Lecture Notes
on
Under Graduate Demography

Prof. Sudhanshu Sekhar Maiti
Department of Statistics
Visva-Bharati

June 29, 2012

Contents

1 Source of obtaining vital statistics data

- Census
- Vital statistics registers
- Hospital records
- Adhoc surveys

2 Measurement of mortality

2.1 Crude death rate (C.D.R)

Crude death rate is defined as the ratio of total number of death (from all possible causes) and total population in a given region during a given period and it is given by,

$$m = \frac{D}{P} \times 1000$$

where m = C.D.R per thousand of population.

D = Total number of deaths (from all causes) which occurred in the given population during the given period.

P = Total population of the given region during the given period.

2.1.1 Merits

1. Easy to calculate and understand.
2. It is widely used vital statistics rate. In numerous demographic and public health problems it is used as an index of mortality.
3. Since the entire population of the region is exposed to the risk of mortality, C.D.R can be treated as the probability that a person belonging to the given population will die in the given period.

2.1.2 Demerits

1. Serious drawback of crude death rate is that it completely ignores the age and sex distribution of the population. Experience shows that mortality is different in different segments of the population. Children in the early ages of their life, and the older generation are exposed to higher risk of mortality as compared to the younger people. Moreover,

mortality rate is also different for females irrespective of their age group, than their male counterparts.

2. It is not suitable for comparing the mortality situation in two places or same place in two different periods unless,

   (i) The population of the places being compared have more or less the same age and sex distribution.

   (ii) Two periods are not too distant, since in a stable large community age-sex structure of the population shows very little change.

2.2 Specific death rate (S.D.R)

It is calculated for a specific section of the population and is defined as,

$$S.D.R = \frac{D_{sp}}{P_{sp}} \times 100$$

where D_{sp} = Number of deaths in the specified section of the population in a given region during a given period.

P_{sp} = Number of persons in the specified section of the population in the given region during the given period.

Note: Specificity is made on age, sex, race, occupation etc..

2.3 Age specific death rate (A.S.D.R)

A.S.D.R,

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

where ${}_nD_x$ = Number of deaths in the age group $(x, x+n)$ i.e. number of deaths between the ages x and $x+n-1$ l.b.d(last birth day) and ${}_nP_x$ = Total number of people in the age group $(x, x+n)$ in that given region during the given period.

Taking $n = 1$, we get the annual A.S.D.R given by,

$$m_x = \frac{D_x}{P_x} \times 1000$$

To be more specific, the age specific death rates for males and females are given by,

$${}_n^m m_x = \frac{{}_n^m D_x}{{}_n^m P_x} \times 1000$$

and

$${}_n^f m_x = \frac{{}_n^f D_x}{{}_n^f P_x} \times 1000$$

2.3.1 Merits

The death rates specific to age and sex overcome the drawback of C.D.R, since they are computed by taking into consideration the age and sex composition of the population. By eliminating the variation in the death rates due to age-sex distribution of the population S.D.R's provide more appropriate measures of the relative mortality situation in the regions.

2.3.2 Demerits

1. S.D.R's are not of much utility for overall comparison of mortality conditions prevailing in two different regions, say, A and B. For example, it might happen that for certain age groups the mortality pattern for region A is greater than that for B, but for the others the case may be opposite. Hence it will not be possible to draw general conclusion regarding the overall mortality pattern in region A as compared to the region B. In overall, to draw some valid conclusions, the different age or/and sex specific death rates must be combined to give a single figure, reflecting the true picture of mortality in the region.

2. In addition to age and sex distribution of the population social, occupational and topographical factors come into operation causing what is called *differential mortality.* Inclusion of these factors, make S.D.R more complicated.

2.4 Standardized death rate (S.T.D.R)

We desire to find a single index of mortality viz. weighted average of age specific death rate for each of A and B. The weights being same in both the cases. This is done by considering a third population, known as *standard population.* Very often we use the life table stationary population as standard. If A and B be two places the usual procedure is to take as standard the actual population of a bigger community of which A and B are parts. For comparing the mortality rates of West Bengal and Uttar Pradesh, the population of India may be taken as standard.
Let,
$^{L}m_x$ = Specific death rate at age x in the local population.
$^{S}m_x$ = Specific death rate at age x in the standard population.
$^{L}P_x$ = Number of persons at age x in the local population.
$^{S}P_x$ = Number of persons at age x in the standard population. We may calculate the following four weighted averages viz.,

(*i*) Age specific death rates of the local population are weighted with the numbers of persons at the corresponding ages in the standard population, i.e. crude death rate of the standard population subjected to age specific death rates of the local population

$$= \frac{\sum {}^SP_x \times {}^Lm_x}{\sum {}^SP_x}$$

(*ii*) Age specific death rates of the standard population are weighted with the numbers of of persons at the corresponding ages in the local population, i.e. crude death rate of the local population subjected to age specific death rates of the standard population

$$= \frac{\sum {}^LP_x \times {}^Sm_x}{\sum {}^LP_x}$$

(*iii*) Age specific death rates of the local population are weighted with the numbers of persons at the corresponding ages in the same local population, i.e. crude death rate of the local population

$$= \frac{\sum {}^LP_x \times {}^Lm_x}{\sum {}^LP_x}$$

(*iv*) age specific death rates of the standard population are weighted with the numbers of persons at the corresponding ages in the same standard population

$$= \frac{\sum {}^SP_x \times {}^Sm_x}{\sum {}^SP_x}$$

(*i*) i.e. $\frac{\sum {}^SP_x \times {}^Lm_x}{\sum {}^SP_x}$ is the standardized death rate. Some times it is called *Direct Standardization.*
(*ii*) is not suitable as it doesn't consider the age specific death rates of the local population.
(*iii*) is not suitable for comparison purposes because, as we have already stated, the age distributions of different classes of people are likely to be different.
(*iv*) is not at all suitable because it has nothing to do with any local population.

The use of formula for direct standardization (i.e.(*i*)) requires S.D.R's (specific death rate) for all segments of the given population (local) be known. In some cases, however, we may have a population classified according to age for instance, but the S.D.R's for the individual age groups may not be available, only the total number of deaths and hence the C.D.R may be known.

In such a case, we may approximately assume,

$$\begin{aligned} (i) \times (ii) &= (iii) \times (iv) \\ \Rightarrow \frac{(i)}{(iii)} &= \frac{(iv)}{(ii)} \\ \Rightarrow (i) &= \frac{(iv)}{(ii)} \times (iii) \\ &= c' \times (iii) \end{aligned}$$

where $c' = \frac{(iv)}{(ii)}$
i.e. S.T.D.R of local population = $c' \times$ C.D.R of local population.

Here c' is called adjustment factor.

$$c' = \frac{\sum^S P_x \times^S m_x / \sum^S P_x}{\sum^L P_x \times^S m_x / \sum^L P_x}$$

The calculation of S.T.D.R by adjusting the C.D.R in this manner is called *Indirect Standardization* of specific death rates.

2.5 Mortality table or Life table

What is a life table?
Life table is a device for presenting in a compact form, the mortality situation prevailing in a community. It contains the values of several functions of age in years. From these values we get answers to questions of the type: If 1,00,000 babies were born at the same time and if they experience through out their life time the prevailing mortality, how many would reach age 20, age 43 etc.; what would be the average numbers of persons in the age interval 13 to 14; what is the probability that a person of age 69 would meet his/ her next birth day; how many years more can a man of age 60 expect to survive; what is the average longevity of the person etc..

- **Stationary population:** A population is said to be stationary if it is of constant size and constant age and sex composition over time, such a population is conceived of under the following conditions:

  (i) If every year the number of births is exactly l_0 (say) and is equal to the number of deaths and these are distributed uniformly through out the year.

  (ii) If the population is not affected by emigration or immigration.

- **Stable population:** concept of a stable population is due to A.J Lotka is very much akin to (inclined to) that of stationary population. A population is said to be stable if

  (i) it has a fixed age and sex distribution.

  (ii) constant mortality and fertility rates are experienced at each age.

  (iii) the population is closed to emigration or immigration.

In other words, for a stable population, the over all rates of births and deaths remain constant and consequently such a population increases at a constant rate, thus supporting the Mathu's law (or compound interest law) of population growth.

In particular, if the constant over all birth and death rates are equal, then the population size remains fixed and in this case stable population becomes stationary population.

Types of life table: Life tables are of two types i.e. (a) Complete life table and (b) Abridged life table.

Abridged life tables are also of two types i.e. First type **(due to G.King)** and Second type **(due to Greville, Reed and Merrel)**

2.5.1 Description of a complete life table

Age(x)	$\mathbf{l_x}$	$\mathbf{d_x}$	$\mathbf{q_x}$	$\mathbf{L_x}$	$\mathbf{T_x}$	$\mathbf{e_x^0}$
0	l_0	d_0	q_0	L_0	T_0	e_0^0
1	l_1	d_1	q_1	L_1	T_1	e_1^0
2	l_2	d_2	q_2	L_2	T_2	e_2^0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
x	l_x	d_x	q_x	L_x	T_x	e_x^0
$x+1$	l_{x+1}	d_{x+1}	q_{x+1}	L_{x+1}	T_{x+1}	e_{x+1}^0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

$\mathbf{l_x}$: Number of persons who attain (or rather are expected to attain) the exact age x out of an assumed number of births, say l_0, called *cohort* or *radix* of the life table.

$\mathbf{d_x}$: Number of persons among the l_x persons reaching at age x, who die before reaching the age $x + 1$.So,

$$d_x = l_x - l_{x+1}$$

$\mathbf{q_x}$: The probability that a person of exact age x will die before reaching age $x + 1$. So,

$$q_x = \frac{d_x}{l_x}$$

Some tables also use another function p_x which is the probability that a person of precise age x will survive till his/her next birth day. Therefore,

$$p_x = \frac{l_{x+1}}{l_x} = \frac{(l_x - d_x)}{l_x} = 1 - \frac{d_x}{l_x} = 1 - q_x$$

$\mathbf{L_x}$: Number of years lived in the aggregate by the cohort of l_0 persons between ages x and $x+1$. So,

$$L_x = \int_0^1 l_{x+t} dt$$

Note: Each of l_{x+1} persons live one complete year in the age interval $(x, x+1)$, d_x persons contribute varying fraction of one year in the interval. If a_x be the averages of these fractions, then

$$L_x = (l_{x+1} \times 1) + (a_x \times d_x) = l_x - d_x + a_x d_x = l_x - d_x(1 - a_x)$$

If $a_x = 0.5$ which is equivalent to assume that d_x deaths are uniformly distributed in the age interval $(x, x+1)$ or that l_{x+t} is a linear function of t for $0 \leq t \leq 1$. Therefore we have,

$$L_x = l_x - 0.5 d_x = \frac{l_x + l_{x+1}}{2}$$

L_x can be interpreted in another way. Since $(x+1) - x = 1$, then

$$\int_0^1 l_{x+t} dt = \frac{1}{(x+1) - x} \int_x^{x+1} l_t dt$$

can be taken as the average number of persons in the age group $(x, x+1)$.
Another interpretation of L_x: Suppose in a community there are l_0 births every year distributed uniformly over the year and the mortality rates at different ages (as shown by the q_x column) remains the same year after year, and there is no migration. Then after 100 years or so, the size of the population will be constant over the years and the age composition of the population will be constant. So, L_x column represent the age distribution of the life table.
$\mathbf{T_x}$: Total number of years lived in the aggregate by the cohort after reaching at age x (it is nothing but the total future life time of the l_x persons). Therefore,

$$T_x = \int_0^\infty l_{x+t} dt$$

another interpretation of T_x is that it is the number of persons with age x or more in a stationary population.

$\mathbf{e_x^0}$: The average number of years lived after age x by each of the l_x persons who attain that age. It is called the (complete) expectation of life (or life expectancy) at age x and is obtained from the relation,

$$e_x^0 = \frac{T_x}{l_x}$$

e_0^0, expectation of life at age 0, is the average number of years lived at birth or the average longevity of a person belonging to the given community.

A closely related concept is that of the ***curtate expectation of life****, denoted by e_x, which represents the average number of complete years of life lived after age x by any of the l_x persons who attain age x. Thus we have,*

$$e_x = \frac{\sum_{t=1}^{\infty} l_{x+t}}{l_x} = \frac{1}{l_x}\sum_{t=1}^{\infty} l_{x+t}$$

so that,

$$e_x^0 = e_x + \frac{1}{2}$$

2.5.2 Construction of a complete life table

The pivotal column (most important column) of a life table is the q_x column. If the q_x column is known then starting with suitable cohort of l_0 births we can compute the life table proceeding as shown below:

$$\begin{aligned} l_0 q_0 &= d_0 \\ l_0 - d_0 &= l_1 \\ l_1 q_1 &= d_1 \\ l_1 - d_1 &= l_2 \\ l_2 q_2 &= d_2 \\ &\vdots \end{aligned}$$

Thus l_x and d_x column can be filled up.

When l_x column is known we can fill up L_x column by using the approximate relation

$$L_x \simeq \frac{1}{2}(l_x + l_{x+1})$$

It is to be noted that this approximate formula does not hold for every years of life, specially at age 0,1. For the early years, we have to use more elaborate formula.

We fill up T_x column by using the relation,

$$T_x = L_x + T_{x+1}$$

So, T_x values are cumulative sums of L_x values from bottom of the table.

Lastly, we compute e_x^0 by using the relation,

$$e_x^0 = \frac{T_x}{l_x}$$

The unknown q_x values are estimated from the age-specific death rates (m_x).

$$\begin{aligned} m_x &= \frac{D_x}{P_x} \\ &\simeq \frac{d_x}{L_x} \\ &\simeq \frac{d_x}{l_x - \frac{1}{2}d_x} \\ &= \frac{\frac{d_x}{l_x}}{1 - \frac{1}{2}\frac{d_x}{l_x}} \\ &= \frac{q_x}{1 - \frac{q_x}{2}} \\ &= \frac{2q_x}{2 - q_x} \end{aligned}$$

$$\Rightarrow q_x = \frac{2m_x}{(2 + m_x)}$$

For the early years of life, the values of m_x are usually not so reliable owing to defects in census records. Besides, the assumption that **deaths are distributed uniformly over the years of age** is not valid for the early ages, especially for age 0: mortality is generally very high in the first few weeks after birth and then it diminishes sharply. It is, therefore, necessary to have alternative formulae for q_x for say $x = 0, 1, 2$. We shall consider an alternative formula for q_0 based on registration data alone. here the assumption will be made that the effect of migration is negligible, which is probably legitimate at age 0. This formula is due to Kuczynski.

Note that in order to survive the first year of age, a child must survive till the end of the calender year in which it is born and then live long enough in the next calender year to attain the exact age 1. Hence, denoting the probabilities of these two events by p' and p'', respectively, we have

$$p_0 = p'p''$$

The probabilities p' and p'' are estimated by,

$$\hat{p'} = \frac{B_0 - D_0^0}{B_0}$$

and

$$\hat{p''} = \frac{B_{-1} - D_{-1}^{-1} - D_{-1}^{0}}{B_{-1} - D_{-1}^{-1}}$$

respectively, where

$\mathbf{B_{-1}}$ = number of children born in the preceding calender year,

$\mathbf{B_0}$ = number of children born in the current calender year,

$\mathbf{D_0^0}$ = number of children born and deceased in the current calender year.

$\mathbf{D_{-1}^{-1}}$ = number of children born and deceased in the preceding calender year,

$\mathbf{D^0{}_{-1}}$ = number of children born in the preceding calender year and deceased in the current calender year before reaching age 1,

A formula for q_x, due to Chiang, which is applicable for all the age groups is $q_x \simeq \frac{m_x}{1+(1-a_x)m_x}$. Out of l_x persons of the cohort alive at age x, l_{x+1} live for one full year within the age interval $(x, x+1)$; the remaining d_x, who die within the age interval $(x, x+1)$, live for varying fractions of one year. Suppose a_x is the average of these fractions.

In computing L_x, the formula

$$L_x \simeq \frac{1}{2}(l_x + l_{x+1})$$

is used. But this is not applicable for early years of life. Therefore we can use

$$\begin{aligned} L_x &= l_{x+1} + a_x d_x \\ &= l_x - (1 - a_x)d_x \end{aligned}$$

It is applicable for all cases.

Chiang has shown, on the basis of his study of U.S mortality data that for $x \geq 5$, $a_x = 0.5$ irrespective of race, age and sex, and $a_1 = 0.43$, $a_2 = 0.45$, $a_3 = 0.47$ and $a_4 = 0.49$ irrespective of race and sex, $a_0 = 0.10$ for whites and $a_0 = 0.14$ for coloureds.

Note that the last age interval in a complete life table will be an open interval (ω, ∞). As such, the L value for the interval, say ${}_{\infty}L_{\omega}$, will have to be computed by a formula other than usual one. The general approach is to make use of the observed A.S.D.R for the age interval together with l_{ω}. For a life table, the central death rate for the interval will be

$$\begin{aligned} {}_{\infty}m'_{\omega} &= \frac{{}_{\infty}d_{\omega}}{{}_{\infty}L_{\omega}} \\ &= \frac{l_{\omega}}{{}_{\infty}L_{\omega}} \end{aligned}$$

since the persons dying after age ω are precisely those who were alive at that age. hence replacing ${}_{\infty}m'_{\omega}$ by the observed A.S.D.R ${}_{\infty}m_{\omega}$, we have

$${}_{\infty}L_{\omega} \simeq \frac{l_{\omega}}{{}_{\infty}m_{\omega}}$$

2.5.3 Abridged life table

As opposed to complete life table, which considered the age interval as a year through out the table and the various functions are calculated for every year of age, there are abridged life tables. The abridgment may be of two kinds. In the first type of abridgment the functions are evaluated for single years of age, as in a complete life table, but these are now given for the greatest part of the table at intervals of 5 years or 10 years. In the second form of abridgment the functional values are stated for the major part of the table for 5 years or 10 years age groups and hence this type is obtained through a condensation of a complete life table rather than through the omission of some of its rows.

2.5.4 Construction of abridged life table

King's Method:

Suppose the life table functions q_x, l_x and e_x^0 are to be given at 5 year intervals in the abridged life table.

Age(x)	0	1	2	3	4	5	10	15	...	$x-5$	x	$x+5$	......
$\mathbf{l_x}$	l_0	l_1	l_2	l_3	l_4	l_5	l_{10}	l_{15}	...	l_{x-5}	l_x	l_{x+5}	......
$\mathbf{q_x}$	q_0	q_1	q_2	q_3	q_4	q_5	q_{10}	q_{15}	...	q_{x-5}	q_x	q_{x+5}	......
$\mathbf{p_x}$	p_0	p_1	p_2	p_3	p_4	p_5	p_{10}	p_{15}	...	p_{x-5}	p_x	p_{x+5}	......
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	......

Then the first step would be to compute the probabilities of death, q_x at the pivotal ages by the usual procedure. Next one has to form $p_x = 1 - q_x$ for the pivotal ages.

To compute the next life table function l_x for the pivotal ages, we note that

$$l_{x+5} = l_x \times_5 p_x$$

$$\log l_{x+5} = \log l_x + \log {}_5p_x$$

So, it is necessary to compute ${}_5p_x$ from the available p_x values. For the first pivotal age, ${}_5p_x$ is evaluated from Newton's forward formula as follows:

Ignoring differences higher than the third, we have

$$\log p_{x+1} = \log p_x + 0.2\Delta \log p_x - 0.08\Delta^2 \log p_x + 0.048\Delta^3 \log p_x,$$

$$\log p_{x+2} = \log p_x + 0.4\Delta \log p_x - 0.12\Delta^2 \log p_x + 0.064\Delta^3 \log p_x,$$

$$\log p_{x+3} = \log p_x + 0.6\Delta \log p_x - 0.12\Delta^2 \log p_x + 0.056\Delta^3 \log p_x,$$

$$\log p_{x+4} = \log p_x + 0.8\Delta \log p_x - 0.08\Delta^2 \log p_x + 0.032\Delta^3 \log p_x,$$

Hence we get

$$\begin{aligned}\log {}_5p_x &= \sum_{i=0}^{4} \log p_{x+i} \\ &= 5\log p_x + 2\Delta \log p_x - 0.4\Delta^2 \log p_x + 0.2\Delta^3 \log p_x \\ &= 2.4\log p_x + 3.4\log p_{x+5} - \log p_{x+10} + 0.2\log p_{x+15}\end{aligned}$$

noting that

$$\Delta^r \log p_x = (E^5 - 1)^r \log p_x$$

For the remaining pivotal ages, one uses Newton's forward formula based on p_{x-5}, and the differences corresponding to p_{x-5}, as follows:

$$\log p_x = \log p_{x-5} + \Delta \log p_{x-5},$$

$$\log p_{x+1} = \log p_{x-5} + 1.2\Delta \log p_{x-5} + 0.12\Delta^2 \log p_{x-5} - 0.032\Delta^3 \log p_{x-5},$$

$$\log p_{x+2} = \log p_{x-5} + 1.4\Delta \log p_{x-5} + 0.28\Delta^2 \log p_{x-5} - 0.056\Delta^3 \log p_{x-5},$$

$$\log p_{x+3} = \log p_{x-5} + 1.6\Delta \log p_{x-5} + 0.48\Delta^2 \log p_{x-5} - 0.064\Delta^3 \log p_{x-5},$$

$$\log p_{x+4} = \log p_{x-5} + 1.8\Delta \log p_{x-5} + 0.72\Delta^2 \log p_{x-5} - 0.048\Delta^3 \log p_{x-5},$$

Hence

$$\begin{aligned}\log {}_5p_x &= 5\log p_{x-5} + 7\Delta \log p_{x-5} + 0.6\Delta^2 \log p_{x-5} - 0.2\Delta^3 \log p_{x-5} \\ &= 0.2\log p_{x-5} + 3.2\log p_x + 2.2\log p_{x+5} - 0.2\log p_{x+10}\end{aligned}$$

Having obtained these, one forms the sum

$$N'_{x5)} = \sum_{i=1}^{5} l_{x+i}$$

for each pivotal age x. These sums are similar to those involved in eqn. (7) and eqn. (9). The formula corresponding to (7), for the first pivotal age, is

$$\begin{aligned} N'_{x5)} &= 5l_x + 3\Delta l_x - 0.4\Delta^2 l_x + 0.2\Delta^3 l_x \\ &= 1.4l_x + 4.4l_{x+5} - l_{x+10} + 0.2l_{x+15} \end{aligned}$$

and for formula corresponding to (9), for the other pivotal ages, is

$$\begin{aligned} N'_{x5)} &= 5l_{x-5} + 8\Delta l_{x-5} + 2.6\Delta^2 l_{x-5} - 0.2\Delta^3 l_{x-5} \\ &= -0.2l_{x-5} + 2.2l_x + 23.2l_{x+5} - 0.2l_{x+10} \end{aligned}$$

In case the formula gives a negative value (this will happen for very high values of x), $N'_{x5)}$ will be taken to be zero.

By taking cumulative totals of $N'_{x5)}$ starting from the end of the table, the values of

$$N'_x = \sum_{i=1}^{\infty} l_{x+i} = N'_{x5)} + N'_{x+5}$$

are obtained.

Lastly, one evaluates e_x^0 for the pivotal ages by using the fact that

$$\begin{aligned} e_x^0 &= \frac{\int_0^\infty l_{x+t}dt}{l_x} \approx \frac{\frac{1}{2}l_x + N'_x}{l_x} \\ &= 0.5 + \frac{N'_x}{l_x} \end{aligned}$$

Greville's method and method of Reed and Merrell:

Age(x)	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_nL_x}$	$\mathbf{T_x}$	$\mathbf{e_x^0}$
0	l_0	⋮	⋮	⋮	⋮	⋮
1	l_1	⋮	⋮	⋮	⋮	⋮
2	l_2	⋮	⋮	⋮	⋮	⋮
3	l_3	⋮	⋮	⋮	⋮	⋮
4	l_4	⋮	⋮	⋮	⋮	⋮
5-10	l_4	⋮	⋮	⋮	⋮	⋮
10-15	l_4	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮

$\mathbf{l_x}$: Number of persons, out of a cohort of l_0 persons, living at the beginning of the interval.

$\mathbf{{}_nq_x}$: Probability that a person of age x will die before reaching age $x+n$.

$$ {}_nq_x = \frac{l_x - l_{x+n}}{l_x} = 1 - \frac{l_{x+n}}{l_x} $$

$\mathbf{{}_nd_x}$: Number of deaths in the age group $(x, x+n)$.

$$ {}_nd_x = l_x \times_n q_x $$

$\mathbf{{}_nL_x = \int_0^n l_{x+t}dt}$: Number of years lived in the aggregate by the cohort in the age group $(x, x+n)$ or number of members of the life table stationary population belonging to the age group $(x, x+n)$.

$\mathbf{T_x}$: Total future life time of the cohort after reaching the age x or number of members of the life table stationary population of age x or above.

$$ T_x = \int_0^\infty l_{x=t}dt =_n L_x + T_{x+n} $$

$\mathbf{e_x^0}$: Expectation of life at age x.

$$ e_x^0 = \frac{T_x}{l_x} $$

Computation:

$$ \begin{aligned} {}_nm_x = \frac{{}_nd_x}{{}_nP_x} &\approx \frac{{}_nd_x}{{}_nL_x} \\ &\approx \frac{2.{}_nd_x}{n(2.l_x -_n d_x)} \\ &= \frac{2.{}_nd_x/l_x}{n(2 -_n d_x/l_x)} \\ &= \frac{2.{}_nq_x}{n(2 -_n q_x)} \end{aligned} $$

i.e.

$$ {}_nq_x \approx \frac{2n.{}_mq_x}{2 + n.{}_nm_x} $$

But Greville uses more precise formula

$$ {}_nq_x = \frac{2n.{}_nm_x}{2 + n.{}_nm_x + \frac{n^2}{6}{}_nm_x^2 - \frac{d}{dx}({}_nm_x)} $$

$[{}_nm_x = \frac{{}_nD_x}{{}_nP_x} \simeq \frac{{}_nd_x}{{}_nL_x}$

Now,

$$ {}_nL_x = \int_0^n l_{x+t}dt = \int_x^{x+n} l_t dt $$

$$ \begin{aligned} \frac{d}{dx}({}_nL_x) &= \frac{d}{dx}\int_x^{x+n} l_t dt \\ &= \int_x^{x+n}\left\{\frac{d}{dx}l_t\right\}dt + \frac{d}{dx}(x+n).l_{x+n} - \frac{d}{dx}(x).l_x \\ &= 0 + l_{x+n} - l_x \\ &= -{}_nd_x \end{aligned} $$

Using Leibnitz's rule for differentiation of integrals

$$ \frac{d}{d\alpha}\int_{\phi_1(\alpha)}^{\phi_2(\alpha)} F(x,\alpha)dx = \int_{\phi_1(\alpha)}^{\phi_2(\alpha)} \frac{\partial F}{\partial \alpha}dx + F(\phi_2,\alpha)\frac{d\phi_2}{d\alpha} - F(\phi_1,\alpha)\frac{d\phi_1}{d\alpha} $$

$$ {}_nm_x = -\frac{1}{nL_x}\frac{d}{dx}({}_nL_x) = -\frac{d}{dx}(\ln {}_nL_x) $$

Therefore,

$$ \begin{aligned} -\ln {}_nL_x &= \int {}_nm_x dx + c' \\ \Rightarrow_n L_x &= e^{-\int {}_nm_x dx - c'} \\ &= c.e^{-\int {}_nm_x dx} \ldots\ldots\clubsuit \end{aligned} $$

Now using Euler-Maclaurin's formula

$$ \sum_{r=0}^{n-1} f(a+i\omega) = \frac{1}{\omega}\int_a^{a+r\omega} f(x)dx - \frac{1}{2}\left[f(a+r\omega) - f(a)\right] + \frac{\omega}{12}\left[f'(a+r\omega) - f'(a)\right] - \ldots\ldots $$

we have,

$$ \begin{aligned} T_x &= \sum_{i=0}^{\infty} {}_nL_{x+in} \\ &= \frac{1}{n}\int_x^{\infty} {}_nL_t dt - \frac{1}{2}\left[0 - {}_nL_x\right] + \frac{n}{12}\left[0 - \frac{d}{dt}({}_nL_t)_{t=x}\right] + \ldots \\ &= \frac{1}{n}\int_x^{\infty} {}_nL_t dt + \frac{1}{2}{}_nL_x - \frac{n}{12}\frac{d}{dt}({}_nL_t)_{t=x} + \ldots \\ &= \frac{1}{n}\int_x^{\infty} c.e^{-\int {}_nm_t dt}dt + \frac{c}{2}e^{-\int {}_nm_x dx} - \frac{n}{12}\frac{d}{dt}\left(c.e^{-\int {}_nm_t dt}\right)_{t=x} + \ldots \\ &= \frac{c}{n}\int_x^{\infty} c.e^{-\int {}_nm_t dt}dt + \frac{c}{2}e^{-\int {}_nm_x dx} - \frac{n}{12}ce^{-\int {}_nm_x dx}(-{}_nm_x) + \ldots \\ &= c\left[\frac{1}{n}\int_x^{\infty} c.e^{-\int {}_nm_t dt}dt + \frac{1}{2}e^{-\int {}_nm_x dx} + \frac{n}{12}{}_nm_x e^{-\int {}_nm_x dx} + \ldots\right] \end{aligned} $$

Differentiating both sides with respect to x, we have approximately,

$$
\begin{aligned}
l_x &= c\left[\frac{1}{n}\left(-e^{-\int {}_nm_x dx}\right) + \frac{1}{2}e^{-\int {}_nm_x dx}(-{}_nm_x) + \frac{n}{12}\left\{\frac{d}{dx}({}_nm_x)e^{-\int {}_nm_x dx} + {}_nm_x.e^{-\int {}_nm_x dx}(-{}_nm_
\right.\right. \\
&= ce^{-\int {}_nm_x dx}\left[-\frac{1}{n} - \frac{{}_nm_x}{2} + \frac{n}{12}\left\{\frac{d}{dx}({}_nm_x) - {}_nm_x^2\right\}\right] \\
&= {}_nL_x\left[-\frac{1}{n} - \frac{{}_nm_x}{2} + \frac{n}{12}\left\{\frac{d}{dx}({}_nm_x) - {}_nm_x^2\right\}\right]
\end{aligned}
$$

Now,

$$
\begin{aligned}
{}_nq_x &= \frac{{}_nd_x}{l_x} \\
&= \frac{{}_nd_x}{{}_nL_x}.\frac{{}_nL_x}{l_x} \\
&\simeq \frac{{}_nm_x}{\frac{1}{n} + \frac{{}_nm_x}{2} - \frac{n}{12}\left\{\frac{d}{dx}({}_nm_x) - {}_nm_x^2\right\}} \\
&\simeq \frac{2n.{}_nm_x}{2 + n.{}_nm_x + \frac{n^2}{6}\left\{{}_nm_x^2 - \frac{d}{dx}({}_nm_x)\right\}}
\end{aligned}
$$

]

Reed and Merrell empirically obtained a relationship between ${}_nm_x$ and ${}_nq_x$ as

$$
{}_nq_x \simeq 1 - e^{-n.{}_nm_x - an^3.{}_nm_x^2}
$$

where a may be taken to be 0.008.

2.5.5 Uses of life tables

1. Life tables are indispensable (absolutely necessary) for the solution of all questions concerning the duration of human life. These tables based on scientific uses of statistical methods, are the key stone or pivots from which the whole science of life hinges. Life tables forms the basis for determining the rates of premium to various amount of life assurance. Life table provides the actuarial science with a sound foundation, converting the life insurance business from a mere gambling in human lives to the ability to offer well calculated safe guard in the event of death.

2. Life tables are needed by the demographers to devise measures such as 'Net Reproduction Rate' to study the rate of growth of population. They have also been used it in preparation of population projections

by age sex, i.e. in estimating what the size of the population will be at some future date.

3. Life tables for two or more different groups of population may be used for for the relative comparisons of various measures of mortality such as death rate, expectation of life at various ages etc. of particular interest, in the comparison of e_0^0 i.e. the average longevity for members of a population.

4. Life tables are as well used

   (i) by the Government and the private establishments for determining the rates of retirement benefits to be given to its employees.

   (ii) for predicting the school going population in connection with school building programmes.

   (iii) for estimating the probable number of future widows and orphans in a community.

   (iv) for computing the approximate size of future labour forces etc.

3 Measurement of fertility

3.1 Crude birth rate(C.B.R)

Crude birth rate is denoted by i' and is given by,

$$i' = \frac{B}{P} \times 1000;$$

where

i' = Crude birth rate per thousand of population;

B = Number of live births which occurs in the given population during the given period;

P = Total population of the given region during the given period.

3.1.1 Merit

It is simple, easy to calculate and readily comprehensible. It is based on the total number of live births and total size of the population and does not necessitate the knowledge of these figures for different sections of the community or the population.

3.1.2 Demerits

1. The crude birth rate, though simple, is only a crude measure of fertility and it is unreliable since it completely ignores the age and sex distribution of the population.

2. C.B.R, is not a probability, since the whole population cann't be regarded as exposed to the risk of producing children. In fact, only the females and only those between the child bearing age group (usually taken as 15 to 49 years) are exposed to the risk and as such whole of the male population and the female population outside the child bearing age group should be excluded. Moreover, even among the females who are exposed to the risk, the risk varies from one age group to another, a woman under 30 is certainly under greater risk as compared to a woman over 40.

3. As a consequence of variations of climatic conditions, the child bearing age groups are not identical in all the countries. In tropical countries the period starts at an apparent earlier date than in countries with cold weather. Accordingly crude birth rate does not enable us to compare the fertility situations in different countries.

4. Crude birth rate assumes that women in all the ages have the same fertility, an assumption which is not true since younger women have, in general higher fertility than elderly women. C.B.R thus gives us an estimate of heterogeneous figure and is unsuitable for comparative studies.

5. The level of crude birth rate is determined by a number of factors such as age and sex distribution of the population, fertility of the population, sex ratio, marriage rate, migration, family planning measures and so on. Thus a relatively high crude birth rate may be observed in a population with a favorable age and sex structure even though fertility is low, i.e. a population with large proportion of the individuals in the age group 15-49 years will have a high crude birth rate, other things remaining same.

3.2 General fertility rate(G.F.R)

By relating the number of live births to the number of females in the *child-bearing ages*, the *general fertility rate (G.F.R)* is obtained. The formula for

G.F.R is thus

$$i = \frac{B}{\sum_{\omega_1}^{\omega_2} {}^{f}P_x} \times 1,000$$

where

i = General fertility rate per thousand of females in child-bearing ages;

B = Number of live births in the given region during the given period;

${}^{f}P_x$ = Number of females of age x l.b.d (last birth day) in the given region during the given period; and

ω_1, ω_2 = Lower and upper limits of the female reproductive period respectively. Births to mothers under 15 and 49 years are so rare that they are not recorded separately but are included in the age groups 15 and 49 respectively.

3.2.1 Merits

1. general fertility rate is a probability since the denominator consists of the entire female population which is exposed to the risk of producing children.

2. G.F.R reflects the extent to which the female population in the reproductive ages increases the existing population through live births. Obviously, G.F.R takes into account the sex distribution of the population and also the age structure to a certain extent.

3.2.2 Demerits

G.F.R gives a heterogeneous figure since it overlooks the age composition of the female population in the child bearing age. Hence it suffers from the drawback of non-comparability in respect of time and country.

3.3 Specific fertility rate(S.F.R)

S.F.R is the ratio of the number of births to the female population of the specified section in a given region to the total number of female population in that specified section, multiplied by 1000.

Specificity is made on age, marriage, migration, state, region etc..

Age-specific fertility rate:

$${}_{n}i_x = \frac{{}_{n}B_x}{{}_{n}^{f}P_x} \times 1000$$

where

${}_nB_x$ = No. of live births to women of age x to $x+n-1$ in the given region during the given period.

${}_n^fP_x$ = No. of women of age x to $x+n-1$ in the region during the given period. If $n = 1$, we get annual age-specific fertility rate as

$$i_x = \frac{B_x}{{}^fP_x} \times 1000$$

Remark 1 *1. In the computation of age specific fertility rate, the female population in the child bearing age group is placed in small age groups so as to put them in common with other of the child bearing capacity. Grouping of women of different ages is necessary since the capacity to bear children varies from age to age e.g. the women in the age group 20 to 25 are more liable to the risk of producing children than the women in the age group 40 to 45.*

2. *fertility data for different countries show that generally specific fertility starts from a low point, rises to a peak some where between 20 and 29 years of age and after that declines steadily. The age-specific fertility curve is, therefore, a highly positively skewed.*

3. *Age-specific fertility rate is a probability rate. It removes the drawback of G.F.R by taking into account the age composition of the women in the child bearing age group and is thus suitable for comparative studies. However, the use of age-specific fertility rates for comparing the fertility is not an easy job. Generally age-specific fertility rate will be higher for certain age groups and lower for the remaining age groups in one region than in the other. Accordingly it is different to say if the fertility is higher or lower in one region as compared to other.*

3.4 Total fertility rate(T.F.R)

To be practically useful, age-specific fertility rates have, therefore, to be combined into a single quantity. For this purpose a standardised fertility rate may be employed, which is to be computed by the same method as is used in computation of a standardised death rate. A much simpler method is to add up the annual age-specific rates and take the sum, called the total fertility rate (T.F.R), as an index of overall fertility of the community. Thus,

$$T.F.R = \sum_{\omega_1}^{\omega_2} i_x$$

If the age-specific fertility rates are given in age-groups $(x, x+n)$,i.e. width of the interval is n, then T.F.R is approximately given by,

$$T.F.R = n \sum_x {}_n i_x$$

In particular, if one deals with quinquennial age group, i.e. $n = 5$ for each class, then

$$T.F.R = 5 \sum_x {}_5 i_x$$

4 Measurement of population growth

Measurements of population growths are as follows:

- Crude rate of natural increase and vital index
- Gross reproduction rate
- Net reproduction rate

4.1 Crude rate of natural increase (C.R.N.I)

C.R.N.I is given by the difference of crude birth rate and crude death rate. Thus,

$$C.R.N.I = C.B.R - C.D.R$$

4.2 Vital index (V.I)

V.I is given by the ratio of C.B.R and C.D.R. Thus,

$$V.I = \frac{C.B.R}{C.D.R}$$

4.3 Gross reproduction rate (G.R.R)

$$G.R.R = 1000 \times \sum_{\omega_1}^{\omega_2} {}^f i_x = 1000 \times \sum_{\omega_1}^{\omega_2} \frac{{}^f B_x}{{}^f P_x}$$

If ${}^f B_x$ be the number of female babies born to the woman in the age group $(x, x+n)$, then,

$$G.R.R = n \sum_{\omega_1}^{\omega_2} {}_n^f i_x \times 1000 = n \sum_{\omega_1}^{\omega_2} \frac{{}_n^f B_x}{{}_n^f P_x} \times 1000$$

In particular for the quinquennial data,

$$G.R.R = 5\sum_{\omega_1}^{\omega_2} {}_5^f i_x \times 1000$$

Remark 2 *1. The computation of G.R.R requires the availability of following data:*

(i) the classification of the births according to the age of the mothers at the time of birth,

(ii) the sex of new born babies. Usually such data are not available. In that case, however an approximate value of G.R.R may be obtained under the assumption that sex ratio at births remains more or less constant at all the ages of the woman in reproductive period.
Now,

$$\textit{sex ratio} = \textit{no. of male births/no. of female births} = \textit{constant}$$

$$\begin{aligned} \Rightarrow \frac{{}^m B_x}{{}^f B_x} &= k \\ \Rightarrow \frac{{}^f B_x}{{}^m B_x + {}^f B_x} &= \frac{1}{k+1} = c(say) \end{aligned}$$

that is ${}^f B_x = c({}^m B_x + {}^f B_x) = c.B_x$ *where* $B_x = {}^m B_x + {}^f B_x$ *is the total number births to woman of age* x *during the given period in the given region.*

Now,

$$\frac{{}^f B_x}{B_x} = c \forall x$$

$$c = \frac{{}^f B_x}{B_x} = \frac{\sum_{\omega_1}^{\omega_2} {}^f B_x}{\sum_{\omega_1}^{\omega_2} B_x} = \frac{{}^f B}{B}$$

where ${}^f B$ *is the total number of female births* B *is the total number of births.*
Hence,

$${}^f B_x = \frac{{}^f B}{B} \times B_x$$

So,

$$\begin{aligned} G.R.R &= \frac{{}^f B}{B} \times \left(\sum_{\omega_1}^{\omega_2} \frac{B_x}{{}^f P_x} \right) \times 1000 \\ &= \frac{{}^f B}{B} \times T.F.R \times 1000 \end{aligned}$$

2. *As a measure of fertility, G.R.R is quite useful for comparing the fertility in different regions or in the same region at different periods of time. Gross reproduction rate may be regarded as measure of the extent to which a sex under consideration (i.e. female sex in this case) is replacing itself, unity being the criteria foe exact replacement. Thus if G.R.R is less than unity, the population would decline no matter how low the death rate may be and if G.R.R is greater than unity then the population would increase no matter how high the death rate may be. Theoretically* $0 \leq G.R.R \leq s$.

3. *the accuracy of G.R.R depends upon the accuracy of the computation of* ${}^f i_x$, *the main sources of error being*

   (i) *under registration of births,*

   (ii) *under statements or inadequate statements of women age at the time of registration,*

   (iii) *errors in enumeration or estimates of female population,* ${}^f P_x$, *by age groups.*

4. *G.R.R is computed based on the hypothesis that none of the newly born female babies is subject to the risk of mortality till the end of the reproductive period of life. This is a very serious limitation of G.R.R, since all the girls born do not survive till the end of the child bearing span. Accordingly G.R.R leads to fallacious conclusion as it inflates the number of potential mothers. The drawback is overcome in net reproduction rate.*

4.4 Net reproduction rate (N.R.R)

To take into consideration the factor of mortality in measuring population growth, we may to begin with, construct a life table for females on the basis of the observed age-specific death rates for females, ${}^f m_x$. The values in the L_x column of the table, denoted by ${}^f_n L_x$, give the mean size of the cohort of ${}^f l_0$ females in the age interval x to $x+n$. In the usual notation, let ${}^f_n B_x$ be the number of female births to the women in the age group x to $x+n$ at any period t (say). Then

$$\frac{{}^f_n L_x}{{}^f l_0} \times {}^f_n B_x$$

gives the average number of female children that would be born to the cohort ${}^f l_0$ in the age-group x to $x+n$. The quantity

$${}^f_n \pi_x = \frac{{}^f_n L_x}{{}^f l_0}$$

gives the life table probability of survival of a female to the age interval x to $x+n$ and is called survival rate.This implies that out of newly born female babies ${}_n^f\pi_x \times 1000$ will enter into the child bearing age-interval x to $x+n$; ${}_n^f\pi_x \times 1000$ into the age group $x+n$ to $x+2n$ and so on.

Hence, instead of multiply $\frac{{}_n^fB_x}{{}_n^fP_x}$ by 1000 alone as in G.R.R, we multiply it by the factor ${}_n^f\pi_x \times 1000$ for each interval x to $x+n$. Finally a new measure of (female) population growth known as *net reproduction rate*, is given by,

$$\begin{aligned} N.R.R &= 1000 \times \sum_x n \left[\frac{{}_n^fB_x}{{}_n^fP_x} \times_n^f \pi_x\right] \\ &= 1000 \times \sum_x n \times_n^f i_x \times_n^f \pi_x \end{aligned}$$

summation being taken over all the age-group of reproductive span.

Remark 3 *1. Since N.R.R takes into account the mortality of the new born (female) babies, we have*

$$N.R.R \leq 1000 \times \sum_x n.\frac{{}_n^fB_x}{{}_n^fP_x}; \left[Since,_n^f \pi_x \leq 1\right]$$

$$\Rightarrow N.R.R \leq G.R.R$$

The sign of equality holds iff all the new born girls survive at least till end of the reproductive period. Thus G.R.R provides an upper limit to N.R.R and hence, in theory, N.R.R also ranges from 0 to 5 per annum.

2. *It may be pointed out that out of a number of girls born to 1000 women, some die in infancy and some do not marry at all. Of the married women some become widows and it is only the balance who pass through the fertility period and thus add to the population growth. Thus N.R.R may be interpreted as the rate of replenishment of that population.*

3. *If N.R.R = 1, we may conclude that if the current fertility and female mortality rates prevail in the future, then a group of new born girls will exactly replace itself in the next generation, i.e. the present female generation will exactly maintain itself. Thus in this case the population has a tendency to remain more or less constant. On the other hand if N.R.R is greater than unity then the population has a tendency to increase while N.R.R less than unity indicates a declining population.*

4. *It should be clearly borne in mind that the use of N.R.R for population projections, i.e., for forecasting future population changes is not desirable at all because of the following two reasons:*

(*i*) *It assumes that current mortality and fertility rates prevail in future, an assumption which is not true since in practice both these rates go on changing from time to time.*

(*ii*) *It overlooks the factor of migration. The population of a given region in any given period may be depleted more by emigration rather than by declining birth rate or it may increase as a result of fresh stock of immigrants who might be more virile.*

5 Population estimation and projection

Estimates of the population inhabiting a region may be

1. **An inter censal estimate:** The estimate of the population corresponding to a time point between two past censuses,

2. **A post censal estimate:** The estimate of the population corresponding to a time point in the past but subsequent to the latest census,

3. **A projection:** The estimate of the population corresponding to a time point in the future.

5.1 Methods of estimation

- **Mathematical method:** takes into account the population at time t (P_t) to be a simple mathematical function of t.

- **Component method:** It needs not only census data on the population size but also registration data on births, deaths and migration.

5.2 Inter censal and post censal estimates by mathematical method

Let $t = 0$ and $t = 1$ be, in suitable units and with a suitable chosen origin, the time points at which the last two censuses were taken place.

Under the assumption of *linear growth* for the population, we take

$$P_t = a + bt$$

Taking $t = 0$ and $t = 1$, we then have $P_0 = a$ and $P_1 = a + b$, so that the estimates of a and b are $a = P_0$, $b = P_1 - P_0$.

The fitted equation is then

$$P_t = P_0 + t(P_1 - P_0)$$

$$or, P_t = P_0 + t(P_1 - P_0) \ldots\ldots (A)$$

On the other hand, if one assumes *exponential growth* then one has to write

$$P_t = ab^t$$

Taking $t = 0$ and $t = 1$, we have $P_0 = a$ and $P_1 = ab$, the estimates of a and b as $a - P_0$, $b = \frac{P_1}{P_0}$.

the fitted equation is then

$$P_t = P_0(P_1/P_0)^t$$

$$or, P_t = P_0^{1-t} P_1^t \ldots\ldots (B)$$

(A) and (B) gives inter censal estimates if $0 < t < 1$, while they give post censal estimates if $t > 1$.

5.3 Inter censal and post censal estimates by component method

Let $B^{(0-t)}$, $D^{(0-t)}$, $I^{(0-t)}$ and $E^{(0-t)}$ denote the number of births, the number deaths, the total immigration and the total emigration occurring between time 0 and t $(0 < t < 1)$. An inter censal value is,

$$P_t = P_0 + B^{(0-t)} - D^{(0-t)} + I^{(0-t)} - E^{(0-t)};$$

A post censal value is

$$P_t = P_1 + B^{(1-t)} - D^{(1-t)} + I^{(1-t)} - E^{(1-t)}$$

with $t > 1$.

5.4 Projection by mathematical method

Problem: To predict on the basis of the size (and perhaps also the composition) of the current population, what the size (and perhaps also the composition) of the population will be at some future date.

Basic approach: Mathematical method is based on an assumed form of the population at time t, say P_t, as a function of t.

5.4.1 Logistic curve

The population model: Suppose a population has the size P at time t and the size $P+\Delta P$ at time $t+\Delta t$. The rate of increase of the population at time t is $\frac{dP}{dt} = \lim_{\Delta t \to 0} \frac{\Delta P}{\Delta t}$. We may consider the relative growth rate, which is

$$\frac{1}{P} \times \frac{dP}{dt}$$

and examine its behavior as a function of time.

Let us first assume that,

$$\frac{1}{P} . \frac{dP}{dt} = r; (r > 0)$$

$$or, \frac{d \log P}{dt} = r$$

where r is a constant.
On integration,

$$\log P = \int r dt = rt + a$$

$$or, P = Ae^{rt}; (say)$$

where A is a positive constant.
Therefore, with constant value of relative growth rate, population follows compound interest law.

We observe that if $t \to -\infty$, $P \to 0$ and $t \to \infty$, $P \to \infty$.
The second result appears to be unrealistic, because in a region of fixed area having limited means of sustenance, the population can not grow without limit. So the assumption of constant value of relative growth rate has to be changed.

A reasonable assumption is that relative growth rate decreases as population increases with time. Taking a very simple decreasing function of P, viz. $r(1 - kP)$, where r and k are positive constants, we can write

$$\frac{1}{P} \times \frac{dP}{dt} = r(1 - kP)$$

$$or, \left(\frac{1}{P} + \frac{k}{1 - kP} \right) \frac{dP}{dt} = r$$

On integration, it gives,

$$\ln P - \ln (1 - kP) = rt + a; say$$

$$or, \frac{P}{1-kP} = e^{a+rt} = A.e^{rt}$$

where A is a positive constant.

$$or, P(1+kAe^{rt}) = A.e^{rt}$$

$$or, P = \frac{A.e^{rt}}{1+kAe^{rt}} = \frac{1}{k+\frac{1}{A}e^{-rt}}$$

when $t \to -\infty$, $P \to 0$ and $t \to \infty$, $P \to \frac{1}{k}$. Denoting this ultimate population size by L, we have

$$P = \frac{L}{1+\frac{L}{A}e^{-rt}}$$

Let $t = \beta$ when $P = \frac{L}{2}$.
So,

$$\frac{L}{2} = \frac{L}{1+\frac{L}{A}e^{-r\beta}}$$

$$\Rightarrow A = Le^{-r\beta}$$

$$\Rightarrow P = \frac{L}{1+e^{r(\beta-t)}}$$

This is the form in which the equation to *Logistic Curve* is generally written. here r is initial relative growth rate.

5.4.2 Properties of logistic curve

1. The differential equation is

$$\begin{aligned} \frac{dP}{dt} &= rP(1-kP) \\ &= rP\left(1-\frac{P}{L}\right) \end{aligned}$$

   since, r, P, and $\left(1-\frac{P}{L}\right)$ are all positive, $\frac{dP}{dt}$ is positive. So, according to the logistic law, the population always increases with time.

2. From,

$$\frac{dP}{dt} = rP\left(1-\frac{P}{L}\right)$$

we have,

$$\begin{aligned}\frac{d^2P}{dt^2} &= r\frac{dP}{dt}\left(1-\frac{P}{L}\right)+rP\left(-\frac{1}{L}\frac{dP}{dt}\right)\\ &= r\left(1-\frac{2P}{L}\right)\frac{dP}{dt}\end{aligned}$$

Therefore, $\frac{d^2P}{dt^2} >=< 0$ according as $P <=> \frac{L}{2}$.
P takes value $\frac{L}{2}$ when $t=\beta$. So the curve has point of inflection at $t=\beta$ and the curve is concave upwards for $t<\beta$ and convex downwards for $t>\beta$.

3. From,

$$\frac{dP}{dt} = rP\left(1-\frac{P}{L}\right)$$

we observe that $\frac{dP}{dt}=0$, when $P=0$ and $P=L$. These values are attained when $t\to-\infty$ and $t\to\infty$. So the curve has two asymptotes viz. $P=0$ and $P=L$.***The shape of the curve is like an elongated 'S'***.

4. Let,

$$\begin{aligned}\frac{dP}{dt} &= rP\left(1-\frac{P}{L}\right)=\phi(P), say\\ \Rightarrow \frac{d\phi(P)}{dt} &= r\left(1-\frac{P}{L}\right)\frac{dP}{dt}+rP\left(-\frac{1}{L}\frac{dP}{dt}\right)\\ &= r\left(1-\frac{2P}{L}\right)\frac{dP}{dt}\end{aligned}$$

Now,

$$\begin{aligned}\frac{d\phi(P)}{dt} &= 0\\ \Rightarrow P &= \frac{L}{2}\end{aligned}$$

again,

$$\begin{aligned}\frac{d^2\phi(P)}{dt^2} &= r\left(-\frac{2}{L}\frac{dP}{dt}\right)\frac{dP}{dt}+r\left(1-\frac{2P}{L}\right)\frac{d^2P}{dt^2}\\ &= -\frac{2r}{L}\left(\frac{dP}{dt}\right)^2+r^2\left(1-\frac{2P}{L}\right)^2\frac{dP}{dt}\end{aligned}$$

at $P = \frac{L}{2}$,

$$\frac{d^2\phi(P)}{dt^2} = -\frac{2r}{L}\left(\frac{rL}{4}\right)^2$$

Therefore, $\phi(P)$ i.e. $\frac{dP}{dt}$ is maximum at $t = \beta$ i.e. Rate of Change of population is maximum at $t = \beta$.

5. The meaning of the constants in the formula are as follows:

   L = Ultimate size of population attained when $t \to \infty$.

   β = The time point when population attains the size $\frac{L}{2}$.

   r = Value of $\frac{1}{P}.\frac{dP}{dt}$ when $t \to -\infty$. So it is the initial growth rate of population.

5.4.3 Fitting of Logistic Curve

A. Pearl and Reed Method:

Denoting the population size at time point t by P_t, the equation of logistic curve can be written as

$$P_t = \frac{L}{1 + e^{r(\beta - t)}}$$

There are three constants, viz. L, r, and β, which are to be estimated from the observed data.

Suppose we are given the population for N equidistant points of time $t = 0, 1, 2, \ldots, (N-1)$. Since there are three constants, they can be so chosen as to make the curve pass through three chosen points. The three points (t, P_t) should be so chosen as to cover the entire range of time more or less evenly. supposing that the selected points are equidistant on the time scale, we can denote them by (t, P_t), $(t+n, P_{t+n})$, $(t+2n, P_{t+2n})$, or by suitable change of origin for t, by $(0, P_0)$, (n, P_n), $(2n, P_{2n})$. Since the curve passes through these points, we have,

$$\frac{1}{P_0} = \frac{1}{L} + \frac{e^{r\beta}}{L}$$

$$\frac{1}{P_n} = \frac{1}{L} + \frac{e^{r(\beta-n)}}{L}$$

$$\frac{1}{P_{2n}} = \frac{1}{L} + \frac{e^{r(\beta-2n)}}{L}$$

Let,

$$d_1 = \frac{1}{P_0} - \frac{1}{P_n} = \frac{e^{r\beta}\left(1 - e^{-rn}\right)}{L}$$

$$d_2 = \frac{1}{P_n} - \frac{1}{P_{2n}} = \frac{e^{r(\beta-n)}\left(1-e^{-rn}\right)}{L}$$

So,

$$\frac{d_1}{d_2} = e^{rn}$$

$$or, r = \frac{1}{n}\left(\ln d_1 - \ln d_2\right) \ldots\ldots (1)$$

Again,

$$1 - \frac{d_2}{d_1} = 1 - e^{-rn} = \frac{Ld_1}{e^{r\beta}}$$

$$\Rightarrow \frac{d_1^2}{d_1 - d_2} = \frac{e^{r\beta}}{L} = \frac{1}{P_0} - \frac{1}{L}$$

$$\Rightarrow \frac{1}{L} = \frac{1}{P_0} - \frac{d_1^2}{d_1 - d_2} \ldots\ldots (2)$$

r and L can be estimated from (1) and (2). Using these estimates and the relations

$$\frac{L}{P_0} - 1 = e^{r\beta},$$

we get the estimate of β as

$$\beta = \frac{1}{r}\log\left(\frac{L}{P_0} - 1\right)$$

Obviously, the estimates obtained by the "method of 3 selected points" are rough ones. Pearl and Reed suggest a method based on least square principles, for improving upon the rough estimates.

Regarding P as a function of r, L and β, we can write

$$P = f(r, L, \beta) = \frac{L}{1 + e^{r(\beta - t)}}.$$

Therefore,

$$\begin{aligned} P &\simeq f(r_0, L_0, \beta_0) + \delta_r\left(\frac{\partial f}{\partial r}\right)_0 + \delta_L\left(\frac{\partial f}{\partial L}\right)_0 + \delta_\beta\left(\frac{\partial f}{\partial \beta}\right)_0 \\ &= f_0 + \delta_r x + \delta_L y + \delta_\beta z \end{aligned}$$

where

$x = \left(\frac{\partial f}{\partial r}\right)_{r=r_0, L=L_0, \beta=\beta_0}$

$y = \left(\frac{\partial f}{\partial L}\right)_{r=r_0, L=L_0, \beta=\beta_0}$

$z = \left(\frac{\partial f}{\partial \beta}\right)_{r=r_0, L=L_0, \beta=\beta_0}$
by least square principle, the normal equations for determining δ_r, δ_L, δ_β are

$$\sum_i x_i \left(P_i - f_{0i}\right) = \delta_r \sum_i x_i^2 + \delta_L \sum_i x_i y_i + \delta_\beta \sum_i x_i z_i$$
$$\sum_i y_i \left(P_i - f_{0i}\right) = \delta_r \sum_i x_i y_i + \delta_L \sum_i y_i^2 + \delta_\beta \sum_i y_i z_i$$
$$\sum_i z_i \left(P_i - f_{0i}\right) = \delta_r \sum_i x_i z_i + \delta_L \sum_i y_i z_i + \delta_\beta \sum_i z_i^2$$

the solution of these equations give values of δ_r, δ_L and δ_β. So we now get better estimates of r, L and β. This procedure can be repeated to get further better estimates.

B. Method of Rhodes:

If the logistic curve $P_t = \frac{L}{1+e^{r(\beta-t)}}$ passes through the observed points, (t, p_t), $t = 0, 1, 2, 3, \ldots, N-1$, we then have for $t = i-1$ and $t = i$,

$$\frac{1}{P_{i-1}} = \frac{1}{L} + \frac{e^{r(\beta-i+1)}}{L}$$

and

$$\frac{1}{P_i} = \frac{1}{L} + \frac{e^{r(\beta-i)}}{L}$$

so that

$$\frac{1}{P_i} = \frac{1-e^{-r}}{L} + e^{-r}\frac{1}{P_{i-1}}$$

The relationship may be put in the form

$$y_i = A + Bx_i$$

where $y_i = \frac{1}{p_i}$, $x_i = \frac{1}{P_{i-1}}$ and $A = \frac{1-e^{-r}}{L}$ and $B = e^{-r}$. Thus the two variables x and y should be exactly linearly related if the population precisely follows the logistic law. The problem is to estimate the constants A and B, assuming that the deviations of the points (x_i, y_i) from an exact linear relationship arise from errors in both x_i and y_i. The proper estimates of A and B are taken to be

$$\hat{B} = \sqrt{\sum_{i=1}^{N-1} \left(y_i - \bar{y}\right)^2 / \sum_{i=1}^{N-1} \left(x_i - \bar{x}\right)^2}$$

and

$$\hat{A} = \bar{y} - \hat{B}\bar{x}$$

where $\bar{x} = \frac{1}{N-1}\sum_{i=1}^{N-1} x_i$, $\bar{y} = \frac{1}{N-1}\sum_{i=1}^{N-1} y_i = \bar{x} + \left[\frac{1}{P_{N-1}} - \frac{1}{P_0}\right]$.
We can get the estimates of r and L from those of B and A as follows:

$$\hat{r} = -\ln \hat{B}$$

and

$$\hat{L} = \frac{1-\hat{B}}{\hat{A}}$$

For finding the estimate of β we observe that

$$e^{r(\beta-t)} = \frac{L}{P_t} - 1$$

$$\Rightarrow \beta = \frac{1}{r}\ln\left(\frac{L}{P_t} - 1\right) + t$$

Taking $t = 0, 1, 2, \ldots, N-1$ and adding, we get

$$\beta = \frac{1}{Nr}\sum_{t=0}^{N-1} \ln\left(\frac{L}{P_t} - 1\right) + \frac{N-1}{2}$$

Knowing the estimates of r and L, we can get the estimate of β from the above relation.

6 Force of mortality or Instantaneous death rate

Let l_x be the number of persons at precise age x, and let $-\Delta l_x$ be the number of persons among them who die between the age x and $x + \Delta x$. Then the force of mortality at age x is given by,

$$\mu_x = \lim_{\Delta x \to 0} \frac{1}{l_x}\frac{-\Delta l_x}{\Delta x} = -\frac{1}{l_x}\frac{dl_x}{dx}$$

6.1 Relation between m_x and μ_x

$m_x = \frac{D_x}{P_x} \simeq \frac{d_x}{L_x}$, symbols have their usual meanings.
Now,

$$L_x = \int_x^{x+1} l_t dt$$

So, $\frac{dL_x}{dx} = l_{x+1} - l_x = -d_x$.
Using Leibnitz's rule for differentiation of integrals

$$\frac{d}{d\alpha}\int_{\phi_1(\alpha)}^{\phi_2(\alpha)} F(x,\alpha)dx = \int_{\phi_1(\alpha)}^{\phi_2(\alpha)} \frac{\partial F}{\partial \alpha}dx + F(\phi_2,\alpha)\frac{d\phi_2}{d\alpha} - F(\phi_1,\alpha)\frac{d\phi_1}{d\alpha}$$

So,

$$\begin{aligned} m_x &\simeq -\frac{1}{L_x}\frac{dL_x}{dx} \\ &\simeq -\frac{1}{l_{x+\frac{1}{2}}}\frac{dl_{x+\frac{1}{2}}}{dx} \\ &= \mu_{x+\frac{1}{2}} \end{aligned}$$

Therefore,

$$m_x = \mu_{x+\frac{1}{2}}$$

7 Graduation of mortality rates

The observed age specific death rates (m_x) as calculated from the census data and registration data are found to be subject to various irregularities. For any mathematical work involving these rates, it is necessary to remove these irregularities. Possible way is to get an expression for m_x as a function of x.

First step is to derive an expression for μ_x as a function of x and then, using the relationship between m_x and μ_x, find an expression for m_x as a function of x. Makeham's attempt in this aspect is successful one.

Makeham assumed that death can occur from one of the two general causes:

(i) accident, the effect of which may be supposed to be constant through out the life,

(ii) decrease in the capacity to resist disease.

7.1 Makeham's graduation formula for m_x

If $g(x)$ denotes the power to resist disease at age x, the force of mortality may be supposed to vary inversely as $g(x)$, provided the factor accident is absent.

Thus according to Makeham

$$\mu_x = A + \frac{B}{g(x)},$$

where $A > 0$,$B > 0$ and $g(x)$ is a decreasing function of x.

Makeham further assumed that in a short interval of time a person looses a constant proportion of the force of resistance as he or she still has.

Thus he assumed

$$\frac{1}{g(x)}\frac{dg(x)}{dx} = -r; (r > 0)$$

Hence, we get,

$$\ln g(x) = -\int r dx = -k_1 - rx$$

$$\Rightarrow g(x) = e^{-k_1 - rx} = k_2 e^{-rx}; (k_2 > 0)$$

So,

$$\begin{aligned} \mu_x &= A + \frac{B}{k_2 e^{-rx}} \\ &= A + B'C^x, say \end{aligned}$$

where $B' = \frac{B}{k_2}$ and $C = e^r$ As we know that,

$$\begin{aligned} m_x = \mu_{x+\frac{1}{2}} &= A + B'C^{x+\frac{1}{2}} \\ &= A + B''C^x \end{aligned}$$

This is known as the Makeham formula for graduation of mortality rates.

Remark 4 *Makeham's formula has been found to be very useful for all ages from 20 onwards.*

7.2 Makeham's graduation formula for l_x

we have found that $\underline{\mu_x = A + B'C^x}$.

Again,

$$\mu_x = -\frac{1}{l_x}.\frac{dl_x}{dx} = -\frac{d \ln l_x}{dx}$$

$$\begin{aligned} \Rightarrow \ln l_x = -\int \mu_x dx &= -\int (A + B'C^x)\, dx \\ &= -\left(R + Ax + \frac{B'C^x}{\ln C}\right) \\ &= -R - Ax - B_1 C^x \end{aligned}$$

Therefore,

$$\begin{aligned} l_x &= e^{-R - Ax - B_1 C^x} \\ &= (e^{-R}).(e^{-A})^x.(e^{-B_1})^{C^x} \\ &= k.s^x.g^{C^x} \end{aligned}$$

7.3 Gompertz's graduation formula for m_x

Prior to Makeham, Gompertz deduced a graduation formula for m_x. He proceeded in the same way as Makeham did but he overlooked the 'accident' factor.

Taking $A = 0$, we get (proceeding as before)

$$\mu_x = B'C^x$$

$$\Rightarrow m_x \simeq \mu_{x+\frac{1}{2}} = B'C^{x+\frac{1}{2}} = B_1C^x$$

7.4 Gompertz's graduation formula for l_x

Gompertz's formula for l_x is obtained from that of Makeham's by putting $s = 1$ and we have

$$l_x = k.g^{C^x}$$

7.5 Identification of a given set of data

Which one will give a good fit??

Gompertz or Makeham !!!

For m_x:-

Gompertz:

$$m_x = BC^x$$

and

$$m_{x+1} = BC^{x+1}$$

$$\Rightarrow \frac{m_{x+1}}{m_x} = C, constant$$

So, if we find that ratios of successive values of m_x are more or less constant, then Gompertz's formula will give a good fit.

Makeham:

$$m_x = A + BC^x$$

$$\Rightarrow \Delta m_x = m_{x+1} - m_x = BC^x(C-1)$$

$$\Rightarrow \Delta m_{x+1} = m_{x+2} - m_{x+1} = BC^{x+1}(C-1)$$

$$\Rightarrow \frac{\Delta m_{x+1}}{\Delta m_x} = C, constant$$

So, if the ratios of successive first order differences of m_x are more or less constant, Makeham's formula will give a good fit.

For l_x:-

Gompertz:

$$l_x = k.g^{C^x}$$
$$\Rightarrow \ln l_x = \ln k + C^x \ln g$$
$$\Rightarrow \Delta \ln l_x = C^x(C-1)\ln g$$
$$\Rightarrow \Delta \ln l_{x+1} = C^{x+1}(C-1)\ln g$$
$$\Rightarrow \frac{\Delta \ln l_{x+1}}{\Delta \ln x} = C, constant$$

So, if the ratios of successive first order differences of $\ln l_x$ values are more or less constant, Gompertz's law will give a good fit.

Makeham:

$$l_x = k.s^x.g^{C^x}$$
$$\Rightarrow \ln l_x = \ln k + x \ln s + C^x \ln g$$

Hence,

$$\Delta \ln l_x = \ln s + C^x(C-1)\ln g$$
$$\Rightarrow \Delta^2 \ln l_x = C^x(C-1)^2 \ln g$$

and

$$\Rightarrow \Delta^2 \ln l_{x+1} = C^{x+1}(C-1)^2 \ln g$$
$$\Rightarrow \frac{\Delta^2 \ln l_{x+1}}{\Delta^2 \ln l_x} = C, constant$$

So, if the ratios of successive second order differences of $\ln l_x$ are more or less constant, Makeham's law will give a good fit.

7.6 Fitting of Makeham's formula for l_x

1. **Method of selected points:**

   $$l_x = k.s^x.g^{C^x}$$
   $$\Rightarrow \ln l_x = \ln k + x \ln s + C^x \ln g$$

   There are four constants and they can be determined from four independent equations. We can find the constants as to make the curve pass through four chosen points. For better result, the point should be so chosen as to cover the entire range of data more or less evenly and for ease of computation they should be equispaced. With suitable change of origin for x, we denote the chosen points as $(0, l_0)$, (n, l_n), $(2n, l_{2n})$, $(3n, l_{3n})$.

Now,

$$
\begin{aligned}
\ln l_0 &= \ln k + \ln g \\
\ln l_n &= \ln k + n \ln s + C^n \ln g \\
\ln l_{2n} &= \ln k + 2n \ln s + C^{2n} \ln g \\
\ln l_{3n} &= \ln k + 3n \ln s + C^{3n} \ln g \ldots\ldots (1)
\end{aligned}
$$

$$
\begin{aligned}
\Delta \ln l_0 &= n \ln s + (C^n - 1) \ln g \\
\Delta \ln l_n &= n \ln s + C^n (C^n - 1) \ln g \\
\Delta \ln l_{2n} &= n \ln s + C^{2n} (C^n - 1) \ln g \ldots\ldots (2)
\end{aligned}
$$

$$
\begin{aligned}
\Delta^2 \ln l_0 &= (C^n - 1)^2 \ln g \\
\Delta^2 \ln l_n &= C^n (C^n - 1)^2 \ln g \ldots\ldots (3)
\end{aligned}
$$

$$\frac{\Delta^2 \ln l_n}{\Delta^2 \ln l_0} = C^n \ldots\ldots (4)$$

we find C from (4). Knowing C, we get g from one of the equations of (3). Similarly we get s from (2) and k from (1).

2. **Method of group average:**

   Suppose we are given the values of l_x for N equidistant values of x, and let $N = 4n + k'$, where $k' = 0, 1, 2$ or 3. We reject k' values from the beginning or the end and divide the remaining $4n$ values (which we may conveniently denote by $l_0, l_1, \ldots, l_{4n-1}$) in four groups of n values each.

   Let,

$$S_1 = \sum_{x=0}^{n-1} \ln l_x = n \ln k + \frac{n(n-1)}{2} \ln s + \frac{(1 - C^n)}{(1 - C)} \ln g$$

$$S_2 = \sum_{x=n}^{2n-1} \ln l_x = n \ln k + \frac{n(3n-1)}{2} \ln s + \frac{C^n(1 - C^n)}{(1 - C)} \ln g$$

$$S_3 = \sum_{x=2n}^{3n-1} \ln l_x = n \ln k + \frac{n(5n-1)}{2} \ln s + \frac{C^{2n}(1 - C^n)}{(1 - C)} \ln g$$

$$S_4 = \sum_{x=3n}^{4n-1} \ln l_x = n \ln k + \frac{n(7n-1)}{2} \ln s + \frac{C^{3n}(1 - C^n)}{(1 - C)} \ln g \ldots\ldots (1)$$

$$
\begin{aligned}
\Delta S_1 &= n^2 \ln s + \frac{(C^n-1)^2}{(C-1)} \ln g \\
\Delta S_2 &= n^2 \ln s + \frac{C^n(C^n-1)^2}{(C-1)} \ln g \\
\Delta S_3 &= n^2 \ln s + \frac{C^{2n}(C^n-1)^2}{(C-1)} \ln g \ldots\ldots (2)
\end{aligned}
$$

$$
\begin{aligned}
\Delta^2 S_1 &= \frac{(C^n-1)^3}{(C-1)} \ln g \\
\Delta^2 S_2 &= \frac{C^n(C^n-1)^3}{(C-1)} \ln g \ldots\ldots (3)
\end{aligned}
$$

and lastly,

$$\frac{\Delta^2 S_2}{\Delta^2 S_1} = C^n \ldots\ldots (4)$$

Thus we find C from (4), knowing C, we get g from one of the equations in (3), s from one of the equations in (2) and lastly k from one of the equations in (1).