deaths and death rates in any population have universal appeal and approval, the same is not the case with the number of births in a population.

The rapid growth of population in developing countries since the 1950s has actually resulted from a decline in death rates. As a method of reducing the growth rates in a population no one will suggest an increase in the death rate or even a reduction in the efforts of communities and governments towards prolonging life. The only sensible and socially acceptable alternative is to reduce the number of births and the birth rates in the population. Various recommendations on policies and programmes, including a strong programme of family planning, have been made towards achieving this objective and have been implemented with varying degrees of success in developing countries during the past three decades. The solution to the problems of rapid or slow rate of growth of population lies largely in the domain of manipulation of fertility.

NOTES

5.2 Concepts

It is important to distinguish between certain terms used in the study of fertility before going into the definition of basic measures of fertility. First, we must differentiate between the terms **fertility** and **fecundity** as used in demography. The term ‘fertility’ is used to connote the actual reproductive performance of a woman or the number of children a woman has or the average number of children for a group of women. Sometimes the term ‘natality’ is used to refer to the more general analysis of childbearing, though this usage is becoming less common, and the term fertility is commonly used to cover all aspects of reproduction. The term ‘fecundity’ is used to connote the physiological capacity to bear children and is the opposite of the term 'sterility'. No direct measurement of fecundity is possible, whereas fertility can be studied from the statistics of births. Fertility is possible only when a woman attains adulthood (or menarche) and usually ends with her attaining menopause, around the age of 45. The fertility of an individual woman is thus limited by her own fecundity. Although demographers distinguish between fecundity and fertility, these terms are used quite loosely in medical literature and are sometimes treated as being synonymous. A woman's fecundity varies according to a monthly cycle, depending on the time of ovulation within her menstrual cycle. Also immediately following a birth, it is not possible for a woman to conceive because there are a few months of amenorrhoea following the birth, and even after the commencement of menstruation the first few cycles are anovulatory. The length of this post-partum amenorrhoea and the anovulatory period has been found to be strongly related to the duration of breastfeeding of the child by the mother. Hence, the factors of fecundity, duration of breastfeeding of the child by the mother and the consequent period of amenorrhoea following a birth are of great significance in the analysis of fertility in a population.

NOTES

There is always a big difference between the actual fertility and the maximum number of children that it is physiologically possible for a woman to bear. Theoretically, if a woman gave birth to one child every 10 months over a period of 31 years (15 to 45), she could produce 37 children during the physiologically limited childbearing period. Even if she gave birth to a child every 15 months throughout her reproductive period, she would produce a total of 25 children. Of course, no population ever reaches that maximum, and there is great variability between the reproductive productivity of women. According to *The Guinness Book of World Records*, the largest number of children born to a woman and officially recorded is 69 by a woman who lived near Moscow during the eighteenth century (16 pairs of twins, 7 sets of triplets and 4 sets of quadruplets). A religious sect living on the borders of the United States of America and Canada, called *Hutterites*, has recorded the highest average for any community, which is about 11 births per woman. In this community, culture places a high positive value on couples having children and any form of birth control is considered sinful.

Populations which do not take explicit measures to limit the number of births are said to experience 'natural fertility'. The experience of the Hutterite community is an example of the natural fertility in a population. In such populations fertility is considered to be an essentially biological phenomenon, and its level varies mainly due to social customs such as varying age at marriage and different breastfeeding and weaning practices and not due to any contraceptive measures adopted for the spacing of children. The concept of natural fertility was first put forward by Louis Henry (1961).

Some women are, for various reasons, unable to bear any children. Such women are said to be 'sterile' or infertile. The term sterility may be used in connection with individuals or groups, consisting of either men or women or both. In general, as in the case of fertility, sterility measures are also computed in demography only for women. Sterility is of two types: 'primary sterility' and 'secondary sterility'. Primary sterility refers to women who have not produced any live births at all. Secondary sterility refers to the women who have become sterile after the birth of one or more children.

The frequently used concept for the statistical estimation of fertility within a menstrual cycle is ‘fecundability’. It denotes the probability of a woman conceiving within a monthly menstrual cycle. The term was first introduced by an Italian statistician, Corrado Gini, in 1924, as a way of improving the understanding of the birth interval or the interval between two successive births. Subsequently this was used to a great extent by Louis Henry in the analysis of birth intervals in historical demography (see Henry, 1972). The idea has received wider use in the models of reproduction developed in recent years. Originally, fecundability was defined as the probability of conception in a menstrual cycle among women who menstruate regularly but do not practice contraception. However, in recent years fecundability is defined as the probability of conception in a menstrual cycle, including women who are using contraception, and is called *residual fecundability*. The term used for fecundability among women not using contraception is ‘natural fecundability’. Fecundability is normally calculated only for married women. Effective fecundability refers to those conceptions which lead to live births. Comparisons of natural and residual fecundability have been used to assess contraceptive effectiveness.

Even though childbearing is basically a biological phenomenon, it is generally argued that variations in the level of fertility are not primarily due to the differences in physiological capacity but are more often produced by responses of individuals and couples to the social systems in which they live. The biological and behavioural factors which directly influence fertility are called *proximate determinants* of fertility. Various social, economic and other factors influence childbearing through these proximate determinants. The distinguishing feature of proximate determinants is their direct effect on fertility. Scholars differ on the importance of different various factors that contribute to a given level of fertility in a population and of the contributions of family planning programmes towards reduction of these fertility levels in a population. No single theoretical model has yet been developed which captures all aspects of fertility behaviour. Empirical work conducted in developing countries has suggested that the key variables that contribute to variations in fertility behaviour at the aggregate or individual level are education and labour force participation of women, marital patterns, the duration of breastfeeding, general and child mortality levels and contraceptive use. The transition in the fertility patterns of countries from high to low levels of fertility has accompanied their levels of industrialization and development. Fertility is considered to be closely linked to socio-economic development and the process of industrialization. In their theories of fertility, since the early 1950s, economists, utilizing various concepts of economics, viewed fertility performance simply as ‘economic behaviour’, a rational ‘economic response’, and the sociological significance of reproduction was not included in the economic analysis of fertility. This has apparently stimulated Easterlin (1975, 1985 with Crimmins) to propose a framework combining the sociology and economics of human fertility.

NOTES

Earlier, Davis and Blake (1956) had convincingly deliberated that socio-economic variables could not have a direct effect on fertility, and they have to operate through other variables, 11 in all – which they termed *intermediate variables* – that included age at marriage, permanent or temporary contraception and induced abortion. Subsequently, Bongaarts (1978) argued that of these 11 intermediate variables, the four most important proximate variables are marriage, post-partum infecundability due to variations in breastfeeding, contraception and induced abortion.

Hobcraft and Little (1984) have proposed a form of exposure analysis, which relies on individual data permitting a tabular as well as regression analysis of the proximate determinants and of the social factors of interest. The basic aim of these frameworks is to explain fertility in terms of the proximate determinants and transfer the study of the important social factors into an assessment of their impact on the proximate variables. Various aspects of human fertility have been studied in different cultural settings and the literature is really vast in the field of demography.

The scope of this chapter, however, is limited to providing some general information regarding important measures used in fertility analysis. The rest of this chapter is devoted to concepts and definitions of the measures generally used in this field of study.

5.3 Types of Analysis: Period and Cohort Measures

NOTES

The analysis of fertility is basically carried out in two ways: one is in a period perspective and the other in a cohort perspective. In the period perspective, the events that occur in a given period of time (calendar years) are studied in relation to the durations of exposure of the population during that period. In the cohort perspective, the events and duration of exposure are studied for well-defined cohorts as they move over time. The term 'cohort' indicates a group of people who have a similar experience at the same time. Two types of cohorts are generally used in demography: birth cohorts, that is, those born in the same year or period and marriage cohorts, those who are married in the same year or period. Period measures generally look at fertility rates in a cross-sectional way. The fertility measures considered in a longitudinal way are called *cohort measures*. By definition, the period measures are, in general, calculated for a period of one calendar year, whereas the cohort measures are generally calculated from the experience of a group of women born during a certain calendar year or married during a certain calendar year. Thus, period measures are the total experience of different cohorts of women (born at different periods of time) in any given calendar year. The exact relationship between cohort and period fertility is difficult to grasp, especially since it is rarely possible to look at both using the same data source. The period measures are simpler and used frequently. Hence, these are discussed first.

5.3.1 Crude birth rate (CBR)

Crude birth rate is the most widely used measure of period fertility. It is defined as follows:

$$\text{CBR} = \frac{\text{Number of births during a year}}{\text{Population at mid-year}} * 1000$$

Crude birth rate is considered a good measure of overall changes due to births in a population and is simple in concept and measurement. The fact that every birth is counted directly as an addition to the population should be recognized. Hence, the denominator increases and the rate will not reflect the actual chance of childbearing. Because of this bias, the CBR tends to underplay changes in fertility. The rate is called crude because it includes all ages and both sexes in the denominator. As we know, only women can bear children and that too within certain age limits. Thus, in computing the CBR we do not explicitly consider the population at risk of having those births. The measure is greatly affected by the age–sex composition and other characteristics of the population. Any comparison of two populations with this measure may be misleading, because they may vary widely in their age–sex composition.

Table 5.1: Basic Measures of Fertility for Kerala, 2001

Age Group (i)	Number of Births (B)	Female Population (W)	ASFR (Per 1000 Women) (2/3) *1000
1	**2**	**3**	**4**
15-19	36748	1499920	24.5
20-24	223656	1543523	144.9
25-29	198224	1489290	133.1
30-34	63738	1330656	47.9
35-39	15083	1311576	11.5
40-44	1685	990887	1.7
45-49	292	974123	0.3
Total	539427	9139975	363.9
Total Population		31841374	

Source: Registrar General of India (2004), *Sample Registration System*, Statistical report 2001

Note: ASFR of SRS, 2001 is applied to the 2001 Census population to obtain the number of births
CBR = (539427/31841374)*1000 = 16.94
GFR = (539427/9139975)*1000 = 59.02
TFR = Total of ASFR* 5 (5-year period)/1000 = (1819.5/1000) = 1.82
GRR = TFR*(1/2.05) = 0.89 where (1/2.05) is the proportion of female babies in total births, with the assumption of sex ratio at birth of 105 male births to 100 female births

For example, the CBR (derived using data from Table 5.1) for Kerala in the year 2001 is

$$CBR = (539427/31841374) * 1000 = 16.94$$

5.3.2 General fertility rate (GFR)

As mentioned, the CBR uses the entire population in the denominator. A more meaningful measure will be to use only women of the reproductive age groups in the denominator. The rate thus obtained is called *general fertility rate* (GFR) and is computed as follows:

$$GFR = \frac{\text{Number of births during a year}}{\text{Mid-year female population aged } 15-49} \times 1000$$

The child–woman ratio (CWR) discussed earlier can be taken as a good approximation of the GFR with an appropriate multiplier. This measure partially controls for the age composition

NOTES

of the population. As far as the risk of conception is concerned, the denominator used in the calculation of this measure is not a homogeneous group. It uses a wide range of age groups in the denominator consisting of women of different fecundity status. Thus it does not control entirely for age structure, and comparisons using GFR may also be misleading though not to the same degree as those using the CBR. To overcome this it is necessary to consider fertility rates for each age or age group. The following is the estimate of GFR obtained for the Kerala population in the year 2001 (derived by using data from Table 5.1):

$$GFR = (539427 / 9139975) * 1000 = 59.02$$

5.3.3 Age-specific fertility rate (ASFR)

While the denominator of GFR takes into account all the women of reproductive age, in the age-specific fertility rate (ASFR) both the numerator and denominator pertain to births and number of women of a specific age or in a specific age group. The ASFR is defined as follows:

$$\text{ASFR at age x} = \frac{\text{Births to women aged x in a year}}{\text{Mid-year female population aged x}} \times 1000$$

Or

$$\text{ASFR}\left(\text{in age group } x, x+n\right) = \frac{\text{Births to women aged x to }(x+n)\text{ in a year}}{\text{Mid-year female population aged x to }(x+n)} \times 1000$$

NOTES

Generally, 5-year age groups of women are used for calculating the ASFR, resulting in seven numbers, one for each age group – 15–19, 20–24 and 45–49. The ASFRs are found to be useful in fertility analysis since there are wide variations in fertility and fecundity by age. For the calculation of this measure, it is necessary to have births classified by the age of the mother as well as all women by same age or age group. The ASFR obtained for the women in the age group 30–34 for Kerala, 2001 is given below:

$$ASFR (30–34) = (63738 / 1330656) \times 1000 = 47.9$$

The ASFRs obtained for the other age groups are provided in Table 5.1.

The rate can be computed for a single woman or for 1000 women as given in the equation above.

The general pattern of the ASFR is that the rate increases to a maximum between ages 20–29 and then decreases slowly to reach zero by age 50. The modal value of the rate depends to a certain extent on the age at marriage of women of the population. It is to be noted that the ASFRs are unaffected by any variations in the age structure of the population. The rates are generally affected by the quality of the birth statistics classified by age of mother and the reporting of age by the woman. Since the ASFRs generally produce seven values by 5-year age groups, it is very difficult to use this measure for comparison purposes. A composite measure discussed below is a refined one and generally used for the comparison of fertility in different populations.

5.3.4 Total fertility rate (TFR)

The sum of the ASFRs over different ages 15 to 49 or 15 to 44 is known as total fertility rate (TFR). Thus,

TFR = Sum of ASFRs

If the rates refer to intervals greater than one, then the ASFRs represent only the average for the interval and it is necessary to multiply them by the width of the interval.

Total fertility rate is the most widely used measure of fertility by demographers. The TFR is generally expressed as number per woman. This measure can be thought of as the number of children a woman would have if she survived to age 50 and throughout her reproductive life span she is subjected to a fertility schedule described by the ASFR in question. The importance of TFR is that it is a single figure and is independent of age structure.

However, the computation of TFR requires a lot of information, such as the births by age of the mother and age of the women. If the ASFR used for the computation of TFR is expressed per 1000 women, it is necessary to divide the TFR by 1000.

The TFR obtained for Kerala's population in the year 2001 (derived by using data from Table 5.1) is

TFR = [24.5 + 144.9 + 133.1 + 47.9 + 11.5 + 1.7 + 0.3) ∗ 5 / 1000] = 1.82

5.3.5 Child–Woman ratio (CWR)

The child-woman ratio (CWR) is the simplest measure of fertility based on the age distribution of a population and used in the analysis of fertility. It is defined as given below:

$$CWR = \frac{{}_5P_0}{{}_{35}W_{15}} \times 1000$$

${}_5P_0$ = number of children under 5 years of age at a particular time

${}_{35}W_{15}$ = number of women in the age group 15–49 (reproductive span) at the same time

The subscript 15 on the right of W indicates the beginning of the age interval and the subscript 35 on the left of W indicates the duration of the interval beginning at age 15. The denominator may sometimes be women aged 15–44. The usefulness of this ratio is that it requires only limited information. A census generally gives the number of children under age 5 and in 5-year age groups for other ages and thus this measure is particularly useful when using census data. Basically, it is a crude measure and if fertility is high the ratio will

NOTES

be high and if fertility is low, the ratio will be low. Also, since ${}_5P_0$ is the number of total survivors of the births of five consecutive years, the ratio depicts the fertility trend in the recent past adjusted for mortality among children. It is not a convenient measure to use to compare populations with substantially different levels of infant and child mortality, or where under-reporting of children is a major problem.

The CWR for the state of Kerala for the year 2001 (derived by using census data given in Table 4.4) is CWR = 1000 ∗ (2765164/9139775) = 303

NOTES

Techniques have been developed to convert the CWR into a standard fertility measure such as CBR or the GFR. This is because the children under 5 years of age are the survivors of the births to the women that occurred in the previous five years. Hence if we have an idea of the mortality experiences of these children during the past five years they can be converted into births during the past five years and conventional fertility measures can then be worked out. The details of this method and the package programme that can be used to get these estimates are described in Chapter XII.

5.3.6 Specific fertility rates

It is also possible to derive fertility rates that are specific to many other characteristics. The most commonly used specific rates are related to marriage. The ASFRs defined earlier take the number of women in the age group. However, in a society where births can occur only within wedlock, the denominator used in the calculation may not be the actual population at risk. As we said earlier, GFR is actually an ASFR for the age group 15–49. If we use the number of married women rather than all the women in the denominator, we will get a new rate called *general marital fertility rate* (GMFR). The data pertaining to fertility measures of married women are given in Table 5.2.

Table 5.2: Basic Measures of Fertility for Kerala, 1981, 1991 and 2001

Age Group	1980-82			1991			2001		
	Number of Births (B)	Female Population (W)	ASFR (per 1000 Women) (2/3)∗1000	Number of Births (B)	Female Population (W)	ASFR (per 1000 Women) (5/6)∗1000	Number of Births (B)	Female Population (W)	ASFR (per 1000 Women) (8/9)∗1000
1	2	3	4	5	6	7	8	9	10
15-19	63768	1499325	42.50	46474	1593505	29.16	36748	1499920	24.5
20-24	271667	1424099	190.80	239237	1602390	149.30	223656	1543523	144.9
25-29	194837	1111524	175.30	175342	1390614	126.09	198224	1489290	133.1
30-34	83878	857313	97.80	53027	1127005	47.05	63738	1330656	47.9
35-39	37678	700377	53.80	14554	967062	15.05	15083	1311576	11.5
40-44	10902	635527	17.20	3408	783424	4.35	1685	990887	1.7
45-49	2832	605696	4.70	461	644214	0.72	292	974123	0.3
Total	665563	6833861	582.10	532503	8108214	371.72	539427	9139975	363.9
Total Population		25539093	2.91		29098516	TFR=1.86		31841374	TFR= 1.82
CBR		26.06			18.3			16.94	

Source: Registrar General of India (2004), *Sample Registration System*, Statistical report 2001

$$GMFR = \frac{\text{Live births in a year}}{\text{Married women aged } 15-49} \times 1000$$

GMFR calculated for married women in the age range 15–49 for Kerala, 2001, using data from Table 5.1 and Table 5.6 is given by

GMFR = (539427/6497495) ∗ 1000 = 83.0

Another set of rates used mostly in developed countries where a considerable proportion of births take place outside marriage consists of the legitimate and illegitimate fertility rates. As is evident from the names of the measures, the former refers to births that have occurred to currently married women, whereas the latter refers to the births that have occurred to women who are not currently married, viz., single, widowed, separated and divorced. In order to compute the rates, the denominators are also to be appropriately modified. The general legitimate fertility rate (GLFR) and the general illegitimate fertility rate (GILFR) are defined as follows:

$$GLFR = \frac{\text{Legitimate births in a year}}{\text{Married women aged } 15-49} \times 1000$$

and

$$GILFR = \frac{\text{Illegitimate births in a year}}{\text{Single, widowed, divorced, separated women aged } 15-49} \times 1000$$

NOTES

The ASFRs which also control for marriage by using the married women in the denominator are called *age-specific marital fertility rates* (ASMFR). It is also possible to compute the *total marital fertility rate* (TMFR), which is equivalent to the TFR for a married woman. However, there are problems in the interpretation of the TMFR which make it less attractive than the TFR as a measure of period fertility. By definition, TMFR is the average experience of a woman who gets married by age 15 and stays in married life until age 50. In reality, age at marriage and duration of stay in the marital state vary widely among populations. It is exaggerated by the high fertility rates of few married women in the age group 15–19. Hence, this measure is not used frequently in demographic analysis. The differentials in age at marriage can be resolved to some extent by simply leaving out the ASMFRs for the age group 15–19 while calculating the TMFR.

As previously mentioned, ASFRs specific to many other characteristics can be calculated. Specialized analysis may call for rates specified for birth order, parity, religion or education. (See Shryock and Siegel, 1971: Ch. 16). It is always important to remember that in such structures the denominator should include all those at risk of having the births included in the numerator. The source of the data and the measures used should be precisely explained.

5.3.7 Gross reproduction rate (GRR)

GRR is another summary measure of period fertility rate. This rate is essentially a TFR, with the modification that it is computed only for female births. GRR is the average number of daughters that would be born to a woman during her lifetime if she passed through the childbearing ages experiencing the average age-specific fertility pattern of a given period. If the data on births by sex are available for each group in the childbearing age group, the ASFRs may be computed for female children and the GRR can be computed by summing up these ASFRs, multiplying the sum by 5, and finally dividing the product by 1000. The data on births by sex is often not available. In this case GRR is related to the TFR in the following manner, assuming that the sex ratios at birth do not vary with the age of the mother:

$GRR = TFR \times (1/1 + S)$

where S = sex ratio at birth

Thus the GRR for Kerala, 2001, is derived as GRR = [1.82 ∗ (1/2.05)] = 0.89 (assuming that S, the sex ratio at birth [number of boys to girls] is 105 to 100).

NOTES

It may be noted that if the sex ratio at birth is 1:1, then the value of GRR will be one-half of the TFR. If we have no knowledge of sex ratios at birth, we can assume it to be 105 boys to 100 girls. This rate is generally used in comparing the fertility of various population groups. While computing the measure, no attention is paid to mortality.

A measure which uses both the age-specific schedules of fertility and mortality is called net reproduction rate (NRR). In general, NRR measures the extent to which a cohort of newly born girls will replace their mothers under predetermined schedules of fertility and mortality. Both these measures are normally calculated for the female population only. This rate is discussed in Chapter XI.

5.4 Levels, Trends and Patterns of Fertility

All the above measures of fertility are used to measure the levels of fertility at any point of time, for example, for Kerala in the year 2001. However, they do not tell how fertility has changed in the past and how the pattern of fertility among women within their reproductive ages changes. The data given in Table 5.2 on the age patterns of fertility in Kerala in the years 1981, 1991 and 2001 illustrates the need to study trends and patterns of fertility while studying the dynamics of fertility. The state of Kerala underwent a fertility transition (the third stage of demographic transition) mentioned in Chapter II beginning in the early 1970s, almost two to three decades ahead of other states in India. In 1951 its CBR was close to 40, almost the same as that of India as a whole; in 1971 it declined to 32; in 1981 to 26; in 1991 to 18.3; and in 2001 it was 16.94 as calculated earlier. The TFR values declined from 2.91 in 1981 to 1.86 in 1991 and to 1.82 in 2001. Thus between 1981 and 1991 there was more than one child decline per woman and it has considerably slowed down after 1991. During 1981 and 1991 the ASFRs of women in the age groups

Table 5.3: Fertility Differential in Kerala in 2006

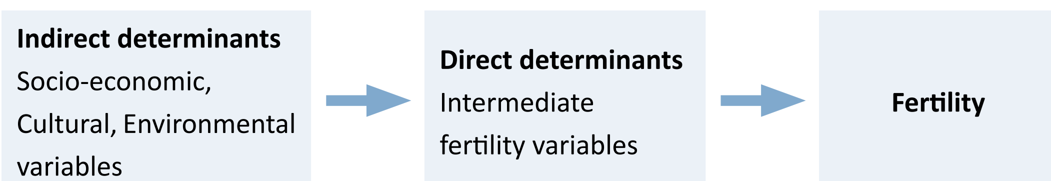
Category	CBR	GFR	TFR
Total	14.9	52.9	1.7
Rural	15.0	53.4	1.7
Urban	14.6	51.2	1.7
Graduate and above	-	81.8	1.9
Below primary		22.8	2.0

30–34, 35–39 and 40–44, that is, older women within the reproductive ages, declined very sharply from 98 to 47, 54 to 17 and 17 to 4 in these age groups, respectively. It is a common pattern observed in most of the fertility transition in developing countries that declines in fertility occur rapidly among older women and trickle down to younger women. In the earlier transition of fertility declines occur mostly due to limitation of births at two, three or four children and the spacing patterns change only at a later stage. Table 5.2 shows the patterns of fertility for the state of Kerala for the years 1981, 1991 and 2001, illustrating how even with widely differing total fertility rates and levels of fertility, there is a commonality in the age pattern of fertility. Table 5.3 shows the different levels of fertility by place of residence and educational attainment.

NOTES

5.5 Davis and Blake Intermediate Determinants of Fertility

Various studies on fertility levels and trends in different population groups revealed that there were substantial differences in fertility levels among various socio-economic groups. Analysing such differentials and trying to find out the reasons, Kingsley Davis and Judith Blake (Davis and Blake, 1956) described the concept of 'intermediate variables' as a set of "factors through which and only through which, social, economic, and cultural conditions can affect fertility". The model that Davis and Blake developed was as follows:



Davis and Blake identified a set of 11 intermediate variables as follows:

(a) Factors affecting exposure to intercourse (intercourse variables):

- Those governing the formation and dissolution of (sexual) unions in the reproductive period:
  1. Age of entry into sexual union
  2. Permanent celibacy
  3. Amount of reproductive period spent after or between unions:
     – When unions are broken by divorce, separation, or desertion
     – When unions are broken by death of husband

- Those governing the exposure to intercourse within unions
  4. Voluntary abstinence
  5. Involuntary abstinence (from major illnesses, unavoidable but temporary separation)
  6. Coital frequency (excluding periods of abstinence)

(b) Factors affecting exposure to conception (Conception Variables)

  7. Fecundity or infecundity, as affected by voluntary causes
  8. Use or non use of contraception
  9. Fecundity or infecundity, as affected by voluntary causes (sterilization, sub-incision, medical treatment etc.)

(c) Factors affecting gestation and successful parturitions (Gestation Variable)

  10. Fetal mortality from involuntary causes
  11. Fetal mortality from voluntary causes

NOTES

5.6 Bongaarts' Model of Proximate Determinants of Fertility

Bongaarts developed a model (1978) to study the impact of proximate determinants of fertility. He identified only the following eight as important:

a. The exposure factor
   1. Proportion married

b. The deliberate marital fertility control factors
   2. Contraception
   3. Induced abortion

c. Natural marital fertility control factors
   4. Lactational infecundability
   5. Frequency of intercourse
   6. Sterility
   7. Spontaneous abortion
   8. Duration of the fertile period

Bongaarts (1978) further narrowed down this list to only four factors using which he carried out an empirical analysis of their impact on fertility using a framework now called *Bongaarts Model*.

1. **The proportion of married women** is the proportion of women who are sexually active in their reproductive age and living in stable sexual unions (marriage). Using three indices, that is, index for overall fertility (TFR), marital fertility (TM) and proportion married (Cm) on 59 countries for the period 1970–75, he found that countries with

higher fertility (TFR > 5) typically have high proportion of married and countries with TFR below 5 typically have less values on their index of proportion married.

2. **Contraception** is defined as any deliberate parity-dependent practice including abstention and sterilization undertaken to reduce the risk of conception. Calculating values for total marital fertility (TM), total natural marital fertility (TNM) and index of non-contraception (Cc) for 30 countries he found that marital fertility was less in the countries where contraception were commonly practiced.
3. **Lactational infecundability** refers to the duration of the period of infecundity after gestation, which usually exists until the normal pattern of ovulation and menstruation is restored and this depends on the duration and intensity of lactation. By comparing average birth interval lengths in the presence and absence of lactation Bongaarts found that prolonged lactation has an inhibitory effect on ovulation and thus increases the birth interval and reduces natural fertility. He used the index Ci to denote the effect of lactation.
4. **Induced abortion** is defined 'as any practice that deliberately interrupts the normal course of gestation'. By calculating the index Ca which refers to the proportion by which fertility is reduced as the consequence of the practice of induced abortion, he found that Ca declines with increasing incidence of induced abortion.

NOTES

The following equations summarize the basic structure of the Bongaarts model by relating the fertility measures to the proximate determinants:

$$TFR = Cm * Cc * Ca * Ci * TF \quad (1)$$

$$TM = Cc * Ca * Ci * TF \quad (2)$$

$$TNM = Ci * TF \quad (3)$$

where TFR is the total fertility rate; TM is the total marital fertility rate, TNM is the total natural marital fertility rate and TF is the total fecundity rate.

Cm, Cc, Ca and Ci are the indices of marriage, contraception, induced abortion and post-partum infecundability respectively. The indices can only take values between 0 and 1. When there is no fertility-inhibiting effect of a given intermediate fertility variable, the corresponding index equals 1; if the fertility inhibition is complete, the index equals 0.

Estimation of the index of proportion married (Cm): It has to be noted that the higher the level of marriage in the population the less the inhibiting effect and vice versa. The index of marriage is estimated using the formula

$$Cm = \sum m(a)\, g(a) \,/\, \sum g(a)$$

where

Cm = Index of proportion married

m(a) = Age-specific proportion of married females and is derived by dividing the number of married women of a particular age group by the number of women in the same age group

g(a) = Age-specific marital fertility rate and is derived by dividing the births of a particular age group by the number of women in the same age group.

Estimation of the index of contraception (Cc): The index of contraception in the model measures the inhibiting effect of contraception on fertility in the population. It is estimated using the formula

$$Cc = 1 - 1.18 * u * e$$

where

u = Proportion using contraception among married women of reproductive age (15–49 years)

e = Average use effectiveness of contraception. The coefficient 1.18 represents an adjustment for the fact that women do not use contraception if they know that they are sterile.

The indices of use effectiveness (e) proposed for particular contraceptives are: pill = 0.90, IUD = 0.95, sterilization = 1.00 and others = 0.70.

Thus, Cm is the weighted average of the proportions married in each age group weighted by the age-specific fertility rate of that age group.

Estimation of the index of abortion (Ca): The index of abortion measures the inhibiting effect of abortion on fertility in the population. The index of abortion is estimated using the formula

NOTES

$$Ca = TFR/TFR + A$$

where

A = b * TA

TA is the total abortion rate (Average number of induced abortions per woman at the end of the reproductive period if induced abortion rates remain at prevailing levels throughout the reproductive period)

b = Average number of births averted per induced abortion

b = 0.4 (1+u), where u = Prevalence of contraceptive use

Hence b = 0.4 when u = 0 and b = 0.8 when u = 1.0

Ca = 1.0 if TA is 0

Estimation of the index of postpartum infecundability (Ci): The index of post-partum infecundability measures the inhibiting effect of breastfeeding or abstinence on fertility in the population. The index of postpartum infecundability in the model is estimated using the effect of breastfeeding (lactation amenorrhoea) or post-partum abstinence. The ratio of natural fertility in the presence and absence of post-partum infecundability therefore equals the ratio of the average birth interval without and with post-partum infecundability. If no breastfeeding and post-partum abstinence are practised, the birth interval averages about 20 months, which is the sum of

i. 1.5 months of minimum post-partum anovulation.

ii. 7.5 months of waiting time to conception.

iii. 2 months of time added by spontaneous intrauterine mortality.

iv. 9 months for a full-term pregnancy.

Bongaarts and Potter (1983) state that in the presence of breastfeeding and post-partum abstinence, the average birth interval equals approximately 18.5 months (7.5 + 2 + 9) plus the duration of post-partum infecundability.

The index of postpartum infecundability (Ci) is estimated as

$$Ci = 20/(18.5+i)$$

where

i = Average duration of post-partum infecundability caused by breastfeeding or post-partum abstinence.

We have the final model as

$$TFR = Cm \times Cc \times Ca \times Ci \times TF$$

where TF is the total fecundity of the population, that is, the total number of children that would have been born to a woman, if she is married at age 15, does not use any contraception, has no induced abortion and does not breastfeed the child.

Tables 5.4A, 5.4B and 5.4C give worked out examples of estimating Cm, Cc and Ci for Kerala using data available from NFHS-3 (2005–06). Since no data on TA (total abortion rate) is available we could not estimate Ca. Assuming that there is no abortion i.e. Ca=1 we can estimate TF as TF = TFR/(Cm ∗ Cu ∗ Ci).

Table 5.4A: Estimation of Index of Marriage (Cm) for Kerala using NFHS-3 Data

Age Group	Women	Currently Married Women	Proportion Married	ASFR (per Woman)	ASMFR (per Married Woman)	Cm
15-19	520	56	0.1080	0.0380	0.3519	
20-24	520	293	0.5630	0.1600	0.2842	
25-29	544	484	0.8900	0.1230	0.1382	
30-34	530	499	0.9420	0.0540	0.0573	
35-39	566	513	0.9060	0.0170	0.0188	
40-44	500	443	0.8860	0.0060	0.0068	
45-49	386	329	0.8520	0.0010	0.0012	
15-49	3566	2617	0.7340	2.0000	4.2900	0.4649

Table 5.4B: Estimation of Index of Contraceptive Use (Cc) for Kerala using NFHS-3 Data

Method	Percent of Married Women Aged 15-49 Using Method	Use Effectiveness	
Female sterilization	48.7	1.00	
Male sterilization	1.0	1.00	
IUD	2.3	0.95	
Pill	0.4	0.90	u = 68.6
Condom	5.5	0.70	e = 0.8957
Traditional methods	10.7	0.50	Cc = 1–1.08∗u∗e
All methods	68.6	0.8957	0.3364

NOTES

Table 5.4C: Index of Post-partum Infecundability (Ci) for Kerala using NFHS-3 Data

Median duration of breast feeding i = 25.2 months
Ci = 20 / (18.5 + i) = **0.4577**
TFR of Kerala during 2002–05 was 1.8.
Since we have no data on abortion rate for Kerala in NFHS-3 assuming Ca=1 survey we can estimate TF by the equation TF = TFR / (Cm∗Cc∗Ci) = 25.0 assuming Ca=1 or no abortion. This is abnormally high for any population.

The above results show that the unknown factors of TF can vary from population to population making the applications of Bongaarts' model difficult in the estimation of parameters. On the other hand, the model can be used in the same population to study changes in a given parameter over time since we have

$$TFR(t) / TFR(0) = (Cm * Cc * Ca * Ci)^t / (Cm * Cc * Ca * Ci)^0$$, since the TF term cancels.

5.7 Coale's Indices

NOTES

In the late 1960s, Coale developed four indices of fertility. Three of these attempt to measure the overall marital and non-marital (illegitimate) fertility in a population and the fourth one measures the prevalence of marriage. The indices are symbolically represented as follows:

I_f = Index of overall fertility

I_g = Index of marital fertility

I_h = Index of non-marital fertility

I_m = Index of proportion married.

These indices were developed essentially to study the fertility transition in Europe. The four indices are theoretically based on the technique of indirect standardization discussed in detail in a later chapter but the principle is simple.

For computing these indices, Coale selected the fertility schedule of Hutterites, known to have the maximum age-specific fertility rates, as the standard. Thus, the indices computed for any given population express the level of its fertility as a proportion of the fertility the population would have had if it had experienced the highest fertility pattern on record, which is that of the Hutterites. Coale's indices show how close the fertility being measured comes to the theoretical maximum of the Hutterites. The index of overall fertility (I_f) is derived by expressing the observed number of births as a proportion of the number of expected births if the population experiences the Hutterite ASMFRs. The expected number of births is obtained by applying the standard Hutterite marital fertility rates by age (given in column 4 of Table 5.5) to the observed age distribution of women in the given population.

$$I_f = \frac{B}{\sum F_i * W_i}$$

Table 5.5: Coale's Index of Overall Fertility (I_f) for Kerala, 2001

Age	Female Population	ASFR (per 1000 Women)	Hutterite Fertility	Number of Births	Expected Number of Births
i	W	f	F	(col. 2 × col. 3) / 1000	(col. 4 × col. 2) / 1000
1	2	3	4	5	6
15-19	1499920	24.5	300	36748	449976
20-24	1543523	144.9	550	223656	848938
25-29	1489290	133.1	502	198224	747624
30-34	1330656	47.9	447	63738	594803
35-39	1311576	11.5	406	15083	532500
40-44	990887	1.7	222	1685	219977
45-49	974123	0.3	61	292	59422
Total				539427	3453239
I_f = (5394727/3453239) = 0.1562091					

The expected births worked out by applying the Hutterite marital fertility schedule to women in Kerala in 2001 was 3453239 and the actual births were 539427 and the ratio of actual to expected. Thus we have I_f = B / B (expected) = 0.1562.

The same procedure is repeated to get index of marital fertility I_g but by using married women and legitimate births B_L.

$$I_g = \frac{B_L}{\sum F_i * M_i}$$

where B_L is the legitimate births in Kerala and the denominator is the expected births worked out assuming that the Hutterite fertility schedule is applied to Kerala women.

Table 5.6: Coale's Index of Marital Fertility (I_g) for Kerala, 2001

Age	Married Female Population	ASMFR (per 1000 women)	Hutterite Fertility	Number of Births	Expected Number of Births
i	M	g	F	(col. 2 × col. 3) / 1000	(col. 4 × col. 2) / 1000
1	2	3	4	5	6
15-19	169346	217.0	300	36748	50804
20-24	832055	268.8	550	223656	457630
25-29	1245129	159.2	502	198225	625055
30-34	1202612	53.0	447	63738	537568
35-39	1187648	12.7	406	15083	482185
40-44	886583	1.9	222	1685	196821
45-49	974123	0.3	61	292	59422
Total	6497495			539427	2409484
The I_g value for Kerala is 539427 / 2409484 = 0.2239					

NOTES

Similarly, I_h is also derived as the sum of illegitimate births (born to unmarried women) and taking the distribution of unmarried women. Since there are no illegitimate births reported in Kerala in 2001 this value is zero.

$$I_h = \frac{B_I}{\sum F_i * U_i}$$

However, I_m provides a measure of proportions married in each age, weighted by the Hutterite fertility rate and is expressed as follows:

$$I_m = \frac{\sum F_i * M_i}{\sum F_i * W_i}$$

NOTES

I_m for Kerala = 0.6977

B = number of births in the population

B_L = number of legitimate births

B_I = number of illegitimate births

W_i = number of women in the i^{th} age interval

M_i = number of married women in the i^{th} age interval

U_i = number of unmarried women in the i^{th} age interval

F_i = age-specific marital fertility rate in i^{th} age interval in the standard population

The following identity between I_f, I_m, I_g and I_h can be proved:

$$I_f = I_m * I_g + (1 - I_m) * I_h$$

If in the population there are no illegitimate births, then I_h is 0 and the relationship becomes

$$I_f = I_m * I_g$$

If I_g takes a value of one, it indicates that the marital fertility of the study population is the same as the Hutterite population, and if I_g = 0.5 it indicates that the marital fertility of the study population is half of the marital fertility of the Hutterites.

The calculation of Coale's indices is illustrated by using the data from the state of Kerala. The expected number of births from the Hutterite schedule for the total women as well as the married women is computed first and is provided in Tables 5.4 and 5.5 respectively.

A useful shortcut is that I_m can be calculated as the ratio of the expected births that would occur to married women and to all women if they experienced the standard Hutterite fertility rates. We have

Expected births to married women = 2409484, Expected births to all women = 3453239

I(m) = 0.6977

It is important to understand that the calculations are straightforward. The standard rates are the Hutterite ASMFRs, and the index population in the example is of Kerala, 2001. Indeed, in many cases it would be preferable to use a different standard from the Hutterite one. For example, if one was comparing the fertility levels of the regions of a country, then the appropriate standard would be national rates.

5.8 Cohort Rates

As we have noted, the fertility analysis can be approached in two ways: period and cohort. The basic measures of period fertility have been discussed earlier. Most of the period measures use census and vital statistics data. It is impossible to measure both cohort and period fertility rates using the same data source. It is also very difficult to fully understand the nature of relationship between cohort and period fertility rates. Basically, cohort fertility studies the fertility behaviour of a group of women who had experienced a specific event at a given time period in the past, subsequent to this experience at future intervals of time. The analysis usually covers women who were born or married during a particular period, considered birth or marriage cohorts. This type of analysis considers the experience of a selected group of people over time. It is in contrast to period analysis in which births occurring during a particular time period are studied. The major problem in cohort analysis is that it needs a consistent time series data set. The experiences of the young or more recent cohorts are unknown to the analyst since they are yet to fructify. Generally, cohort measures are affected by the death of the individual and migration of the individuals belonging to that particular cohort of women. Although in many cases information is directly available on period fertility and they can be computed more easily, cohort measures have significant advantages; they are often less variable over time and they come closer to describing the sequential nature of fertility behaviour.

NOTES

A commonly calculated measure of cohort fertility is cumulative fertility, which is the summation of the cohort's childbearing experience from the beginning of exposure to risk until some later date. The terms 'completed fertility' or 'lifetime fertility' are used to refer to cumulated fertility at the end of the reproductive age span. It should be evident that because childbearing is spread over more than 30 years of life, the completed fertility rate for any given cohort cannot be assigned to any particular calendar year as a measure of fertility in that year. A timing change is known as a tempo change. An example of this effect can be obtained from Japan in 1966. The TFR for 1966 in Japan fell to 1.6 compared with 2.2 in 1965 and 2.1 in 1967 (Frejka, 1973). The year 1966 was recorded by the Japanese as the year of the 'White Horse', and it was believed that a girl born in this year would grow up ill tempered and no one would marry her. So many Japanese postponed childbearing in this year. Clearly, tempo effects have a dramatic effect on period fertility rates, but have comparatively little effect on cohort rates. The cohort fertility rates are affected by quantum effects. The quantum is usually the completed fertility of a cohort, that is, the average births per woman. While analysing births of each order separately, however, the quantum indicates the proportion of the women who, having reached a given parity, will proceed to have another child. Measures of timing of births of different birth order for actual birth cohorts give a more realistic indication of the timing or spacing of births than those calculated from births in a single calendar year. The cohort measures include the median age of the mother, cumulative fertility at different ages and median intervals between births.

The period measures are affected by the tendency of couples to advance or postpone their births (like the Japanese experience in 1966), whereas the cohort fertility rates are not. If the couples have their children earlier the period rate will rise, while if they have them later it will fall. However, the cohort rates may not change.

While computing the TFR, we have assumed that a woman will survive up to age 50 and produce children as in the given ASFRs. However, no real group of women will actually experience any particular set of period ASFRs. The hypothetical group supposed to do so is called a *synthetic cohort*. Thus there are three different types of cohorts used in formal demography: birth, marriage and synthetic.

NOTES

5.8.1 Parity progression ratio (PPR)

Parity progression ratio (PPR) is simply the probability of a woman having another child given that she has already had a certain number. This ratio is normally calculated for marriage or birth cohorts of women who have completed their childbearing. The ratio can answer a question such as “Among the women who had their third child what is the proportion who have a fourth child?” The PPR is a sensitive indicator of family-building patterns since it reflects the sequential nature of fertility decisions. The ratio is also a very useful indicator for the study of the reproductive strategies followed in a population. The rate is generally represented by a series of a_n values where a_0 means the proportion of women in the cohort who become mothers and a_n is the proportion of women who proceed from parity n to n+1. It can be shown that the completed fertility rate of a cohort (CFR), which is the equivalent of TFR in the period measure, can be expressed as an arithmetic series of products of PPRs as shown below:

$$CFR = a_0 + a_0\, a_1 + a_0\, a_1\, a_2 + a_0\, a_1\, a_2\, a_3 + \ldots\ldots\ldots\ldots$$

The products $a_0\, a_l$, represent the proportion of women who have had at least two children.

5.8.2 Birth intervals

The PPRs only give the extent to which the women have had their first, second or higher order births, not their timing. As already discussed, the timing of births has an important influence on period fertility rates. Also, the spacing can affect the intrinsic growth rate as well as the mean generational length of any population. The analysis of birth intervals, as opposed to the conventional measures, can provide insights into the mechanism underlying the fertility behaviour of a population by disaggregating the reproductive process into a series of stages, beginning with marriage and followed by first, second and successive births. The long-term effect of the spacing pattern is to widen the length of a generation which slightly depresses the intrinsic rate of growth, and the short-term effect causes a temporary shift in the period fertility measures. These are discussed separately.

Any birth interval between two successive births of women of specified birth order, say between 2 and 3, aggregated over a group of women in a population can be used in a number of ways to study the fertility patterns in the population. First we distinguish between 'closed birth interval' (between two successive births) and the 'open birth interval', which is the interval between the survey date and the last birth of women who have not completed their reproduction. When women stop childbearing after a specific birth order, say 3, then there will be no closed interval after the third child and the open interval keeps on increasing with time. Thus the ratio of the average open interval to the last closed interval is an indicator of the extent to which the women have stopped childbearing with the last child. Many models have been developed to estimate PPRs from a distribution of the open and the closed birth intervals (Srinivasan, 1980).

NOTES

Second, any closed birth interval between two successive live births can be considered to be made up of four components: (1) The duration of infertile period (mostly due to post-partum amenorrhoea) following the earlier birth; (2) the periods of menstruating intervals following resumption of menstruation following the last birth; (3) the periods of infertile periods associated with abortions (induced or spontaneous between the two live births); and (4) the period of pregnancy associated with the latter child (usually assumed as 9 months). Models have been developed for each of these four components and used to estimate the fecundability parameter, incidence of abortions and post-partum amenorrhoea (Srinivasan 1975, 1985). The extent to which family planning methods have limited childbearing and postponed the interval can thus be studied by analysing data on closed and open birth intervals.

5.8.3 Lexis diagram

The data required for birth interval analysis usually come from survey data compiled in the form of birth histories. Analyses of birth histories are complicated because of the fact that many women would not have completed their childbearing at the time of survey. The survey data usually cover the birth history of women aged 15–49. Thus the data include selected groups of women from different age groups. These problems, called *censoring* and *selectivity* in the literature, make the analysis of birth intervals one of the more difficult and statistically complex areas of formal demography. One of the basic and useful tabulations of the birth interval is to provide the distribution of women by different intervals at which they have delivered a child following marriage, or successive births, and the respective mean or median by age or marriage duration of women. Such tabulation will serve as useful baseline data to study changes in the birth timing at various time points in the same population and also for comparing the levels of birth intervals in different populations.

The diagram was first devised by the German statistician, Wilhelm Lexis, in 1875 to be used in the computation of period and cohort rates. It provides a convenient way of displaying data with the aim of demonstrating the relationship between periods and cohorts. The diagram is composed of a grid made of units of calendar time on the horizontal axis (x) and age or the intervals of time from marriage or any event on which a *cohort* is defined on the vertical axis (y). Diagonal lines running upwards from the horizontal axis separate particular cohorts. By locating data within the appropriate square, parallelogram or triangle of the diagram it is possible to give a more immediate representation of the way in which information on periods and cohorts is related. Using this diagram we can estimate three types of rates, namely, cohort age-specific rate, age-period–specific rate and cohort-period–specific rate. The cohort-age–specific rate is defined as the Lexis diagram with hypothetical data on the number of births to two marriage cohorts, 1991 and 1992, during the years 1991–95 as given in Figure 5.1. From this data we compute the marital fertility rates of women according to the above three procedures.

NOTES

In Table 5.7, the rows represent the years of marriage duration in completed years: 0, 1, 2, 3, 4 and 5 and the columns represent the calendar years 1991 to 1995. Let us assume that in a city, 10,000 women married in 1991 and 10,000 married in 1992 were followed up until the end of 1996 to determine how many marriages survived without any break or attrition due to death of a marriage partner or divorce or separation and how many births took place in these marriages in different periods of time. The survivors of the marriage cohort of 1991 would have completed one year of marriage duration in 1992, two years in 1993 and so on and 5 years in 1996.

Figure 5.1: Lexis Diagram

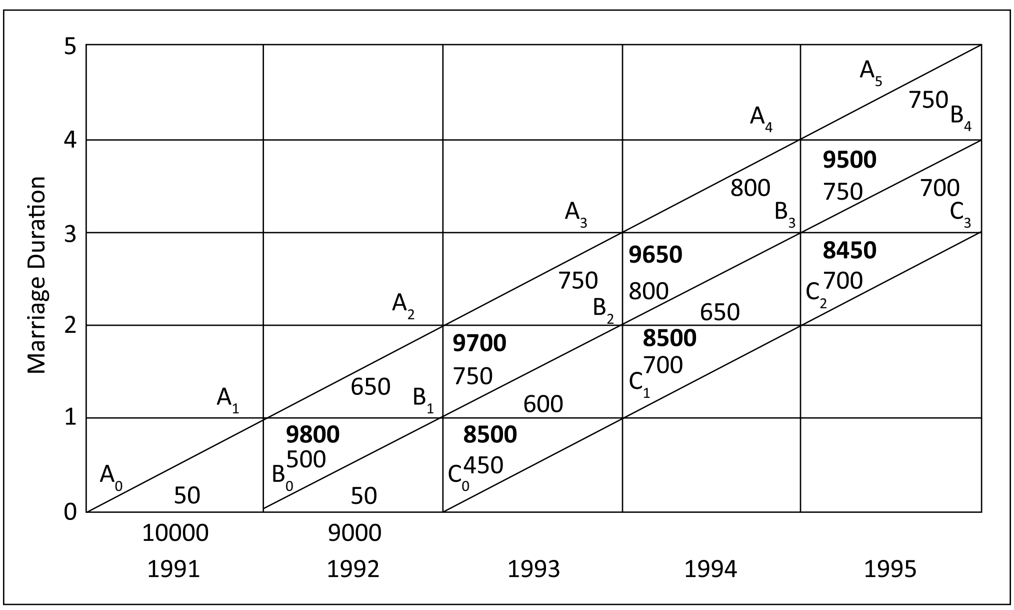


Table 5.7: Fertility of Marriage Cohorts of 1991 and 1992 to the end of 1995 (Hypothetical)*

Marriage Duration (Years)	Births and Number of Women (1991 Marriage Cohort)					Births and Number of Women (1992 Marriage Cohort)				
	Calendar Year					Calendar Year				
	1991	1992	1993	1994	1995	1991	1992	1993	1994	1995
0	**10000**						**9000**			
	(50)	(500)					(50)	(450)		
1		**9800**						**8500**		
		(650)	(750)					(600)	(700)	
2			**9700**						**8500**	
			(750)	(800)					(650)	(700)
3				**9650**						**8450**
				(800)	(750)					(700)
4					**9600**					
					(750)					

* Assuming that 10,000 women got married on 1 January 1991 and 9000 on 1 January 1992, the survivors of each of these cohorts and the births to them until 1995 are shown separately.

The number of survivors of the cohort in successive years is given in bold. The number of births that occurred to the cohorts during different periods is given in the brackets below. For example, 10,000 women were married in the city in 1991 (assuming uniformly throughout the year); among them 9800 continued in the married state in 1992 at the end of one year (the remaining 200 becoming widowed, died, separated or divorced within 1 year of marriage); 9700 surviving in the married state in 1993; 9650 in 1994; and 9600 in 1995. These numbers are indicated in bold in the first diagonal column. The follow-up study of the 1991 cohort has revealed, according to the table, that for this cohort of women there were 5800 births before the end of 1995 and these 5800 births are distributed in different calendar year-marriage duration segments. These are given in the diagonal starting with the number 50 in the first cell. The births in each calendar year for each cohort are given in Table 5.7 in parentheses. Similarly, in the marriage cohort of 1992 there were 9000 marriages, out of which 8500 survived the first year, all the 8500 the second year and 8450 the third year. These are indicated in the second diagonal column. In the Lexis diagram, each marriage is denoted by a diagonal line; for instance, those who are married in 1991, say on 1 January, would have completed one year of married life by 1 January 1992, two years on 1 January 1993 and so on. The line A_0, A_1, A_2, A_3, A_4 can be considered as the marriage line for each individual. The line terminates as soon as the woman dies or is widowed or her marriage ends in a separation or divorce. The number of births that occurred to these women in different periods of time is given within each triangle in the diagonal column: B_0C_0 to B_3C_3. From these data, we can compute three types of rates: cohort marriage duration-specific fertility rates, marriage duration period-specific fertility rates and cohort period-specific fertility rates. The computation of these is as given below.

NOTES

Cohort marriage duration-specific fertility rate

We can calculate the cohort marital duration-specific rate between exact durations of marriage, for example, 2 and 3 years after marriage for the 1991 marriage cohort, that is, the parallelogram $A_2A_3B_3B_2$ in the above Lexis diagram.

$$\text{Cohort period fertility rate} = \frac{\text{Number of births by a given cohort of women during a year}}{\text{Person-years lived by the cohort during the year}}$$

$$\text{Cohort marriage duration} - \text{specific rate (between exact ages x to x + n of the specified cohort)}$$

$$= \frac{\text{No. of births between marital durations x to x + n of the given cohort}}{\text{Person-years lived between exact ages x to x + n of the same cohort}}$$

NOTES

Married women surviving at 2 years of marriage duration in marriage cohort 1991 to calendar year 1993	9700
Married women surviving at 3 years of marriage duration in marriage cohort 1991 to calendar year 1994	9650
Person-years of married life lived between durations of 2 and 3 years after marriage (9700 + 9650)/2	9675

Births for this cohort during this period are 1550 with 750 births in 1993 and 800 in 1994.

Hence ASFMR of 1991 marriage cohort between their second and third year of married life is

$1000 \times (1550/9675) = 160.21$

Marriage duration period-specific marital fertility rate

We can estimate the marital fertility in the calendar year 1994 for the marriage cohorts of 1991 and 1992 as follows:

Total births for married women with marital duration of two completed years during 1994 come partly from the marriage cohort of 1991 and partly from the marriage cohort of 1992. These are given in the rectangle $B_2C_2B_3A_3$. There are a total of 1450 births in this rectangle with 800 births from the 1991 cohort and 650 from the cohort of 1992.

The average number of married women living at the beginning of 1994 from the marriage cohort of 1991 is

$(9700 + 9650) / 2 = 9675$

The average number of married women at the end of 1994 from the marriage cohort of 1992 is

(8500 + 8450) / 2 = 8475

Hence, the average number of years of married life lived by women with completed two years of marriage duration in the calendar year 1994 is

(9675 + 8475) / 2 = 9075

Hence, the cohort period-specific marital fertility rate for marriage duration of two and three years will be

1000 × (1450/9075) = 159.78

Cohort period-specific fertility rate

Here we consider the fertility rate experienced by a specified cohort in a given calendar year of time.

For example, if we consider the 1991 cohort and their marital fertility experience during 1994, we have to consider the births within the parallelogram A3A4B2B3 and divide the same by the appropriate number of married women years of exposure.

The births are 800 + 800 = 1600.

The number of women estimated as on 1 January 1994 from this cohort is

(9700 + 9650)/2 = 9675.

The number as on 1 January 1995 is (9650 + 9600)/2 = 9625.

Hence the married-woman-years of exposure for this cohort during 1994 is

(9675 + 9625)/2 = 9650.

Hence the cohort period-specific fertility rate for the 1991 cohort in 1994 is

(1600/9650) × 1000 = 165.80.

Thus, we can see that the three rates can vary from each other because of differences in the size of different cohorts and the variations in the fertility performance of different cohorts over time.

NOTES

Exercises

Exercise 5.1: Compute basic measures of fertility for Rajasthan 2001 and 2011. (Use data from Day 1 Exercise 2.1e. Calculate ASFR, CBR, GFR, TFR, GRR.)

Exercise 5.2: Make a graphical presentation of ASFR in Rajasthan 2001 and 2011.

Exercise 5.3: Compute Bongaarts measures of fertility: estimate C_m, C_c, C_i and TF.

Exercise 5.4: Compute Coale's indices of overall fertility (I_f), marital fertility (I_g) and proportion married (I_m).

CHAPTER VI

Basic Measures of Mortality and Life Table Construction

6.1 Introduction

Mortality analysis is one of the most important branches of demographic studies and is the one with which demographers have been engaged from the very beginning of any systematic study of human populations. This branch deals with the demographic event of death. Measurement of mortality received much attention from actuaries during the eighteenth and nineteenth centuries. Thus the techniques for analysing mortality have a longer history and are more developed than those for analysing fertility. Since death is a biological phenomenon that occurs just once to each individual, the analysis is simpler than, say, the study of fertility wherein the event of birth can occur with varying frequency among women. In this chapter we discuss the basic measures generally used in mortality analysis. The main objective of the chapter is to introduce the concepts, notations, computational procedures and the application of the life tables. A more detailed discussion about the mortality analysis can be obtained from Shryock and Siegel (1976: Ch 14 and 16). The various techniques developed by demographers in the analysis of mortality data, such as the technique of life table analysis, have been extended and have found applications in a wide array of other situations.

NOTES

6.2 Crude Death Rate (CDR)

A simple and direct measure of deaths in a population is the crude death rate (CDR). This rate relates the number of deaths during a year to the mid-year population of that year.

$$\text{CDR} = \frac{\text{Deaths in a year}}{\text{Population at mid-year}} \times 1000$$

Note that the denominator uses the total population with varying risks to death and hence CDR does not measure the risk of dying for a person in the population in the probabilistic sense. This measure is generally a poor indicator of mortality as it does not take age structure into account. The number of deaths is largely a function of the size and the age–sex structure of the population. The risk of dying in general is high in the infant and childhood ages, declines drastically thereafter up to age 20 and then rises slowly but steadily, increasing sharply at ages above 50. This type of curve, known as the reverse J-shaped curve, is characteristic of age patterns of mortality in all populations. Many developing countries with higher mortality levels at every age show a lower CDR than developed countries because the former have a much larger proportion of their population at younger ages due to high fertility. A crucial factor in the study of mortality is to take account of its variations by age.

The CDR for the state of Kerala in the year 2001 derived by using data from Table 6.1 is

$$CDR = (210160 / 31841374) \times 1000 = 6.6$$

6.3 Age-Specific Death Rate (ASDR)

The CDR is the weighted average of the age-specific rates, the weights being the population size at each age. The death rates may vary not only by age, but differently for the two sexes. The age-specific rates for males and females, separately, are the commonly used specific rates. The ASDR for the age group x to x + n is defined as follows:

NOTES

$$_nM_x = \frac{_nD_x}{_nP_x} \times 1000$$

$_nD_x$ = number of deaths between x and x+n in the year

$_nP_x$ = number of persons aged between x and x+n at the middle of the year

6.4 Useful Measures of Mortality

The groups of population which are at greater risk of dying are newborn infants and mothers in the reproductive age groups. Conventional infant mortality rate (IMR) is defined as the number of deaths under age 1 during a specified period divided by the number of live births in the same period and usually expressed per thousand births.

$$IMR = \frac{\text{Number of deaths under 1 year in a year}}{\text{Number of live births in the same year}} \times 1000$$

The measure is not a proper rate but a ratio, as the denominator is not wholly the population at risk of the events in the numerator. Some of the deaths under age 1 in the given year may be among the births which occurred during the previous year and some of the newborns during the year may die in the next year before reaching their first birthday.

The IMR for Kerala state in the year 2001 is as given below (using data from Table 6.1)

$$IMR = (5619/539427) \times 1000 = 10.4$$

Also, the risk of dying during the first year of life is not uniform in the interval. The risk is the maximum soon after birth and decreases slowly. During the early weeks, the causes of infant deaths tend to be different from those which occur later. For this reason, IMR is often divided into two parts, based on the intensity of risk of mortality. The first four weeks or one month, called the neonatal period, is found to have a very high risk and the number of infant deaths during this period is termed neonatal mortality rate (NMR).

Neonatal mortality rate (NMR) is defined as

$$NMR = \frac{\text{Deaths of babies less than 4 weeks old}}{\text{Number of live births in the same year}} \times 1000$$

In the case of Kerala in the year 2001, NMR = (4597 / 539427) × 1000 = 8.7

The death of infants after four weeks is termed *post-neonatal mortality* (PNMR) and is defined as follows:

$$PNMR = \frac{\text{Deaths of babies aged 4-52 weeks during a year}}{\text{Number of live births during the same year}} \times 1000$$

The PNMR for Kerala in the year 2001 is given below (using data from Table 6.1):

PNMR = (1022 / 539427) × 1000 = 1.7

NOTES

For the control of diseases in any population it is important to know the cause of death. Thus cause-specific measures of mortality are useful in understanding the relative contributions of different causes to the overall mortality rates in the population and resources to be mobilized for controlling the specific disease. The most commonly used measure is the *cause-specific death rate* (CSDR) which is defined as follows:

$$CSDR = \frac{D_c}{P} \times 1000$$

where

D_c = number of deaths in a year due to a particular cause ‘c’;

P = mid-year total population

CSDRs can be separately computed by age and sex. Another measure, which expresses the number of deaths due to a particular cause as a proportion of all deaths is called *cause-specific death ratio* and is also commonly used in mortality analysis. The cause-specific death rates and ratios computed for the state of Kerala in the year 1991 are provided in Table 6.2. Moreover, distribution of deaths in India by top ten major causes are given in Table 6.2A for the period of 2001-2003.

Table 6.1: Age-Specific Death Rates for Kerala, 2001

Age Group	Total Population	Number of Deaths	ASDRs (per 1000 Population (col.3/col.2) * 1000	Male			Female		
				Population	Number of Deaths	ASDRs (per 1000 Population) (col.6/col.5) * 1000	Population	Number of Deaths	ASDRs (per 1000 Population) (col.9/col.8) * 1000
1	2	3	4	5	6	7	8	9	10
<1	530847	5619	10.59	269052	3602	13.39	261797	1999	7.64
0-4	2236616	106	0.05	1141718	886	0.78	1094907	299	0.27
5-9	2546296	1444	0.57	1296858	755	0.58	1249448	795	0.64
9-14	2989684	1130	0.38	1525229	888	0.58	1464466	133	0.09
15-19	2986988	1412	0.47	1485937	1009	0.68	1501056	409	0.27
20-24	2986471	2823	0.95	1441778	1678	1.16	1544692	983	0.64
25-29	2788512	3953	1.42	1298085	2267	1.75	1490418	1761	1.18
30-34	2518555	3571	1.42	1186886	2533	2.13	1331664	1211	0.91
35-39	2468405	4899	1.98	1155829	3140	2.72	1312569	1671	1.27
40-44	1952906	6829	3.50	961271	4943	5.14	991637	1983	2.00
45-49	1927746	6741	3.50	952887	4437	4.66	974861	2393	2.45
50-54	1438715	10199	7.09	725360	7248	9.99	713359	3113	4.36
55-59	1131184	10798	9.55	542161	7469	13.78	589022	3373	5.73
60-64	1032994	16402	15.88	480782	11054	22.99	552209	5422	9.82
65-69	902765	22355	24.76	400035	14049	35.12	502724	8181	16.27
70-74	613932	26924	43.86	273542	14781	54.04	340387	12099	35.55
75-79	399421	26690	66.82	173716	13988	80.52	225702	12516	55.45
80-84	217403	26836	123.44	90501	12222	135.04	126899	14305	112.73
85+	171934	31428	182.79	66987	13706	204.60	104944	17354	165.36
Total	31841374	210160	6.60	15468614	120656	7.80	16372760	90000	5.50

Death of babies			
0-4 weeks	4597	NMR	9.0
4-52 weeks	1022	PNMR	2.0
Live Births	539427		

Table 6.2: Cause-Specific Death Rates and Ratios for Kerala by Major Cause Groups, 1991

Major Cause Groups	Number of Deaths (D)	CSDR (D/P) × 1000	Cause-Specific Death Ratio (D/TD) × 100
1	2	3	4
Senility	12645	0.43	7.2
Cough	30383	1.04	17.3
Diseases of circulatory system	38812	1.33	22.1
Causes peculiar to infancy	7376	0.25	4.2
Accidents and injuries	20021	0.69	11.4

Contd...

Major Cause Groups	Number of Deaths (D)	CSDR (D/P) × 1000	Cause-Specific Death Ratio (D/TD) × 100
1	2	3	4
Other clear symptoms	33017	1.13	18.8
Fevers	1054	0.04	0.6
Digestive disorders	5620	0.19	3.2
Disorder of central nervous system	25641	0.88	14.6
Childbirth and pregnancy	1054	0.04	0.6
Total deaths (TD) 175623			
Total population (P) 29098518			

Sources: Registrar General of India (1995,) *Survey of Causes of Death (Rural), India, Annual Report 1991*

Table 6.2A: Cause-Specific Death Ratios for India by Major Cause Groups, 2001-03

Major Cause Groups	Number of Deaths (D)	Cause-Specific Death Ratio (D/TD) × 100
Cardiovascular diseases	21374	18.8
COPD, asthma, other respiratory diseases	9905	8.7
Diarrhoeal diseases	9246	8.1
Perinatal conditions	7162	6.3
Respiratory infections	7051	6.2
Tuberculosis	6810	6.0
Malignant and other neoplasms	6476	5.7
Senility	5807	5.1
Unintentional injuries: Other	5559	4.9
Symptoms, signs and ill-defined conditions	5435	4.8
Total deaths (TD) 113692		

Source: Office of the Registrar General, India (2009) Report on Causes of Death in India 2001-2003

NOTES

The *maternal mortality ratio* (MMR) is a widely used type of cause-specific mortality rate representing approximately the risk of dying as a result of complications of pregnancy, childbirth and puerperium. This ratio is generally defined as the number of deaths due to puerperal causes per 100,000 live births.

$$MMR = \frac{D_p}{B} \times 100000$$

where D_p = Deaths due to puerperal causes (deaths that occur to mothers within 42 days of delivery due to causes associated with pregnancy and childbirth) during a year

B = Number of live births in the same year

The MMR estimated in SRS (2011) for the state of Kerala in the year 2007–09 is given below:

$$MMR = (12 / 14624) \times 100000 = 81$$

If the denominator uses the number of married women, then the measure is called a maternal mortality rate (MMRT).

$$MMRT = \frac{D_P}{\text{Number of married women aged } 15-49} \times 1000$$

where D_p = Deaths due to puerperal causes (deaths that occur to mothers within 42 days of delivery due to causes associated with pregnancy and childbirth) during a year

The MMRT observed in SRS (2011) for the state of Kerala in the year 2007-09 is given below:

$$MMRT = (12/287854) \times 100000 = 4.1$$

Since MMRT is defined per 100,000 married women and MMR is defined per 100,000 live births, hence MMRT and MMR are connected by the following formula:

$$MMRT = MMR \times GMFR$$

NOTES

6.5 Life Tables

The statistical model which combines the mortality rates of a population at different ages into a single set-up is called a *life table*. Life tables are principally used to measure the level of mortality of a population at different ages in a probabilistic concept. The life table is a powerful tool for the analysis of mortality and provides the most complete and exact way of comparing the mortality of different populations or groups. It is generally used by public health workers, demographers, actuaries and many others in studies of longevity and population growth, as well as in making projections about population size and also in studies of widowhood, orphanhood, length of married life, length of working life etc.

Life tables are generally constructed for two types of age categories. If life table information is given for each year of age it is referred to as *complete*. If it uses age groups instead of a single year of age, it is called an *abridged life table*. Abridged life tables are commonly used by demographers. If one constructs a life table based on the occurrence of deaths within a cohort of individuals born in the same year or groups of years, it is called a *cohort life table*. A *period life table* assumes a hypothetical cohort and is based on the experience of age-specific rates of a particular period and is normally used in mortality analysis. The great advantage of a life table is that it provides measures which are not affected by differences in the age structure.

Whatever the type of life table, the definitions of various life table functions remain the same. The basis of the table is a set of probabilities of dying, ${}_nq_x$, which gives the proportion of individuals alive at age x who die before reaching age x + n. The usual life table consists of 11 columns which are defined as follows:

1. Age (x) which is the initial age of the age interval (x, x+n) where x is the initial age and n is the length of the interval. In abridged life tables the interval length is usually 5 years,

with the exception of infancy (1 year), early childhood years (4-year age group) and last interval (open ended). The prefix n is used only for age groups other than single-year. For abridged life tables the age groupings are 0, 1–4, 5–9, 10, 15, up to 85+; this can be indicated in the first column as 0, 1, 5, 10, 15 --- 85+

2. m(x, n): central death rate for the age interval (x, x+n), usual notation is ${}_nm_x$;
3. q(x, n): probability of an individual age x dying before the end of the age interval (x, x+n), usual notation is ${}_nq_x$;
4. l(x): number of survivors at age x in a life table with radix (starting population) of 100,000 persons, usual notation is l_x;
5. d(x, n): number of deaths in age interval (x, x+n), usual notation is ${}_nd_x$;
6. L(x, n): number of person-years lived in age interval (x, x+n), usual notation is ${}_nL_x$;
7. S(x, n): the proportion of the life table population in age group (x, x+n) who are alive n years later, usual notation is ${}_nS_x$;
8. T(x): number of person-years lived at ages x and older, usual notation is T_x;
9. e(x): expectation of life at age x, usual notation is e_x;
10. a(x, n): average number of years lived in the age interval (x, x+n) by those dying during that age interval, usual notation is ${}_na_x$.
11. P(x, n): probability of survival between exact ages x and x+n, defined as ${}_nP_x = 1 - {}_nq_x$.

NOTES

The conventional life tables are constructed on the basis of certain assumptions. The life table is assumed to trace the life history of a cohort and thus the membership of the cohort cannot be increased or decreased due to other determinants of population change, namely, migration. A life table population is considered as closed to migration and resembles a stationary population, in which the number of the deaths and births are equal. The death schedule is also assumed to operate in the given set pattern and periodic variation due to random causes is not anticipated. Thus the life table is a deterministic model.

The work of Graunt in the seventeen century was a precursor of the modern life table, though a more systematic approach was taken by Halley. Several further attempts to produce a life table were made in the eighteenth century. The first official complete life tables for the United States were prepared in 1900–1922.

On the basis of more reliable life tables, several series of model life tables have also been developed recently and these are discussed in Chapter XI.

Various assumptions for the computation of life table functions are evolved by different researchers. As stated, the life table is generally constructed on the basis of the probability of dying, ${}_nq_x$, between exact ages x to x+n. The main problem is in the conversion of central death rates (${}_nd_x$) values into probabilities of dying (${}_nq_x$). For any population, the only information available to the researcher is the ASDR, that is, the m_i values. The crucial problem is in computing ${}_nm_x$ values; they are computed from the ASDR available for a population. It has to be recognized that while ${}_nm_x$ as death rates are obtained by relating deaths to the mid-period population, ${}_nq_x$ as a probability is obtained by relating deaths to population at the beginning of the period. In order to convert the ASDR into probability of dying between exact ages x to x+n, the distribution of deaths in the age interval is assumed to be linear or curvilinear. If we assume that the deaths are uniformly or linearly distributed over x to x + n, it can be shown that ${}_nm_x$ and ${}_nq_x$ are functionally related as follows:

As per the definition of ${}_nm_x$ derived earlier, we have

$$ {}_nm_x = \frac{{}_nd_x}{{}_nP_x} = \frac{{}_nd_x}{n.l_x - \left(\frac{n}{2}\right){}_nd_x} $$

where ${}_nP_x$ = mid-year population between ages x and x+n

NOTES

Since at the beginning of the interval the number of persons living is l_x, ${}_nP_x$ will be equal to n ∗ l_x if there are no deaths between x and x+n. However, there are ${}_nd_x$ deaths and since they have occurred uniformly (assumed to be) during the interval, on an average each death will remove n/2 person-years and hence (n/2) ∗ ${}_nd_x$ has to be removed from the person-years of exposure, leading to the above equation. Dividing the numerator and denominator in the above equation by l_x leads to

$$ {}_nm_x = \frac{{}_nd_x/l_x}{n - \left[\left(\frac{n}{2}\right).\frac{{}_nd_x}{l_x}\right]} = \frac{2\,{}_nq_x}{2n - n.{}_nq_x} $$

where

$$ {}_nq_x = \frac{2n.{}_nm_x}{2 + n.{}_nm_x} $$

It can be seen that the linear assumption is not quite appropriate, since at older ages the risk of dying increases sharply with age and also during infancy when it is much higher in the first few weeks of life. If we assume that in such cases l_x could be exponential in form, then,

$$ l_x = \exp\,(a + {}_nm_x) \quad \text{leading to} \quad {}_nq_x = 1 - \exp\,(-\,n * {}_nm_x) $$

At some ages such simplifying assumptions of deaths by linear or exponential curves is not found to be empirically acceptable. Greville's method is generally the most used for ages 5 to 69 when the exponential curve of l is found to be valid; other methods are used for ages

0–4 and 70+ populations. The method assumes that values follow a modified exponential curve. Greville derived the $_nq_x$ values as follows:

$$_nq_x = \frac{_nm_x}{(1/n) + {_nm_x}\left[(1/n) * \left\{(n/12) + \left({_nm_x} - \log_e c\right)\right\}\right]}$$

Empirically, the value of c has been found to be between 1.08 and 1.10; $\log_e$ c could be assumed to be about 0.095 as an intermediate value. The other functions of the life table can be obtained as above, with the exception of the $_nL_x$ values.

The first value L_o is calculated as follows:

$L_o = 0.276l_0 + 0.724l_1$ using the general relationship

$$_nL_x = [{_na_x} * l_x + (1 - {_na_x}) * l_{x+n}] * n$$

where $_na_x$ = proportion of the period x to x+n lived by those who die during this period

NOTES

In many states in India, the IMRs and NMR are still high and a large number of children die not only in the first few weeks of life due to unavoidable genetic and other causes but also in the post-neonatal period, between 1 and 12 months, due to a variety of infectious diseases. Thus the proportion of infant deaths that occur in the post-neonatal period is higher in the developing countries than in the developed countries, where the majority of infant deaths occur in the first four weeks of life. Hence, the average duration of life lived by those who die within one year (a_0) is greater in the developed countries than in the developing countries. Studies have shown that the a_0 values at different levels of IMR are as follows:

Developed countries:

- $a_0 = 0.1$
- $_4a_1 = 0.4$
- $_na_x = 0.5$ for all other ages

Developing countries: (IMR above 30)

- $a_0 = 0.3$ or 0.2
- $_4a_1 = 0.4$
- $_na_x = 0.5$ for all other ages

The value of $_4a_1$ is assumed to be 0.4 and at all other ages 0.5 for all populations. In Greville's method, the death rates in the life table and the population are assumed to be the same, and the desired value of $_nL_x$ is calculated using the following equation:

$$L_x = ({_nd_x}/{_nm_x})$$

This equation can also be used for the 0–4 age group without any serious loss of accuracy. In the Kerala exercise, we have used the above formula. For the last age interval, that is, the interval with the indefinite upper age limit, the usual approximation for L_x is

$$L_x = (l_x/m_{x+})$$

Another simple method for the construction of life tables is that suggested by Chiang (1974). According to him the values are calculated as follows:

$${}_nq_x = \frac{n \cdot {}_nm_x}{1+\left(n-{}_na_x\right) \cdot {}_nm_x}$$

where ${}_na_x$ = average fraction-years of life lived by the persons who die in the age interval x to x+n

This is an exact life table where we not only use the data on the death rates but also the age at death of those dying in any age group.

NOTES

Similarly, the ${}_nL_x$ values are obtained by

$${}_nL_x = n\ (l_x - {}_nd_x) + {}_na_x \cdot {}_nd_x$$

The main problem or choices are faced in the conversion of ${}_nm_x$ into ${}_nq_x$ values, that is, central death rates into probabilities of dying. The rest of the life table can be completed with the usual relationships explained earlier. To recapitulate,

$$l_{x+n} = l_x\ (1 - {}_nq_x)$$
$${}_nd_x = l_x - l_{x+n}$$
$${}_nL_x = (n/2)\ [l_x + l_{x+n}]$$
$$L_{x+} = l_x\ /m_{x+}$$
$$T_x = \sum {}_nL_x$$
$$e_x = T_x\ /\ l_x$$

In any life table the average age at death, that is, the mean age of persons dying in a year is given by

$$\frac{{}_nL_x \cdot {}_nM_x}{\text{Total deaths}} = T_x = \frac{e_x}{l_x}$$

Thus in a stationary population the average age at death is the expectation of life at birth. In a growing population, that is, where $l_0(t)$ is greater than $l_0(t\text{-}1)$ then the average age at death in a population will underestimate the expectation of life.

It is not necessary to go through all the steps given in the boxes above in the actual construction of a life table. The software package MORTPAK developed by the UN or the PAS package developed by the US Census Bureau can also be used in the computation of all life table functions from a given set of ASDRs using different formulae. The impact of various causes of death can also be analysed using the life table approach by omitting from the deaths those due to a particular cause, recomputing the death rates and using these to construct cause-deleted life tables. This enables the researcher to extract the effects of one cause of death from overall death rates and examine the resultant gain in life expectancy.

NOTES

Tables 6.3 and 6.4 provide an illustrative example of the various columns of the life table for the state of Kerala in 2001, for males and females computed by using Greville's method.

The input values are the ASDRs for males and females given in Table 6.1 in the age groups less than 1 (0 –1), 1–4, 5–9, 10–14 until 80–84 and 85+. We constructed the life table using the LIFETAB programme under MORTPAK. The expectation of life for males is 69.5 and for females 72.6.

A line graph depicting the survivorship values of males and females at different ages (l_x) for Kerala in 1991 is given in Figure 6.1. Generally, the survivorship values of females at all ages are higher than for males as observed for Kerala in this figure.

Table 6.3: Abridged Life Table Based on Deaths and Population for Kerala, 2001 (Males)

Age x	Width n	$_nM_x$	$_na_x$	$_nq_x$	l_x	$_nd_x$	$_nL_x$	$_5P_x$	T_x	e_x
0	1	0.01613	0.088	0.0159	100,000	1,590	98,550	0.98305	6,954,834	69.55
1	4	0.00078	1.811	0.0031	98,410	305	392,972	0.99652	6,856,284	69.67
5	5	0.00058	2.500	0.00291	98,105	285	489,812	0.99709	6,463,311	65.88
10	5	0.00058	2.500	0.00291	97,820	284	488,389	0.99685	5,973,499	61.07
15	5	0.00068	2.500	0.00339	97,536	331	486,851	0.99541	5,485,110	56.24
20	5	0.00116	2.500	0.0058	97,205	564	484,615	0.99276	4,998,259	51.42
25	5	0.00175	2.500	0.00869	96,641	840	481,104	0.99035	4,513,644	46.71
30	5	0.00213	2.500	0.01061	95,801	1,017	476,462	0.98795	4,032,540	42.09
35	5	0.00272	2.500	0.01349	94,784	1,279	470,722	0.9806	3,556,079	37.52
40	5	0.00514	2.500	0.02538	93,505	2,374	461,592	0.97579	3,085,356	33.00
45	5	0.00466	2.500	0.02301	91,132	2,097	450,414	0.96427	2,623,765	28.79
50	5	0.00999	2.500	0.04874	89,034	4,340	434,321	0.94256	2,173,350	24.41
55	5	0.01378	2.500	0.06659	84,694	5,640	409,373	0.91308	1,739,029	20.53
60	5	0.02299	2.500	0.10871	79,055	8,594	373,788	0.86645	1,329,656	16.82
65	5	0.03512	2.500	0.16142	70,461	11,374	323,868	0.80364	955,868	13.57
70	5	0.05404	2.500	0.23802	59,087	14,064	260,273	0.71998	632,000	10.70
75	5	0.08052	2.500	0.33514	45,023	15,089	187,390	0.5971	371,727	8.26
80	5	0.13505	2.500	0.50481	29,934	15,111	111,891	0.39301	184,336	6.16
85	+	0.20461	4.887	1.00000	14,823	14,823	72,445		72,445	4.89

Table 6.4: Abridged Life Table Based on Deaths and Population for Kerala, 2001 (Females)

Age x	Width n	$_nM_x$	$_na_x$	$_nq_x$	l_x	$_nd_x$	$_nL_x$	$_nP_x$	T_x	e_x
0	1	0.00764	0.073	0.00758	100,000	758	99,297	0.99203	7,680,362	76.80
1	4	0.00027	1.721	0.00109	99,242	108	396,720	0.99771	7,581,065	76.39
5	5	0.00064	2.500	0.00318	99,133	315	494,880	0.99818	7,184,344	72.47
10	5	0.00009	2.500	0.00045	98,819	45	493,981	0.99909	6,689,464	67.69
15	5	0.00027	2.500	0.00136	98,774	134	493,532	0.99773	6,195,484	62.72
20	5	0.00064	2.500	0.00318	98,639	313	492,413	0.99547	5,701,951	57.81
25	5	0.00118	2.500	0.00589	98,326	579	490,181	0.99478	5,209,538	52.98
30	5	0.00091	2.500	0.00454	97,747	443	487,625	0.99456	4,719,357	48.28
35	5	0.00127	2.500	0.00635	97,303	617	484,973	0.99186	4,231,732	43.49
40	5	0.00200	2.500	0.00995	96,686	962	481,025	0.98893	3,746,759	38.75
45	5	0.00245	2.500	0.01220	95,724	1,168	475,700	0.98314	3,265,735	34.12
50	5	0.00436	2.500	0.02158	94,556	2,041	467,679	0.97513	2,790,034	29.51
55	5	0.00573	2.500	0.02823	92,515	2,612	456,048	0.96207	2,322,355	25.10
60	5	0.00982	2.500	0.04792	89,904	4,308	438,749	0.93732	1,866,307	20.76
65	5	0.01627	2.500	0.07819	85,596	6,692	411,248	0.88103	1,427,558	16.68
70	5	0.03554	2.500	0.16322	78,903	12,879	362,321	0.8002	1,016,310	12.88
75	5	0.05545	2.500	0.24351	66,025	16,078	289,930	0.67199	653,989	9.91
80	5	0.11273	2.500	0.43972	49,947	21,963	194,829	0.46484	364,059	7.29
85	+	0.16536	6.047	1.00000	27,985	27,985	169,230		169,230	6.05

The following equations were used in the computation of various columns where

$_nM_x$ = Age-specific central death rate (given)

$_na_x$ = Average person-years lived by those who die between ages x and x+n

$_nq_x$ = Probability of dying between exact ages x and x+n (age-specific mortality rate)

l_x = Number of survivors at age x

$_nd_x$ = Number of deaths occurring between ages x and x+n

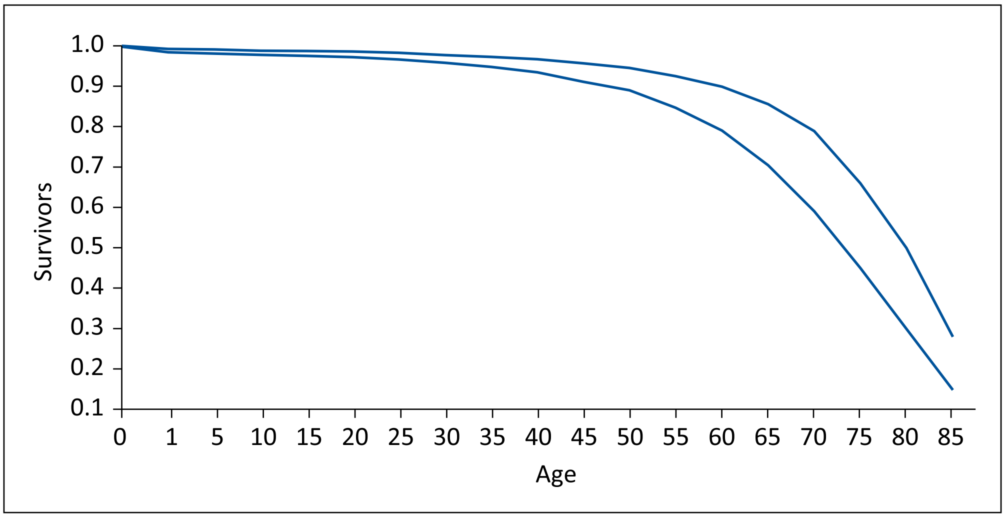
$_nL_x$ = Number of person-years lived between ages x and x+n

$_5P_x$ = Survival ratio for persons aged x to x+5 surviving 5 years to ages x+5 to x+10 = $_5L_{x+5}/_5L_x$ (first $_5P_x = {_5L_0}/_5l_0$, second $_5P_x = {_5L_5}/_5L_0$, last $_5P_x = T_{x+5}/T_x$

T_x = Number of person-years lived after age x. e_x = Life expectancy at age x

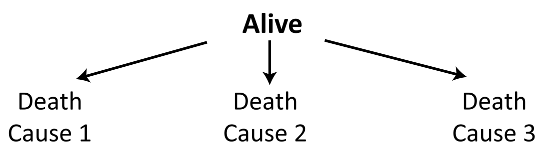
In the above life table death is an ultimate event that removes a person from the cohort and the initial cohort can only continuously decline, that is, $l_{(x+1)}$ has to be less than l_x for all x. Hence they are called *decrement life tables*. In a life-death situation it is impossible for a dead person to come alive! There are some forms of life table where there can be addition to the cohort age 'x' and such life tables are called *increment-decrement* life tables. For example, if we take the initial cohort as the number of newly identified cases of tuberculosis (TB) in a year, l_0 being the number of new cases of TB diagnosed in a year and all started on some treatment regime, then the possible outcomes in any successive month are (1) the person continues to suffer from TB; (2) the person recovers from TB;

Figure 6.1: Male and female survivors at different ages in Kerala, 2001



(3) the person who earlier recovered is reinfected with TB; and (4) the person dies during the period. The probabilities of each of these four outcomes change from month to month. If we take the persons suffering from the disease, as the main outcome the probability can go down in a month and increase in a succeeding month because a person who recovered earlier gets reinfected with the disease.

Then there are other forms of life tables where deaths due to different causes are studied using life table techniques. They are called *multiple-decrement* life tables. If there are three causes of death then the decrements due to each of these three causes can be studied separately.



In ‘multi-state’ life tables a person’s status in any month depends on the status in the previous month. For example, if we consider the probabilities of a cohort of single women getting married in any future month ‘t’, we have the following options: she can remain single until ‘t-1’ and can get married in ‘t’; she can get married before ‘t’, widowed before ‘t’ and get remarried in ‘t’; she can get divorced before ‘t’ and get married in ‘t’. Her getting married in time ‘t’ depends on whether she is alive at time ‘t’, single, widowed or divorced at the beginning of the month. The computation of all the values of such life tables essentially depends on the computation of probabilities of occurrence of marriage at time ‘t’ for a woman who is ‘single’, ‘widowed’ or ‘divorced at time ‘t-1’.

NOTES

Exercises

Exercise 6.1: Compute CDR, ASDR, NMR and PNMR for Rajasthan, 2001.

Age Group	Total Population	Number of Deaths	Males		Females	
			Population	Number of Deaths	Population	Number of Deaths
1	2	3	4	6	8	9
<1	1220962	107078	645900	55741	575062	51296
0-4	6012258	146098	3135155	72736	2877103	73654
5-9	8068070	13716	4242507	5091	3825563	8416
9-14	7241941	7966	3847418	3847	3394523	4073
15-19	5490214	7137	2987604	4481	2502610	3003
20-24	4749366	9499	2470943	3954	2278423	5240
25-29	4197223	9234	2127889	4469	2069334	4759
30-34	3821151	11081	1917187	5752	1903964	5521
35-39	3453860	8635	1788941	5904	1664919	2997
40-44	2760275	9937	1476036	7085	1284239	2954
45-49	2269651	9533	1183740	6866	1085911	2715
50-54	1775046	15265	974037	10422	801009	4966
55-59	1349315	18890	646505	11702	702810	6817
60-64	1361274	26136	676223	17447	685051	9248
65-69	982874	28110	457287	14999	525587	13035
70-74	757998	38658	376076	25348	381922	14437
75-79	315930	19904	145849	10982	170081	9031
80-84	237485	21849	107561	9487	129924	12382
85+	154711	23315	65929	14069	88782	9810
Total	54998642	424963	29272787	234639	26371755	193060
Death of babies						
0-4 weeks	78209					
4-52 weeks	54913					
Live Births	1664029					

Exercise 6.2: Construct life table for males and females using ASDR data from SRS for 2001.

The following equations are to be used in the computation of various columns:

$_nM_x$ = Age-specific central death rate (given: use data from Day 1 Exercise 2.1f)

$_na_x$ = Average person-years lived by those who die between ages x and x+n

$_nq_x$ = Probability of dying between exact ages x and x+n (ASMR)

$$_nq_x = \frac{n * {}_nm_x}{1+\left(n-{}_na_x\right) * {}_nm_x}$$

l_x = Number of survivors at age x

l_0 = 100,000

$l_{x+n} = l_x * (1 - {}_nq_x)$

${}_nd_x$ = Number of deaths occurring between ages x and x+n

${}_nd_x = l_x - l_{x+n}$

${}_nL_x$ = Number of person-years lived between ages x and x+n

${}_nL_x = n\,(l_x - {}_nd_x) + {}_na_x \times {}_nd_x$

$L_{x+} = l_x / m_{x+}$

${}_5P_x$ = Survival ratio for persons aged x to x+5 surviving 5 years to ages x+5 to x+10
$= {}_5L_{x+5} / {}_5L_x$

(first ${}_5P_x = {}_5L_0/{}_5l_0$, second ${}_5P_x = {}_5L_5/{}_5L_0$, last ${}_5P_x = T_{x+5}/T_x$).

T_x = Number of person-years lived after age x.

$T_x = \sum {}_nL_x$

e_x = Life expectancy at age x.

$e_x = T_x/l_x$

Exercise 6.3: Show age pattern of ${}_nm_x$, l_x for males and females in a graph.

CHAPTER VII

Migration and Urbanization – Basic Concepts and Measures

A. MIGRATION

7.1 Introduction: Importance of Migration and Concepts

The term 'migration' is derived from the Latin word 'migrare', which means 'to move or shift'. In demography it is usually considered the third component of population change, the other two being 'fertility' and 'mortality' discussed in the earlier chapters. Demographers view migration as the most important component of population dynamics in the future, since the effects of the forces of mortality and fertility are nearing completion and hence plateauing. The study of demography initially focused on the component of mortality, and then turned to fertility and in the coming decades it is expected that migration will occupy the centre stage of demographic analysis.

NOTES

In demography the term 'migration' has a specific connotation and is defined in the multilingual demographic dictionary of the UN as "spatial mobility between one geographical unit and another, involving a permanent change of residence". Thus it involves **a change of residence and crossing of political and/or administrative geographic units within a country or across countries**. It excludes 'tourism', which does not involve change of residence, short-term and cyclical migrations where people change their locations temporarily, even for extended time periods, but return to their permanent residence eventually. Thus a 'migrant' is defined as a person who has changed his/her usual place of residence from one migration-defining area to another at least once during the migration interval (usually, the interval may be 1 year, 5 years, 10 years, or intercensal period). This migration-defining area is defined for each country separately. In India, in the censuses the areas studied so far are the districts, states and the country as a whole. In surveys the areas can be defined by the researchers. In countries where there are 'National Population Registers' that record on a continuing basis the movements of peoples across defined boundaries, especially when there is a change of residence, as is being done in the Scandinavian countries, migration can be studied at any desired level of disaggregation of geographical units during any period of time.

The importance of migration and its effect on the growth and rise of a population has been well recognized and researched. It may be regarded as an important tool in the growth of a population and labour force of an area. The analysis and measurement of migration finds its use in the preparation of population estimates and projections at national or sub-national levels. Data on age, sex, mother tongue, duration of residence/place of birth or citizenship, occupation etc. of the migrants help us to understand the nature and magnitude of the problem. The social and cultural assimilation of immigrants in the country of destination are major concerns. There may be social and psychological effects of migration on the migrants and on the population of both sending and receiving areas. Due to the supply of skilled and unskilled labour, growth of industry, occupational and employment status of migrants, the business cycle is also affected. Keeping in view the extent of migration, policies and laws regarding migration can be formed. International migration from developing to developed countries has become a major political issue of the modern era.

7.2 Concepts and Measures

NOTES

Migration is classified into two types: **international migration**, which indicates movement of residence across countries and **internal migration** or movement within a country across states, districts etc. The internal migrants are called *in-migrants* and *out-migrants* when moving into or out of a predefined administrative area within a country contrasted with *immigrants* and *emigrants* when the movements take place across countries.

By definition, migration involves two areas: **place of origin** and **place of destination**. Hence it has to be studied between well-defined areas. Place of origin is the place or area from which a move is made and this can be defined in terms of a specified interval of time or place of birth. Similarly, the place of destination indicates the end point of movement as specified in a time interval or at the time of the study. **Migration stream** is defined as the total number of moves made during a given migration interval with a common area of origin and of destination. In practice, however, it is a flow of migrants having a common area of origin and destination and graphically expressed by the thickness of arrows between two areas, thickness based on the volume of migration.

The Indian population censuses usually study migration using two approaches: **'lifetime migration'** and **'place of last residence'**. There is a third approach termed **'fixed duration residence'** approach that is not used in the Indian Census, but is adopted in many other countries. A lifetime migrant is a person whose place of residence at census/survey date is different from his/her place of birth. The number of such persons in a population is referred to as lifetime migrants. This definition grossly underestimates both the number of moves and the number of migrants, as it excludes all moves that occurred between departure from place of birth and arrival in the area of residence as reported on a census date, and does not include migrant persons who moved out and subsequently returned to the place of birth. This type of analysis is done using the data on 'place of birth' collected from each individual at the time of the census, and whether the place of birth is the same as the place of enumeration, 'within the same district', 'outside the district but within the state', 'outside the state' and if so from which state, and 'from outside the country'.

The **'place of last residence'** data is based on the question posed to each individual in the place of enumeration as to from where he moved most recently if not living there since birth, and when he moved and studies the movement of persons during specified time periods such as less than 1 year, 1–5 years, 5 years, 10 years and over 10 years. A **'return migrant'** is defined as a person who has moved back to his/her place of birth after other moves. The **'fixed duration residence'** approach bases its data on the place where an individual resided for a specified time in the past, prior to the enumeration: 1 year, 5 years, 10 years etc. Such information requires good memory and recall information on the part of the respondent as to where he resided at the specified time in the past.

The term **'gross migration'** either in terms of lifetime migration or place of last residence status connotes the total volume of migration, that is, the sum of in-migrants and out-migrants between the two defined places. On the other hand, **'net migration'** connotes the net effect of these two factors in a specified area, that is, in-migrants minus the out-migrants. A positive sign connotes a net in-migration and a negative sign a net out-migration. Migration rates are computed, as in the case of birth and death rates, as the ratio of the number of migrants divided by the number of persons exposed to migration in an area during the specified period of time.

NOTES

7.3 Estimations of Migration Rates from Place of Birth Data

When the information is collected by place of birth and place of enumeration for each person in the census, the analysis of data is carried out by cross-classifying the persons in the census by the place of enumeration and place of birth. This can be done by taking different units such as the state, rural and urban areas within a state (rural-urban), to places within the same district (intra-district), to places in other districts within the state (inter-district) and to outside the state but within the country (inter-state) and to outside the country (international). The estimation of lifetime in-migrants and lifetime out-migrants is illustrated using 'place of birth data' relating to Kerala as obtained in the 2001 Census.

First we give a model tabulation plan of such data for any country in Table 7.1

Assume that there are four sub-regions A, B, C and D in a country and in any census the distribution of enumerated population is as given in Table 7.1.

In the classification, the sum of the values in each row excluding the diagonal elements gives the out-migrants from the regions and the sum of entries in any column excluding

Table 7.1: Place of Birth Tabulation in Estimating Migrants and Non-migrants

Place of Birth	Place of Enumeration			
	A	B	C	D
A	a_{11}	a_{12}	a_{13}	a_{14}
B	a_{21}	a_{22}	a_{23}	a_{24}
C	a_{31}	a_{32}	a_{33}	a_{34}
D	a_{41}	a_{42}	a_{43}	a_{44}

diagonal element would give the in-migrants to the region. The number of out-migrants from region A is equal to $a_{12}+a_{13}+a_{14}=x_1$. Similarly the number of out-migrants from regions B, C and D are equal to x_2, x_3 and x_4, respectively. The number of in-migrants in region A is the column total omitting the diagonal and equal to $a_{21}+a_{31}+a_{41}=y_1$. Similarly y_2, y_3 and y_4 and are the number of in-migrants to regions B, C and D respectively. (Note: the total number of in-migrants is equal to the total number of out-migrants). The diagonal elements a_{11}, a_{22}, a_{33} and a_{44} denote the non-migrants in region A, B, C and D respectively as their place of birth and place of enumeration are the same. The net lifetime migrants to any region can be obtained as the difference of in-migrants and out-migrants. For example, net lifetime migrants to region A = $(y_1 - x_1)$.

Taking into account the number of migrants (both in and out) we can calculate the percentage of lifetime migrants in a country, that is, the number of persons who are enumerated at different places other than their place of birth divided by the total population of that country and expressed as a percentage. A worked out example of place of birth data for females in the state of Kerala from the 1991 Census is given in Table 7.2. The place of birth data are not reported for a substantial number of respondents and the totals including these at different states of enumeration are given in the last column of the table.

Table 7.2 shows the number of females born in different states cross-classified by the states of enumeration. The actual table extends to all the states and union territories, but for the purposes of illustration we have taken the country as a whole to be made up of five states, Andhra Pradesh, Karnataka, Kerala, Maharashtra, Tamil Nadu and 'other states' considered as a single unit. The penultimate row total shows the number of females enumerated in different states other than the state of births, whereas the column

Table 7.2: Distribution of Population by State of Birth and State of Enumeration, Females, India, 1991

State of Birth	Place of Enumeration							
	Andhra Pradesh	Karnataka	Kerala	Tamil Nadu	Maha-rashtra	Other States	Total Females Born in the State	Total Out-migrants
Andhra Pradesh	32127228	290511	3660	113400	152167	159262	3287702	719000
Karnataka	211144	21053942	33880	84273	487118	38281	2195072	854696
Kerala	22210	107541	11458898	185286	81110	94557	1512169	490704
Tamil Nadu	132098	241097	126080	26855142	76750	60107	2763323	636132
Maharashtra	100345	228013	9410	13660	3605327	655211	3712063	1006639
Other states	135902	40760	11460	39078	744342	23971398	2434424	971542
Total in-migrants	601699	907922	184490	435697	1541487	1007418		
Total in-migrants including place of birth not stated	639939	934022	200500	510494	1957060	9018066		

Source: Census of India, 1991, Migration Tables D_1

Table 7.3: Lifetime Female In-Migrants, Out-Migrants and Net Migrants, Andhra Pradesh, 1991

Place of Origin and Destination	Lifetime In-migrants	Lifetime Out-migrants	Lifetime Gross Migrants	Lifetime Net Migrants
Andhra Pradesh	601699	719000	1320699	–117301
Karnataka	907922	854696	1762618	53226
Kerala	184490	490704	675194	–306214
Tamil Nadu	435697	636132	1071829	–200435
Maharashtra	1541487	1006639	2548126	534848
Other states	1007418	971542	1978960	35876
Total	4678713	4678713	9357426	0

totals indicate the number of persons born in different states. It is found that 11,458,898 females have their state of birth as Kerala and were also enumerated in Kerala at the time of the 1991 Census. They are considered as non-migrants for Kerala. However, this is an underestimate, since many of them might have gone to many other places but returned to Kerala by 1991. As per the 1991 Census, the total number of females born in Kerala but enumerated in Andhra Pradesh was 22,210, in Karnataka 107,541, in Maharashtra 8,110, in Tamil Nadu 185,286 and in other states of India 94,557. The out-migrants from Tamil Nadu to Kerala are more than from 'all other states' category. It is also seen from this table that 3660 females have their place of birth as Andhra Pradesh but are enumerated in Kerala. Similarly, 33,880 born in Karnataka, 126,080 born in Tamil Nadu, 9,410 born in Maharashtra and 11,460 born in other states but enumerated in Kerala. These are lifetime in-migrants and total 184,490 and the lifetime out-migrants total 490,704. The net migration is –306,214 and the gross is 684,194. The negative sign of net migration indicates that Kerala is a net out-migrating state.

NOTES

Table 7.3 provides the estimates of lifetime out-migrants and in-migrants, gross and net migrants in different states based on Table 7.2. It has to be noted that at the national level the net lifetime migrants is zero since the people born in one area are moving to another area of the country. Here we do not include persons born in other countries and those enumerated Indians born in India but living abroad.

7.4 Estimation of Intercensal Migration from Place of Birth Data from Two Censuses

Intercensal net migration can also be estimated from the place of birth data if it is available for the same region at the two consecutive censuses. It is estimated by subtracting the survivors of lifetime migrants counted in the first census from the lifetime migrants counted in the second census. Since all those who were counted as migrant in the first census will not survive up to the second census, it is necessary to apply an appropriate survival ratio to the migrants counted in the first census and get the number of migrants who survive up to the second census. This indirect estimate of intercensal net migration, assuming that the census is conducted after 'n' years, can be given as follows:

$NM = (I_{t+n} - O_{t+n}) - (S_I I_t - S_o O_t)$ ...(7.1)

where

NM = net intercensal migration

I_t = number of lifetime in-migrants at time t in a particular area

I_{t+n} = number of lifetime in-migrants at time t+n in that particular area

O_t = number of lifetime out-migrants at time t from that particular area

O_{t+n} = number of lifetime out-migrants at time t+n from that particular area

S_I and S_O are intercensal survival ratios indicating what proportions of I_t and O_t will survive during the intercensal period.

We can rewrite the formula 7.1 as

$NM = (I_{t+n} - S_I I_t) - (O_{t+n} - S_o O_t)$ ...(7.2)

Separating the in-migrants and out-migrants

$NM = M_1 - M_2$

where

$M_1 = I_{t+n} - S_I I_t$ and

$M_2 = O_{t+n} - S_o O_t$

NOTES

The net balance of intercensal migration is also the net balance in its two components, that is, net migration among persons born outside the area (M_1) and enumerated in Kerala and among the persons born inside the area (M_2) and enumerated outside.

In applying this method, the main problem is in the estimation of survival ratios S_I and S_O. The precise estimation of survival ratios of migrants requires a considerable amount of data on their age–sex distributions and computation because S_I and S_O are different. In-migrants and out-migrants differ in their age structure. There are several procedures to estimate these two parameters; some are elaborate and more complex and some are simple but sufficient for practical purposes when the total migration rate is small. A simple method is given below.

7.5 Estimation of Survival Ratio

If the data on age distribution of in-migrants is not available, it is almost impossible to estimate the survival ratios accurately. In such a situation, we can take both S_I and S_O as equal to overall census survival ratio, that is, ratio of persons aged n years and above in the country at the second census to persons of all ages in the first census (where n is the interval between the two censuses in terms of years).

Let P_t be the total population at time t, $P_{n+, t+n}$ be the population aged n and above at time t+n (n is the interval between the two censuses), then the overall survival ratio

$$S = \frac{P_{n+,\, t+n}}{P_t}$$

The overall survival ratio can also be approximately estimated from an appropriate life table covering the intercensal period, n, as

$$S = T_n / T_0$$

where, T_n and T_o are stationary population above age n and 0 respectively.

The above formulae assume that the age distribution of migrants, in and out, is the same as the age distribution in the first census. Though the census survival ratios may not measure the accurate probability of survival and there may be some error in the migration estimates, it is expected to be more accurate than what is estimated by ignoring the mortality factor completely. If we ignore the mortality factor, the expression 7.2 is reduced to

$$NM^1 = (I_{t+n} - I_t) + (O_t - O_{t+n})$$
$$= M^1_1 + M^1_2$$

NOTES

Table 7.4: In- and Out-migration, Kerala, 1981 and 1991 (based on Birthplace Data)

Category	Total		Males		Females	
	1981	1991	1981	1991	1981	1991
In-migration	341,180	398,470	170,232	197,970	170,928	200,500
Out-migration	1,131,598	1,123,006	639,309	589,500	492,289	533,506
Net migration	–790,438	–724,538	–469,077	–391,530	–321,361	–333,006

Taking the data for Kerala for 1981 and 1991 given in Table 7.4 on lifetime migrants we have the following:

The ratio of population aged 10 and above in 1991 to total population of all ages in 1981 works out to 0.93188.

Ignoring the survival ratios, the net migration can be calculated as follows:

$M = (I_{t+n} - I_t) + (O_t - O_{t+n})$

$I_{t+n} - I_t = 398470 - 341180 = 57290$

$O_t - O_{t+n} = 1,123,006 - 1,131,598 = -8592$

$M = 57290 + (-8592) = 48698$

where t is 1981 and t+n is 1991

If we apply the survivorship to the 1981 component then the difference works out to 41,269. The estimate of net migration by ignoring the survival ratios is about 18% higher than the estimate when survival ratios are accounted. The volume of net migration during the decade 1981–91 is very small compared to total lifetime migrants; this is partly because the number of lifetime emigrants who have gone abroad from Kerala, which is of substantial volume during the decade 1981–91, is not captured by this method.

7.6 Estimation of Migration from Place of Last Residence Data

Estimates of migration based on place of birth data grossly underestimate the total volume of migration. A better way of estimating migration statistics is the use of 'place of last residence' data; and additional information on migration can be obtained when these data are cross-classified by the duration of residence in the place of enumeration.

The 'place of birth' data does not give any information about persons who have migrated more than once, that is, those who have made more than one move. It does not give any information about the residence at the time of the last move of in-migrants. In order to get information on direct moves, it is necessary to ask questions about the place of last residence of the respondent and when he/she moved into the present location. From this data, we can identify persons as migrants whenever their place of last residence and place of current residence are different. Hence, the *'migrant'* group will include all *'lifetime migrants'* and *'return migrants'*, that is, all persons who have ever migrated or who have at any time lived outside the place of birth. According to this concept, non-migrants are those persons who have never moved outside their places of birth.

NOTES

Data tabulated according to the place of last residence can be utilized in the same way as the place of birth data for estimating migration. By cross-classifying the place of last residence data with the place of current/present residence, that is, place of enumeration, the places of origin of the in-migrants to an area and the place of destination of the out-migrants from an area an estimate of net migration can be obtained. The requirement of data and the method of estimation of migration by this approach are very much the same as the place of birth data, except that the place of last residence is referred to here and not the place of birth. Table 7.5 provides a classification of migrants from Kerala by their duration of residence in the place of enumeration and by their place of birth and Table 7.6 provides the percentage distribution of migrants with different duration of residence: less than 1 year, 1–4, 5–9, 10–19 and 20 and above years.

From these tables it can be seen that more than 60% of the lifetime migrants to Kerala are from Tamil Nadu, but it has declined over the years. It was 70.3% among those living in Kerala for 20 years or more and coming down to 59.3% among those who have recently moved, that is, within a year.

Table 7.7 shows that the bulk of the migrations of females in Kerala, more than 95% (86,84,199 out of a total 91,38,458) has been within the state, mostly due to marriage with persons within the state. Among those who moved within the state, most of them, 75% (63,43,612 out of 86,84,199), reported their last place of residence within the same district.

Table 7.5: Distribution of Lifetime Migrants from and to Kerala, 2001

State of Birth	Duration of Residence in Years					Total
	Less than 1	1-4	5-9	10-19	*20+	
Andhra Pradesh	729	2314	1016	904	1126	6089
Karnataka	2532	8238	6177	8292	11450	36689
Maharashtra	1335	4660	2947	2801	2750	14493
Tamil Nadu	9413	28751	21839	28669	43848	132520
Other states	1859	5442	2806	2817	3225	16149
Total	15868	49405	34785	43483	62399	205940

* Includes duration not stated

Table 7.6: Percent Distribution of Lifetime Migrants in Kerala by Duration of Residence

State of Birth	Duration of Last Residence					Total
	Less than 1	1-4	5-9	10-19	*20+	
Andhra Pradesh	4.6	4.7	2.9	2.1	1.8	3.0
Karnataka	16.0	16.7	17.8	19.1	18.3	17.8
Tamil Nadu	59.3	58.2	62.8	65.9	70.3	64.3
Maharashtra	8.4	9.4	8.5	6.4	4.4	7.0
All other major states	11.7	11.1	8.1	6.5	5.2	7.9
Total	100.0	100.1	100.1	100.0	100.0	100.0

NOTES

Table 7.7: Distribution of Lifetime Migrants in Kerala by Duration of Residence, 2001

Category of Migrants	Duration of Residence (in years)					Total
	Less than 1	1-4	5-9	10-19	20+	
Last residence elsewhere in India	3,57,955	15,70,500	14,01,602	19,76,660	26,15,256	91,38,458
Within the state of enumeration but outside the place of enumeration	3,15,706	14,53,345	13,25,919	18,87,612	25,18,207	86,84,199
Elsewhere in the district of enumeration	2,28,771	10,35,283	9,75,551	13,91,693	16,95,360	63,43,612
In other districts of the state of enumeration	86,935	4,18,062	3,50,368	4,95,919	8,22,847	23,40,587
States in India beyond the state of enumeration	42,249	1,17,155	75,683	89,048	97,049	4,54,259
Total	3,66,771	15,83,995	14,11,407	19,85,926	26,22,034	91,90,481

Source: Census of India, 2001, Series I, India Part II D (1), Migration Tables

The effects of spatial and temporal origins of in- and out-migrants and the effect of variation of one or another can be studied with the help of this type of tabulation.

7.7 Limitations

Data on the place of last residence also suffer from absence of a definite time reference. The place of last residence does not indicate a definite period of in-migration. So persons who have migrated 30 years ago or even earlier and persons who have migrated recently will be grouped together and called *migrants*. An important advantage of the place of last residence data is that direct movement between the places can be estimated while the place of birth data does not give an idea of the intervening movement; it gives an idea only of the first place of residence and the last place of residence where the person has arrived. It has to be noted that the sum of migrants of all durations may not be equal to the total migrants as migrants with an unspecified duration are not shown. The category of all migrants includes international migrants who are not shown in Table 7.7. From the table, it can also be seen that recent migrants (within the past 5 years) constitute about one-fifth of the total migrants both among inter-state and intra-state migrants, indicating that movements are continuing to take place in recent years as in the past and Kerala is a state of continuous movement of people from one area to the other.

NOTES

7.8 Accuracy of Data

The accuracy of data on place of birth and place of last residence and duration of residence in the place of enumeration should always be checked before the migration data are subjected to in-depth analysis.

a. **On the data on place of last residence**, the respondent may not be aware of the duration of residence of all the members of a household. There may, therefore, be a certain number of persons whose duration of residence is not known and may be reported as 'duration of residence not known'. The duration of residence data may also suffer from a bias in digit preference. Duration of stay of 4 years or 6 years may be reported as 5 years, while duration of stay of 9 years or 11 years may be reported as 10 years. Digit preference has been almost universally manifested in age reporting. Zachariah, in his study found that the number of migrants reporting duration of residence as 10 years was very much greater than single-year estimates. Similarly, the number in the case of 15 years duration was greater than for 13 or 14 years. The digit preference quotients discussed in the earlier chapters can also be applied to migration data.

b. **On the data on place of birth**, though one would expect that such a simple question as the place of birth will be easily understood and answered correctly with accuracy and completeness there are still large response errors. The census question is generally asked to the head of the household or a responsible member of the household. The respondent may not be aware of the exact place of birth of all the members of the household, particularly daughters-in-law. A person living at a certain place for quite long time may report it as his/her place of birth. Thus, the chances of unintentional reporting of wrong place of birth are quite high. A person can correctly identify his/her place of birth as rural or urban. A person born in a less known place, maybe a village, may specify his place of birth as a nearby town or city. In such cases, the data will reflect the place of birth as urban instead of rural. Sometimes, the boundary of areas may also change, and the respondent may not be familiar with such changes and, due to ignorance, he may report his place of birth incorrectly.

Sometimes, the place of birth data appears to be extremely inadequate for migration analysis. In India it is customary for a woman to deliver her first child, and sometimes, her second child also, at her father's household. Hence this question provides spurious data for the estimation of migration from the place of birth data. Another problem associated with the use of place of birth data for migration analysis is that the timing of migration is not known.

It is not certain whether the place of birth is more likely to be remembered than the place of last residence but this may be true for people who have made many moves. Hence there may not be much improvement in the data collected if the place of birth data is replaced by the place of last residence.

Cross-classification of data in such a manner provides a lot of information about the patterns and character of internal migration. However, such details for various timings of migration require a lot of data collection and tabulations. The duration by place of last residence data would give a somewhat higher number of migrants for a given interval because it would count circular migrants.

NOTES

7.9 Estimation of Migration from Place of Residence at a Fixed Prior Date

Sometimes a question on the place of residence at a fixed prior date is included in the censuses/surveys for estimating migration. In this section, we learn how to obtain an estimate of migration using this concept. Advantages and limitations of data will also be examined. We also learn about the additional information obtained when the data are cross-classified according to the place of birth.

When the concept of place of last residence is used for estimating migration all the migrants, irrespective of their migration intervals, are pooled together. For example, all migrants, whether they migrated 20 years back or a day prior to the census, are counted as migrants. However, using the concept of place of residence at a fixed prior date, the migration interval is definite. A person whose place of residence at a fixed prior date (fixed by a researcher according to the research needs) is different from his/her place of enumeration is considered as a migrant. Similarly, all those persons whose place of residence at a fixed prior date was the same as their place of enumeration are treated as non-migrants by this concept. With this approach it is possible to get surviving migrants for a fixed duration. All those migrants who were alive at that fixed prior date and who subsequently migrated to other places but died afterwards cannot be counted using this approach. Similarly, this approach does not take into account those migrants who moved during the migration interval, but subsequently returned to the place where they were residing at the fixed prior date; this holds good for all those children who were born outside the place of enumeration too.

7.9.1 Estimation of primary, secondary and return migrants

From the data on place of birth and place of residence t years ago, it is possible to estimate primary, secondary and return migrants. Primary migrants are those who were living in their place of birth t years ago and in another place in t-n years ago. Secondary migrants would be those who were living outside their place of birth t years ago and in another place in t-n years ago. Return migrants are classified as people who were living outside their birth place t years ago and had returned to birth place t-n years ago. Let A, B and C be areas of a hypothetical country where the census was conducted on 1 March 2001. The information about place of birth and place of residence as on 1 March 1991 was also collected from the persons residing in these areas in 2001. All the information is presented below in Tables 7.8A, 7.8B and 7.8.C. These tables classify the persons enumerated in 2001 according to their place of residence in 1991 and their place of birth and Table 7.8D provides the manner in which the primary, secondary and return migrants are estimated in this scheme of data collection.

Table 7.8A: Distribution of Persons Enumerated in Area A on 1 March 2001, According to Their Place of Residence on 1 March 1991 and Place of Birth

Place of Birth on 1 March 1981	Place of Residence on 1 March 1991		
	A	B	C
A	X_{11}	X_{12}	X_{13}
B	X_{21}	X_{22}	X_{23}
C	X_{31}	X_{32}	X_{33}

Table 7.8B: Distribution of Persons Enumerated in Area B on 1 March 2001, According to Their Place of Residence on 1 March 1991 and Place of Birth

Place of Birth on 1 March 1981	Place of Residence on 1 March 1991		
	A	B	C
A	Y_{11}	Y_{12}	Y_{13}
B	Y_{21}	Y_{22}	Y_{23}
C	Y_{31}	Y_{32}	Y_{33}

Table 7.8C: Distribution of Persons Enumerated in Area C on 1 March 2001, According to Their Place of Residence on 1 March 1991 and Place of Birth

Place of Birth on 1 March 1981	Place of Residence on 1 March 1991		
	A	B	C
A	Z_{11}	Z_{12}	Z_{13}
B	Z_{21}	Z_{22}	Z_{23}
C	Z_{31}	Z_{32}	Z_{33}

NOTES

Table 7.8D: Estimates of In- and Out-Migrants (Primary, Secondary and Return) for Areas A, B and C

Areas	In-Migrants				Out-Migrants			
	Primary	Secondary	Return	Total	Primary	Secondary	Return	Total
A	$X_{22} + X_{33}$	$X_{23} + X_{32}$	$X_{12} + X_{13}$	I_A	$Y_{11} + Z_{11}$	$Y_{31} + Z_{21}$	$Y_{21} + X_{31}$	O_A
B	$Y_{11} + Y_{33}$	$Y_{13} + Y_{31}$	$Y_{21} + Y_{23}$	I_B	$X_{21} + Z_{22}$	$X_{32} + Z_{12}$	$X_{12} + Y_{32}$	O_B
C	$Z_{11} + Z_{22}$	$Y_{12} + Y_{21}$	$Z_{31} + Z_{32}$	I_C	$X_{33} + Y_{33}$	$Y_{13} + X_{23}$	$X_{13} + Z_{23}$	O_C

Where $I_A = X_{21} + X_{33} + X_{23} + X_{32} + X_{12} + X_{13}$

$O_A = Y_{11} + Z_{11} + Y_{31} + Z_{21} + Y_{21} + Z_{31}$

Similarly I_B, I_C, O_B, and O_C are computed.

From the above hypothetical data, in-migrants, out-migrants and net migrants by primary, secondary and return migration status during the period 1 March, 1991 to 1 March 2001 are estimated as follows:

a. In-migrants to area A during the period 1 March 1991 to 1 March 2001 are categorized as primary, secondary and return migrants.

   i. *Primary immigrants to area A*

   = Persons enumerated in area A on 1 March 2001 but enumerated elsewhere on 1 March 1991 in their place of birth

   = Persons enumerated in area A on 1 March 2001, also enumerated in area B on 1 March 1991 and born in area B + persons enumerated in area A on 1 March 2001 and also enumerated in area C on 1 March 1991 and born in area C

   = $X_{22} + X_{33}$ (from Table 7.8A)

   ii. *Secondary in-migrants to area A*

   = Persons enumerated in area A on 1 March 2001, enumerated in area B on 1 March 1991 and born in area C + persons enumerated in area A on 1 March 2001, enumerated in area C on 1 March 1991 and born in area B

   = $X_{32} + X_{23}$

   iii. *Return in-migrants to area A*

   = Persons enumerated on 1 March 2001 in area A, enumerated elsewhere on 1 March 1981 but born in area A

   = Persons enumerated on 1 March 2001 in area A, enumerated in area B on 1 March 1991 and born in area A + persons enumerated in area A on 1 March 2001, enumerated in area C on 1 March 1981 and born in area A

   = $X_{12} + X_{13}$

b. Out-migrants from area A during the period 1 March 1981 to 1 March 2001 are persons who were enumerated in area A on 1 March 1991 and also enumerated elsewhere (i.e. in area B or C) on 1 March 2001. These out-migrants from area A are further categorized as follows:

NOTES

i. *Primary out-migrants from area A*

= Persons born in area A and who were enumerated in area A on 1 March 1991. However, they were subsequently enumerated in area B or C on 1 March 2001

$= Y_{11} + Z_{11}$

ii. *Secondary out-migrants from area A*

= Persons born in area B, enumerated in area A on 1 March 1991 and enumerated in area C on 1 March, 2001 + persons born in area C, enumerated in area A on 1 March 1991 and also enumerated in area B on 1 March 2001

$= Y_{31} + Z_{21}$

iii. *Return out-migrants from area A*

= Persons born in area B but enumerated in area A on 1 March 2001 and also enumerated in area B on 1 March 2001 + persons born in area C, enumerated in area A on 1 March 1991 and further enumerated in area C on 1 March 2001

$= Y_{21} + Z_{31}$

Following a similar procedure, in-, out- and net migrants for each of the areas A, B and C can be estimated. Details of such migration are given below.

NOTES

7.10 Estimation of Net Migration from Vital Statistics Data

We can use data on the registered births and deaths in an area in a given period of time and the population change that has occurred during that time in order to estimate the net migration to the area. Here migration is estimated as the difference between the observed population growth and the growth that is expected because of natural increase, that is, the difference between the births and deaths. Since Kerala has had a fairly complete system of registration of births since the 1980s this approach can give a fairly good estimate of net migration at even the district level. Table 7.9 gives the increase in the population of Kerala during the decades 1951–61 to 1991–2001 and the births and deaths registered in the state during these decades.

From the table it can be seen that, in every decade since 1951, migration had a negative effect on population growth. In general, every succeeding decade since independence witnessed an increasing volume of net out-migration (an exception being the 1961–1971 decade). These need not entirely be of net migration between Kerala and the other states

Table 7.9: Components of Population Growth, Kerala, 1951-2001 (in ’000)

Decade	Population Growth	Birth	Deaths	Natural Migration	Net
1951–61	3,355	6,684	3,000	3,684	–330
1961–71	4,444	7,095	2,333	4,762	–318
1971–81	4,106	6,398	1,819	4,579	–473
1981–91	3,620	6,071	1,707	4,364	–744
1991–01	2,765	5,482	1,827	3,655	–890

in India. In most decades this was basically the case, but in recent decades considerable emigration has taken place from the state to the Gulf region and to other countries. These flows will affect the net migration estimates obtained by the balancing equation method. The figures should not, therefore, be interpreted to mean that Kerala has lost an increasing number of migrants to the rest of the country. The loss could as well be to countries outside India. The origins and destinations of the migrants within the country are largely obscure.

B. URBANIZATION

7.11 Concepts and Definitions

The term **urbanization** connotes the restructuring of human settlements that occurs because of industrialization and modernization of a society and is considered a hallmark of development. Urban ways of thinking and living and urbane are indicative of the changes that occur within the minds of people as much as in their settlement patterns. Sociologically, urbanization is a change in human settlement patterns arising out of industrialization and modernization. The term is used to denote both the process of urbanization and the levels and tempo of urbanization. It is a social phenomenon related to changes in the economic structure of a population when it moves from a predominantly agricultural society to an industrial society. There are many conceptual and definitional problems, varying from country to country. Unlike the precision with which a birth or death is defined, urbanization is a somewhat vague concept. The UN has recommended that the member countries should develop their own definition keeping in view the following three components of a location or settlement before it is defined as an 'urban' area:

a. Type of local government

b. Number of inhabitants, and

c. Proportion of population engaged in non-agricultural activities.

Keeping in view the above recommendation, the Census of India has defined an area as urban if it has statutory status with local administration like towns/municipalities/corporation/notified area committee/cantonment board/estate office etc.

In addition, where there is no local administration the settlements are treated as urban if they fulfil the following three conditions:

i. A population of more than 5000

ii. More than 75% of the male working population is engaged in non-agricultural activities

iii. Density of population is more than 400 persons per square kilometre.

The categories of occupations or economic activities which are to be included in 'non-agricultural' activities have changed from census to census. Until the 1981 Census the occupations of fishing, livestock, hunting, planting and orchards were included in agricultural workers; in 1981 these were included in the category of non-agricultural workers. Thus two types of places are included as urban areas: the first notified by the

NOTES

state or central governments as 'urban' and second those that satisfy the second criteria (b) given above. The Census of India has also defined certain urban centres as **'urban agglomerations (UAs)'** and certain areas as **'standard urban areas'**.

The term **'urban agglomeration'** is defined as

> *A continuous urban spread constituting a town and its adjoining urban outgrowths (OGs) or two or more physically contiguous towns together and any adjoining urban OGs of such towns.*

Examples of OGs are a railway colony, university campus, port area etc. that may come up near a city or statutory town outside its statutory limits but within the revenue limits of a village or villages contiguous to the town or city. In 2001 Census, for the purpose of delineation of UAs the prerequisite criteria were (a) The core town or at least one of the constituent towns of an UA should necessarily be a statutory town; and (b) the total population of all the constituents (i.e. towns and OGs) of an UA should not be less than 20,000 (as per 1991 Census).

The term **'Standard Urban Area'** is defined from 1981 Census as follows:

NOTES

1. *It should have a core town with a minimum population of 50,000;*
2. *The contiguous other urban areas as well as rural administrative units should have mutual socio-economic links with the core; and*
3. *The probability is that this entire area will get fully urbanized in a period of two to three decades.*

The urban populations are also studied by urban locality size class in order to study the type of urbanization taking place in a country or region. In India the urban areas are classified into six classes as follows:

Class	Population Size
I	>100,000
II	50000 to 99999
III	20000 to 49999
IV	10000 to 19999
V	5000 to 9999
VI	<5000

Table 7.10 gives the percentage of population of the urban areas living in different class towns in India from 1961 to 2001 as obtained from various censuses.

The table reveals that the percentage share of urban population in Class I cities has increased from 51.4% in 1961 to 68.5% in 2001. The percentage share of all other smaller cities and towns (Class II to VI) has systematically declined over the period. Class V and VI towns are practically disappearing. When urbanization takes place because of the natural processes of industrialization as it occurred in western countries, first smaller towns are formed, then bigger towns which become bigger over time and ultimately bigger cities

Table 7.10: Percent of Urban Population of Different Size Class, 1961–2001

Size Class	1961	1971	1981	1991	2001
I (>100000)	51.4	57.2	60.4	65.2	68.5
II (50-100000)	11.2	10.9	11.6	11.0	9.6
III (20-50000)	16.9	18.0	14.4	13.2	12.4
IV (10-20000)	12.8	10.9	9.5	7.8	7.0
V (5-10000)	6.9	4.5	3.6	2.6	2.4
VI (<5000)	0.8	0.4	0.5	0.3	0.2
Total	100.0	100.0	100.0	100.0	100.0

and metropolises come into existence when large populations shift from agriculture to industry. This is not the case in developing countries, especially India. People move from villages to big cities bypassing small towns because of better educational and employment opportunities in these cities and not due to the natural processes of push and pull factors of area economic development that exist in developed countries. Now the large-scale movements are from village to metropolis, largely from villages to slums in cities. This type of movement has raised concepts such as 'suburbs' defined as a cluster of communities immediately surrounding a central city and residential in character; 'semi-urban' or 'urban' areas which are areas in the process of transition from a rural to urban mode of settlement, and slums which have come to be accepted as a part of the modern urbanization process in developing countries. Slums constitute almost 50% of the metropolitan city of Mumbai and other big cities.

NOTES

7.12 Measurement

There are two aspects by which urbanization of a country or a sub-national area is measured: **level and tempo of migration**. By 'level of migration' we mean the percentage of population that is considered 'urban'. By 'tempo of migration' we mean the speed or rate at which this proportion is changing over time or how fast the migration levels are changing. The experiences of urbanization in many countries, developing as well as developed, reveal that the percentage of urban population increases very slowly in the early stages of urbanization, and picks up pace after it reaches a level of over 30% of the total population. This continues until the level of urbanization reaches about 70% and then slows down. In many developed countries 90% or more of the population is currently living in urban areas.

The level of urbanization is measured by one or more of the following indices:

i. Percent of population urban

ii. Urban/rural ratio

iii. Size of locality of residence of median inhabitant in urban areas and

iv. Mean city population size

The first two relate the urban population to the total or rural population and the next two relate the urban areas among themselves in terms of their size.

The tempo or speed of urbanization is measured over a period of time as

i. Annual growth rate of urban population size using linear or exponential curve

ii. Annual change of urban percentage points using linear or exponential rates

iii. Annual change in urban/rural ratios using linear or exponential curves

7.13 Urbanization in Kerala

Table 7.11 presents the data for Kerala on the total and urban populations in 1981, 1991 and 2001 for illustration of the indices of level and tempo of urbanization.

The table reveals that during the decade 1991–2001, in terms of urban/rural ratio and the rates of growth in urban population, percent urban and urban/rural ratios there has been a significant decline from the earlier decade. Both the level and tempo of urbanization in Kerala has slowed down. According to 2001 Census, the state had five municipal corporations and 53 municipalities, 101 towns and 25.96% of the population as urban, which is slightly lower than the national average of 27.81%. This is surprising since by all standards of human development, health, education and other parameters Kerala is considered to be one of the most advanced states, if not the best. The reality is that urbanization in Kerala is not limited to the designated cities and towns but is spread all across the state. Barring a few panchayats in the hilly tracts and a few isolated areas, the entire state depicts the picture of an urban-rural continuum. Thus, Kerala society by and large can be termed as urbanized.

NOTES

In 1991 there were 197 census towns with a population of 7.68 million or 26.44% of the population declared urban and according to 2001 Census there were only 159 census towns with 8.27 million population or 25.97% declared urban. Thus there was a declining trend of urbanization during the decade 1991–2001; the percentage decennial growth of urban population in the state was 60.89 during 1981–91 but declined sharply to 7.04% during 1991–2001. This was mainly due to change in jurisdiction in statutory urban areas when the number of census towns declined from 197 to 159.

Table 7.11: Urbanization in Kerala: Indicators of Level and Tempo of Urbanization

Year	Population in Thousand					Rate of Change (Exponential)		
	Total	Urban	Rural	Percent Urban	Urban/ Rural Ratio	Urban Population	Percent Urban	Urban/ Rural Ratio
1981	24454	4771	19683	19.51	24.24			
1991	29099	7680	21419	26.39	35.86	0.0476	0.03	0.039
2001	31841	8267	23574	25.96	35.07	0.0074	–0.002	–0.002

Kerala does not have the problems of primate city development and metropolitan city development found in other states. The massive rural to urban migration, which accentuates urban problems and urban poverty in other states, is only marginally present in urban Kerala. This is because of lack of development of any large industrial base in the state.

There are no prescribed criteria for constitution of cities. Municipalities were elevated to the status of corporations on considerations of their importance, pace of urbanization in the area, need for integrated development of the urban core and its neighbourhood, density of population, income and demand for more progressive civic administration.

7.14 Indices of Concentration and Dispersion in Urban Areas

Among the urban areas, cities, municipalities and towns and notified urban areas, the manner in which the populations are distributed and dispersed throws much light on the tempo of urbanization. In addition to the indices of rates of change in percent urban and the urban/rural ratio discussed above, three other indices namely (a) mean city population size (MC); (b) index of city distribution (ICD); and (c) primacy index (PI) are used and are defined as follows:

NOTES

a. Mean city population size

This index takes into account the size of cities, unlike percent urban or urban/rural ratio. It tells the size of the city in which the average urban dweller lives. This index has the same concept as the mean age of a population and the calculation is similar. It is defined as

$$MC = \frac{\text{Sum of } C_i^2}{P}$$

where

C_i is the population of city I, and summation is done over all cities and

P is the total population of the country.

The calculation of this index requires data on every settlement in the country. If instead of all cities, cities over a certain population size (over 50,000 or 100,000) are taken, this index is not significantly affected. This index can be considered to be a product of two parts: the mean city population size and the proportion of population in cities. Hence change in this index over time can be decomposed due to changes in these two dimensions. The lower limit of this index is very close to 1 when the population of the country is completely dispersed without any city and the upper limit is the total population when the whole population lives in one city.

b. Index of city distribution

Before defining this index, it is appropriate to discuss the well-known ‘rank size rule’ used in urban studies. Studies of cities and towns around the globe have revealed that there is a close association between the size of a city and the rank it holds when all the cities of a country are arranged in the rank order of largest (in terms of population size) to smallest. The population size of any city is found to be functionally related to the largest city with its rank size as follows:

$$C_K = C_1 \, . \, K^{-z}$$

where, C_1 is the largest city or rank one in the country and C_K is the city of rank K.

Though this is a simple rule, the value of z is found to be a very useful index of distribution of population among cities. Given the population size of different cities and towns, this index can be computed by least square method for which there is a software programme. If z = 1, it means that the second city's population (C_2) will be $C_1/2$ or one-half of the population of the largest city, the third city will be one third and so on.

z = 0 implies that all cities and towns are of the same population size and that urban population of the country is equally distributed across all cities. The higher the value of 'z' the higher is the dispersion of population among the cities and the higher the standard deviations. The greater the index value, the greater is the concentration of urban population in the largest city. The comparison can also be made with the expected distribution under the rank size rule.

NOTES

c. Primacy Index

The primacy index (PI) is also based on the z value which shows the dominance of the largest city with respect to the second largest or subsequent cities. It is evaluated on the basis of data of the first two cities, four cities or even 10 cities. As we stated earlier, if z = 1 we have the population of any city equal to the population of the largest city divided by its rank. The size of the largest city C_1 is equal to the sum of the second and third largest cities and a fraction of the fourth city, that is,

$$C_1 = (C_2 + C_3 + (1/6) * C_4)$$

In other words, the ratio of the population of the first city to the sum totals of the population of the other cities of lower ranks can be used as an index of primacy of the largest city in urbanization and is called the *primacy index* (PI). The four-city PI is given by

$$PI_4 = \frac{C_1}{C_2 + C_3 + C_4}$$

If there are 11 cities the denominator will be the sum of populations of the second to eleventh city. Taking the state of Kerala, using the census population of 2001 and restricting the analysis to class I towns, that is, those with a population of over 100,000 we computed the various indices of urbanization using the PAS package and the results are given in Table 7.12. There were 14 class I towns with Kochi having the highest population. The various indexes are also given in the table.

Table 7.12: Estimation of Indices of City Population for Kerala, 2001

Total and Urban Population			Indices	
Total population		31,841,274	Index	Value
Urban population (Class I cities)		8,267,135	Summary Indices	
City Populations Ranked by Size			Total population	31,841,274
Rank	Name of City	Population	Urban population	8,267,135
1	Kochi	1,355,972	Rural population	23,574,139
2	Tiruvanthapuram	889,635	Percent urban	25.96
3	Kozhikode	880,247	Urban/rural ratio	0.35
4	Kannur	498,207	City population	5,692,104
5	Kollam	380,901	Non-city population	26,149,170
6	Thrissur	330,122	Number of cities	14
7	Alappuzha	282,675	**Primacy Indices**	
8	Palakkad	197,369	First 4 cities	0.598
9	Kottayam	172,878	First 11 cities	0.688
10	Mallapuram	170,409	Index of city distribution	0.876
11	Cherthala	141,558	City concentration ratio	0.441
12	Guruvayoor	138,681	Index of mean city pop size	130,535
13	Kanhangad	129,367	Index of city concentration	0.4
14	Vadakara	124,083		

NOTES

7.15 Slums: Definition and Concepts

UN-HABITAT defines a slum household as a group of individuals living under the same roof in an urban area who lack one or more of the following:

1. Durable housing of a permanent nature that protects against extreme climate conditions.
2. Sufficient living space which means not more than three people sharing the same room.
3. Easy access to safe water in sufficient amounts at an affordable price.
4. Access to adequate sanitation in the form of a private or public toilet shared by a reasonable number of people.
5. Security of tenure that prevents forced evictions.

7.15.1 Slums in India

For the purpose of Census of India, 2001, the slum areas broadly constitute

1. All specified areas in a town or city notified as 'slum' by State/Local Government and UT Administration under any Act including a 'Slum Act'.
2. All areas recognized as 'Slum' by State/Local Government and UT Administration, Housing and Slum Boards, which may have not been formally notified as slum under any act.
3. A compact area of at least 300 population or about 60–70 households of poorly built congested tenements, in unhygienic environment usually with inadequate infrastructure and lacking in proper sanitary and drinking water facilities.

Exercises

Exercise 7.1: Compute indicators of level and tempo of urbanization for Rajasthan.

Exercise 7.2: Estimate the indices of city population for Rajasthan.

Exercise 7.3: Calculate number of migrants (in and out) for Rajasthan.

CHAPTER VIII

Nuptiality Analysis

8.1 Introduction

In most countries marriage is the accepted social sanction required for a man and woman to live together and to enter into a legitimate sexual relationship for the biological continuity of the population. Thus, an analysis of marriage is an important area for the study of the demographic process, especially of levels and trends in fertility, of a population. It is considered an important proximate determinant of fertility. Sociologically, a study of marriage goes beyond its impact on fertility. It is a way of child rearing, child support, gender equity and equality and organization of the society itself. Since the family constitutes the primary unit of a society and marriage forms the basis of organization of families, the study of society at large depends to a large extent on the study of marriage as an institution. In this chapter, we confine our attention to the role of marriage as it affects fertility and will not go into its larger sociological implications.

In the studies on population dynamics, marriage and its dissolution are found to be important factors in determining fertility levels in most of the societies. The analysis of marriage, separation, widowhood and remarriage are collectively called *nuptiality analysis* in demography. Marriage is considered to affect all the three demographic components, but more particularly births and migration. In this chapter we discuss some commonly used measures of nuptiality. If the main interest is to study the effects of marriage on the fertility pattern, then the interest on nuptiality will focus on age at marriage, remarriage, the duration spent by women outside the state of marital union because of widowhood, non-marriage, separation, divorce etc., when pregnancies are not expected. It is also important to study at what ages in the reproductive period the women have their sexual unions within marriage in order to assess their impact on fertility. We will be mainly interested in examining when a woman is exposed to the risk of childbearing which would correspond to her age at first marriage. Since biologically fertility is highest in the late teens and early 20s for any woman, early marriage has a greater impact on fertility than later marriage. If the research interest extends to the general areas of family formation and its break-ups, then separation, divorce, widowhood and remarriage are of as much interest as marriage. In recent years, developed countries are facing a rapid change in their nuptiality patterns as a consequence of social change. With increased employment of women, greater autonomy and increased expectation of life of men and women, the demographic process has been affected considerably. A large proportion of births in the developed countries occur outside marriage, with men and women living in a

NOTES

consanguineous union without formal marriage. In Britain in 2008 almost 50% of all births occurred to women who were not formally married. Such births in the developed world have led to a theory of second demographic transition briefly discussed in Chapter II.

Marriage is considered as the union of man and woman by social consent, whatever be the form, contract and obligations of such a marriage. When this contract is broken, again with social sanction, whether it is through legal procedures, simple consent or other means, it is identified as **divorce**. When one of the spouses dies, the marital status is called **widowhood**. **Separation** is the result of the man and woman living away from each other without any union. Once the marriage is dissolved by any of the said ways and either of the spouses marries again, it is called **remarriage**. Marital status also includes persons who are not married by any social sanction, that is, single persons.

8.2 Crude Marriage Rate (CMR)

NOTES

Many measures of marriage are used in nuptiality analysis. The most common and the simplest measure is the crude marriage rate (CMR). If M is the number of marriages among residents in an area during a year, and P is the mid-year population in the area during the same year, then the CMR is defined as

$$CMR = \frac{M}{P} * 1000$$

This measure also suffers from the usual defects of a crude rate. The numerator of the rate uses the number of marriages. However, marriage involves two persons and if we use the number of persons getting married during the year, the rate becomes twice as large. Further, the entire population cannot be considered as the population exposed to the risk of marriage, as the very young and the persons who are already married do not belong to the exposed category.

8.3 General Marriage Rate (GMR)

As stated above, the denominator of the CMR is not actually the population exposed to the risk of marriage. A more refined measure can be calculated by taking the persons of marriageable ages as the denominator. The rate thus obtained is called general marriage rate (GMR) and is represented by

$$GMR = \frac{M}{P_{15+}} * 1000$$

Here, the total number of marriages in the year is again employed in the numerator and the population aged 15 and above is used in the denominator. Further refinement can be done by taking account of unmarried persons in the marriageable ages and sex composition of the population.

Hence, an alternate definition is

$$GMR = \frac{M}{\text{Unmarried persons aged } 15+} * 1000$$

This rate can be calculated separately for males and females. Since rate of marriage among males and females would differ depending upon the sex ratio of the population, it calls for the calculation of marriage rates separately for males and females. As remarriage is also possible, it may be necessary to identify the numerator by the order of the marriage as first, second etc., and the denominator adjusted accordingly.

8.4 Age-Specific Marriage Rate (ASMR)

The GMR defined above partially controls the age distribution of the population, but the internal variations in the age composition within the population of those marriageable are not taken into these rates. It should also be noted that since generally husbands tend to be older than wives, the age-specific rate will be different for the sexes. Hence it is more advisable to have an age-sex specific marriage rate (ASMR), which in general is defined as

$$ASMR = \frac{{}_nM_x}{{}_nP_x} * 1000$$

where

${}_nM_x$ = number of marriages in a year between ages x and x + n

${}_nP_x$ = number of persons in the age group (x, x + n)

The ASMR can be computed separately for males and females. Instead of the number of persons in the age group (x, x + n), the number of unmarried persons can also be used in the denominator. As discussed earlier, remarriages are possible and hence it is possible to re-enter the married state. Thus it is important to have an order-specific marriage rate. The measure generally used is the age-specific first marriage rate (ASFMR).

$$ASFMR = \frac{{}_nM_x^1}{{}_nP_x^0} * 1000$$

where

${}_nM_x^1$ = number of first marriages in a year in the age group (x, x + n)

${}_nP_x^0$ = number of single or unmarried persons in the same age group and eligible for marriage for the first time

NOTES

In general, age order-specific marriage rate (AOSMR) can be defined as the number of marriages of a given order i in a specified age category, per thousand persons of the same age group at the order i-1 during a period of time.

$$AOSMR = \frac{{}_nM_x^i}{{}_nP_x^{i-1}} * 1000$$

where

${}_nM_x^i$ = number of i^{th} order marriages in the age group x to x + n

${}_nP_x^{i-1}$ = persons in the same age group whose $(i-1)^{th}$ marriage remains dissolved and they are eligible for the i^{th} marriage

8.5 Total Marriage Rate (TMR)

Like the TFR which is used in fertility analysis, it is also possible to derive a total marriage rate (TMR) from the ASMR. The value of TMR indicates the total number of marriages in a synthetic cohort passing through life together. Thus,

NOTES

$$TMR = \sum(ASMR)$$

In this measure, the rates for all ages are given equal weights.

The ASFMR can be used to construct a nuptiality table, which is similar to the life table discussed in Chapter VI. From such a table, one can determine what proportion of a cohort of single persons would be married at various ages, assuming that the marriage rates used in constructing the table continued to prevail. It is also possible to construct a table taking both marriage and mortality into account. This type of table also indicates the pace at which the number of single persons is decreased by marriage and death. Information on the average age at marriage, proportion of single persons who remain single at each age and the proportion of persons who will eventually marry can be obtained from this table.

In general, the analysis of marriage is entirely different from the analysis of deaths and births. Deaths and births are biological events; on the other hand, marriage is an event which is affected by social customs. Whereas death is not a renewable event and all individuals have to face it, marriage is a renewable event and any individual can live without marrying. Thus the analyst should be cautious about the definitions of marital status and the data available for analysis.

8.6 Mean Age at Marriage

If data are available on the distribution of marriages of persons by single-year age, it is easy to calculate the mean age at marriage. This is one of the most widely used measures for comparing the marriage pattern of different populations. It is also important to measure the mean age at marriage of different orders of marriages. Generally, the age at first marriage is used in nuptiality analysis.

8.7 Singulate Mean Age at Marriage (SMAM)

In most populations of developing countries, information on age at marriage of persons getting married in a year or a period of time is often not available, and if available, grossly incomplete and poor in quality. In such situations the method adopted by Hajnal (1953) has been used to estimate the average age at marriage provided data on persons remaining unmarried at each age are available. The measure obtained by this method is called *singulate mean age at marriage* (SMAM). Subsequently, several other indices for measuring the mean age at marriage indirectly have been derived by researchers. This method, however, is the one which is widely used in the literature on marriages. This methodology generally ignores the effect of migration and mortality. The SMAM is based on the fact that the average age at marriage is actually the same as the average number of years lived in the single state by women who ever-marry and on the assumption that the experiences of a cohort of women are the same as those observed in the synthetic cohort.

NOTES

If S(x) is the proportion of women single at age x and if 'd' and 'D' are the lower and upper ages at which marriages occur then it can be proved that

$$SMAM = d + \frac{(d-D)S(D) + \sum S(x)}{1 - S(D)}$$

And if age intervals of n years are taken, then

$$SMAM = d + \frac{1}{1 - S(D)}\left[n \,.\, \sum {}_nS_x - (D-d)S(D)\right]$$

where S(D) is approximated by

$$\frac{{}_nS_D + {}_nS_{D-n}}{2}$$

Using the data given in Table 8.1, the SMAM is computed as follows:

1. First calculate the years of singleness lived by 100 persons up to age 15:
   Single years up to age 15 = 15 × 100 = 1500
2. Then compute single years lived between 15 and 50 years of age (using proportion single given in Table 8.1)
   (86.74 + 41.61 + 12.97 + 5.75 + 3.89 + 3.50 + 3.16) × 5 = 788.1
   Total years of singleness lived before = 788.1
3. Total years of single life lived by 100 women = 2288.1

Table 8.1: Computation of SMAM for Kerala, Females, 2001

Age Group	Females	Singles	Single (%)
15-19	1499920	1301066	86.74
20-24	1543523	642194	41.61
25-29	1489290	193100	12.97
30-34	1330656	76570	5.75
35-39	1311576	50966	3.89
40-44	990887	34665	3.50
45-49	974123	30788	3.16
50-54	712819	24518	3.44

NOTES

4. Now, calculate the percentage still single at the exact age of 50. This is usually the average of single in age group 45–49 and 50–54
   Percentage single at the exact age of 50 = (3.16 + 3.44) / 2 = 3.3
5. The total number of years of singleness lived before age 50, as calculated above, includes some that were experienced by those never marrying. These thus have to be calculated and subtracted from the total.
   The single-years lived by those not marrying by exact age 50 = 3.3 × 50 = 165
6. Total years of singleness lived by those who married before age 50 = 2288.1 – 165 = 2123.1
7. SMAM = 2123.1 / (100–3.3) = 21.96

Exercises

Exercise 8.1: Computation of SMAM for Rajasthan for Males and Females, 2001

Data for computation of SMAM

Age	Percentage Singles	
	Males	Females
15-19	85.12	58.83
20-24	43.66	10.06
25-29	12.67	1.40
30-34	4.00	0.55
35-39	2.29	0.42
40-44	1.88	0.31
45-49	1.62	0.23
50-54	1.76	0.30

CHAPTER IX

Standardization of Rates and Ratios

9.1 Introduction

One of the most important problems in demography is comparison of rates and ratios among different populations or for the same population over time. Such comparisons are essential for an understanding of underlying factors contributing to demographic differences between populations and in the same population over time. In such situations, inferences based on a comparison of the differences between two crude rates, such as CDRs, can be misleading, as the changes between the rates are brought about by several factors that may confound the main concerns or causative factors underlying the differences. Broadly speaking, the factors contributing to differences between two crude rates, such as the CDRs, could be classified into two groups: those arising from changes in the composition of the population (i.e. the differences in the size of the subgroups of the population) and those arising out of the changes in the specific rates. If the populations are not similarly constituted with regard to age distributions, direct comparison of the rates may be misleading. The approach which controls the variation in one of the factors extraneous for the purpose of comparison of rates or ratios is called the technique of **standardization**. The index arrived at by standardization is called a *standardized rate*. If the variations in the age composition are controlled, then the rate is called *age-standardized rate*. If more than one factor is controlled, say, age and sex, then it would be defined as *age-sex standardized rate*.

Generally two types of standardization are used in demographic literature: *direct* and *indirect* standardization.

NOTES

9.2 Direct Standardization

In direct standardization, age structure of a standard population is combined with observed age-specific rates for each population to be compared to obtain comparable estimates. The age-standardized rate (R^S) is obtained as follows:

$$R^S = \sum \left(C_i^S r_i \right)$$

where

C_i^S = proportion of the population at age i in the standard population

r_i = age-specific rate at age i in the population to be compared

If the difference between the age-standardized rate and the crude rate of standard population is greater than zero, then it is presumed that age-specific rates r_i are higher than r_i^S in their aggregate effect. If the first few subgroups show $r_i > r_i^S$ and $r_i < r_i^S$ for the rest, then the difference between standardized rate and the rate observed for the standard population may become negligible. Thus, standardization can mask the difference between age-specific rates. Standardized rates present only a general summary picture. However, the advantages of standardization often far outweigh its disadvantage, viz., that a single index is more convenient for comparison than a large set of specific rates. If adjustment is made for more than one variable, say for three, then direct standardized rate for A, B, C will be as follows:

$$R^S_{A,B,C} = \frac{\sum\left(P^S_{ijk} r_{ijk}\right)}{\sum\left(P^S_{ijk}\right)}$$

where

i, j, k = subgroups of the variable A, B, C

P^S = corresponding standard population

r_{ijk} = rate in the study population in the subgroup

An illustration of the direct standardized estimate of CDR for the state of Kerala taking the Indian population as the standard is given in Table 9.1. The CDRs of India and Kerala for 2001 were 8.4 and 6.6 respectively and the difference in death rate was 1.8 points. The direct standardized death rate for Kerala taking the age distribution of India as the standard works out to 4.87 and the gap between death rates of India and Kerala widens to 3.53. This is because Kerala's population has a higher proportion in old ages where the death rates are biologically higher and hence, without taking account of the differences in their age distributions, the differences are artificially narrowed. The result thus obtained suggests that comparing CDRs between India and Kerala may be misleading.

Table 9.1: Computation of Standardized CDR for Kerala, 2001 Using Direct Standardization Method

Age Group	Kerala		India	
	Total Population	ASDR (per 1000)	Total Population (SP)	Expected Number of Deaths (col.3 × col.4/1000)
1	2	3	4	5
0-4	2236616	2.60	110741993	287929
5-9	2546296	0.57	128659320	73336
10-14	2989684	0.38	125180125	47568
15-19	2986988	0.47	100483407	47227
20-24	2986471	0.95	90003749	85504
25-29	2788512	1.42	83645082	118776
30-34	2518555	1.42	74472312	105751
35-39	2468405	1.98	70762476	140110

Contd...

Age Group	Kerala		India	
	Total Population	ASDR (per 1000)	Total Population (SP)	Expected Number of Deaths (col.3 × col.4/1000)
1	2	3	4	5
40-44	1952906	3.50	55887085	195605
45-49	1927746	3.50	47535530	166374
50-54	1438715	7.09	36685226	260098
55-59	1131184	9.55	27727165	264794
60-64	1032994	15.88	27590233	438133
65-69	902765	24.76	19859828	491729
70-74	613932	43.86	14747907	646843
75-79	399421	66.82	6568713	438921
80-84	217403	123.44	4578122	565123
85+	171934	182.79	3482054	636485
Total	31841374	6.60	1028610329	5010306
CDR DCDR	6.60 4.87		8.4	

Note: DCDR is computed as the total number of expected deaths divided by the total standard population. India's population is used as the standard

9.3 Indirect Standardization

Suppose that for the standard populations of which we want to compare rates, we have no information on the age-specific rates but have information on crude rate and age distribution for the event under consideration and we do not have data on their age distribution or on the distribution of the extraneous factor that we have to control, then the technique used for standardization is called *indirect standardization*. In indirect standardization, the actual age structure of the study population to be compared is combined with the standard age-specific rates to estimate an expected number of events, which is compared with the observed number of events to provide a standard index. Then, the indirectly standardized rates for the population to be compared are obtained by multiplying the crude rate in the standard population by the standard index.

The indirect standardized estimate of the CDR for Kerala using the Indian ASDR as the standard is given in Table 9.2. The results suggest that the indirect estimate of CDR is 5.64 while the observed rate is 6.60.

Measures such as the Coale's index and TFR discussed in Chapter V can be considered as standardized rates. Coale's indices are obviously indirectly standardized rates, using the age-specific fertility schedules for the Hutterite population. Similarly, the TFR is also considered a standardized rate, adjusted for age by directly standardizing on a population with equal number of women in each age group. Similarly, life table estimates are also the results of direct standardization and adjusted for both age and sex of the population under consideration.

Table 9.2: Computation of Standardized CDR for Kerala, 2001 Using Indirect Standardization Method

Age Group	Kerala		India		
	Total Population	ASDR (per 1000)	Number of Deaths	ASDR (per 1000)	Expected Number of Deaths (col.5×col.2)/1000
1	2	3	4	5	6
0-4	2236616	2.60	5815	19.3	43167
5-9	2546296	0.57	1451	1.9	4838
10-14	2989684	0.38	1136	1.3	3887
15-19	2986988	0.47	1404	1.7	5078
20-24	2986471	0.95	2837	2.3	6869
25-29	2788512	1.42	3960	2.7	7529
30-34	2518555	1.42	3576	2.9	7304
35-39	2468405	1.98	4887	3.6	8886
40-44	1952906	3.50	6835	4.4	8593
45-49	1927746	3.50	6747	6.2	11952
50-54	1438715	7.09	10200	9.9	14243
55-59	1131184	9.55	10803	15.9	17986
60-64	1032994	15.88	16404	22.3	23036
65-69	902765	24.76	22352	38.6	34847
70-74	613932	43.86	26927	51.3	31495
75-79	399421	66.82	26689	81.6	32593
80-84	217403	123.44	26836	99.5	21632
85+	171934	182.79	31428	171.5	29487
Total	31841374		210290		313419
CDR	6.60			8.4	
ICDR	5.64				

Note: ICDR is computed by dividing the total number of deaths experienced in the given population by the expected number of deaths if the schedule of specific rates of standard prevails in the given population, and then multiplying the result thus obtained by the CDR of the standard population. India's ASDR is used as the standard.

The problem that would naturally be faced by the analyst is the choice of the standard population. In fact, there is no population that can be considered as ideal for a 'standard'. The reason is that no two populations can be expected to have an identical population composition in a real situation. Thus the selection of a standard population is arbitrary. When the rates of two populations are to be compared, it is also customary to consider one of them as the standard population. Sometimes the larger population of which the two populations to be compared forms subgroups, like states of a country, could also be considered as standard.

9.4 Decomposition of Rates

One can decompose the differences between crude rates of two populations or a population at different time periods by an extension of the method of standardization. This method can be useful in estimating the extent to which the differences in the rates and proportions are attributable to different factors influencing the composition of the population. As an illustration the decomposition of CBR is discussed in this section.

The first assumption the researcher makes when performing this procedure relates to the additivity of the results of contributory factors on the dependent variable. The second assumption is that of the functional independence of components of CBR. Another assumption in this specific case is that illegitimate fertility is negligible. CBR can be described as

$$CBR = \frac{\left[\sum(A_i M_i F_i)\right] \times W}{P}$$

where

- W = the number of women in reproductive ages
- P = the total population
- A_i = proportion of women in age group i among all women of reproductive ages
- M_i = proportion of married women of age group i among all women in the age group
- F_i = age-specific marital fertility rate for age group i

The change in CBR between two time periods of a population, time 1 and time 2, which has to be decomposed, can be expressed as a sum of the four components of CBR in the above model. The four component factors are proportion of women in reproductive ages (W/P); proportion of women in a given age i within the reproductive ages, 'A_i'; proportion married in each age i 'M_i' and marital fertility rate 'F_i'. The effect of each factor is estimated by changing only that factor and keeping the others constant at the base value. For example, the effect of change in the proportion of women in reproductive ages W/P is obtained by

$$d(W) = \sum(A1_i\, M1_i\, F1_i)\,(W2/P2 - W1/P1) \qquad \text{...(1)}$$

The effect of age distribution within the reproductive ages A can be estimated by keeping this factor alone as a variable between two points of time as observed and keeping the other two factors, proportion of women in the reproductive ages and the proportion married at the two points of time, unchanged, at the levels observed at the first point of time.

Thus, the effect due to age factor is derived as follows

Effect due to age factor = $(\Sigma A2_i\, M1_i\, F1_i - \Sigma A1_i\, M1_i\, F1_i)\, W1/P1$

or $\Sigma\, ((A2_i - A1_i)\, M1_i\, F1_i)$ ...(2)

Similarly effect due to marital structure factor is

$\Sigma\, (A1_i\, (M2_i - M1_i)\, F1_i)$ ...(3)

and due to marital fertility factor is

$\sum\, (A1_i\, M1_i\, (F2_i - F1_i)\,)$ ...(4)

where

W1 = number of women in reproductive ages of 15–49 at time 1

P1 = total population at time 1

$A1_i$ = proportion of women in i^{th} age group to all women in the reproductive ages of 15–49 at time 1

$A2_i$ = proportion of women in i^{th} age group to all women in the reproductive ages of 15–49 at time 2

$M1_i$ = proportion of married women in i^{th} age group at time 1

$M2_i$ = proportion of married women in i^{th} age group at time 2

$F1_i$ = age-specific marital fertility rate of i^{th} age group at time 1

$F2_i$ = age-specific marital fertility rate of i^{th} age group at time 2

The total change in fertility, or CBR, between the two points of time is given by

$$d(CBR) = dW + dA + dM + dF$$

where

d(CBR) = total difference in CBR between two time periods

dW = difference in CBR due to change in the proportion of women in reproductive ages

dA = difference attributable to age structure within reproductive ages between two time periods

dM = difference attributable to marriage composition between two time periods

dF = difference attributable to specific marital fertility rates between two time periods

This relationship is obtained by taking the population A1 in the first period as the standard and the assumptions stated above. Theoretically, the sum of the contribution of the components should equal the total change as resulting from CBR2 – CBR1. This might not be the case partly because of cumulative small rounding off errors but mostly due to the presence of joint effects of two or more of these factors which are not taken into account in the above model. These joint effects are also called *interactions*.

The decomposition of changes in CBR and GFR between 1991 and 2001 due to the changes in the four factors viz., proportion of women in reproductive ages, age structure within reproductive ages, marital status and marital fertility of women in the reproductive ages during the decade 1991 and 2001 in Kerala are provided as an example. The decomposition is done first by taking 1991 as standard (Tables 9.4, 9.5 and 9.6) and later 2001 as standard (Tables 9.7, 9.8 and 9.9). The final results of the decomposition are provided in Table 9.10.

Table 9.3 provides the input data used for the decomposition. Tables 9.4 through 9.9 provide the computations on the effects of various said factors on the changes in the CBR by taking the population characteristics of 1991 and 2001. The birth rates and general fertility rates for Kerala in the two years 1991 and 2001 were obtained by applying the ASMFR estimated by SRS for the same year to the age–sex–marital status distributions of the corresponding census populations. The differences between SRS rates and the rates obtained by this procedure indicate the differences that exist between SRS age–sex–marital status distribution of the populations surveyed by the SRS and the census populations.

Table 9.3: Input Data for the Decomposition of CBR and GFR for Kerala during the Period 1991–2001

Age Group	1991			2001		
	Number of Women	Percentage Married (%)	ASMFR (per 1000 Women)	Number of Women	Percentage Married (%)	ASMFR (per 1000 Women)
1	2	3	4	5	6	7
15-19	1593505	11.25	259.2	1499920	13.06	187.59
20-24	1602390	55.50	269.0	1543523	57.41	252.39
25-29	1390614	83.55	150.9	1489290	84.91	156.76
30-34	1127005	89.34	52.7	1330656	90.29	53.05
35-39	967062	89.05	16.9	1311576	89.68	12.82
40-44	783424	85.14	5.1	990887	86.49	1.97
45-49	644214	80.92	0.9	974123	81.72	0.37
Total	8108214 = W1			9139975 = W2		
	1991 Total Population P1 = 29098518 W1/P1 = 0.2786 CBR = 18.30 GFR = 65.67			**2001** Total Population P2 = 31841374 W2/P2 = 0.2870 CBR = 16.94 GFR = 59.02		

Total observed change in CBR between 1991 and 2001 = 18.30 – 16.94 = –1.36

The extent of CBR decline explained by the effect due to changes in the proportion of women in reproductive ages in 1991 is computed by using the formula

$$\left(\frac{W2}{P2} - \frac{W1}{P1}\right) \cdot \sum \left(A1_i \cdot M1_i \cdot F1_i\right)$$

In this equation, the second factor is GFR at time 1 which from Table 9.3 can be seen to be 65.67. Thus the effect of this factor works out to (0.2870 – 0.2786) × 65.67 = 0.5516.

If population P_2 (2001) had been chosen as the standard, the formula would be

$$\left(\frac{W2}{P2}-\frac{W1}{P1}\right)\cdot\sum\left(A2_i \cdot M2_i \cdot F2_i\right)$$

and the second factor in the equation is GFR at time 2 which is 59.02 from Table 9.3 and this value works out to (0.2870 – 0.2786) × 59.02 = 0.4958.

Similarly, the changes in the CBR attributable to the factors A, M and F taking 1991 as the base is computed in Tables 9.4, 9.5 and 9.6 respectively and taking 2001 as base in Tables 9.7, 9.8 and 9.9, respectively. The figures are given in Table 9.10. The percentage changes (relative change in CBR) are computed relative to the observed change in CBR, not to explained change.

In Kerala, between 1991 and 2001 the CBR has declined from 18.30 to 16.94 or just 1.36 points. Two of the four factors – (W/P), proportion of women in the reproductive ages and M, the marital structure – contributed significantly to rise in the CBR value during this period by 0.55 and 0.66 points respectively. The two other factors A, the proportion of women in different age groups within the reproductive ages and F, the marital fertility rates have contributed significantly to declines in the CBR values by –1.55 and –0.91 points respectively. This led to a net decline of 1.25 points in the CBR value during 1991 and 2001. The actual decline was 1.36 points and the unexplained portion of 0.11 points in the CBR can be attributed to the effects of interaction among the four factors. It is gratifying to note that we were able to explain about 91% of the decline in CBR due to these four factors and factoring the effect of each factor.

Table 9.4: Computation of the Role of Age Structure in Changes in CBR and GFR, Kerala (Base Population Year 1991)

Age Group	Age Distribution of Women Aged 15-49, 1991, A_1 (%)	Age Distribution of Women Aged 15-49, 2001, A_2 (%)	Change in Age Structure A_2-A_1	Married Women 1991, M_1 (%)	Marital Fertility 1991 F (per 1000)	Change Due to Age Structure (col.4 × col.5 × col.6)/10000 (per 1000)
1	2	3	4	5	6	7
15-19	19.653	16.411	–3.242	11.25	259.2	–0.9455
20-24	19.763	16.888	–2.875	55.50	269.0	–4.2921
25-29	17.151	16.294	–0.856	83.55	150.9	–1.0798
30-34	13.900	14.559	0.659	89.34	52.7	0.3103
35-39	11.927	14.350	2.423	89.05	16.9	0.3646
40-44	9.662	10.841	1.179	85.14	5.1	0.0512
45-49	7.945	10.658	2.713	80.92	0.9	0.0198
Total						–5.5715
Change in GFR due to change in age structure = (Sum of col. 7) = 5.571						
Change in CBR due to change in age structure = (Sum of col. 7 × (W1/P1) = –1.552						

Table 9.5: Computation of the Role of Marital Status in Changes in CBR and GFR, Kerala (Base Population Year 1991)

Age Group	Married Women 1991 M_1 (%)	Married Women 2001 M_2(%)	Change in Marital Structure M_2–M_1	Age Distribution of Women Aged 15–49, 1991, A_1 (%)	Marital Fertility 1991 F_1 (per 1000)	Change Due to Marital Structure (col.4 × col.5 × col.6)/10000 (per 1000)
1	2	3	4	5	6	7
15-19	11.25	13.06	1.81	19.653	259.2	0.922023
20-24	55.50	57.41	1.91	19.763	269.0	1.015380
25-29	83.55	84.91	1.36	17.151	150.9	0.351973
30-34	89.34	90.29	0.95	13.900	52.7	0.069588
35-39	89.05	89.68	0.63	11.927	16.9	0.012699
40-44	85.14	86.49	1.35	9.662	5.1	0.006652
45-49	80.92	81.72	0.80	7.945	0.9	0.000572
Total						2.378888
Change in GFR due to change in marital structure = (Sum of col. 7) = 2.379						
Change in CBR due to change in marital structure = (Sum of col. 7) × (W1/P1) = 0.663						

Table 9.6: Computation of the Role of Marital Fertility in Changes in CBR and GFR, Kerala (Base Population Year 1991)

Age Group	Marital Fertility 1991 F_1 (per 1000)	Marital Fertility 2001 F_2 (per 1000)	Change in Marital Fertility F_2–F_1	Age Distribution of Women Aged 15–49, 1991, A_1 (%)	Married Women 1991 M_1 (%)	Change Due to Age Structure (col.4 × col.5 × col.6)/10000 (per 1000)
1	2	3	4	5	6	7
15-19	259.2	187.59	–71.61	19.653	11.25	–1.5833
20-24	269.0	252.39	–16.61	19.763	55.50	–1.8218
25-29	150.9	156.76	5.86	17.151	83.55	0.8397
30-34	52.7	53.05	0.35	13.900	89.34	0.0435
35-39	16.9	12.82	–4.08	11.927	89.05	–0.4333
40-44	5.1	1.97	–3.13	9.662	85.14	–0.2575
45-49	0.9	0.37	–0.53	7.945	80.92	–0.0341
Total						–3.2468
Change in GFR due to change in marital fertility = (Sum of col. 7) = –3.247						
Change in CBR due to change in marital fertility = (Sum of col. 7/1000) × (W1/P1) = –0.905						

Table 9.7: Computation of the Role of Age Structure in Changes in CBR and GFR, Kerala (Base Population Year 2001)

Age Group	Age Distribution of Women Aged 15-49, 1991, A_1 (%)	Age Distribution of Women Aged 15-49, 2001, A_2 (%)	Change in Age Structure $A_2–A_1$	Married Women 2001 M_2 (%)	Marital Fertility 2001 F_2 (per 1000)	Change Due to Age Structure (col.4 × col.5 × col.6)/10000 (per 1000)
1	2	3	4	5	6	7
15-19	19.653	16.411	–3.242	13.06	187.59	–0.7944
20-24	19.763	16.888	–2.875	57.41	252.39	–4.1657
25-29	17.151	16.294	–0.856	84.91	156.76	–1.1400
30-34	13.900	14.559	0.659	90.29	53.05	0.3157
35-39	11.927	14.350	2.423	89.68	12.82	0.2786
40-44	9.662	10.841	1.179	86.49	1.97	0.0201
45-49	7.945	10.658	2.713	81.72	0.37	0.0082
Total						–5.4775
Change in GFR due to change in age structure = (Sum of col. 7) = –5.477						
Change in CBR due to change in age structure = (Sum of col. 7) × (W2/P2) = –1.572						

Table 9.8: Computation of the Role of Marital Status in Changes in CBR and GFR, Kerala (Base Population Year 2001)

Age Group	Married Women 1991 M_1 (%)	Married Women 2001 M_2 (%)	Change in Marital Structure $M_2 – M_1$	Age Distribution of Women Aged 15–49, 2001 A_2 (%)	Marital Fertility 2001 F_2 (per 1000)	Change Due to Marital Structure (col.4 × col.5 × col.6)/10000 (per 1000)
1	2	3	4	5	6	7
15-19	11.25	13.06	1.81	16.411	187.59	0.557200
20-24	55.50	57.41	1.91	16.888	252.39	0.814092
25-29	83.55	84.91	1.36	16.294	156.76	0.347383
30-34	89.34	90.29	0.95	14.559	53.05	0.073372
35-39	89.05	89.68	0.63	14.350	12.82	0.011590
40-44	85.14	86.49	1.35	10.841	1.97	0.002883
45-49	80.92	81.72	0.80	10.658	0.37	0.000315
Total						1.806836
Change in GFR due to change in marital fertility = (Sum of col. 7) = 1.807						
Change in CBR due to change in marital fertility = (Sum of col. 7 × (W2/P2) = 0.519						

Table 9.9: Computation of the Role of Marital Fertility in Changes in CBR and GFR, Kerala (Base Population Year 2001)

Age Group	Marital Fertility 1991 F_1 (per 1000)	Marital Fertility 2001 F_2 (per 1000)	Change in Marital Fertility F_2-F_1	Age Distribution of Women Aged 15-49, 2001 A_2 (%)	Married Women 2001 M_2 (%)	Change Due to Age Structure (col.4 × col.5 col.6)/10000 (per 1000)
1	2	3	4	5	6	7
15-19	259.2	187.59	–71.61	16.411	13.06	–1.5348
20-24	269.0	252.39	–16.61	16.888	57.41	–1.6104
25-29	150.9	156.76	5.86	16.294	84.91	0.8108
30-34	52.7	53.05	0.35	14.559	90.29	0.0460
35-39	16.9	12.82	–4.08	14.350	89.68	–0.5251
40-44	5.1	1.97	–3.13	10.841	86.49	–0.2935
45-49	0.9	0.37	–0.53	10.658	81.72	–0.0462
Total						–3.1531
Change in GFR due to change in marital fertility = (Sum of col. 7) = –3.153						
Change in CBR due to change in marital fertility = (Sum of col. 7) × (W2/P2) = –0.905						

The effects of different factors are summarized in Table 9.10.

Table 9.10: Extent of Changes in CBR Explained by Different Factors

	Change in CBR between 1991 and 2001	–1.36 with 1991 as base	–1.36 with 2001 as base
1	Change due to proportion of women in reproductive ages (W)	0.55	0.50
2	Age structure within reproductive ages (A)	–1.55	–1.56
3	Marital structure (M)	0.66	0.52
4	Marital Fertility (F)	–0.91	–0.91
	Total explained	–1.25	–1.35

An interesting point emerges from the above analysis. Though the actual difference in CBR between 1991 and 2001 to be explained is small (only 1.36 points), the effect of changes in the proportion of women in the reproductive ages (W) is to increase the CBR by 0.55 points. On the other hand, the effect of changes in marital fertility is negative, –0.905. When fertility reaches a low level two opposing forces come into play on the CBR: the effects of a decline in ASFR that contributes to a decline in CBR and an opposite effect of an increase in the proportion of women in reproductive ages contributing to an increase in the CBR. The second factor is the momentum factor on fertility discussed in Chapter II and the above results are an empirical demonstration of the effects of **demographic momentum** wherein increase in the proportion of women in reproductive ages could even contribute an increase in CBR if not compensated by higher declines in MFR and increases in age at marriage. It is also interesting to note that the substantive findings are not altered whether we take 1991 as the base or 2001 as the base.

Mari Bhat (2006) estimated the magnitude of the demographic dividend on the per capita income of the country at the national level and at the state level using the above described method of decomposition. He argued that the per capita income rises just because of the higher growth rates of the working age population compared to the total population even if the productivity per worker remains the same. This is a result of the savings rate at the household level, declines in fertility and consequent rise in investments. It is also due to greater participation of women in the labour force as a result of their spending less time in reproduction and rearing of children and also due to a possible rise in productivity because of the increase in the educational levels of the population.

If P is the total population, P(w) is the population in working ages and Y is the total economic output then the per capita output denoted by y is

$$y = Y/P = Y/P(w) \cdot P(w)/P$$

Multiplying and dividing by P(w) on the right hand side,

$$y = g \cdot (P(w)/(P))$$

The first factor on the right-hand side, g, is the productivity per worker and the second factor is the proportion of the working age population to the total. Taking logarithms and differentiating the above with time we get

$$dy/y = dg/g + (dP(w)/P(w) - dP/P)$$

$$g(1) = g(2) + g(3) - g(4)$$

where

g(1) = rate of change in the per capita income

g(2) = rate of change of output per worker

g(3) = growth rate of working age population

g(4) = growth rate of population as a whole

Hence the demographic dividend can be considered as the rate of growth of workers minus the rate of growth of the population as a whole.

Exercises

Exercise 9.1: Compute direct standardized estimate of death rate for Rajasthan using the data given below for 2001

Age Group	Rajasthan		India
	Total Population	ASDR (per 1000)	Total Population
1	2	3	4
0-4	6012258	24.3	110741993
5-9	8068070	1.7	128659320
10-14	7241941	1.1	125180125
15-19	5490214	1.3	100483407
20-24	4749366	2.0	90003749
25-29	4197223	2.2	83645082
30-34	3821151	2.9	74472312
35-39	3453860	2.5	70762476
40-44	2760275	3.6	55887085
45-49	2269651	4.2	47535530
50-54	1775046	8.6	36685226
55-59	1349315	14.0	27727165
60-64	1361274	19.2	27590233
65-69	982874	28.6	19859828
70-74	757998	51.0	14747907
75-79	315930	63.0	6568713
80-84	237485	92.0	4578122
85+	154711	150.7	3482054
Total	54998642		1028610329
CDR		7.7	

Exercise 9.2: Compute indirect standardized estimate of death rate for Rajasthan, 2001 using the given data

Age Group	Rajasthan		India ASDR (per 1000)
	Total Population	ASDR (per 1000)	
1	2	3	4
0-4	6012258	24.3	19.3
5-9	8068070	1.7	1.9
10-14	7241941	1.1	1.3
15-19	5490214	1.3	1.7

Contd...

Age Group	Rajasthan		India
	Total Population	ASDR (per 1000)	ASDR (per 1000)
1	2	3	4
20-24	4749366	2.0	2.3
25-29	4197223	2.2	2.7
30-34	3821151	2.9	2.9
35-39	3453860	2.5	3.6
40-44	2760275	3.6	4.4
45-49	2269651	4.2	6.2
50-54	1775046	8.6	9.9
55-59	1349315	14.0	15.9
60-64	1361274	19.2	22.3
65-69	982874	28.6	38.6
70-74	757998	51.0	51.3
75-79	315930	63.0	81.6
80-84	237485	92.0	99.5
85+	154711	150.7	171.5
Total			54998642
CDR	7.7		8.4

Exercise 9.3: Decompose GFR and CBR for Rajasthan with the given data set.

Input data for the decomposition of CBR and GFR for Kerala during the period 1991-2001

Age Group	1991			2001		
	Number of Women	Percentage Married (%)	ASMFR (per 1000 Women)	Number of Women	Percentage Married (%)	ASMFR (per 1000 Women)
1	2	3	4	5	6	7
15-19	1593505	11.25	259.2	1499920	13.06	187.59
20-24	1602390	55.5	269.0	1543523	57.41	252.39
25-29	1390614	83.55	150.9	1489290	84.91	156.76
30-34	1127005	89.34	52.7	1330656	90.29	53.05
35-39	967062	89.05	16.9	1311576	89.68	12.82
40-44	783424	85.14	5.1	990887	86.49	1.97
45-49	644214	80.92	0.9	974123	81.72	0.37
Total	8108214 = W1			9139975 = W2		
	1991 Total Population P_1 = 29098518 CBR = 18.30 GFR = 65.67			**2001** Total Population P_2 = 31841374 CBR = 17.30 GFR = 59.20		

CHAPTER X

Population Projections

10.1 Introduction

The topic of population projections deals with computations of future population size and their characteristics based on a knowledge of the past trends and realistic assumptions about the future trends in the components of its change, fertility, mortality and migration. Since no one can predict the future trends in fertility, mortality and migration with 100% accuracy, it is also not possible to predict the future size and characteristics of a population accurately. Projections are merely intelligent exercises as to what would happen to the current population under specified assumptions of fertility, mortality and migration in future years. In this context, it is important to distinguish between the terms 'forecast' and 'projections'. **Forecast** has an element of prediction into the near future using the current data and most plausible assumptions. For instance, we have forecasts for the weather conditions in an area. They are the best estimates of what would happen in the immediate future with regard to rainfall, weather and humidity conditions in an area. On the other hand, **projections** are usually carried out based on a number of alternative assumptions, which are all considered in the realm of possibility. In practice, the distinction between forecasts and projections is often with regard to the probability of the assumptions coming true and hence containing an element of arbitrariness.

The computation of any measure related to the demographic process from the available data for a time period for which it is not available in the past is called *demographic estimation*. Estimation is a term that refers to a past date while projection refers to a future date. The reliability of the projected values will generally depend upon the validity of the assumptions used and the accuracy with which the assumptions are translated into quantitative terms. Hence, the insight the researcher has of the population he/she is dealing with is as important for population projection as the technique of projection involved. The formulation of a set of realistic assumptions regarding the future trajectory of a population as well as of the rates that would determine its growth and changes are important issues in the projection exercise. Population projections produce the probable size and composition of the future population and there is no means of checking their credibility until the actual counts at a later date are available. In this chapter commonly used techniques are discussed and only certain guidelines used in population projection are provided without details on the mathematics of the computations.

NOTES

10.2 Uses of Projections

The topic of population projections has in the recent years assumed enormous national and global significance. The rapid dwindling of the earth's non-renewable natural resources, the exponentially rapid phase with which these resources are getting exploited with the acceleration of industrial development in large developing countries and the aspirations of all to achieve high material standards have once again brought into focus the population size in different countries. The main cause of global warming is the various human activities aimed at rapid but unsustainable industrial development. Population projections are needed not only to generate concern for the demographic future and for generating plans for development but also to stimulate decision makers to formulate appropriate policies and programmes designed to alter the rate of population growth and its distribution to facilitate sustainable socio-economic development. Population projections can also bring to focus the structural consequences of socio-economic development.

NOTES

Thus projections are used in two very different types of planning: first, in the planning of various services of education, health and employment needed for different areas in future years in the context of projections made and second, in altering these trends by appropriate programmes of family planning, contraceptive services, laws on age at marriage, abortions, migration etc., to have sustainable development. Thus, population projections serve the dual function of assisting the policy makers in formulating policies which serve as the basis for development planning and also help in the planning for altering the course of population trends in the desired direction. The impact of efforts to influence population trends may be felt only after some time. Many developing countries such as China, India, Indonesia and Singapore have used the results of their population projection exercises to a considerable extent in the formulation of population policies, such as the one-child family in China, realizing the sterilization targets necessary to achieve the two-child norm in India, provision of subsidized housing to encourage small families in Singapore in the 1960s and monitoring the use of contraception by couples in Indonesia. The government and segments of the public usually evaluate the current and probable future growth and distribution patterns of their populations in order to decide whether to attempt particular modifications of the assumed trends via population policies.

In India, one important criterion for the allocation of central finances to the states is the total population of the state. Similarly, the number of seats in parliament is decided on the basis of population. The same is true at state assembly, district and even panchayat levels. In a democratic nation, the political strength of a community is determined by the number of voters. There is, therefore, a fear that some communities may multiply faster and so increase their political strength as opposed to other communities. Hence projections, by religious groups, racial communities and other groups are undertaken in order to make an objective assessment of the problem. Another important factor in Indian politics is that a majority of the political parties choose their candidate according to the strength of the caste in a particular area. Thus, population projections by community, and if possible, by caste, are also important for political reasons. The recently conducted caste-census in India may help in this type of projection.

A minimum requirement for the development planning process is the projection of the size and age structure of the population. All development plans, economic or social, imply a judgement about the size and characteristics of the future population. For estimating the future production of goods and services, it is necessary to have a projection of the labour force, which also enables the establishing of employment targets. The changing levels of education, experience and skills of the workforce can be obtained from an educational projection. The requirements of future school enrolments can be obtained from population projections by age. This will enable planning of investments in school buildings, teacher training and production of educational material. Projections of the future age and sex composition of the population are essential for estimating the incidence and prevalence of various diseases and planning for the number of hospitals, hospital beds and specialized facilities, as well as training programmes for medical specialists, paramedical and auxiliary personnel. The probable surpluses and shortages of different types of labour can be assessed by comparing the projections of labour demand with that of manpower in the occupational and industrial sector, by sex, age and educational level. Employment and educational programmes can thus plan for necessary modifications. The future consumption requirement of various goods and services can be obtained from the projection of population by age and sex in different socio-economic strata, presuming that consumption patterns vary according to socio-economic conditions. For example, if the proportion of population under age 6 increases, this will imply an increased demand for products for babies and children, such as baby foods, children's clothes and toys. The future food requirements also depend on the future age–sex structure of the population and will serve as a basis for investment in agriculture and related sectors. The rate of growth of urban and rural areas, in terms of population, households and family-size composition are essential to plan for the general amenities needed, such as water supply, electricity and sanitation. The provision of various social services requires projections of age–sex composition as well as the size of future population. Projections are also widely used in the fields of law and order, energy allocation and market research. For reducing regional disparities within a country, it is essential to have sub-national projections. They are useful for planning transport, communications and electric power systems as well as for planning new industrial towns and rural communities. The implications of alternative assumptions of population projections on the population size and their social and economic implications can also be worked out.

NOTES

10.3 Techniques of Projection

10.3.1 Algebraic methods

In the 1920s and 1930s the techniques of population projections using mathematical curves were quite popular and were based on the assumption that population growth, in the long run, follows some mathematical laws. This was found to be true in the population sizes of different animal species. These methods assume a certain autonomy in population growth, some degree of determinism. Among the autonomous methods of projection, the curves used for fitting the past trends in population are the linear, exponential, logistic and the Gompertz curves. The simplest method is to relate the future population with the current population, algebraically, through a mathematical function.

Algebraically, the population at time t+1 can be considered as a function of population at time t, that is,

$$P_{t+1} = f(P_t)$$

The mathematical function used can range from the very simple to the very complex. If we use mathematical relationships to make projections it is always better to fit a simple descriptive model to the existing data and to use the model to estimate the future population for short durations of time in the future. Population is considered a function of time. The simplest mathematical model is that which says that population grows arithmetically or geometrically. Thus the population at time t is arithmetically related with that at time '0' as under:

$$P_t = P_0 (1 + rt)$$

where t = difference between the two periods and r = annual rate of growth

NOTES

If we assume that the population grows geometrically, then

$$P_t = P_0 (1+r)^t$$

The most frequently used mathematical model assumes that population growth will follow an exponential distribution, which is a generalization of the geometric function when time t is considered to be a continuous variable, in which

$$P_t = P_0 e^{rt}$$

In all the above cases, the growth rate 'r' to be used in the future projections is estimated from the observed rates of growth in the recent past.

Fitting of more complex curves to the past data on population trends and using them to project the future sizes has been commonly used in the past. The most widely used is the logistic curve (Pearl and Reed, 1920). This method is adopted when observations have been made at equal intervals of time in the recent past and the projections assume that the population growth will be high in the initial periods when the mortality rates start declining and slow down later when fertility rates fall, the curve being of sigmoid shape.

This method has an underlying Malthusian logic. When the population size is small relative to the resources available it tends to increase rapidly but the growth rate will decline when the population size increases and the resources become limited. This is simply the resources putting a check on population growth. Ultimately, the population size stabilizes at a particular level when the growth rate becomes zero. The population size eventually reaches an upper limit or asymptotic value when there is a balance between its size and the resources available. This method, behind a logistic curve, has been found to give a good fit for populations growing under conditions of limited food and space but without any restrictions on reproduction, populations that can be categorized as 'Malthusian'. The curve is also found to fit remarkably well with the growth of organisms in culture, with limited availability of food or growth media. The curve can only be fitted over short intervals of time and gives erroneous results when fitted over long time periods. Although the logistic curve is no longer accepted as the best method of projection, it has important uses in demographic analysis and projections of the components of population growth, fertility, mortality and migration. In its simplest form the curve is defined as follows:

$$Y_t = \frac{1}{A + BC^t}$$

In a growing population, the curve is also defined as

$$\log \frac{K - P_t}{P_t} = a + bt$$

Or

$$\log \frac{1 - P_t / K}{P_t / K} = a + bt$$

where K = maximum value or asymptote reached by the population

When p is a proper fraction, log (1 – p)/p is called logit of p. In a logistic curve the logit of P_t/K is a linear function. Solving the above equation we get

$$P_t = \frac{K}{1 + e^{a+bt}}$$

However, though the above procedures are still used by researchers, they are not very useful to planners and policy makers. From a knowledge of the population size and using logit transformations we can estimate the parameters K, a and b and use the equation in the projections of future populations. The Gompertz curve is a double exponential curve with three parameters a, b, and k expressed as follows:

$$y = ae^{-be^{kt}}$$

The fitting of these curves to past population sizes serves a dual purpose: first to help estimate the population in the intermediate years between observed time points in the past and, second, to predict the population sizes in the future, using the curve and time as the predictor. We have tried to demonstrate the usefulness of these curves in predicting

NOTES

Table 10.1: Projected population (000's) in year 2011 by different curve fitting and deviation from the actual counts in Census 2011, Kerala

	Linear fit	Exponential	Logistic	Gompertz curve
Taking data from 1921-2001	34237	40242	41610	36838
Taking data from 1981-2001	35851	36844	36222	NA
Observed-expected (taking data from 1921-2001)	–849	–6854	–8222	–3450
Taking data from 1981-2001	–1463	–3456	–2834	NA

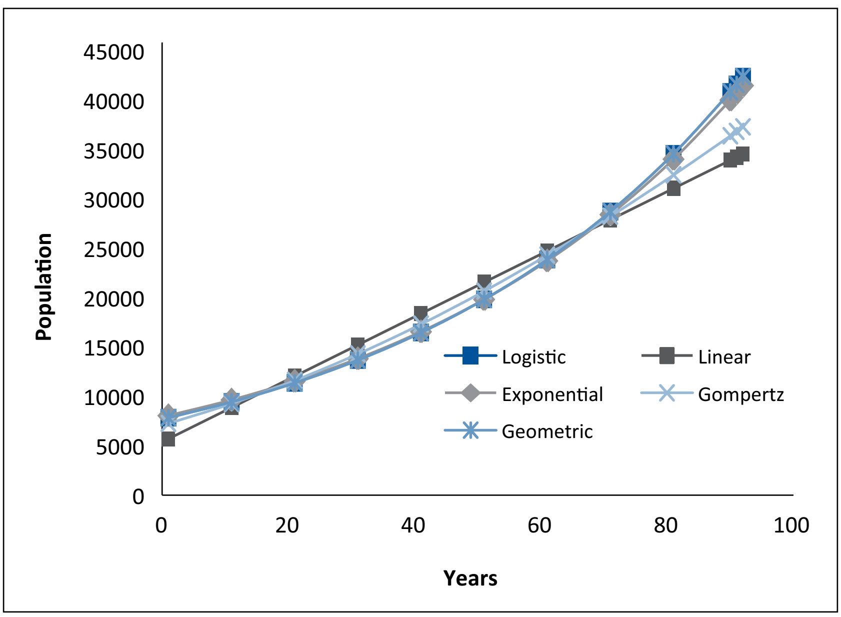
the population of Kerala for 2011 using the knowledge of population sizes in the census years 1921 to 2001 (nine points). The parameters of four curves fitted through these nine points by linear, exponential, logistic and Gompertz curves are given in Table 10.1. A graph of these four curves is given in Figure 10.1.

The population of Kerala as per the 2011 Census count (provisional figures) is 33,387,677 or 33,388 thousand while the predicted values are as given in Table 10.1 (The curve fitting through the nine points was done on the basis of the least square method, population in thousands).

NOTES

It can be seen that knowledge of the population sizes of Kerala even over an 80-year period (1921–2001) has not helped much to predict the population of the state for 2011 with any accuracy, even compared to knowledge from 1981. The techniques of using various mathematical formulae for projections, used widely in the nineteenth and early twentieth centuries, are no longer favoured by demographers, except in rare situations when only limited data on population size is available for the past periods. The use of mathematical curves highlights the point that population changes in the future years are more influenced by the current population size and recent trends in its components than its historical trends over a long period of time.

Figure 10.1: Different curve fitting on population of Kerala (1921–2011)



10.3.2 Component methods

The algebraic methods are, in a way, crude and illogical on two grounds. First, future population growth depends only on the current population size, its distribution and components of change viz., current levels and likely future trends in fertility, mortality and migration. It does not depend on what the population was in 1912 or 1951. To project for the future, the latest available data on population should be taken as the base. Second, the future changes in its components of fertility, mortality and migration can be more realistically assumed based on their immediate past trends and related policies.

Hence, since the 1950s, the component method has become the most commonly used method of population projection in all countries of the world. This method can be considered as an extension of the balancing equation discussed in Chapter II. Any population change between year t and t+1 can be broken down into its components as follows:

$$P_{t+1} = P_t + (\text{Births} - \text{Deaths}) + (\text{Immigrants} - \text{Emigrants})$$

NOTES

This seems theoretically and conceptually simple, but it is a very powerful tool and can become complex if this technique is used for different segments of the population, such as an age group, marital status category and a specified socio-economic stratum.

Component projection basically requires three things: first, a base population from which the projection starts which should be the most recent one having data on age–sex distributions of the population; second, a set of assumptions about the course of events on births, deaths and migration during the period covered by the projection; and third, a method by which the assumptions are applied to the base population.

The base population that is usually used for projections is the age–sex distribution of the population obtained from the most recent census. Though we may have estimates of the population for a more recent year, based on earlier projections, it is better to use the recent census population as the base because (a) any recent projections are only best estimates and have been obtained on the basis of assumptions, some of which may not be consistent with those to be made in the current projections and (b) census-based projections are universally accepted as a sound basis for planning and resource allocation purposes by all concerned since censuses have a statutory base in every country.

As indicated earlier, the components of population growth are fertility, mortality and migration. Thus, any projection of population implies separate projections of mortality, fertility and migration by age–sex groups. In addition to a set of assumptions on these components that are considered to be the most likely on the basis of extrapolation of (recent) past trends, it is also customary to assume some plausible variations around the most plausible set so that the effects of alternative levels of fertility, mortality and net migration on population size and characteristics can be worked out. The selection of alternatives will depend heavily on the purpose of projection and the availability of data of adequate quality.

Of the three components, assumptions on the future course of fertility are the most unpredictable since fertility levels in most countries depend on the voluntary choice of family size by millions of couples. While there is a certain degree of universal certainty about mortality, the same cannot be said about fertility and migration. Changes in fertility have got a greater impact on the size and structure of a population, compared to changes in mortality or migration, as discussed in the earlier chapters.

Three techniques are usually utilized for the projection of fertility rates. They are the **period-fertility approach, the cohort-fertility approach and marriage-parity-interval progression approach**. The period-fertility method consists of projecting the current age-specific birth rates progressively over time either as unchanged and remaining constant or changing in a predefined manner over successive periods of time (usually at 5-year intervals) and then applying the projected schedule of rates to the mean 'exposed-to-risk' female population in each projection period to determine projected number of births. The addition of marital status may be desirable in period-fertility projections, especially if illegitimate fertility is negligible. The cohort-fertility approach examines trends in fertility for separate birth or marriage cohorts of women, usually in terms of age-specific or duration-specific birth rates, and cumulative and completed fertility rates are derived for these cohorts. The cohort method of fertility projection may be adapted to take into account marital status. The marriage-parity-interval progression method involves a greater disaggregation of data. Probabilities of marriage and of giving birth by parity of women are employed in this method to determine the number of births of various orders from the number of women who are exposed to the risk of giving birth to a child of a given order. The method is applied by a series of attrition rates. In this method, a good deal of data regarding marriage and child-spacing patterns of women is necessary. Of these three alternative approaches to assumptions on the future course of fertility, the easiest and the most widely used method is the period-fertility approach.

NOTES

Mortality can be projected using either a constant or a progressively changing set of age–sex specific death rates over the projection period. These age–sex specific death rates are used to compute the survivorship ratios or the probabilities of surviving from one age group to the next during the period of projection. When reliable data on ASDRs are not available, mortality models representing generalized schemes for projecting the mortality rates applicable to the population with a known or assumed level of life expectancy at birth can be employed.

Net external migration can be projected either in absolute terms or as a percentage of the base population based on recent trends from the last two censuses. The projected age–sex composition of migrants should be derived from recent historical migration records. Some adjustments are necessary in age–sex-specific net migration figures to allow for mortality, fertility and ageing within the respective cohorts during each projection period.

The estimation of the future trends and patterns of the components of population change will largely depend on the demographic analysis of the past trends. This has been greatly assisted in recent years by social and economic models explaining the role of external forces in the changes of demographic measures over time as well as variations of such measures among population aggregates. In developed countries, fluctuations in nuptiality and fertility rates have been closely linked to short-term variations in such economic variables as employment, income and prices. Hence, assumptions on future fertility levels are based on future courses in employment, prices and income. In developing countries where fertility levels are still in the declining phase, such assumptions of linkage of their levels with socio-economic conditions may not be appropriate. As fertility starts to decline to lower levels in developing countries, the relationship between fertility and economic cycles may come to resemble that of the developed countries. The relationship between these variables is, however, not constant over time and space. For example, the relationship between income and fertility may be positive in the pre-transition stage, negative in the transition stage and positive again in the post-transition stage. Attempts to explain fertility differentials between population aggregates by means of social and economic variables are yet to reach a point at which they could provide a tool for improved projections.

The factors related to the standard of living and those related to the provision of health services are found to affect mortality trends and patterns in most populations, but the nature of their relationships is not yet so fully unravelled as to be used in the projection of mortality patterns in the future on the basis of inputs in these services. It is certain that in the not too distant future, projections of the future course of fertility, mortality and migration will largely be based on the projections of the future course of social and economic conditions of the population. Their linkages and two-way models in which social and economic conditions and demographic parameters influence each other will be used. For the present, demographic projections in developing countries are carried out without any assumed linkages with social and economic developments.

NOTES

10.4 An Illustrative of the Component Projection Method

In the following paragraphs, the basic principles of component projection are explained by a simple illustration of population projection for Kerala during the period 1991 to 2001.

The 1991 female population of Kerala will be projected over two 5-year periods until 2001.

The projections are done first from 1991 to 1996 and then from 1996 to 2001.

Step 1: Choosing the base population

The smoothened 5-year age distribution of the female population of Kerala, as given by the Registrar General of India, has been chosen as the base. The figures in thousands are given in column 2 of Table 10.2. It is better to choose the latest census population, rather than any other estimates for recent years, since the estimates are themselves projections on the basis of some assumptions and these will get compounded with assumptions on the new projections. It is better to choose 5-year age groups since this will offer a convenient mode for projections at 5-year intervals.

Step 2: Surviving the 1991 population to 1996

In the next step we survive the 1991 population in each age group to 1996. It has to be recognized that the survivors of any age group in 1991 will move to the next age group in

Table 10.2: Input Data Used for the Component Projection of Female Population, Kerala

Age Group	1991 Female Population (in '000s)	Assumed Survivorship Ratios $_nS_x$		Assumed Fertility Rates	
		1991–96	1996–2001	1991–96	1996–2001
1	2	3	4	5	6
0-4	1349	0.99729	0.99750		
5-9	1398	0.99858	0.99870		
10-14	1516	0.99830	0.99847		
15-19	1594	0.99741	0.99765	0.01068	0.00985
20-24	1602	0.99665	0.99697	0.07317	0.06950
25-29	1191	0.99580	0.99620	0.13117	0.12503
30-34	1127	0.99426	0.99470	0.08337	0.08000
35-39	967	0.99160	0.99210	0.03250	0.03100
40-44	783	0.98702	0.98764	0.00816	0.00750
45-49	644	0.97988	0.98074	0.00095	0.00080
50-54	561	0.96932	0.97059		
55-59	504	0.95196	0.95387		
60-64	459	0.92166	0.92470		
65-69	364	0.86798	0.87237		
70-74	233	0.77551	0.78112		
75+	117	0.51859	0.52202		
Total	14809				
e_0				74.5	75
TFR				1.70	1.62
GRR				0.82725	0.78754
MACB				28.70765	28.71262

e_0 – Expectation of life at birth
TFR – Total fertility rate
GRR – Gross reproduction rate
MACB – Mean age of mothers at childbearing

NOTES

1996. For example, the group aged 0–4 will survive to age group 5–9 in 1996. In order to survive the population from one age group in 1991 to the next one in 1996 we need to have the survivorship ratios.

$S(x) = {}_5L_5/{}_5L_0$ using the standard life table notations discussed earlier which give the proportion surviving from the age group (x to x+4) to (x+5 to x+9). In order to estimate the survivorship values we need a life table that describes the average mortality situation during the period 1991 to 1996 in Kerala. Based on previous trends in the expectation of life at birth, e_0 value for females for Kerala for the period 1991–96 is assumed as 74.5 years. It is possible to estimate the life table survivorship values by age groups for any assumed value of expectation of life using a model life table. The model life table usually adopted for India is the South Asian Pattern of the UN. However, the survivorship values for the assumed expectation of life for females as 74.5 years for Kerala during 1991–96 were estimated from 'West Model Life Tables' of the Coale-Demeney model life tables and are given in column 3 of Table 10.2. To obtain the population in 1996 one has to simply multiply the 1991 base population by the survivorship ratios, moving the results one row down. Thus, in the first row of the table 0.99729 of the 0–4 age group survive to become the 5–9 age group. Thus 1349 × 0.99729 = 1345000 survive and this figure is inserted in the 5–9 rows in 1996 of Table 10.3. This process is repeated for each age group to produce the 1996 population.

Table 10.3: Projected Female Population in 1996 and 2001 for Kerala

Age	1991 Female Population (in '000s)	1991–96 Survivorship Ratios	1996 Female Population (in '000s)	1991–96 Survivorship Ratios	2001 Female Population (in '000s)
1	2	3	4	5	6
0-4	1349	0.99729	1139	0.99750	1166
5-9	1398	0.99858	1345	0.99870	1136
10-14	1516	0.99830	1396	0.99847	1343
15-19	1594	0.99743	1513	0.99765	1394
20-24	1602	0.99665	1590	0.99697	1509
25-29	1391	0.99580	1597	0.99620	1585
30-34	1127	0.99426	1385	0.99470	1591
35-39	967	0.99160	1121	0.99210	1378
40-44	783	0.98702	959	0.98764	1112
45-49	644	0.97988	773	0.98074	947
50-54	561	0.96932	631	0.97059	758
55-59	504	0.95196	544	0.95387	612
60-64	459	0.92166	480	0.92470	519
65-69	364	0.86798	423	0.87237	444
70-74	233	0.77551	316	0.78112	369
75+	317	0.51859	345	0.52202	427
Total	14809		15557		16290

Note: The female population in the age group 0–4 is computed by multiplying the number of births during the projection period by 0.98347, which is computed from life tables for females (i.e. $_4l_0/(5 \cdot l_0)$)

NOTES

Table 10.4: Projection of Number of Female Births during the Period between Two Projection Years

Age	1991 Female Population (in 000's)	1996 Female Population (in 000's)	Female ASFR 1991–96	Number of Female Births	2001 Female Population (in 000's)	Female ASFR 1996–2001	Number of Female Births 1996–2001
1	2	3	4	5	6	7	8
15-19	1594	1513	0.00520	40	1394	0.00479	35
20-24	1602	1590	0.03660	284	1509	0.03382	262
25-29	1391	1597	0.06383	477	1585	0.06084	484
30-34	1127	1385	0.04057	255	1591	0.03893	290
35-39	967	1121	0.01581	83	1378	0.01508	94
40-44	783	959	0.00397	17	1112	0.00365	19
45-49	644	773	0.00046	2	947	0.00039	2
Total	8108	8938		1158	9516		1186

Note: Female ASFR is computed by multiplying the ASFR given in Table 10.2 by sex ratio at birth which is taken as 105.5 males per 100 females

NOTES

More importantly, there are no figures yet for the 0–4-year-olds in 1996 or 0–9-year-olds in 2001. For these, it is necessary, first, to use the fertility assumptions to estimate the number of births there will be between 1991 and 1996 and then to survive these since some will die before 1996. The estimation of the births is shown in Table 10.4.

Step 3: Estimation of births in the projection period and their survivors

First, the average number of females aged 15–49 during 1991–96 is calculated by simply taking the average of the 1991 base population figures and the new 1996 figures calculated in Table 10.2. The female ASFRs, which constitute the fertility assumptions being used, are then applied to the average population to give the projected female births in one year. These figures then have to be multiplied by 5 because five years of births are required (1991–96). They are then summed up, giving a total of 1458 thousand births occurring during 1991–96. As stated above, some of these births will now have to be struck off as not all will survive until 1996. The Coale-Demeney life tables provide a survivorship ratio giving the probability of 5 years of births surviving from birth to become the 0–4 age group. The probability of 5 years of births surviving from birth to become the 0–4 age group is computed from the female life tables for Kerala and is estimated to be 0.98347. Thus, the projected female population aged 0–4 in 1996 is 1158 × 0.98347 = 1139000, and this figure can hence be inserted in Table 10.3 to complete the projected 1996 population. These 0–4-year-olds can then be 'projected' to 2001. The survivorship ratio given in Table 10.2 is 0.99750; so there will be 1139 × 0.99750 = 1136000 females aged 5–9 in 2001. The whole process could also be repeated to find the 0–4-year-olds in 2001, using, perhaps, different fertility assumptions.

Before the availability of personal computers in the 1980s the above procedure had to be carried out manually step by step and it took time. Now with the easy availability of personal computers and software programmes for projections the whole process can be done easily. Further, all the assumptions can be changed at will and the effects of alternative assumptions on population growth and its age–sex distributions can immediately be checked. The most popular among these programmes is the SPECTRUM (DemProj component within the package) developed by the Futures Group. We give below an application of the programme for population projection of Kerala as used by the Government of India official projections published in 2001.

10.5 Official Projections

Since 1958, Office of the Registrar General, India (ORGI) has been undertaking the exercise of population projections on behalf of the Planning Commission of India. After every census, the Expert Committee/Technical Group under the Chairmanship of Registrar General, India (RGI) has been presenting data on population projections on the basis of the latest available census data. The present projections based on the age–sex distribution of population and migration data of the 2001 Census, latest available levels and trends of fertility and mortality data available from the Sample Registration System (SRS) is an effort in the same direction.

Following this tradition, after the 2001 Census a Technical Group was constituted by the National Commission on Population (NCP) in 2001 under the Chairmanship of the RGI to prepare population projections for the period 2001–2025. The mandate given to the group was, among other tasks, to review the methodology of population projections adopted in the past and prepare fresh projections up to the year 2025 with age, sex, urban–rural and state/UT break-ups for 5-year intervals.

NOTES

After the release of the age–sex distribution data of the population for India/states and migration data based on Census 2001, with the latter released only in November 2005, the projection exercise was carried out and the report submitted in 2006. The assumptions behind the projections, input values and outputs for Kerala are briefly discussed here. Most of the contents are from the report.

a. Base level population

It is well known that the age–sex data in India are subject to errors of coverage and quality to different degrees across the states. Kerala is found to have a better quality compared to other states because of its higher literacy levels over a long period of time. In many cases, the erroneous reporting of age is attributable to ignorance of respondents but in many areas it may be culturally biased to digits ending with 0 and 5. Accordingly, there has been a need to smoothen age–sex distribution. The various indices which measure the accuracy of age–sex data are the sex ratio score, the age ratio score, Whipple's Index and Myers' Index. A number of methods are available for smoothening of the data pertaining to the age–sex distribution. A summary of the indices of the different smoothening procedures for all India on the basis of the Census 2001 age–sex data is given in Table 10.5A. The index values for Kerala are already presented in Tables 4.7 and 4.8 and discussed in Chapter IV.

Table 10.5A: Summary of Indices Measuring the Accuracy of Data – Census 2001

	Smoothened Indices						
	Reported	Smooth-ened	Carrier Farrag	K-King Newton	Arriaga	United Nations	Strong
Sex ratio score	7.43	3.83	4.03	4.33	4.07	4.75	1.79
Male age ratio score	6.00	1.66	3.8	3.76	3.48	2.28	0.81
Female age ratio score	7.51	2.45	4.11	3.78	3.84	3.39	1.06
Accuracy index	35.80	15.60	20.00	20.52	19.54	19.94	7.25

Note: The accuracy index is the sum of the male and female age ratio scores plus three times the sex ratio score, all calculated using data for ages 10–14 through 65–69

Although the Strong method of smoothening gave the lowest values of sex and age ratio scores, the Technical Group recommended another method for age smoothening. The methodology is briefly described below.

NOTES

If W_1, W_2, W_3, ……., W_n are respectively the n quinquennial age groups, 0-4,5-9,10-14 and so on up to 75+, then

$$S(W_2) = 0.25 \times O(W_1) + 0.50 \times O(W_2) + 0.25 \times O(W_3)$$

where S is the smoothened population and O is the observed population

Similarly, $S(W_3) = 0.25\times O(W_2) + 0.50 \times O(W_3) + 0.25 \times O(W_4)$

In this way, smoothening of all the n-2 quinquennial age groups (except the first group W_1 and last group W_n) has been carried out. For smoothening W_1 and W_n this formula cannot be applied since there are no preceding and succeeding age groups respectively in these two cases. So, W_1 has been smoothened as under:

$$S(W_{0-4}) = O(W_{0-14}) – [S(W_{5-9}) + S(W_{10-14})]$$

Similarly, $S(W_{75+}) = O(W_{15-75+}) – [S(W_{15-19}) + …+ S (W_{70-74})]$

The age and sex ratio scores based on the above smoothening procedure are mentioned in the Table 10.5A showing the accuracy index as 15.60. Smoothening of age–sex distribution for India and all the major states including Himachal Pradesh, Delhi and the North Eastern Region have been carried out by using the above-mentioned procedure. The smoothened population of 2001 by age and sex for Kerala (for which component method has been applied) is presented in Table 10.5B.

Table 10.5B: Smoothened Age Distribution for 2001 for Kerala Taken for Population Projections 2001–2026

Age Group	Population ('000)		
	Males	Females	Persons
0-4	1392	1321	2713
05-09	1382	1330	2712
10-14	1458	1420	2878
15-19	1485	1503	2988
20-24	1417	1520	2937
25-29	1306	1464	2771
30-34	1207	1367	2574
35-39	1115	1237	2352
40-44	1008	1068	2075
45-49	898	914	1812
50-54	736	748	1484
55-59	573	611	1184
60-64	476	549	1025
65-69	389	475	863
70-74	320	410	730
75-79	200	262	462
80+	107	175	282
Total	15469	16373	31841

NOTES

b. Assumptions regarding fertility

SRS and NFHS provide estimates of TFR at the state level for the country. The SRS data have been used as base level estimates, these being available as a time series. The SRS data show that in most of the states, the level of total fertility has been falling since 1971. Although annual estimates of TFR are available from SRS from 1971 till 2002 for India and major states, reliable estimates are available from 1981 onwards. It may be seen from the SRS data that there has been consistent decline in fertility during the last two decades.

The trend analysis of TFR in terms of log linear, logistic and Gompertz models for projecting the future levels of fertility were discussed at length by the Technical Group. It was decided that the Gompertz model may be used for projecting the future levels of TFR in all the states as well as for the country as a whole. The mathematical form of the model used is as follows:

Gompertz Curve

$$\frac{TFR - Lt}{U - L} = (a)^{b^t}$$

Or alternatively,

> Ln(–(Ln(TFR–L) / (U–L))) = Ln(–Ln(a)) + t . Ln(b)
>
> where U and L are the upper and lower limits of TFR respectively and a and b are constants.

The lowest threshold of TFR was assumed to be 1.8 for all the states except Himachal Pradesh, the observed values of TFRs from 1981 to 2000 were considered for projecting the future levels of TFR as reliable estimates from SRS are available from 1981 onwards. For Himachal Pradesh, the TFRs were considered from 1991–2000. In fitting the Gompertz model, three types of upper asymptotes have been taken for the states depending upon whether the particular state is a high, medium or low TFR state. The upper asymptote (U) has been taken as 6 for southern states and 7 for the northern states. For western and eastern states, U has been taken as 6.5.

The projected values of TFR for India and Kerala for the different quinquennia are as follows:

NOTES

SRS	2000	2001–06	2006–11	2001–16	2016–21	2012–26
India	3.2	2.82	2.48	2.21	2.01	1.89
Kerala	2.9	1.80	1.80	1.80	1.80	1.80

Kerala achieved the replacement level of fertility, that is, TFR of 2.1 by 1988, reached the lower threshold of 1.8 by 2001–06 and remains at that level until 2026. A graph showing the logistic fit of TFR values for Kerala is given in Table 10.6 and Figure 10.2. India is projected to achieve this level by 2016.

Table 10.6: Interpolation and Extrapolation of TFR Using a Logistic Function, Kerala

Item or Year	Index Value	Year	Index	Year	Index	Year	TFR
Asymptotes		1960.5	6.50	1960.5	6.50		
		1961.5	6.50	1965.5	6.50	1965.5	6.2722
Lower	6.50	1962.5	6.50	1970.5	6.50	1970.5	5.8922
Upper	1.80	1963.5	6.50	1975.5	6.50	1975.5	5.0793
		1964.5	6.50	1980.5	6.50	1980.5	3.8765
Input values		1965.5	6.50	1985.5	6.50	1985.5	2.8032
		1966.5	6.50	1990.5	6.50	1990.5	2.2001
1981.50	2.91	1967.5	6.50	1995.5	6.50	1995.5	1.9453
1991.50	1.86	1968.5	6.50	2000.5	6.50	2000.5	1.8509
2001.50	1.82	1969.5	6.50	2005.5	6.50	2005.5	1.8176
		1970.5	6.50	2010.5	6.50	2010.5	1.8060
	1.00	1971.5	6.50	2015.5	6.50	2015.5	1.8021
	1.00	1972.5	6.50	2020.5	6.50	2020.5	1.8007
	1.00	1973.5	6.50	2025.5	6.50	2025.5	1.8002

Contd...

Item or Year	Index Value	Year	Index	Year	Index	Year	TFR
	1.00	1974.5	6.50	2030.5	6.50	2030.5	1.8001
	1.00	1975.5	6.50	2035.5	6.50		1.8000
	1.00	1976.50	6.50	2040.50	6.50		1.8000
	1.00	1977.50	6.50	2045.50	6.50		1.8000
	1.00	1978.50	6.50	2050.50	6.50		
	1.00	1979.50	6.50	2055.50	6.50		
	1.00	1980.50	6.50	2060.50	6.50		
	1.00	1981.50	6.50	2065.50	6.50		
	1.00	1982.50	6.50	2070.50	6.50		
	1.00	1983.50	6.50	2075.50	6.50		
		1984.50	6.50	2080.50	6.50		
		1985.50	6.50	2085.50	6.50		
		1986.50	6.50	2090.50	6.50		
		1987.50	6.50	2095.50	6.50		
		1988.50	6.50	2100.50	6.50		
Beginning date for results	1960.50	1989.50	6.50	2105.50	6.50		

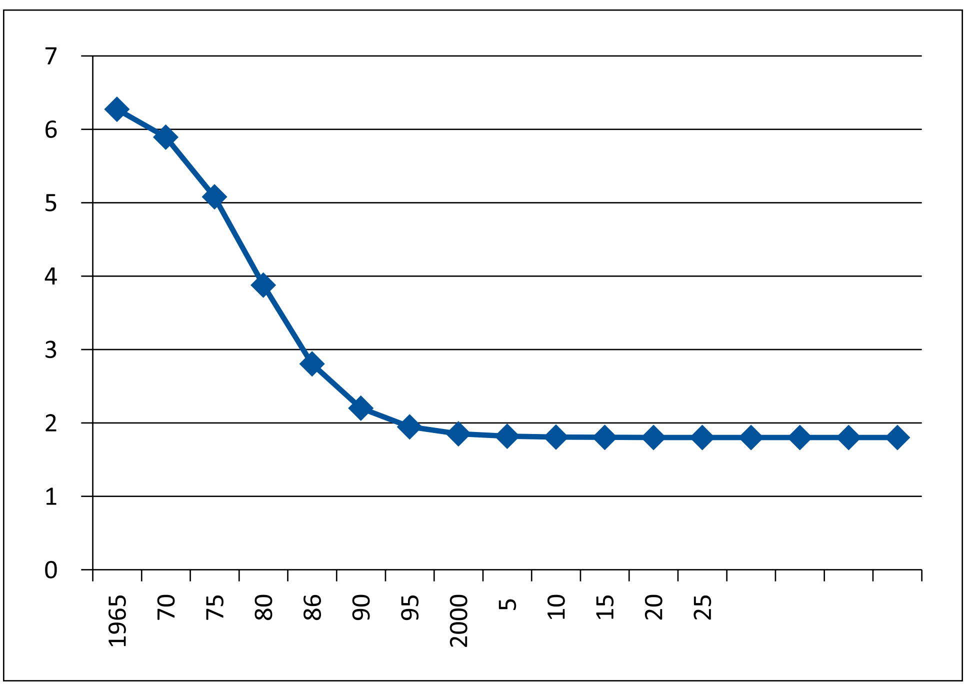
NOTES

The ORGI Expert Committee report presents similar values of projected TFR for all the states.

According to the projections, the year by which TFR of 2.1 will be achieved in India is 2016; Chhattisgarh will achieve it by 2022.

It was assumed that the sex ratio at birth (SRB) will remain at the 2000 level 111 (male births to 100 females) until 2025 while in Kerala it will be at 108.

Figure 10.2: Logistic Fit of TFR Values for Kerala



c. Assumptions regarding mortality

For projecting the likely levels of expectation of life at birth (e_0), working models developed by the UN have been adopted. The UN uses a standard pattern of improvement depending on whether the expectation is likely to improve slowly, moderately or quickly based on historical patterns. It has been assumed that increase in life expectancy becomes relatively slower as it reaches higher levels. The age-specific survival rates were calculated by assuming that the age pattern of mortality observed in the SRS for 1999–2001 would slowly converge to the West model life table pattern as levels of life expectancy increase.

Starting with the average SRS-based ASDR for 1999–2001 centred at 2000 which was taken as base year, separate life tables for males and females for each state have been constructed without taking into consideration the omission factor due to death reporting. A three year average was chosen, since the year-wise mortality rates particularly at the state level would be subject to higher variability. With the help of the West model life table and the expectation of life at the base level, the values of ${}_nL_x$ for each of the 5-year age groups were projected up to 2021–25 by interpolation. The West model life tables were adopted as the standard life tables for this purpose.

NOTES

To decide whether high, medium or low improvement should be assumed for each state, the e_0 values for the periods 1990–94 and 1995–99 obtained from SRS life table have been examined and the pattern thus obtained for e_0 has been assumed to continue in future for India and all the states. The assumed e_0 values for future years for Kerala are as follows:

e_0	2001–06	2006–11	2001–16	2016–2021	2012–26
Males	70.8	72.0	73.2	74.2	75.2
Females	76.0	76.8	77.6	78.1	78.6

d. Assumptions regarding migration

Based on the migration data of Census of India 2001, inter-state net migrants during the decade 1991–2001 have been assumed to remain constant throughout the projection periods for all the states (except Goa), where the component method of projections was used. The component of international migration has been assumed to be negligible, so it has not been considered for the projection exercise.

For Kerala, it was assumed that for the period 2001–26 the annual migration rate per 1000 population will be –0.08 for males and –0.08 for females, that is, for 100,000 population there will be 8 out-migrants per year each for males and females.

e. Population projection for urban population

For projecting the urban population, the urban-rural growth differential (URGD) method as given in Manual –VIII of UN has been adopted. This method is based on the fact that the urban–rural growth differentials follow a logistic pattern. The URGD for the period 1991–2001 has been assumed to be the same in the future as well, up to 2026. The projected urban population by sex for India, states and union territories as on 1 March, 1 July and 1 October are provided in the report.

Based on the above assumptions and using the Spectrum programme, the population projections were carried out and the major results are presented in Table 10.7. In the report of the Expert Committee the outputs are given in terms of the age–sex distribution of the population of each state for the years 2001, 2006, 2011, 2016, 2021 and 2026 in addition to the population sizes for individual years from 2001 to 2026.

The projected population of the state as on 1 March in 2006... 2026 are given in Table 10.7.

Table 10.7: Outputs of Population Projections for Kerala 2001–26

Projected Population Characteristics as on 1 March, 2001–26, Kerala						
Indicator	2001	2006	2011	2016	2021	2026
1	2	3	4	5	6	7
Population ('000)						
Total	31841	33265	34563	35677	36569	37254
Male	15469	16191	16859	17440	17917	18297
Female	16373	17074	17704	18236	18652	18956
Sex ratio	1058	1055	1050	1046	1041	1036
Population density (Sq.Km.)	819	856	889	918	941	959
Population by broad age groups ('000)						
18 years and above	21747	23562	25120	26491	27722	28760
0-14	8303	8005	7849	7593	7319	6988
15-59	20176	21530	22458	23100	23417	23462
60+	3362	3729	4256	4983	5833	6804
Proportion (per cent)						
0-14	26.1	24.1	22.7	21.3	20	18.8
15-59	63.4	64.7	65	64.7	64	63
15-49 (Female population)	55.4	55.5	54.5	52.8	50.7	48.3
60+	10.6	11.2	12.3	14	16	18.3
Median age (years)	28.05	29.87	31.81	33.79	35.79	37.67
Dependency ratio						
Young (0-14)	412	372	349	329	313	298
Old (60+)	167	173	189	216	249	290
Total (Young and Old)	578	545	539	544	562	588

Demographic Indicators: 2001–25, Kerala					
Indicator	2001–05	2006–10	2011–15	2016–20	2021–25
1	2	3	4	5	6
Population growth rate	0.9	0.8	0.6	0.5	0.4
Crude birth rate (CBR)	16.3	15.4	14.2	13.1	12.3
Crude death rate (CDR)	6.8	7.0	7.1	7.4	7.8
Infant mortality rate (IMR)	12.1	11.1	10.0	9.2	8.4
Under-5 mortality rate (q5)	14.1	13.0	11.8	11.0	10.1
Total fertility rate (TFR)	1.8	1.8	1.8	1.8	1.8
Life expectancy of males	70.8	72.0	73.2	74.2	75.2
Life expectancy of females	76.0	76.8	77.6	78.1	78.6

NOTES

10.6 Projection of Population at Sub-national Levels or for Different Groups

A. Ratio method for estimating the size at sub-state levels

The Expert Committee on Population Projections appointed by the RGI projected only the population at the state level by age and sex and the size by sex for the urban areas for the years 2006, 2011, 2016, 2021 and 2026. It does not provide any estimate of the population at the district level. Recently UNFPA and IIPS have made these projections at the district level by age and sex for the total population following the procedures outlined in the ensuing sections.

Based on knowledge of the projected population at the state level for any future year, the estimation of the population at the sub-state level such as a district, tehsil or even a village can be obtained by a method popularly known as the ratio method. In this method it is assumed that the ratio of the population of a district to the total population of the state follows the same trend as observed in the recent past and projected into the future to estimate the district population based on knowledge of the state population. For example, if we want to estimate the population of a district 'd' of Kerala in 2011 from a knowledge of the projected population of Kerala for 2006 already estimated by the component method described above, we proceed as follows:

NOTES

1. First, we estimate the ratios of the population of different districts of Kerala in the last two censuses for which data are available, namely, 1991 and 2001. For each of the years we estimate the ratios of the population of the districts to the total state population. If for any district d, the ratio of the population of the district in 2001 and 1991 is assumed to be R(1,d) and R(0,d) respectively, then **R(1,d) – R(0,d)** is the change in the ratio of the population of the district between 2001 and 1991 and the annualized change in r is given by

   $$r = (R(1,d) - R(0,d))/10$$

2. This is projected for any future year assuming that the same trend of change in the ratio will be observed in the future. If we want to estimate the population of the district 'd' in 2006 the estimated ratio will be

   $$R(1,d) + 5 \times r$$

   and for 2016 will be

   $$R(1,d) + 15 \times r$$

3. After estimating these ratios for the future years for all the districts, the state population projected for any future year is multiplied by these projected ratios to get the estimate of the district population (see Table 10.8). It must be noted that the sum of all the district populations projected in this manner may not add up to the state population and hence pro rata adjustments are required.

   This method, though used commonly for estimating the geographical sub-units within a state, can also be used to project population sizes in different subgroups, such as by religion, educational status etc., and hence has wide applications in demography.

Table 10.8: Projected Population in Districts of Kerala by Ratio Method

State and District	Population		Population Ratio		Annual Difference in Ratios	Projected Population Ratio (2016)	Projected Population 2016
	2001	2011	2001	2011			
Kerala	**31841374**	**33406061**					36582000
Kasaragod	1204078	1307375	0.038	0.039	0.000132	0.0398	1455830
Kannur	2408956	2523003	0.076	0.076	–0.000013	0.0755	2760496
Wayanad	780619	817420	0.025	0.024	–0.000005	0.0244	894279
Kozhikode	2879131	3086293	0.090	0.092	0.000197	0.0934	3415672
Malappuram	3625471	4112920	0.114	0.123	0.000926	0.1277	4673287
Palakkad	2617482	2809934	0.082	0.084	0.000191	0.0851	3112025
Thrissur	2974232	3121200	0.093	0.093	0.000002	0.0934	3418381
Ernakulam	3105798	3282388	0.098	0.098	0.000072	0.0986	3607572
Idukki	1129221	1108974	0.035	0.033	–0.000227	0.0321	1172936
Kottayam	1953646	1974551	0.061	0.059	–0.000225	0.0580	2121155
Alappuzha	2109160	2127789	0.066	0.064	–0.000254	0.0624	2283531
Pathanamthitta	1234016	1197412	0.039	0.036	–0.000291	0.0344	1258007
Kollam	2585208	2635375	0.081	0.079	–0.000230	0.0777	2843833
Thiruvananthapuram	3234356	3301427	0.102	0.099	–0.000275	0.0975	3564997

Note: Projected population for Kerala state is taken from ORGI report (2006)

B. Estimating the age–sex distributions at the district or sub-national levels

Having worked out the population size of a district by the ratio method described above, the age distribution can be worked out for each sex for any given year using the so-called contingency table method or estimating the cell frequencies in a two-way table based on knowledge of the marginal totals. The population of different age groups of the state is the marginal total of the rows and the total population of the districts is the marginal total of the columns. The estimation of the size of the population of any district 'd' is obtained as follows.

1. First we assume that all the districts – for any given sex, say, male – follow the same age pattern of the state and the cell figures are estimated.
2. Adding the cell figures for all ages for the districts will not add up to the estimated district population totals and hence pro rata adjustments are made.
3. Having made these adjustments across the districts for any given age group, the sum will not add up to the total population of the state for that age group, and hence pro rata adjustments are made.
4. The above two procedures are repeated, first adjusting the cell values to match the column totals and then readjusting to match the row totals. This procedure is called *iteration* and is repeated a few times.
5. Usually in such types of iteration the cell values converge to marginal totals of both row and columns and there is theoretical support for such iteration.

6. Usually the convergence to the state totals in each age group and the total population of the district is obtained in three or four iterations. Though there are only a few steps to be followed, in practice, it involves a number of procedures with the Excel sheet. Taking the projected population by age and sex for eight states for 2006, 2011, 2016 from Office of the Registrar General of India (ORGI, 2006) report, UNFPA India (2009) projected district-wise population by age and sex for these years. An example of the output, the final projected age-sex distribution of Ajmer District for year 2016 is given in Table 10.9.

Table 10.9: Projected Population by Age and Sex for Ajmer District of Rajasthan, 2011 and 2016

Age Group	2011			2016		
	Persons	Male	Female	Persons	Male	Female
0 (<1 yr)	54791	28700	26091	48918	26147	22771
1	52636	27817	24819	49791	26456	23334
2	51102	27197	23905	50416	26709	23707
3	50126	26813	23313	50833	26912	23921
4	49638	26632	23006	51079	27073	24006
5	49572	26627	22945	51191	27199	23992
6	49859	26766	23093	51208	27297	23910
7	50434	27021	23413	51167	27376	23791
8	51228	27361	23867	51107	27442	23665
9	52174	27757	24417	51066	27503	23563
10	53270	28224	25046	50983	27563	23420
11	54511	28777	25734	50802	27627	23174
12	55510	29161	26349	51043	27717	23325
13	56070	29255	26815	51938	27846	24092
14	56317	29166	27151	53236	28012	25224
15-19	283459	145691	137768	281521	143775	137746
20-24	274747	148439	126308	289435	154693	134742
25-29	221301	120132	101169	271150	146438	124712
30-34	184969	97348	87621	221290	119133	102156
35-39	165690	83336	82354	186925	96690	90235
40-44	157680	79193	78487	167422	84393	83029
45-49	146361	75254	71107	159119	80060	79059
50-54	118306	61164	57142	141218	71303	69915
54-59	90779	47407	43372	111486	57530	53956
60-64	65165	31647	33518	81358	39548	41810
65-69	51173	23536	27637	60893	29022	31871
70-74	39360	16752	22608	43276	18887	24389
75+	47302	19099	28203	60548	23804	36743
Total	2633530	1366272	1267258	2840419	1474155	1366258

Source: District Level Population Projections in Eight States of India 2006-2016, UNFPA India, 2009

Exercises

Exercise 10.1: Projection of population for Rajasthan using cohort component method.

Step 1. Take five age–sex distributions from first day exercise (col. 1-3).

Step 2. Input survival ratio from life table for year 2001 (col. 4, 5).

Step 3. For age group 5–9 onwards, calculate population survived by multiplying survival ratio with the previous age group population (col. 6, 7).

Step 4. Input data ASFR for year 2001.

Step 5. Calculate mid-year female population aged 15–49 by taking average of columns 3 and 7.

Step 6. Calculate birth to age groups 15–19 to 45–49 in five years
= ASFR * mid-year female population * 5/1000 (col. 10)

Step 7. Calculate total births by adding col. 10 (col. 11).

Step 8. Assume SRB (refer to ORGI Projection Report, 2006 for Rajasthan)

Step 9. Males (0–4) = SRB * total births/(1+SRB)
Females (0–4) = 1 * total births/(1+SRB) or Females (0–4)
= total births – male (0–4)

Step 10. Calculate population (0–4) by multiplying sex-specific survival ratio and male/female population (0–4).

Step 11. By following the above steps, columns 6 and 7 are the projected population. If we have age-specific net migrants, we can adjust and can get the final projected population.

Exercise 10.2: Project district-wise population for Rajasthan using ratio method for 2016 and 2021.

District	2001	2011
Ganganagar	1789423	1969168
Hanumangarh	1518005	1774692
Bikaner	1902110	2363937
Churu	1696039	2039547
Jhunjhunu	1913689	2137045
Alwar	2991552	3674179
Bharatpur	2100020	2548462
Dholpur	983258	1206516
Karauli	1205888	1458248
Sawai Madhopur	1117057	1335551

Contd...

District	2001	2011
Dausa	1323002	1634409
Jaipur	5251071	6626178
Sikar	2287788	2677333
Nagaur	2775058	3307743
Jodhpur	2886505	3687165
Jaisalmer	508247	669919
Barmer	1964835	2603751
Jalor	1448940	1828730
Sirohi	851107	1036346
Pali	1820251	2037573
Ajmer	2178447	2583052
Tonk	1211671	1421326
Bundi	962620	1110906
Bhilwara	2020969	2408523
Rajsamand	982523	1156597
Dungarpur	1107643	1388552
Banswara	1420601	1797485
Chittorgarh	1330360	1544338
Kota	1568705	1951014
Baran	1021473	1222755
Jhalawar	1180323	1411129
Udaipur	2481201	3068420
Pratapgarh	706807	867848
Rajasthan	56507188	68548437

Exercise 10.3: Using URGD method project per cent urban population for 2016, 2021 and 2026.

Year	Rajasthan Population		
	Total	Urban	Rural
2001	56507188	13214375	43292813
2011	68548437	17048085	51500352

CHAPTER XI

Demographic Models and Model Life Tables

11.1 Introduction

A model is an abridged version of reality or phenomenon expressed in a mathematical form. Thus, a sphere can be considered as a model of the earth. In other words, it is the simplified representation of reality where some aspects of truth are omitted or assumed to be unchanging so that other aspects can be observed more clearly. It is that scientific discipline which explains the observations generated by the phenomena with which they are concerned. It is often represented by mathematical relations. Thus demographic models, in general, are the relationships which explain demographic phenomena such as age structure, fertility, mortality and population growth.

There are two types of mathematical models: **descriptive** and **normative**. Descriptive models are the models which try to produce reality as accurately as possible, while normative models describe what would occur under certain precisely specified, often wholly unrealistic conditions. Both types of models can be either **deterministic** or **stochastic**. Deterministic models assume that the parameters included in the model or in the set of equations are invariant with time whereas the stochastic models assume that the parameters vary with time according to preset probability distributions. For example, TFR explains what fertility would be if mortality is zero and is unaffected by migration and age structure. The net reproduction rate **(NRR)** which is a measure that combines a fertility and mortality schedule, tells the number of daughters who will be left behind by a woman in her lifetime if subjected to the same schedule of fertility and mortality in reproductive ages. Similarly, the l_x column of a life table is a representation of what would happen to a population if there were no fertility and migration. Models in mortality are the oldest among the various models used in demography. We first describe how to compute NRR from a given schedule of fertility and mortality.

NOTES

11.2 Computation of Net Reproduction Rate

Net Reproduction Rate is similar to the gross reproduction rate discussed in the chapter on fertility (Chapter V) except that we take into account the life table population of women in each of the reproductive age groups where we are using the fertility schedule. Table 11.1 gives the data on ASFRs for Kerala 2001, used earlier in Table 5.1 and also the

$_nL_x$ column from the life table constructed for women from the ASDRs of 2001 given in Table 6.5. The steps are as follows: first, multiply the $_nL_x$ (life table population of women) in the age group x to x+n with the ASDRs of the same age group and sum them up from ages 15 to 49. This will give the number of children born to women when they are subjected to given mortality levels in the reproductive ages and for Kerala this works out to 1.8049. Next, this is multiplied by the proportion of girls born among total births, which is 0.4808 assuming that there are 108 male births to 100 female births (as was assumed in the official projection) and we get the NRR as 0.8678. This denotes that every woman in Kerala will be replaced by 0.8678 daughters if subjected to the present level of fertility and mortality. Such a situation will naturally lead to a decline in the population over time and if continued indefinitely can even lead to extinction of the population. NRR of 1 is considered the replacement level of fertility when eventually the population will stop growing and will remain stationary. This is the goal of the population policy of India since the late 1980s. NRR of 1 implies a TFR of roughly 2.1.

11.3 Stationary Population

NOTES

Stationary population is the population which results from a zero population growth that lasts for a long period of time, where the number of births and deaths in any given period of time are the same. Thus stationary population can be regarded as a normative model which assumes the growth of the population to be zero. The population which has an unchanging growth rate for a long period of time with unchanging but different birth and death rates is defined as a stable population. Thus, the stationary population where birth and death rates are equal and the growth rate is zero can be considered as a particular case of the stable population. However, both stable and stationary populations are hypothetical model populations based on unchanging fertility and mortality rates which are rarely realized in actual populations. The population generated by any given life table is considered to be a stationary population. The $_nL_x$ column of any life table provides a stationary population where the mortality levels given in the life table continue year after year and there are 100,000 births (the radix of the life table or l_0) every year.

11.4 Stable Populations

Alfred Lotka (1922) was the first to give the concept of stable population and can be considered as the pioneer author of demographic modelling. The basic theory of Lotka is based on mathematical reasoning and was first developed only for the female population. Subsequently, several researchers have tried to improve upon the model. Lotka has shown that the age distribution of a closed population, that is, where there is no migration, that is subjected to constant (but not *necessarily* the same) birth and death rates will, over a long period of time, tend to become stable or unchanged. He proved that the final or ultimate age distribution that a population experiencing unchanged birth and death rates will take is dependent on the birth and death rates and not on the initial age distribution of the population. Different age distributions subjected to the same fertility and mortality rates will all converge to the same age distributions independent of the initial age distributions.

This is generally known as Lotka's Stable Population Theory. The CBR and the CDR in a stationary or stable population are known as 'intrinsic birth rate' and 'intrinsic death rate' respectively. In stationary populations the intrinsic death and birth rates will be equal, but in stable populations these rates will not be equal, thus affecting population size, growth or decline. The difference between the intrinsic birth and death rate of a stable population is called the *intrinsic growth rate or intrinsic rate of natural increase* and is called *Lotka's r*. This measure is usually denoted by r.

Now we turn our attention to the computation of a stable population from a given set of fertility and mortality measures. As observed earlier, the difference between the stable and stationary populations is that while the stable population grows by a factor of r, this is zero in the case of the stationary population. Hence, a stable population can be obtained by adjusting the stationary population for the rate of increase. The life table population can be considered as a stationary population and with a value of r it should be possible to compute the corresponding stable population. Lotka has derived a relationship between NRR and the intrinsic growth rate. According to Lotka,

$$NRR = \exp(rT)$$

where T = mean length of a generation

and

$$r = (\log_e NRR/T)$$

In fact the mean length of generation T can be closely approximated by the mean age of childbearing for a stationary population. Once the intrinsic growth rate is computed, it is easy to derive the stable population. To obtain the stable population we have to multiply the stationary population by exp (-ry) where y is the average age of each age group of the population. The calculation should be done separately for males and females.

NOTES

11.4.1 Calculation of the stable age distribution

To obtain the stable population, we have to first calculate the intrinsic growth rate of the population from an estimate of NRR and the mean age of childbearing (Table 11.1). So the first step is to calculate the NRR and the mean age of childbearing by combining the ASFRs and the life table survival probabilities. The stationary population is taken from $_nL_x$ values from the life table estimates.

NRR is computed by multiplying the total of col. 4 by S' = 1/1 + S, where S' is taken as 0.4878. T is obtained by dividing the total of col. 6 by the total of col. 4.

'r' the intrinsic rate of growth = $\text{Log}_e(\text{NRR})/T$ = –0.0054

Since the ASFR includes both males and females births, the total is divided by the sex ratio at birth (i.e. 0.4808) in order to get the NRR. The stationary population is obtained from the life tables for females.

NOTES

Table 11.1: Net Reproduction Rate for Kerala, 2001

Age Group	Age Specific Fertility Rate ASFR	Women in Stationary Population ${}_nL_x/l_0$	Births to Women in Stationary Population Col. 2 × 3	Mid Point of Age Group	Col. 4 × 5
1	2	3	4	5	6
15-19	0.0245	4.94	0.121030	17.5	2.118025
20-24	0.1449	4.92	0.712908	22.5	16.040430
25-29	0.1331	4.90	0.652190	27.5	17.935225
30-34	0.0479	4.88	0.233752	32.5	7.59694
35-39	0.0115	4.85	0.055775	37.5	2.091563
40-44	0.0017	4.81	0.008177	42.5	0.347523
45-49	0.0003	4.76	0.001428	47.5	0.067830
Total	0.3639		1.8049		46.1975
$Total_a$			0.8708		25.5956
			NRR		T

The mean age of childbearing (T) = (Total of col. 6/Total of col. 4)

= 46.1975/1.8049 = 25.5956

Intrinsic rate of natural increase (r) = $(\log_e NRR/T)$

= $(\log_e 0.8678)/25.9281$

= **–0.005554**

The negative sign of ‘r’ indicates that the population of Kerala will start declining if the present levels of fertility and mortality continue for a long time.

Once the intrinsic rate of natural increase and the female stationary population are available it is very easy to calculate the stable population. This is obtained by multiplying the stationary population in each group by exp (-ry) to obtain the stable population in that age group where ‘y’ is the average age of the age group (Table 11.2). One problem, however, is that so far only females have been considered. To include males it is necessary to repeat the calculations separately by using a male stationary population. One small complication is that the radix in this life table should not be the usual 1000 or whatever, but instead should be increased to perhaps 1080 so that the ratio of the radices of the male and female life tables is equal to the sex ratio at birth. The calculation and the age distribution of stable population for Kerala are provided in Table 11.2.

Table 11.2: Calculation of Stable Population for Kerala, 2001

Age	l_x Values		Stationary Population $_nL_x$		Average Age	Factor	Stable Population		Stable Population T = 10000	
	F	M	F	M	Y	exp (-ry)	F	M	F	M
1	2	3	4	5	6	7	8	9	10	11
0	1000	1080	993	1065	0.3	1.0017	995	1067	55	59
1	992	1063	3967	4244	2.6	1.0145	4025	4306	223	239
5	991	1059	4949	5290	7.5	1.0424	5159	5514	286	306
10	988	1056	4939	5274	12.5	1.0717	5293	5652	294	314
15	988	1053	4935	5257	17.5	1.1018	5437	5793	302	322
20	986	1050	4924	5234	22.5	1.1328	5578	5928	310	329
25	983	1043	4902	5196	27.5	1.1646	5709	6051	317	336
30	977	1035	4876	5146	32.5	1.1973	5838	6161	324	342
35	973	1024	4849	5084	37.5	1.2309	5969	6257	331	347
40	967	1010	4810	4985	42.5	1.2655	6087	6309	338	350
45	957	984	4757	4864	47.5	1.3010	6189	6329	344	351
50	946	961	4677	4690	52.5	1.3376	6256	6274	347	348
55	925	915	4560	4420	57.5	1.3751	6271	6079	348	338
60	899	854	4387	4037	62.5	1.4138	6202	5707	344	317
65	856	761	4112	3498	67.5	1.4535	5977	5084	332	282
70	789	638	3623	2811	72.5	1.4943	5414	4201	301	233
75	660	486	2899	2024	77.5	1.5363	4454	3109	247	173
80	499	323	1948	1209	82.5	1.5794	3077	1909	171	106
85+	280	160	1692	782	88.5	1.6328	2763	1277	153	71
Total							92101	87976	5114	4885

Note: F- Females, M- Males, T- Total, and r (–0.00554) is obtained from NRR and mean age of childbearing. l_x and $_nL_x$ values are obtained from the life table

11.5 Model Life Tables

The life tables which give a set of probability of dying at different ages corresponding to a given level of mortality or average expectation of life at birth are called *model life tables*. These models are derived based on the assumption that there is a typical age pattern of mortality irrespective of the mortality level. Model life tables are identical to the real life tables in every way, except that they relate to no particular place or time. Two types of model life tables are available: **empirical** and **relational**. The empirical model life tables are derived based on the life tables constructed from the pattern of mortality found in historical populations. The relational model life tables are derived from mathematical relationships rather than from empirical life tables. In this section only the empirical model life tables are discussed. The more generally used model life tables are the UN and Coale–Demeney model life tables. These are discussed in detail in the ensuing sections.

11.5.1 United Nations model life tables

In 1955, the United Nations Population Division compiled 158 empirical life tables and subjected them to complex regression analysis relating each $_nq_x$ value to its preceding $_nq_{x-n}$ value and developed equations by which, for any given g_0 value, subsequent values could be estimated by a series of recursive equations. A system of model life tables was thus derived corresponding to the values of q_0, 20, 25..., 60 and thereafter q_0 = 110, 120, 330. The expectation of life at birth ranging from 20 to 75 years is covered by the UN model life tables. The mortality pattern corresponding to the expectation of life at birth of 20 years is indexed as level 1. The levels 2, 3 are made corresponding to e_0 equal to 22.5, 25.0 and so on up to age 55. After age 55, levels are marked for a difference over every 5 years.

This first set of model life tables has been criticized on several grounds. The principal objection is that the inputs used were representative of a period which is entirely different from the present population. Hence, these life tables are not used for generating mortality patterns for the present time. The estimation of $_nq_x$ values start from an arbitrary choice of q_0 values, thus it could lead to under-estimation of values at older ages. Some of the 158 empirical life tables used to calculate the regression equation are based on unreliable records of mortality experience. Some of the life tables used had been smoothened or adjusted using some sort of interpretation and so they were not 'real'.

NOTES

Subsequently, in 1982, the UN gave a new set of model life tables and attempted to overcome some of the limitations of the model life tables derived earlier. These model life tables are entirely based on the empirical life tables from developing countries. A total of 72 tables, 36 male and 36 female, were prepared from the developing countries of Asia, Central and Latin America and sub-Saharan Africa. These life tables were then divided into four patterns (groups) using graphical and statistical procedures. The four patterns were labelled **Latin American, Chilean, South Asian** and **Far Eastern**. A fifth pattern which is an average of all the empirical life tables is also computed and is usually known as the **general** pattern. The mortality levels expressed in these model life tables are presented for life expectancies at birth between 35 and 75 years at 1-year increments. The mortality patterns expressed in these models have the following characteristics: the Latin American pattern tables have high infant and child mortality, high adult mortality and relatively low old-age mortality; the Chilean pattern has extremely high infant mortality; the South Asian pattern has high mortality under age 15 and over 55 years of age, but relatively low mortality at adult ages; the Far Eastern pattern has high old-age mortality, particularly among males.

11.5.2 Coale–Demeney model life tables

In 1966, Coale and Demeney published model life tables, which were somewhat similar to the UN life tables. They considered a total of 326 life tables, which are based on accurate data for the preliminary analysis. The final model life tables were obtained from a total of 192 life tables and four patterns of mortality by age were inferred. The four patterns, generally known as four families, are called *north* and *south*, *east* and *west*. The model life tables were generated using regression procedures. According to them the regression takes the following form:

$$l(x) = a + b \times e(10)$$

The above equation was used for each age, sex and region separately. The expectation at age 10 is used because it was considered to be a good indicator of the overall level of mortality and was relatively unaffected by the mortality rate for any other age group. Using this general relationship, 24 life tables were prepared for each family, where the expectation of life at birth ranges from 20 years to 77.5 years. There are certain characteristics that can be observed about the four families. The north pattern is characterized by low infant and old-age mortality, but by high adult mortality. The south model has high mortality under age 5, particularly among infants, low adult mortality and high mortality over age 65. The eastern pattern is characterized by high infant and high old-age mortality relative to childhood and adult rates. The west model is considered as describing an average mortality pattern.

The main opposition to these models is that they use inputs generally from European countries. It has also been shown that there are several populations that have mortality patterns different from the four families generated by Coale-Demeney.

The general use of the model life tables is that life tables can be generated for a population where accurate life tables are not available or where past data are not adequate to visualize the future. The model life tables are also useful for anticipating **future** trends in mortality for population projections of any population. The life tables are used to estimate childhood mortality from limited data. These models are also used for the indirect estimation of mortality for a population where the data are limited or unusable. The general discussion of the use of model life tables in indirect estimation can be obtained from the United Nations Manual IX.

NOTES

Exercises

Exercise 11.1: Calculate NRR, mean length of generation and intrinsic growth rate r.

Step 1. Use ASFR from Day 1 Exercise 2.1e and '$_nL_x/l_0$' from Exercise 2.1f (female life table).

Step 2. Calculate births to women in stationary population (col. 2 × col. 3).

Step 3. NRR = total births/(1+S).

Step 4. Mean length of generation T = sum of col. 6/sum of col. 4.

Step 5. Intrinsic growth rate r = l_n(NRR)/T

CHAPTER XII

Selected Software Packages in Demographic Analysis

12.1 Introduction

In recent years, a number of software packages in demographic analysis of population data have been developed and are being routinely used. Three such packages are currently being used by population analysts and researchers and taught to students in demography courses. They are **MORTPAK** developed by the United Nations, Population Analysis Spreadsheets **(PAS)** developed by the US Census Bureau and **SPECTRUM** developed by the Futures Group, Washington. These packages can be applied even in the absence of detailed knowledge of and assumptions behind each of the demographic techniques. Anyone who enters the required data for a specific demographic method listed in the package can get the outputs or results of the analysis. These packages are briefly described in the ensuing sections.

The US Census Bureau developed the Population Analysis System (PAS) in 1981 to enhance the process of analysing available population data and later expanded it to PASEX in 1994 by Eduardo Arriaga (Arriaga, 1994). The US Census Bureau through the International Program Center has organized a large number of short-term training courses for demographers from developing countries in the application of this package. It is a set of spreadsheets containing frequently used procedures and methods in basic demographic analysis. PASEX spreadsheets include methods for the analysis of the main demographic topics of age structure, mortality, fertility, migration, distribution of population, urbanization and population projections. Originally developed in 1981 for use through Lotus 1-2-3 programs, they were later modified in 1994 for application through Excel programs and hence called PASEX.

PASEX consists of spreadsheets formatted for use in Excel, so users will need to have access to Excel to use PASEX. The spreadsheets and the special program RUP (a population projection program for rural and urban projections) are distributed together with the manuals describing the basic demographic methods and procedures they perform. This package can freely be downloaded online.

The United Nations' MORTPAK is a set of programs for detailed mortality analysis that can be used with PASEX and RUP for analysing population data. The user should obtain this package through the UN. The software package SPECTRUM is a package to analyse

NOTES

the effects of policies and programmes of HIV control and their impact on mortality, age distribution and population growth. It is essentially a population projection package with and without HIV control programmes on the demography of a country. It also estimates the resources needed to implement specific programmes. These demographic analysis tools allow users to produce national and sub-national population estimates and projections along with estimates and rates of fertility, mortality and migration and hence are discussed here in some detail.

12.2 Population Analysis Spreadsheets

There are 45 spreadsheets, each used for a different type of demographic analysis. These are very popular among demographers who wish to carry out basic demographic analysis of a given population data without developing their own programmes for selected widely used demographic techniques.

A brief description of each of these programmes follows.

12.2.1 Age structure of a population

NOTES

For the analysis of **age structure of a population** there are 11 spreadsheets labelled as follows, listed in alphabetical order.

1. ADJAGE

Adjusts any population age structure by sex to a given population total. This spreadsheet takes a population age distribution and proportionally adjusts it to a desired population total. The adjustment can be made independently to given population totals for each sex, or to a total for both sexes combined.

2. AGEINT

Interpolates between two population age structures. The user can specify if the interpolation is to be linear or exponential.

3. AGESEX

Analyses the age reporting in a population age and sex distribution for reporting errors in 5-year age groups. Calculates age and sex ratio indices and the UN age/sex accuracy index. This is widely used in developing countries, particularly in India.

4. AGESMTH

Smoothens a population age distribution using several methods. This is also being used widely in the Indian census.

5. and 6. BASEPOP and BPSTRNG

Prepare the age and sex distribution of a population for making a projection. The difference between the two spreadsheets is the difference in the smoothening process.

7. GRPOP-YB

Makes a graph using population age structures for two or three dates, by year of birth of each cohort.

8. MOVEPOP

Moves the population age distribution pertaining to a specific date to another date, using simple projection techniques.

9. OPAG

Distributes open-ended age groups (younger than 80+) into 5-year age groups up to 80 years and over.

10. PYRAMID

Makes an age pyramid by sex, with absolute numbers and percentages of the population data. This is also being widely used.

11. SINGAGE

Calculates the Whipple, Myers and Bachi indices of age heaping based on enumerated population by single years of age.

12.2.2 Analysis of mortality

There are 15 spreadsheets for analysing the mortality of a population. Some spreadsheets were prepared for populations with full and acceptable data, while others assume that only limited data are available. Although this set of mortality spreadsheets is the largest, the authors did not include the techniques developed and published by the United Nations in the MORTPAK package. Thus the analysis of mortality may require the use of both PASEX and MORTPAK packages. The PASEX spreadsheets related to mortality are as follows:

NOTES

1. ADJMX

Adjusts a pattern of mortality rates by age for both sexes combined or for each sex (ASDRs or central death rates from an empirical or existing life table) for obtaining a user-specified number of deaths in a population.

2. BTHSRV

Estimates infant mortality rates based on information on the number of children born during a year prior to the census and the number still alive at the census date.

3. E0LGST

Fits a logistic function to values of life expectancies at birth for each sex simultaneously, given two or more observed values of life expectancies and the two asymptotes of the logistic.

4. GRBAL

Uses the technique developed by Brass for estimating the completeness of reporting of deaths over 5 years of age in relation to a population (Brass, 1975).

5. and 6. INTPLTM and INTPLTF

Interpolate male and female life tables, respectively, between the values of two given sets of 'pivotal' life tables.

7. and 8. OGITQX and LOGITLX

These two spreadsheets are used for smoothening functions of a life table, using logits of the q(x) and l(x) functions of a life table.

9. LTMXQXAD

Constructs a life table from age-specific death rates or from the probabilities of dying between two specific ages.

10. LTNTH

Selects a Coale–Demeny model life table, region North, that will reproduce a given crude death rate pertaining to a given population age structure.

11. LTPOPDTH

Constructs and smoothens a life table for both sexes or one sex at a time, using population and death data.

12. LTSTH

This is the same as LTNTH, but using region South of the Coale–Demeny model life tables.

13. LTWST

This is the same as LTNTH, but using region West of the Coale–Demeny model life tables.

14. PREBEN

NOTES

Estimates the level of mortality for ages 5 years and above during an intercensal period using Preston-Bennett method, 1983.

15. PRECOA

Uses the technique developed by Preston and Coale for evaluating an available age distribution of deaths in relation to the population.

12.2.3 Analysis of fertility

There are 13 spreadsheets for analysing the fertility of a population. Most spreadsheets are for populations whose registration of births is not complete.

1. ADJASFR

Adjusts a given pattern of *age-specific fertility rates* to reproduce a desired total number of births.

2. and 3. ARFE-2 and ARFE-3

ARFE-2 estimates fertility using the Arriaga method based on the average number of children ever born by 5-year age groups of females for two dates and the pattern of fertility (ASFR) for those two dates. ARFE-3 is similar to ARFE-2 except that is based on the same data at three points of time.

4. ASFRPATT

Provides age-specific fertility rates pertaining to a given total fertility rate.

5. CBR-GFR

Calculates the crude birth rate and the general fertility rate based on a desired total fertility rate.

6. CBR-TFR

Estimates the crude birth rate and the total fertility rate, based on the general fertility rate.

7. PFRATIO

Uses the P/F ratio technique, originally developed by Brass, for adjusting reported age-specific fertility rates to the 'actual' level of fertility.

8. RELEFERT

Uses Rele's technique for estimating the gross reproduction rate of a population for one or two 5-year periods prior to the census date.

9. REL-GMPZ

Uses the technique developed by Brass and associates 'for the evaluation and adjustment of fertility estimates obtained from retrospective reports of birth histories or features of birth histories'.

10. REVCBR

Calculates crude birth rates during two or three 5-year periods prior to the census date, based on the age structure of the population.

11. TFR-GFR

Estimates the total fertility rate and the general fertility rate based on a crude birth rate.

12. TFRLGST

Fits a logistic function to two or more values of total fertility rates, and interpolates and extrapolates. It requires asymptotic values.

13. TFRSINE

Fits a sine function to two values of total fertility rates, and interpolates and extrapolates.

NOTES

12.2.5 Migration

There is one spreadsheet that provides net migration between two areas (IG).

12.2.6 Urbanization and distribution of the population

There is one spreadsheet providing several indices for analysing the urbanization process in a population **(URBINDEX)**.

12.2.7 Other spreadsheets

There are four additional spreadsheets that can be used for several purposes. These are listed below:

CTBL32

Uses a technique for adjusting information in a table to a set of desired marginal totals (totals of rows and/or columns).

FITLGSTC

Fits a logistic function to three (or a multiple of three) observed equidistant values of any index. It does not need asymptotic values.

LOGISTIC

Fits a logistic function to values of any index given two or more observed values of the index and the two asymptotes of the logistic.

SP

Constructs a stable population based on life tables by sex- and age-specific fertility rates (or the intrinsic growth rate).

12.3 MORTPAK

NOTES

The Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat has long conducted demographic estimation and projection activities at the country level, incorporating methodological advances in the construction of model life tables, for example. As a by-product of these activities, this extensive body of computer software has been developed. MORTPAK has already been well tested and is now widely used for analysis of developing country data and in developing country institutions. The design of the applications in MORTPAK as well as the program MATCH has its origins in the United States Census Bureau package, Computer Programs for Demographic Analysis (Arriaga, Anderson and Heligman, 1976). Version 4 of MORTPAK presently being used widely consists of a set of 17 computer programs for undertaking demographic analyses in developing countries, including empirical and model life table construction, graduation of mortality data, mortality and fertility estimation, evaluation of census coverage and age distributions and population projections. The 17 demographic procedures included have been selected by the Population Division as useful for evaluating demographic data from censuses and surveys and preparing reliable estimates of demographic parameters. These procedures incorporate techniques for evaluation and estimation of demographic data, particularly those techniques that incorporate the United Nations model life table system (United Nations, 1982) and generalized stable population equations (Preston and Coale, 1982).

When selecting a new application from the menu, a window in table form presents a brief description of the procedures, categorized according to their major functions: life table and stable population construction, model life table construction, graduation of mortality data, indirect mortality estimation, indirect fertility estimation, other estimation procedures and population projections. The package emphasizes mortality estimation, reflecting the larger number of techniques available and the further advanced mortality estimation compared to that of other demographic components. (Of the nine chapters in the United Nations manual on *Indirect Techniques for Demographic Estimation* (United Nations, 1983), five are dedicated solely, and two partially, to mortality analysis. These are covered in MORTPAK.)

In MORTPAK the LIFTB and STABLE programs calculate empirical life tables and stable populations respectively based on age-specific mortality rates, plus, in the latter case, intrinsic growth rate. The life table method used is based on the approach of Greville (1943), which permits calculation of age-specific separation factors based on the age pattern of the mortality rates themselves. It is, hence, potentially more accurate than methods which assume constant separation factors, and more robust, under developing country circumstances than methods which estimate separation factors based on population age distributions. Although fertility decline is rendering calculation of stable populations less applicable for many countries, for others fertility has changed little and stable population analysis remains useful for evaluation of age distributions and rough approximation of birth and death rates. However, for countries whose fertility decline is recent and mortality change has not greatly altered the adult age distribution, STABLE could be useful for evaluating age distributions and studying population dynamics among adults. In addition, the STABLE program is useful for static simulation of the effects of changed growth rates and/or mortality rates on age distribution.

The applications MATCH, COMPAR and BESTFT construct model life tables and compare or graduate empirical data with respect to a model life table. The procedure MATCH not only generates any United Nations or Coale and Demeny model life table but also enables the entering of a user-designated mortality pattern which then can be adjusted to correspond to any desired level. This user-designated model may be a pattern from a third model life-table system such as the Brass standard (Brass and others, 1968) but perhaps, most importantly, can be an age pattern of mortality for a particular country. In the latter case a demographer can generate a model life-table system specific for the country of interest by using MATCH to construct a series of life tables at different levels of life expectancy, all consistent with the country's average pattern. Comparison of an empirical set of age-specific mortality rates to model life-table patterns, through COMPAR, aids the demographer in the choice of a model life table. However, as data quality improves, the demographer will wish to retain as many characteristics of the original data as possible. COMPAR is then very useful for examining deviations of empirical mortality patterns from the models due to either true difference in age patterns or to data errors. Similarly, BESTFT offers the opportunity to graduate observed age-specific mortality rates with respect to a model life table (standard), either to smoothen a series of observed rates or to estimate consistent rates for age groups in which data are lacking.

NOTES

The procedures UNABR and ICM graduate mortality rates in traditional age grouping into single-year values; UNABR considers the entire age range, and ICM under age 10 only. The procedures are of immediate use when undertaking single-year population projections or special studies of specific age groups such as the school-age population or the elderly.

The next group of programs all relate to indirect estimation of demographic parameters. The five procedures of CEBCS, ORPHAN, WIDOW, COMBIN and BENHR are mortality-specific. CEBCS provides estimates of infant and child mortality based on data of children ever born and children surviving tabulated either by age of mother or duration of her marriage. ORPHAN and WIDOW carry out variations of the maternal orphanhood or widowhood techniques to estimate levels of adult mortality. The procedure COMBIN 'combines' early age mortality estimates (perhaps produced by CEBCS) with adult mortality estimates (perhaps produced by ORPHAN and WIDOW) and produces a full, consistent life table. The technique BENHR is an application of the Bennett-Horiuchi (1981) technique; it exploits the generalized stable population equation to estimate the completeness of death registration using population age distributions from two censuses and intercensal registered deaths.

Two fertility estimation techniques are included. FERTCB estimates age-specific fertility rates based on tabulations of average number of children ever born by age of woman. The essential methodology was developed by G. Mortara (1949). The variation included here was proposed by Arriaga (1983); it has the advantage of providing estimates of fertility change over time. In the same 1983 article, Arriaga presented an extension of the P/F technique originally developed by Brass (Brass and others, 1968). The Arriaga extension, presented in FERTPF, allows the demographer to estimate fertility at two points in time under conditions of fertility change. Children ever born data and the pattern of age-specific fertility are necessary from two enumerations when fertility has not been constant.

NOTES

CENCT and PRESTO provide techniques for evaluating relative coverage and age recording in censuses, as well as estimates of intercensal mortality and fertility. CENCT provides an estimate of the coverage of one census relative to another and hence is an important first step before applying other estimation techniques which assume consistency in coverage between two censuses (such as BENHR and PRESTO). Based on two populations, tabulated by age, and the appropriate model life table, PRESTO enacts the 'integrated method' developed by Preston (1983), providing consistent estimates of the birth rate, life expectancy and intercensal age distributions. Finally a simple and easy-to-use population projection program is included.

PROJCT carries out a single-year projection of a population by age and sex, based on initial male and female populations in 5-year age groups and assumed levels and changes in fertility, mortality and migration.

12.4 The SPECTRUM System

SPECTRUM is a system of policy models that support analysis, planning and advocacy for health programs. It is used to project future needs and examine the effects of policy options. SPECTRUM contains the following five components:

- **DemProj:** *Demography*. DemProj projects the population for an entire country or region by age and sex, based on assumptions about fertility, mortality and migration. A full set of demographic indicators can be displayed for up to 100 years into the future. Urban and rural projections can also be prepared. EasyProj supplies the data needed to make a population projection from the estimates produced by the Population Division of the United Nations.
- **FamPlan:** *Family Planning*. FamPlan projects family planning requirements needed to reach national goals for addressing unmet need or achieving desired fertility. It can be

used to set realistic goals, to plan for the service expansion required to meet program objectives, and to evaluate alternative methods of achieving goals. The program uses assumptions about the proximate determinants of fertility and the characteristics of the family planning program (method mix, source mix, discontinuation rates) to calculate the cost and the number of users and acceptors of different methods by source.

- **AIM:** *AIDS Impact Model*. AIM projects the consequences of the HIV epidemic, including the number of people living with HIV, new infections and AIDS deaths by age and sex, as well as the new cases of tuberculosis and AIDS orphans. AIM is used by UNAIDS to make the national and regional estimates it releases every two years.
- **Goals:** *HIV Vaccine*. This model explores the impact of potential HIV vaccines on the epidemic.
- **LiST:** *Lives Saved Tool (LiST - Child Survival)*. A program to project the changes in child survival in accordance with changes in coverage of different child health interventions.
- **RAPID:** *Resources for the Awareness of Population Impacts on Development*. RAPID projects the social and economic consequences of high fertility and rapid population growth for such sectors as labour, education, health, urbanization and agriculture. This program is used to raise policy makers' awareness of the importance of fertility and population growth as factors in social and economic development.
- **RNM:** *Resource Needs Model*. This model is used to calculate the funding required for an expanded response to HIV/AIDS at the national level.
- **TB:** This model projects the consequences of incident TB, including the number of cases notified for treatment, the number of multiple-drug resistant cases and the number of deaths due to TB. It further projects deaths averted in accordance with increasing case detection.
- **Safe Motherhood Model:** This model estimates the impact of various scores from the Maternal-Neonatal Program Index (MNPI) on a country's maternal mortality ratio. The MNPI is an index of 81 indicators for national efforts to improve maternal and neonatal health services. The model helps managers to gain a better understanding of the impact of policies, budgets and service delivery improvements on maternal health outcomes.
- **Allocate:** Allocate examines the linkages and interactions among three main areas of a representative reproductive health action plan (RHAP): family planning, safe motherhood and post-abortion care. Allocate also shows the interactive impact of changing decisions about levels of funding.

NOTES

12.4.1 Uses of SPECTRUM policy models

Policy models are designed to answer a number of 'what if' questions relevant to entities as small as local providers of primary health care services and as large as international development assistance agencies. The 'what if' refers to factors that can be changed or influenced by public policy.

Models are commonly computerized when analysts need to see the likely result of two or more forces that might be brought to bear on an outcome, such as a population's illness level or its degree of urbanization. Whenever at least three variables are involved (such as two forces and one outcome), a computerized model can both reduce the burden of manipulating those variables and present the results in an accessible way.

Policy issues commonly addressed by the SPECTRUM set of models include:

1. The utility of taking actions earlier rather than later. Modelling shows that little in a country stands still while policy decisions are stalled, and that many negative outcomes can accumulate during a period of policy stasis.
2. The evaluation of the costs versus the benefits of a course of action. Modelling can show the economic efficiency of a set of actions (i.e. whether certain outcomes are achieved more effectively than under a different set of actions), or simply whether the cost of a single set of actions is acceptable for the benefits gained.
3. The recognition of inter-relatedness. Modelling can show how making a change in one area of population dynamics (such as migration rates) may necessitate changes in a number of other areas (such as marriage rates, timing of childbearing etc.).
4. The need to discard monolithic explanations and policy initiatives. Modelling can demonstrate that simplistic explanations may bear little relationship to how the 'real world' operates.
5. The utility of 'door openers'. A set of policies under consideration may not be acceptable to all stakeholders. Modelling can concentrate on favoured goals and objectives and demonstrate how they are assisted by the proposed policies.
6. Few things in life operate in a linear fashion. A straight line rarely describes social or physical behaviour. Most particularly population growth, being exponential, is so far from linear that its results are startling. Modelling shows that all social sectors based on the size of population groups are heavily influenced by the exponential nature of growth over time.
7. A population's composition greatly influences its needs and its well-being. How a population is composed in terms of its age and sex distribution has broad-ranging consequences for social welfare, crime rates, disease transmission, political stability etc. Modelling demonstrates the degree to which a change in age and sex distribution can affect a range of social indicators.
8. The effort required to 'swim against the current'. A number of factors can make the success of a particular program harder to achieve; for example, the waning of breastfeeding in a population increases the need for contraceptive coverage. Modelling can illustrate the need for extra effort, even if simply to keep running in place.

NOTES

APPENDIX I

Key Concepts in Healthcare from Medical Sociology, Epidemiology and Healthcare Administration

Acute condition: Health condition characterized by episodic occurrence, relatively direct causation, relatively rapid onset, rapid progression and short duration and a disposition involving either recovery or death.

Age adjustment: Procedure whereby incidence and prevalence rates are adjusted to consider the age structure of the population being studied. This is one of the more common techniques used to standardize rates. Direct or indirect techniques of standardization are available.

Average hospital daily census: The average number of in-patients (excluding newborns) receiving care in a hospital each day during a particular reporting period.

Average length of stay (ALOS) in hospital: The average number of in-patient days recorded by hospitalized patients during a particular time period. ALOS is calculated by dividing the total number of patient days recorded during the time period by the number of patients discharged from the hospital.

Case finding: The various procedures utilized to determine the numerator data on the number of cases to be utilized in calculating incidence and prevalence rates. Case findings involve both determining what constitutes a 'case' and procedures for calculating the number of cases within the population at risk.

Chronic condition: Health condition characterized by a relatively complex etiology, slow onset and progression, extended (even lifelong) duration, and no clear-cut disposition. Chronic conditions typically cannot be cured, only managed.

Cohort: Refers in its broadest usage to any segment of the population that has some characteristic in common or has experienced a common event in the same time period. In epidemiology, cohorts refer most often to segments of the population that have been exposed to a certain health risk. In demography, cohorts refer most often to those born in the same year or period or those married in the same year or period. In either case, cohorts can be traced over time to determine changes in the composition of the group and the disposition of its members.

Disease: Technically, a scientific construct referring to a medical syndrome involving clinically identifiable and measurable signs and symptoms reflecting underlying biological pathology. The term disease is actually utilized in a much broader sense than this clinical conceptualization, often referring to any condition treatable by the health care system.

Endemic: Situation in which a pathological condition is common to a large portion of a population, to the extent that its presence might be considered 'normal'. The prevalence of endemic conditions does not fluctuate much over time.

Epidemic: Refers to a health condition not normally present within a population but whose appearance represents an 'outbreak' of the particular condition. Generally refers to a condition that is contagious or communicable (which contributes to its abnormally high, but usually short-lived occurrence).

Epidemiology: Literally, the study of epidemics, but has come to mean the study of the etiology, distribution and course of disease within a population.

Etiology: The cause of a health condition. Etiology may be relatively simple and direct as in the case of most acute conditions, or it may be complex and indirect as in the case of most chronic conditions.

Health status: Indicator of the overall state of health of an individual or, more often in health demography, a population. There is no one measure of health status, with existing health status indicators utilizing either outcome measures (e.g. morbidity and mortality) or utilization measures (e.g. physician office visits or hospital admissions).

Incidence: The rate at which the onset of new cases of a particular health condition occurs. Incidence is calculated based on the number of new cases diagnosed during a particular time period (usually 1 year) divided by the population at risk.

Illness: The existence of a clinically identifiable medical syndrome in an individual or a population. Social scientists often distinguish between illness and sickness, with the former referring to the presence of some biological pathology and the latter referring to the presence of some condition recognized by society as a state of ill health.

Morbidity: The level of sickness and disability existing within a population. There is no overall indicator of morbidity, so it is usually looked at in terms of the incidence or prevalence of specific conditions.

Mortality: Refers to the rate at which deaths occur within a population. Mortality rates are calculated by dividing the number of deaths occurring within a particular time period (usually 1 year) by the number of years lived by the population during that time period.

Occupancy rate: The proportion of a hospital's beds (or those of some other health care facility) that are occupied on an average during a particular time period. The occupancy rate is calculated by dividing the average daily census for a particular time period by the number of hospital beds available during that time period.

Population at risk: The portion of a population that has been exposed to a particular health threat or is susceptible to a particular health threat. The population at risk is used as the denominator in calculating incidence and prevalence rates when the condition in question does not affect the total population.

Prevalence: The total number of cases of a particular health condition within a population at a particular point in time. Prevalence is calculated by dividing the number of known cases at a particular point of time by the population at risk at that point of time.

Relative risk: The probability of the occurrence of a particular health condition within a population relative to the risk for some other population. Relative risk is calculated by determining how much more likely a condition is to occur among one population (e.g. smokers) compared to another (e.g. non-smokers). The relative risk is often contrasted to the absolute risk of the occurrence of a condition.

Sickness: The presence of ill health in individuals or populations based on whatever definition the particular population uses for ill health. Social scientists distinguish between sickness and illness, with the former referring to the social construct of 'sickness' and the latter referring to the presence of measurable biological pathology.

Sign: A manifestation of a health condition or disease syndrome that can be identified through clinical tests or through the observation of a health care professional.

Symptom: A manifestation of a health condition that is experienced by the affected individual. Symptoms are often 'internal' (e.g. pain) in the sense that they can only be identified by the individual.

APPENDIX II

Manuals of the United Nations on Population

Manual I. Methods for estimating total population for current dates

United Nations (1952). *Manual I: Methods for estimating total population for current dates* (United Nations Publications, Sales No. 52.XIII.5).

This report is an aid to demographers and population experts to analyse censuses, which became universal operations providing crucial information on national populations. However, the reliability and completeness of census results remain uneven and in several countries only one census has been implemented. This is the main reason why the census-based methods of estimating population size and particularly the methods for adjusting incomplete and deficient census data and extrapolating census results presented in *Manual I*, which has been compiled in the 1950s, remain relevant. The techniques to use incomplete and deficient data for estimating other demographic parameters from censuses and other sources are presented in *Manuals II* and *IV*.

Manual II. Methods of appraisal of quality of basic data for population estimates

United Nations (1955). *Manual II: Methods of appraisal of quality of basic data for population estimates* (United Nations Publications, Sales No. 56.XIII.2).

This manual is an aid to demographers and population experts to appraise the reliability of population data. It overviews the factors determining the quality of data on population size, age and sex distributions and migrations derived from censuses, surveys, vital registration and other sources of data. The manual presents methods for appraising completeness, accuracy and consistency of data from these sources. *Manual II* is closely related to the indirect techniques for demographic estimations presented in the *Manuals I, IV* and *X*.

Manual III. Methods for population projections by sex and age

United Nations (1956). *Manual III: Methods for population projections by sex and age* (United Nations publication, Sales No. 56.XIII.3).

This manual deals with methods of producing population projections, including for countries where the available statistics are incomplete or unreliable. The methods are described in detail and their application under various conditions is illustrated. A specific example is presented and carried through all the successive steps in the making of a population projection. *Manual III* is closely related to other manuals of this cluster, as well as to most manuals of the clusters demographic estimation and demographic models.

Manual IV. Methods of estimating basic demographic measures from incomplete data

United Nations (1967). *Manual IV: Methods of estimating basic demographic measures from incomplete data* (United Nations Publications, Sales No. 67.XIII.2).

This manual presents a variety of methods to estimate fertility and mortality from census and survey data. It combines theoretical exposition with non-technical, do-it-yourself instructions such that all the methods can be employed by persons with limited expertise in demography. Examples of actual application of the most important methods are also provided. *Manual IV* is closely related to other manuals in this cluster, in particular *Manuals II* and *X* as well as to *Manual III* and the publications on stable population and model life tables in the cluster demographic models.

Manual V. Methods of projecting the economically active population

United Nations (1971). *Manual V: Methods of projecting the economically active population*
(United Nations publication, Sales No. E.70.XIII. 2).

This manual is an aid to demographers and economists to analyse levels and past trends of labour supply and demand and prepare projections for 5–10 year period. Projection techniques include extrapolations of past trends and more complex methods which take into account major factors affecting labour supply and demand. The applicability, advantages and drawbacks of each method are described with particular attention to data availability issues in developing countries. *Manual V* is related to the other manuals in this cluster, as well as to selected ones of the cluster demographic estimation and demographic models.

Manual VI. Methods of measuring internal migration

United Nations (1970). *Manual VI: Methods of measuring internal migration*
(United Nations Publications, Sales No. E.70.XIII.3).

This manual is an aid to demographers and population experts to estimate flows of internal migration and stocks of migrants. The manual explains the use of censuses, population registers and sample surveys as sources of data on migration and introduces direct and indirect methods to derive varied indicators. *Manual VI* is related to *Manual II*.

Manual VII. Methods of projecting households and families

United Nations (1973). *Manual VII: Methods of projecting households and families*
(United Nations publication, Sales No. E.73.XIII.2).

This manual is an aid to demographers and population experts to project households and families using data from population and housing censuses. It explains extrapolative and regression projection methods based on headship rate and alternative indicators. Although the illustrative examples relate mostly to developed countries, the applicability of methods for developing countries is discussed. *Manual VII* is related to *Projection Methods for Integrating Population Variables into Development Planning, Modules One and Three*.

Manual VIII. Methods for projections of urban and rural population

United Nations (1974). *Manual VIII: Methods for projections of urban and rural population*
(United Nations publication, Sales No. E.74.XIII.3).

This manual is addressed mainly to population analysts possessing limited technical means especially with a view to its use in less developed countries. While it does not rely on computer technology, the suggested methods may be applied using simple software. The projections of urban and rural population are explained on the assumption that methods of projection of a country's total population, or its total population by groups of sex and age, are already known, and that such projections have in fact already been implemented. Those methods have been dealt with in *Manual III*. It is assumed that the reader is familiar with the appraisal of accuracy in basic statistics, a subject elaborated in *Manual II*. *Manual VIII* is also related to *Projection Methods for Integrating Population Variables into Development Planning*.

Manual IX. The methodology of measuring the impact of family planning programmes on fertility

United Nations (1979). *Manual IX: The methodology of measuring the impact of family planning programmes on fertility* (United Nations publication, Sales No. E.78.XIII.8).

The purpose of *Manual IX* and *Addendum* is to provide demographers and population specialists with methods to estimate family planning programme impact and guidelines for their practical application. Although the publication was compiled two to three decades ago, most methods presented remain relevant. Some methods require limited data and apply simple techniques but they produce crude and not always reliable estimates. Other, more subtle and data-exacting methods yield more robust and nuanced information, but they are more complex and cannot be reduced to an unambiguous application of step-by-step approach. Most chapters include illustrations utilizing the actual statistics of particular countries. Some subjects developed in *Manual IX* have been presented in *Methods of measuring the impact of family planning programmes on fertility: problems and issues*.

Manual X. Indirect techniques for demographic estimation

United Nations (1983). *Manual X: Indirect techniques for demographic estimation*
(United Nations publication, Sales No. E.83.XIII.2).

This report is an aid to demographers and population experts to carry out the best possible evaluation and exploitation of data sources, especially those that are incomplete or deficient. *Manual X* describes a wide range of time-tested techniques to make indirect estimates of demographic parameters. Each of the techniques presented is based on a mathematical model and explained in easy-to-follow examples.

Bibliography

- Arriaga, E.E. (1994). *Population Analysis with Microcomputers Presentation of Techniques*, Vol. 1 and Bureau of Census, USAID and UNFPA, November.
- Bogue, D. (1969). *Principles of Demography*, New York: John Wiley and Sons.
- Bongaarts, J. (1978). 'A Framework for Analyzing the Proximate Determinants of Fertility'. *Population and Development Review*, 4: 105-32.
- Brass, W. (1971). 'On the Scale of Mortality', in Brass, W. (ed.). *Biological Aspects of Demography*, London: Taylor & Francis.
- Caldwell, J.C. (1976). 'Towards a Restatement of Demographic Transition Theory'. *Population and Development Review*, 2: 579-616.
- Chandrasekaran, C. and W.E. Deming (1949). 'On a Method of Estimating Birth and Death Rates and the Extent of Registration'. *Journal of the American Statistical Association*, 44: 101-15.
- Chiang, C.L. (1974). *Life Tables and Mortality Analysis*, Geneva: WHO.
- Coale, A.J. (1971). 'Age Patterns of Marriage'. *Population Studies*, 25: 192- 214.
- -------------(1973) 'The Demographic Transition'. *Proceedings of the 1973 IUSSP Conference*, Liege, 1: 53-72.
- Coale, A.J. and P. Demeny (1966). *Regional Model Life Tables and Stable Populations*, Princeton, NJ: Princeton University Press.
- Coale, A.J. and Trussel (1974). 'Model Fertility Schedules: Variations in the Age Structure of Childbearing in Human Populations'. *Population Index*, 40: 185-258.
- Davis, K. and J. Blake (1956). 'Social Structure and Fertility: An Analytical Framework'. *Economic Development and Cultural Change*, 4: 211-35.
- Dublin, L.I. and A.J. Lotka (1925). 'On the True Rate of Natural Increase, as Exemplified by the Population of the United States, 1920'. *Journal of the American Statistical Association*, 20: 305-39.
- Easterlin, R.A. (1969). 'Towards a Socio-economic Theory of Fertility' in Behram et al. (eds). *Fertility and Family Planning: A World View*, Ann Arbor: The University of Michigan Press, pp. 127-56.
- Easterlin, R.A. and E.M. Crimmins (1985). *Fertility Revolution: A Supply Demand Analysis*, Chicago: The University of Chicago Press.
- ------------(1975). 'Economic Framework for Fertility Analysis'. *Studies in Family Planning*, 6:54-63.
- Euler, L. (1760). 'A General Investigation into the Mortality and Multiplication of the Human Species' (English translation). *Theoretical Population Biology*, 1 (1970): 307-14. Also reprinted in Smith, D.P. and N. Keyfitz (1977). *Mathematical Demography: Selected Papers*, Berlin: Springer-Verlag.
- Frijka, T. (1973). *The Future of Population Growth-Alternative Paths to Equilibrium*, New York: John Wiley and Sons.
- Glass, D.V. (1950). `Graunt's Life Tables'. *Journal of the Institute of Actuaries*, 76:60-64.

- Hajnal, J. (1953). 'Age at Marriage and Proportions Marrying'. *Population Studies*, 7: 115-36.
- Henry, L. (1961). 'Some Data on Natural Fertility'. *Eugenics Quarterly*, 8: 81-91.
- ------------(1972). *On the Measurement of Human Fertility*. Selected Writings of L. Henry translated and edited by M.C. Slaps and E. Lapierre-Adamcyk, Amsterdam: Elsevier Publishing Company.
- Hobcraft, J.N. (1992). 'Fertility Patterns and Child Survival'. *Population Bulletin*, United Nations, 33
- Hobcraft, J.N. and R.J.A. Little (1984). 'Fertility Exposure Analysis: A New Method for Fertility Differentials'. *Population Studies*, 30: 21-45.
- International Institute for Population Sciences, Mumbai. (1994–2002). Various Manuals prepared on different topic for the Course on Master in Population Studies through distance education as teaching materials, Mumbai.
- Keyfiitz, N. (1968). *Introduction to the Mathematics of Population*, Reading, Mass: Addison-Wesley.
- Keyfitz, N. and W.I. Fliger (1971). *Population: Facts and Methods of Demography*, San Francisco: Freeman.
- Klaff, V.Z. (1992). *DEM-LAB: Teaching Demography through Computers*, New Jersey: Prentice-Hall.
- Knodel, J. and E. Van de walle (1979). 'Lesson from the Past: Policy Implications of Historical Fertility Studies'. *Population and Development Review*, 5.
- Kuczynski, R.R. (1935). *The Measurement of Population Growth; Methods and Results*, London: Sidgwick and Jackson.
- Leete, R. (1992). *PEOPLE: A Program for Population Projections*, California: Metagraphica Software Corporation, CA 95066.
- Lotka, A.J. (1922). 'The Stability of the Normal Age Distribution'. *Proceedings of the National Academy of Sciences*, 8: 339-44.
- Malthus, T.R. (1797). Reprint in 1967. *Population: The First Essay* with a foreword by Kenneth E. Boulding, Ann Arbor: University of Michigan Press.
- ------------(1798). Reprint in 1970. *An Essay on the Principle of Population* (ed. A. Flew), London: Penguin.
- Menken, J.A. (1978). 'Current Status of Demographic Models'. *Population Bulletin of the United Nations*, 9: 22-34.
- Notestein, F.W. (1945). 'Population: The Long View' in Schultz, T.W. (ed.) *Food for the World*, Chicago: Chicago University Press.
- Novell, C. (1988). *Methods and Models in Demography*, London: Bellhaven Press.
- Pearl, R. and Reed, L.J. (1920). 'On the Rate of Growth of the Population of the United States since 1790 and its Mathematical Representation'. *Proceedings of the National Academy of Science*, 6: 275-88.
- Registrar General of India (2002). *Census of India 1991, Social and Cultural Tables*, Part II C, New Delhi: Ministry of Home Affairs.
- Census of India (1991). *Final Population Totals: Brief Analysis of Primary Census Abstract*. India Paper 2 of 1992, New Delhi: Ministry of Home Affairs.
- ------------(1994). *Sample Registration System 1992*, New Delhi: Ministry of Home Affairs.
- ------------(1995). *Survey of Causes of Death (Rural), India Annual Report, 1992*, New Delhi: Ministry of Home Affairs.
- ------------(2006). *Population Projections for India and States, 2006-2026*, New Delhi: Ministry of Home Affairs.
- ------------(2007). *SRS based abridged life tables 2001-2005*, Analytical Studies, Report No. 3 of 2007, New Delhi.
- Rowland, D. (2005), *Demographic Techniques*, Oxford University Press.

- Shryock, H.S. and J.S. Sliegel (1971). *The Methods and Materials in Demography* (Vol. I and II), Washington DC, US Bureau of Census.
- Simon, J. (1981). *The Ultimate Resource*, Princeton, NJ: Princeton University Press.
- Srinivasan, K. (1980). 'Birth Interval Analysis in Fertility Surveys'. *World Fertility Survey, Scientific Reports*, No. 7, 1990.
- ------------(1998) 'Basic Demographic Techniques and Applications', New Delhi: SAGE India.
- Stover, J. (1990). *Dem Pro: A Demographic Projection Model for Development Planning*, Washington: The Futures Group International Inc.
- United Nations (1955). 'Age and Sex Patterns of Mortality: Model Life Tables for Underdeveloped Countries'. Department of Social and Economic Affairs. *Population Studies*, 22.
- ------------(1958). *Principles and Recommendations for National Population Censuses*, ST/STAT/Ser/M/27, New York.
- ------------(2008). *Principles and Recommendations for National Population and Housing Censuses Revision 2*, ST/ESA/STAT/SER.M/67/Rev.2, New York.
- ------------(1973). 'The Determinants and Consequences of Population Trends'. Vol. I, ST/SCA/SER./No. 50. *Population Studies*, 50.
- ------------(1982). *'Model Life Tables for Developing Countries'.* Department of International Social and Economic Affairs. *Population Studies*, 77.
- ------------(1983). Manual X 'Indirect Techniques for Demographic Estimation'. Department of International Social and Economic Affairs. *Population Studies*, 81.
- United States Bureau of Census (1994). *PAS: Population Analysis Spreadsheets*, Washington DC: A Computer Diskette.
- Van de walle. (1982). *Multilingual Demographic Dictionary* (English Section), Liege: IUSSP Ordina.
- Whelpton, P.K. (1954). *Cohort Fertility: Native White Women in the United States*, Princeton, NJ: Princeton University Press.
- Wilmoth J, et al. (2012). 'A flexible two-dimensional model for use in indirect estimation', *Population Studies*, 61: 1-28.

Office of the Registrar General &
Census Commissioner, India
Ministry of Home Affairs
Government of India
2/A Man Singh Road, New Delhi-110001
Website: www.censusindia.gov.in