2
The Measurement of Mortality

2.1 Introduction

This chapter introduces the measurement of mortality by considering in detail the various kinds of mortality rate used by demographers. In Section 2.2 the crude death rate is described, and in Section 2.3 the calculation of age-specific death rates is illustrated. Section 2.4 then explains the rationale behind, and the difference between, two types of mortality rate commonly used by demographers: initial rates and central rates. These two types of rate are shown to be manifestations of two different approaches to analysing demographic data: that based on time periods; and that based on birth cohorts. Section 2.5 introduces the Lexis chart as a means of representing and illustrating the difference between initial and central rates, and thereby between the period and cohort approaches. In Section 2.6 the formula which is commonly used to convert age-specific death rates of one type into the other is derived, with the aid of a Lexis chart. Finally, Section 2.7 summarizes the advantages and disadvantages of the two types of mortality rate.

2.2 The crude death rate

The simplest measure of mortality is the number of deaths. However, this is not of much use for practical purposes since it is heavily influenced by the number of people who are at risk of dying.

Because of this, as we saw in Section 1.4, demographers typically measure mortality using *rates*. A death rate is defined as

$$\text{death rate} = \frac{\text{number of deaths in a specified time period}}{\begin{array}{c}\text{number of people exposed to the risk of dying}\\ \text{during that time period}\end{array}}.$$

Thus, in order to measure mortality, data are required about the number of deaths, and about the number of people exposed to the risk of dying. Data on the number of deaths are usually obtained from death registers, and data on the number of people exposed to the risk of dying are typically obtained from a population census. Of course, survey data may also be used, especially in countries where death registration is deficient, or the quality of census data is suspect.

The simplest conceivable death rate is probably the total number of deaths in a given time period divided by the total population. This measure is called the *crude death rate*. The time period used is typically one calendar year. Thus

$$\text{crude death rate} = \frac{\text{total number of deaths in a given year}}{\text{total population}}.$$

An immediate issue arises with the measurement of the total population. During any year, the population will usually change. At what point in the year, therefore, should it be measured? Conventionally, the point chosen is half-way through the year (30 June). The population on 30 June is called the *mid-year population*. Using this definition of the population exposed to the risk of dying, therefore,

$$\text{crude death rate} = \frac{\text{total number of deaths in a given year}}{\text{total mid-year population}}.$$

Denoting the crude death rate in year t by the symbol d_t, the total number of deaths in year t by θ_t, and the total population on 30 June in year t by P_t, we can write

$$d_t = \frac{\theta_t}{P_t}.$$

Now, for simplicity, the subscripts t are usually omitted because, unless otherwise stated, the period of time over which the crude death rate is measured may be assumed to be a single calendar year. Thus

$$d = \frac{\theta}{P}.$$

Since death is a relatively rare event in most populations, the crude death rate is often small. For this reason, it is often expressed as *the number of deaths per thousand of the population*, or

$$d = \frac{\theta}{P} \times 1000.$$

Thus, for example, the population of Peru on 30 June 1989 has been estimated to be 21 113 000 (excluding some Indian people in remote areas). It is estimated that there were 200 468 deaths in Peru in 1989. The crude death rate in Peru in 1989 is therefore equal to 200 468/21 113 000, which is 0.00950, or, multiplying by 1000, 9.5 per thousand.

2.3 Age-specific death rates

The crude death rate does not provide a great deal of information about mortality. In particular, the risk of dying varies greatly with age, and the crude death rate indicates nothing about this variation. Because of this, demographers often find it useful to use *age-specific death rates*. The age-specific death rate at age x years is defined as

$$\begin{array}{c}\text{age-specific death rate at}\\ \text{age } x \text{ years}\end{array} = \frac{\text{number of deaths of people aged } x \text{ years}}{\text{population aged } x \text{ years}},$$

in a given calendar year. When we refer to 'age x years', we mean 'aged x last birthday'. The denominator, as before, is the mid-year population.

Denoting the age-specific death rate at age x years last birthday by the symbol m_x, the number of deaths of people aged x years last birthday by θ_x, and the population aged x

years last birthday by P_x, we can write

$$m_x = \frac{\theta_x}{P_x},$$

or, if preferred,

$$m_x = \frac{\theta_x}{P_x} \times 1000.$$

Note that the subscripts x denote years of age, not calendar years.

Age-specific death rates can be calculated for single years of age, or for age groups, such as 5–9 years last birthday, 10–14 years last birthday, and so on. Because mortality is also known to vary by sex, age-specific death rates are usually calculated separately for males and females. When age-specific rates are calculated for age groups, a special notation is used to denote the precise age group under consideration. The symbol ${}_n\theta_x$ denotes the number of deaths to people between the exact ages x and $x+n$ years. The symbol ${}_nP_x$ is used to denote the mid-year population of people between the exact ages x and $x+n$ years, and the symbol ${}_nm_x$ denotes the age-specific death rate between exact ages x and $x+n$ years. Thus, for example, the age-specific death rate at ages 5–9 years last birthday, ${}_5m_5$, is calculated using the formula

$${}_5m_5 = \frac{{}_5\theta_5}{{}_5P_5}.$$

To take an example, the male population aged 35–44 years last birthday in England and Wales on 30 June 1995 is estimated to have been 3 333 000. The number of deaths reported in England and Wales of males in this age group during the calendar year 1995 was 5860. The age-specific death rate in 1995 for males aged 35–44 years last birthday was, therefore, 5860/3 333 000, or 0.00176. Multiplying this by 1000 gives a rate of 1.76 per thousand.

There is one (and only one) age group for which a different method of calculating age-specific death rates is employed. This is the age group 'under 1 year', or '0 last birthday'. For this age group, the denominator is taken to be the number of live births in the calendar year in question, rather than the mid-year population aged under 1 year. For example, in England and Wales in 1995 there were 648 100 live births, and 3970 deaths to infants under 1 year. The infant mortality rate is therefore equal to 3970/648 100, which is 0.00613 or 6.13 per thousand births. Notice that this rate refers to both sexes. To measure infant mortality, unlike that of other age groups, demographers quite often use a rate referring to both sexes combined.

2.4 The two types of mortality rate

So far, we have been looking at rates in which the denominator is a mid-year population, and the numerator is the number of deaths during the whole of the relevant calendar year.

This procedure violates the principle of correspondence, described in Section 1.4. Why? Two important reasons are as follows:

1 Someone who dies in the relevant year, but before 30 June, will not be alive on that date, and will not be included in the mid-year population, yet that person's death will be included in the numerator.
2 Consider someone whose birthday is on 7 September, and who dies on 9 October. Suppose this person is aged x years last birthday on 30 June. Then when he dies he

will be aged $x + 1$ years last birthday. He will be included in the denominator of the age-specific rate at age x last birthday, but his death will be included in the numerator of the age-specific rate at age $x + 1$ last birthday.

A similar problem affects the calculation of age-specific death rates for infants aged under 1 year when the number of births during the entire year is the denominator. Consider an infant who died on 31 March 1997 aged nine months. This child was born on 30 June 1996. His/her death is included in the numerator for the age-specific death rate at age 0 last birthday in 1997, but his/her birth is included in the denominator for the age-specific death rate at age 0 last birthday in 1996.

What can be done about this? Can rates be obtained in which the numerator and denominator correspond exactly? Yes, they can, but they require additional data. What we really need is to know the exact period of exposure at each age during the given year for each person at risk of dying. Thus, for example, the person whose $(x + 1)$th birthday was on 7 September, and who died on 9 October, would be regarded as contributing 250/365 of a year's exposure during that year at age x last birthday (since there are 250 days between 1 January and 7 September), and 32/365 of a year's exposure during that year at age $x + 1$ last birthday (since there are 32 days between 7 September and 9 October). Summing these fractions of a year over the whole population under investigation for each year of age, and using the result in the denominator, would give a rate in which the numerator and denominator corresponded exactly.

In practice, such detailed information is not usually available except from special (and expensive) investigations designed to elicit it. Therefore demographers rely on mid-year populations as an *approximation* to the correct exposed-to-risk. The approximation is usually quite close in large populations.

There is, however, another approach to measuring mortality rates, which does not lead to violations of the principle of correspondence (at least at the 'person level'). In this second approach, what is done is to calculate the number of people who have their xth birthday during a given period, and then follow them up until either they celebrate their next birthday, or they die (whichever happens first). Dividing the number who die by the original number having their xth birthday gives us an age-specific death rate at age x.

This kind of age-specific death rate is often called a *q-type rate*, and is given the symbol q_x, to distinguish it from the first kind of age-specific rate, m_x, which is called an *m-type rate*. Demographers also use the terms *initial rates* for q-type rates and *central rates* for m-type rates. This is because in q-type rates the exposed-to-risk is defined at the start, or initiation, of the year of age under investigation (that is, when the members of the exposed-to-risk celebrate their xth birthday), whereas in m-type rates the exposed-to-risk is an estimate of the number of persons aged x last birthday at the time the events took place. The average age of these persons is $x + \frac{1}{2}$ years: they are half-way through (in the 'centre' of) the year of age in question.

Strictly speaking, q-type rates do not lead to an exposed-to-risk which is exactly right. Those people who die between exact ages x and $x + 1$ years are actually only 'at risk' of dying for the period between their xth birthday and the point at which they die (since once they have died, they are no longer at risk). The q-type rate assumes that such people are at risk for the entire year between exact ages x and $x + 1$. Thus q-type rates over-estimate the length of time exposed to risk for every person who dies. Nevertheless, they do get the right number of people in the denominator, which m-type rates calculated using mid-year populations cannot be relied upon to do. That is what is meant by saying that q-type rates maintain the principle of correspondence at the 'person level'.

THE DIFFERENCE BETWEEN THE TWO TYPES OF RATE

The two types of mortality rate are examples of two quite different approaches to measuring the components of population change. One approach calculates rates based on a specific calendar time period (m-type rates). This is known as the *period approach.* The other approach calculates rates based on the experience of a specific group of people born during a specific calendar period (q-type rates). Since q-type rates are based on a group of people who celebrate their xth birthday during a given period, it follows that they must all have been born during a period of the same length x years earlier. Such a group of people is known as a *birth cohort*, and this approach is called the *cohort approach.*

2.5 The Lexis chart

The difference between the period and cohort approaches, and hence between m-type and q-type rates, may be illustrated using a diagram called a *Lexis chart.* A Lexis chart has a vertical axis which represents age, and a horizontal axis which represents calendar time. Since people get older as time goes on, the life of any person can be represented on a Lexis chart by a diagonal line running from the horizontal axis until a point which corresponds to the person's age at death measured on the vertical axis (Figure 2.1).

On a Lexis chart, the population alive and aged x last birthday, P_x, at a particular point in time is represented by a vertical line. In Figure 2.2, the line AB represents those alive aged 30 years last birthday on 30 June 1997. Vertical lines like line AB thus represent the denominators of m-type mortality rates.

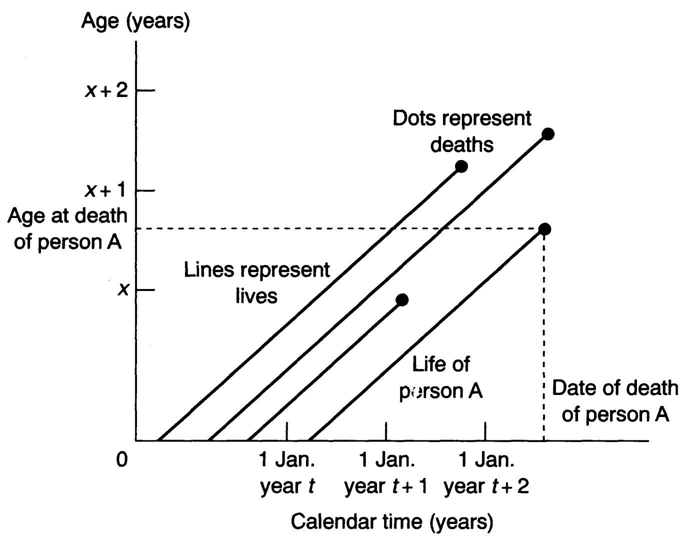

Figure 2.1 Principle behind the Lexis chart. Individual lives are represented by diagonal lines running from the bottom left towards the top right of the chart. As individuals grow older, they move up their 'life lines'. The coordinates of the upper end of each line denote the time of death and the age of the person when he/she died

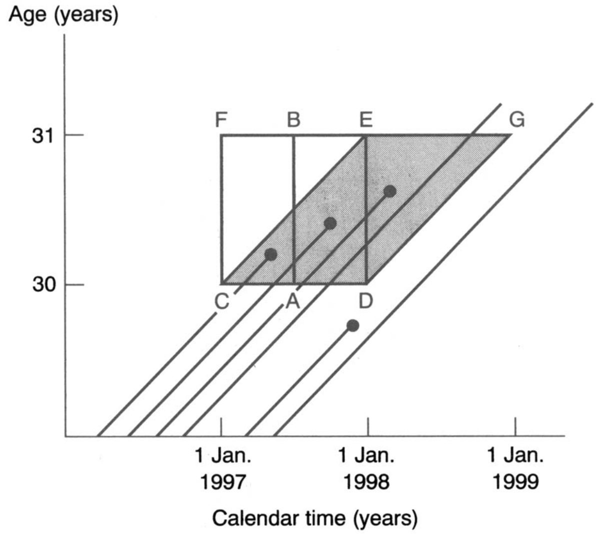

Figure 2.2 A Lexis chart

A set of people who celebrate their xth birthday during a particular time period (which means that they were all born during a particular time period and thus constitute a birth cohort) are represented by a horizontal line. In Figure 2.2, the horizontal line CD represents all the people who celebrated their 30th birthday during the year 1997 (and who were, therefore, all born during 1967). Horizontal lines like line CD represent birth cohorts, or the denominators of q-type mortality rates.

The deaths of people aged x last birthday on their date of death, and who died in a particular time period, are represented by squares. The square CDEF in Figure 2.2 represents all the people who died during the calendar year 1997 who were aged 30 years last birthday when they died. Squares like this represent the numerators of m-type mortality rates.

The deaths of people aged x last birthday on their date of death, and who all celebrated their xth birthday during a particular time period, are represented by parallelograms. In Figure 2.2, the shaded parallelogram CDGE represents all the people who died aged 30 years last birthday when they died, and who celebrated their 30th birthday during the calendar year 1997. Parallelograms like this represent the numerators of q-type mortality rates.

2.6 The relationship between the two types of mortality rate

In principle, there is no necessary relationship between the two types of mortality rate. However, by making a number of assumptions, a theoretical relationship can be derived. Since the assumptions are not too unreasonable, the theoretical relationship works quite well in most practical situations.

Consider the Lexis chart in Figure 2.3. The deaths representing the numerator of the m-type mortality rate at age x last birthday in calendar year t are in the square PQRS. Suppose that there are θ_x of these deaths. The deaths representing the numerator of the q-type mortality rate at age x which most closely overlaps with the numerator of the m-type rate for year t are in the shaded parallelogram TMWN.

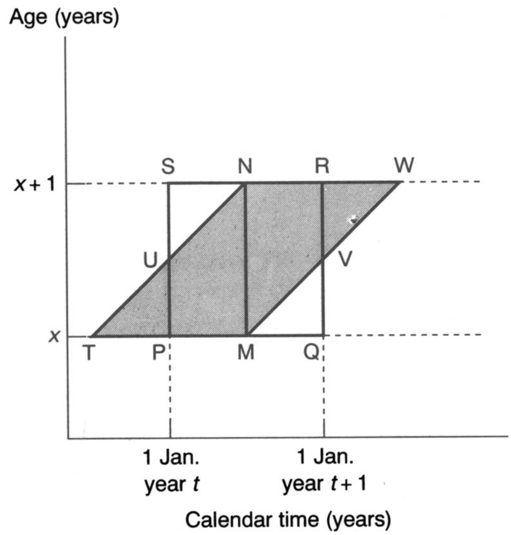

Figure 2.3 Lexis chart illustrating the relationship between the two types of mortality rate

We now make two assumptions. We assume that:

1 mortality only varies with age, and not with calendar time;
2 deaths are evenly distributed across each single year of age.

These two assumptions are made repeatedly in the demographic analysis of mortality. The assumption that mortality only varies with age is easy enough to understand, but the second assumption may need clarification. The assumption that deaths are evenly distributed across each year of age implies that, say, the average age at death of those dying between their 75th and 76th birthdays is 75 years 6 months. It implies that the number of deaths between 75 years exactly and 75 years 6 months is the same as the number of deaths between ages 75 years 6 months and 76 years exactly.

Once we have made these two assumptions, then, on the Lexis chart, within any horizontal band representing the ages between x and $x + 1$, deaths are evenly distributed. This means that the number of deaths is proportional to the area of any portion of the chart. Two sections of the chart which both lie within the horizontal band representing ages between x and $x + 1$ exactly and which have the same area will represent the same number of deaths.

Since the area of the parallelogram TMWN in Figure 2.3 is equal to the area of the square PQRS, then the number of deaths represented by the shaded parallelogram TMWN will also be θ_x.

The m-type mortality rate, m_x, is then given by

$$m_x = \frac{\theta_x}{\text{population represented by the vertical line MN}}, \tag{2.1}$$

and the q-type mortality rate, q_x, is given by

$$q_x = \frac{\theta_x}{\text{population represented by the horizontal line TM}}.$$

But, using the assumptions above, the number of deaths in the triangle TMN must be $\frac{1}{2}\theta_x$, and, since all the lives which cross the line TM also cross the line MN, we have

$$q_x = \frac{\theta_x}{\text{population represented by the vertical line MN} + \frac{1}{2}\theta_x}. \tag{2.2}$$

But, from equation (2.1) above,

$$\text{population represented by the vertical line MN} = \frac{\theta_x}{m_x}. \tag{2.3}$$

Thus, substituting from equation (2.3) into equation (2.2), we have

$$q_x = \frac{\theta_x}{\theta_x/m_x + \frac{1}{2}\theta_x},$$

and the θ_x cancel to leave

$$q_x = \frac{1}{1/m_x + \frac{1}{2}} = \frac{m_x}{1 + \frac{1}{2}m_x},$$

or, as it is often written,

$$q_x = \frac{2m_x}{2 + m_x}.$$

This result is dependent upon the two assumptions we have made. In many practical situations, however, the approximation is satisfactory.

There are certain age groups, though, in which the assumption that deaths are evenly distributed is not valid. This is particularly true of the first year of life. Most deaths to infants during the first year of life take place during the first few weeks of that year. Indeed, in low-mortality populations, it is usual for more than half of all the deaths to infants under the age of 1 year to occur during the first month of life (see Exercise 2.8). Deaths to infants during the first four weeks of life are known as *neonatal deaths*. Neonatal deaths may be measured using the *neonatal death rate*, defined as

$$\text{neonatal death rate} = \frac{\begin{array}{c}\text{number of deaths in a given year to}\\ \text{infants aged 28 days or under}\end{array}}{\text{number of births in the given year}}.$$

For example, in the United Kingdom in 1995, there were 732 000 live births, and 3070 neonatal deaths. The neonatal death rate was therefore equal to 3070/732 000, which is 0.0042, or 4.2 per thousand births.

2.7 Advantages and disadvantages of the two types of mortality rate

The two types of mortality rate have their advantages and disadvantages (Table 2.1). Generally speaking, m-type rates have the advantage of being straightforward to calculate from routinely available data. Their disadvantage is that they do not reflect the experience of 'real' people, and, if calculated using mid-year populations, violate the principle of correspondence. The advantages and disadvantages of q-type rates are in a sense 'mirror-images' of those of m-type rates.

Table 2.1 Advantages and disadvantages of the two types of mortality rate

Type	Advantages	Disadvantages
m-type	Data are readily available Easy to calculate Can be applied to a specific calendar time period	Violate the principle of correspondence if based on mid-year populations Do not reflect the experience of a real group of people
q-type	Reflect the experience of real people Do not violate the principle of correspondence (at least at the 'person level')	Do not apply to a particular calendar year Data are not readily available Awkward to calculate

Exercises

2.1 Table 2E.1 gives the total number of deaths in certain years, together with the estimated mid-year populations for those years, for certain countries in Latin America. Use them to calculate the crude death rate for each of these countries.

2.2 Table 2E.2 gives the estimated mid-year population in certain age groups, together with the number of deaths to people in those age groups, for males and females in Argentina in 1986. Use them to calculate age-specific death rates for the two sexes.

2.3 Table 2E.3 gives the estimated mid-year population in certain age groups, together with the number of deaths to persons in these age groups, in England and Wales in 1995. Use the data in the table to calculate age-specific death rates for the relevant age groups. Comment briefly on your results.

Table 2E.1

Country	Year	Estimated mid-year population	Number of deaths
Argentina	1990	32 322 000	295 796
Brazil	1989	147 404 000	1 164 452
Colombia	1990	32 987 000	201 166
Costa Rica	1991	3 064 000	12 452
Mexico	1991	87 836 000	500 615

Source: Wilkie *et al.* (1996, pp. 101, 102, 167).

Table 2E.2

Age group	Males		Females	
	Mid-year population (thousands)	Number of deaths	Mid-year population (thousands)	Number of deaths
1–4	1 422	1 637	1 380	1 325
5–14	3 062	1 390	2 968	920
15–24	2 430	2 816	2 318	1 437
25–44	4 101	9 690	4 023	5 942
45–64	2 755	36 581	2 753	18 535

Source: Wilkie *et al.* (1996, pp. 179–180).

Table 2E.3

Age group	Mid-year population (thousands)		Number of deaths (thousands)	
	Males	Females	Males	Females
1–4	1403	1335	0.40	0.34
5–14	3394	3219	0.61	0.42
15–24	3348	3172	2.45	0.91
25–34	4252	4076	4.10	1.84
35–44	3523	3480	5.86	3.64
45–64	5630	5900	44.20	27.79
65–74	2078	2477	74.50	52.70
75–84	1032	1702	91.60	96.40
85 and over	240	708	46.60	107.50

Source: *Population Trends* 87 (1997), pp. 47 and 55.

2.4 Table 2E.4 gives the numbers of births, and deaths of infants aged under 1 year, classified by sex, in England and Wales in certain recent calendar years.
(a) Calculate sex-specific infant mortality rates for the years in question.
(b) Calculate the infant mortality rates for both sexes combined for the years in question.
(c) Comment briefly on your results.

2.5 Draw a Lexis chart with axes like the one in Figure 2.1. On the chart draw lines to represent the following groups of people:
(a) people alive aged x last birthday on 30 June in year t;
(b) people alive aged $x + 1$ last birthday on 1 January in year $t + 1$;
(c) people who celebrated their xth birthday during year t;
(d) people who celebrated their xth birthday before the beginning of year t.

2.6 On a Lexis chart with axes like the one in Figure 2.1, mark areas representing the following deaths:
(a) deaths of people who died in year t aged x last birthday when they died;
(b) deaths of people who died in year $t + 1$ aged $x + 1$ last birthday when they died;

Table 2E.4

Year	Number of births		Number of deaths of infants aged under 1 year	
	Males	Females	Males	Females
1971	402 500	380 800	7970	5750
1976	301 200	283 600	4880	3460
1981	327 000	308 500	4120	2900
1986	338 200	323 800	3720	2590
1991	357 830	342 200	2970	2190
1993	344 300	328 600	2410	1840
1994	343 500	324 100	2370	1750
1995	331 900	317 000	2290	1680

Source: *Population Trends* 87 (1997), p. 55.

Table 2E.5

Year	Number of births	Number of deaths	
		at ages under 1 year	at ages under 28 days
1971	901 600	16 200	10 800
1976	675 500	9 790	6 680
1981	730 800	8 160	4 930
1986	755 000	7 180	4 000
1991	792 500	5 820	3 460
1995	732 000	4 520	3 070

Source: *Population Trends* 87 (1997), p. 50.

(c) deaths of people who celebrated their xth birthday in year t and who were aged x last birthday when they died;
(d) deaths of people who celebrated their xth birthday in year $t+1$, who died in year $t+1$, and who were aged x last birthday when they died.

2.7 The equation

$$q_x = \frac{2m_x}{2+m_x}$$

shows how the m-type and q-type mortality rates are related to one another (under certain assumptions) over a single year of age. Derive a similar equation for the more general case of an age group of width n years.

2.8 Table 2E.5 gives the numbers of births, deaths of infants aged under 1 year, and deaths of infants aged under 28 days, in the United Kingdom in selected recent calendar years.
(a) Calculate the percentage of infant deaths in each year which were neonatal deaths.
(b) Calculate the infant and neonatal mortality rates for each year.

2.9 Someone proposes calculating an infant mortality rate using the number of births in a given calendar year t in the denominator, and the average number of deaths of persons aged under 1 year in the two calendar years t and $t+1$ in the numerator, arguing that this would better reflect the mortality experience of this birth cohort than the conventional method of calculating an m-type infant mortality rate.
(a) Use a Lexis chart to illustrate the rationale behind this argument.
(b) Why might the suggestion not work as well in practice as in theory?
(c) Suggest a modification to the proposal which should lead to an infant mortality rate which better reflects the experience of the births occurring in year t.