

Mathematical formulas used in population statistics

Prepared by: Ethel Maasing and Ene-Margit Tiit

This document is an additional material to population statistics methodology. It presents the calculation formulas and explanations of the most common rates and indicators used in population statistics.

Demographic labour pressure index (DLPI) is calculated as follows:

$$DLPI = \frac{\sum_5^{14} R}{\sum_{55}^{64} R},$$

where $\sum_i^n R$ is the number of population in age group i to n (including both end points) and i and n are ages in full years.

Demographic labour pressure index, which is calculated as the ratio of 5–14-year-olds to 55–64-year-olds, shows the ratio of persons who will enter the labour market in the next decade to the number of persons who will exit the labour market. If the index is bigger than one, the number of persons entering the labour market in the next ten years is larger than the number of persons potentially leaving it because of ageing. If the index is smaller than one, there will be more persons exiting than entering the labour market in the next decade.

Dependency ratio (DR) is calculated as follows:

$$DR = \frac{\sum_0^{14} R + \sum_{65}^{120} R}{\sum_{15}^{64} R} * 100,$$

where $\sum_i^n R$ is the number of population in age group i to n (including both end points) and i and n are ages in full years.

Dependency ratio expresses the ratio of the total number of dependants (persons aged 0–14 and 65 and older) to the number of working-age population (aged 15–64) (in percentages). The lower the ratio, the smaller is the burden on working population.

Fertile age. Conventionally, women's fertile age is 15–49 years, i.e. 5-year age groups [15; 20), ..., [45; 50) are taken into account. All fertility rates are calculated taking into account women of fertile age.

Age group. Conventionally, age in full years, i.e. the age on the last birthday, is taken into account. This means that a 5-year age group of 15–19 includes persons aged 15–19.

Age-specific fertility rate ${}_a f_x$ is calculated for age group $[x, x + a)$ as follows:

$${}_a f_x = \frac{{}_a B_x}{{}_a \bar{P}_x},$$

where a is the number of years in age group (usually one or five years) used in the calculation, ${}_a B_x$ is the number of live births to women aged $[x; x + a)$ and ${}_a \bar{P}_x$ is the mean annual number of women aged $[x; x + a)$.

Sometimes age-specific fertility rate is shown as 1,000 ${}_a f_x$, which indicates the number of births per thousand women aged $[x; x + a)$ in the reference year.

Age-specific fertility rate shows the intensity of births by age groups, i.e. the mean number of births per woman of a specific age group in the reference year.

General fertility rate (GFR) is an age-specific fertility rate multiplied by a thousand, where $x = 15$ and $a = 35$. It shows the number of live births per thousand women of fertile age (aged 15–49) in the reference year.

Mean age of mother at childbirth (in the case of 5-year age groups) is calculated as follows:

$${}_5\bar{x}_f = \frac{\sum(x + 2.5) * {}_5f_x}{\sum {}_5f_x},$$

where x stands for starting points of the age ranges and $x + 2.5$ for the midpoints) and ${}_5f_x$ is age-specific fertility rate; $x = 15, 20, \dots, 45$ and a total is calculated across all age groups (both in the numerator and denominator).

Mean age of mother at birth of first child (in the case of 5-year age groups) is calculated as follows:

$${}^1_5\bar{x}_f = \frac{\sum(x + 2.5) * {}^1_5f_x}{\sum {}^1_5f_x},$$

where x stands for starting points of the age ranges, $x + 2.5$ for the midpoints and 1_5f_x is the ratio of the number of women of the specific age group who gave birth for the first time during the year to the mean annual number of women of the specific age group.

Total fertility rate (TFR) is calculated as follows:

$$TFR = a \sum {}_af_x,$$

where a is the number of years in age group (usually one or five years), ${}_af_x$ is age-specific fertility rate and a total is calculated across the entire fertile age.

Total fertility rate shows the average number of children that would be born alive to a woman during her lifetime if she passed through all her childbearing years conforming to the age-specific fertility rates of a given year.

Crude birth rate (CBR) is calculated as follows:

$$CBR = \frac{B}{\bar{R}} * 1,000,$$

where B is the number of live births in the reference period and \bar{R} is mean annual population.

Crude birth rate shows the number of live births in a year per 1,000 population.

Gross reproduction rate (GRR) is calculated as follows:

$$GRR = a \sum \frac{{}_aG_x}{{}_a\bar{P}_x},$$

where a is the number of years in age group (usually one or five years), ${}_aG_x$ is the number of daughters born alive to women aged $[x; x + a)$, ${}_a\bar{P}_x$ is mean annual population of women aged $[x; x + a)$, and a total is calculated across the entire fertile age.

Gross reproduction rate shows the average number of daughters born to a woman during her lifetime if she passed through her childbearing years conforming to the age-specific fertility rates of a given year.

Net reproduction rate (NRR) is calculated on the basis of life table data as follows:

$$NRR = a \sum \left(\frac{{}_aG_x}{{}_a\bar{P}_x} * l_x \right) / 100,000,$$

where a is the number of years in age group (usually one or five years), ${}_aG_x$ is the number of daughters born alive to women aged $[x; x + a)$, ${}_a\bar{P}_x$ is the mean annual population of women aged $[x; x + a)$ and l_x is the number of surviving women in the cohort aged x (see life table). A total is calculated across the entire fertile age.

In developed societies, the gross and net reproduction rates do not differ much, as the mortality of young women is low.

Replacement level is the number of births that ensures a non-negative natural increase. Crude birth rate value is 2.06 (previously also 2.1 was used) and net or gross reproduction rate value is 1.

Crude death rate (CDR) is calculated as follows:

$$CDR = \frac{D}{\bar{R}} * 1,000,$$

where D is the number of deaths during the year and \bar{R} is mean annual population.

Crude death rate expresses the number of deaths in a year per 1,000 mean annual population.

Crude rate of natural increase (CRNI) is calculated as the difference between crude birth and death rates, i.e.

$$CRNI = CBR - CDR.$$

Age-specific first marriage rate in 5-year age groups for men/women is calculated as follows:

$${}_5^1m_x = \frac{{}_5^1M_x}{{}_5\bar{P}_x},$$

where ${}_5^1M_x$ is the number of first marriages of men/women aged $[x; x + 5)$ and ${}_5\bar{P}_x$ is the mean annual number of men/women aged $[x; x + 5)$.

It is calculated for age groups 15–19, 20–24, ..., 45–49.

It is also used as $1,000 * {}_5^1m_x$ to show the frequency of first marriages per 1,000 men/women of a specific age group.

Mean age at first marriage for men/women is calculated as follows:

$${}_5\bar{x}_m = \frac{\sum(x + 2.5) * {}_5^1m_x}{\sum {}_5^1m_x},$$

where ${}_5^1m_x$ is age-specific first marriage rate in 5-year age groups for men/women (ages included: 15–49).

For years 1970–1977, instead of ${}_5^1m_{40}$ and ${}_5^1m_{45}$, the value ${}_{10}^1m_{40}$ has been used.

Total first marriage rate (TFMR) for men/women is calculated as follows:

$$TFMR = 5 \sum {}_5^1m_x,$$

where ${}_5^1m_x$ is age-specific first marriage rate in 5-year age groups for men/women (ages included: 15–49). Total first marriage rate shows the share of first-married men/women aged under 50 in the population, assuming that they conform to the age-specific first marriage rates of a given year.

Life expectancy and life table

For calculating **life expectancy**, life tables are used. A **life table** includes annual data on the mortality of the population. For comparability reasons, the table is prepared for a model population of 100,000.

The basis of the table is death probability q_t^x at a given age t in the reference year x , $t = 1, 2, \dots$

The indicator value is calculated as the number of persons D_t^x who died in year x at age t divided by the total number of persons of age P_t^x . The situation is made slightly more complicated by the fact that persons who died in year x at age t include those who reached the age of t in year x as well as those whose last birthday, when they turned t years of age, was in the previous year $x-1$. If mean annual population is used for the total number of population, it does not include the number of persons who died in this year after t birthday, and to avoid this bias, the population figure must be correspondingly corrected. Thus, the calculation is as follows:

$$q_t^x = \frac{D_t^x}{P_t^x + D_t^{x-1}}$$

where D_t^{x-1} is the number of persons who died at age t during the reference year.

In a life table, it is assumed that 100,000 people are born in year x , i.e. the size of the generation of 0-year-olds is $l_0 = 100,000$. By the end of the first year, their number has decreased due to deaths by $100,000 \cdot q_0$, and there are $l_1 = 100,000 \cdot (1 - q_0)$ survivors. To find the number of **survivors** in the following year, the number of persons who died in a year is subtracted from the number who survived; thus, the cohort of survivors declines each year. The maximum potential duration of life is designated with w , which changes according to the population and over time. Traditionally, the figure 100 has been used. Although the maximum duration of life currently exceeds 100 years in the world, it causes a relatively small error. Recently, also value $w = 120$ has been used in some places.

The number of survivors determines the total life expectancy in years for the whole population:

$$S_0 = \sum_{t=0}^w l_t.$$

Similarly, the total life expectancy in years can be found for all persons who have reached the age x :

$$S_x = \sum_{t=x}^w l_t.$$

To calculate average life expectancy the total number of years of life is divided by the number of persons alive in a given year:

$$e_0 = \frac{S_0}{l_0} \text{ ja } e_x = \frac{S_x}{l_x}.$$

The value e_0 is life expectancy at birth, but it is a value that characterizes the reference year x : it shows the life expectancy of a person born in year x assuming that throughout this person's life the mortality rates remain the same as in year x . In reality, this is not the case, as usually mortality decreases and life expectancy increases. This means that on average a person born in year x lives significantly longer than shown by the figure e_0 .

Why is an integral formula used?

As essentially age changes continuously, not as full years, the formula for life expectancy is often expressed as an integral:

$$e_x^0 = \frac{1}{l_x} \int_x^w l_t dt;$$

however, in practical calculations, a total is always used instead of an integral.

Smoothing

In practical calculations, there is another concern, particularly when the population size is relatively small. It appears that mortality indicators may significantly vary by ages (see the below graphs).

Figure 1. Probability of dying at age 0–100 in Estonia, 2019

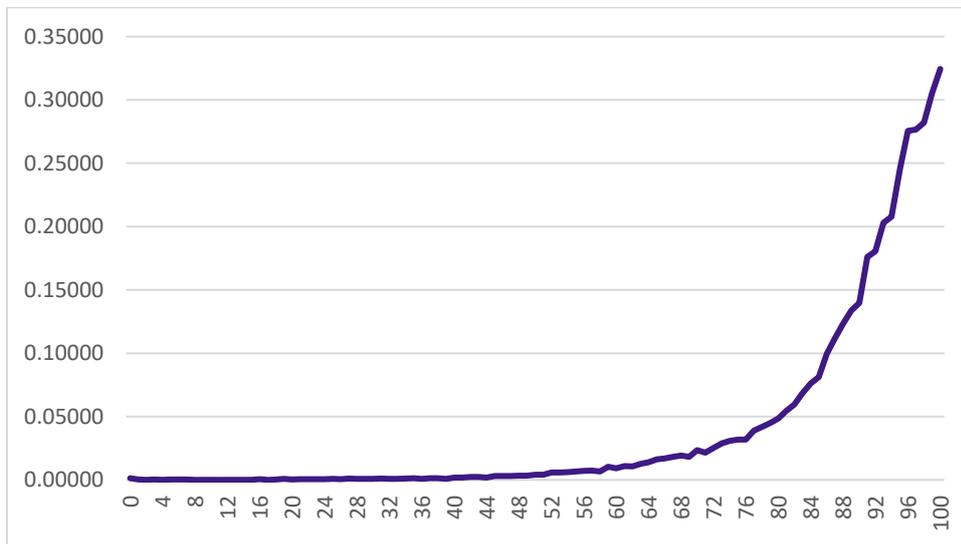


Figure 2. Probability of dying at age 0–15 in Estonia, 2019

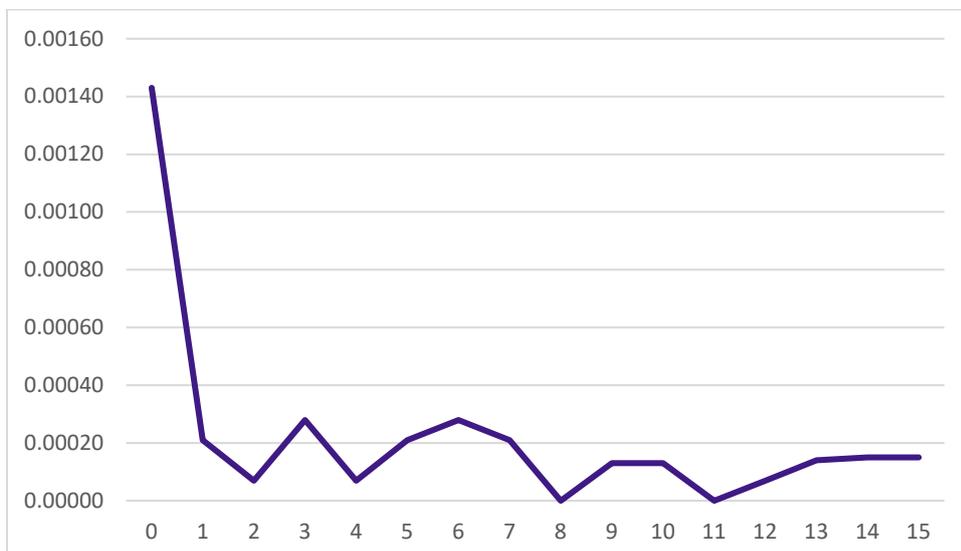
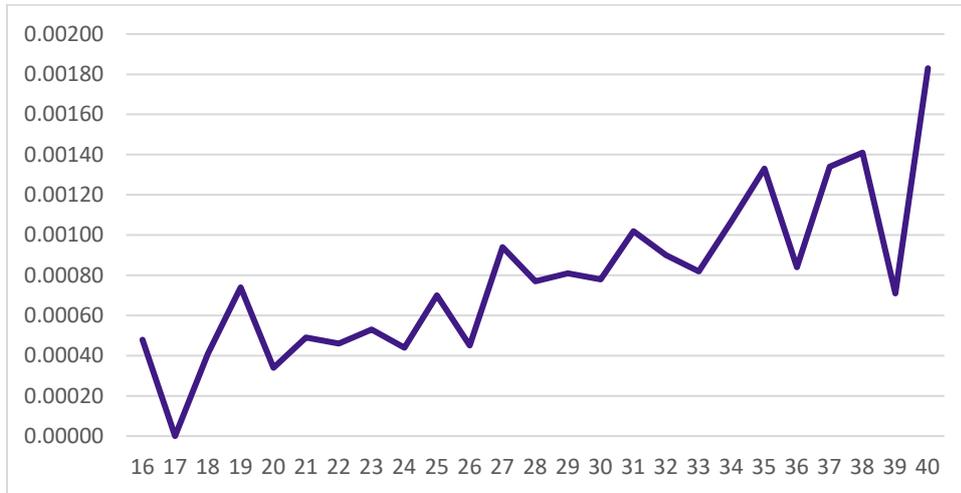


Figure 3. Probability of dying at age 16–40 in Estonia, 2019



As strongly varying death probabilities may cause illogical and difficult to explain effects in calculation results, i.e. in life expectancies, the death probabilities are smoothed. Death probability q_t^x for each year t is replaced by the linear combination of the death probabilities of that year and close years. It is calculated as follows:

$$q_t^x = \sum_{i=t-2}^{t+2} q_i^x a_i,$$

where weights a_i are non-negative and their total is 1. Different weighting methodologies are known and used, and it is practical to apply these differently to death rates or other indicators at different age ranges.

Using 5-year age groups

Death rates are made more stable by using 5-year age groups instead of 1-year age groups. The calculation methods are essentially the same as in the case of 1-year age groups.