INSTITUTO NACIONAL DE ESTADÍSTICA

Life Tables
Methodology

December 2015

Index

1 Introduction

The life table is an instrument of demographic analysis, which enables the analysis of the incidence of mortality regarding individuals from different populations in a given time period, irrespective of the age structure presented by these. More specifically, a **period life table** aims to describe the short-term behaviour of the phenomenon regarding the population studied over a specified period, simulating its incidence on a *cohort* or *fictitious generation* of individuals subject to a pattern of mortality by age identical to the one observed regarding the population studied during the period observed. In this way, the mortality table calculated regarding different populations offers the possibility of establishing comparative analyses regarding the incidence of the phenomenon on each of them, removing the effect of their age composition.

This interpretation of the life table, known as a *period or contemporaries table*, therefore provides a tool for *transversal analysis* of the phenomenon of mortality, in contrast with the *generations mortality tables*, based on a *longitudinal analysis* of a specific generation, from its birth until its complete expiry. This second typology of mortality tables requires a very long period of observation of the phenomenon, making them scarcely operative.

Therefore, the table is composed of a set of *biometric functions* defined regarding a *fictitious cohort* of individuals. It is indispensable to have a clear understanding of each of them, in order for them to be properly interpreted, before detailing how the statistical approximation thereof is carried out, taking observed deaths regarding the population being studied in the reference period. The aforementioned functions are defined and outlined below:

- **Survivors at exact age** x, l_x: this represents the number of individuals in the initial fictitious cohort who live to age x.

- **Theoretical deaths** at age x, d_x: this constitutes the number of deaths in the initial fictitious cohort occurring in individuals who have reached the age of x. It is therefore clear that $l_{x+1} = l_x - d_x$.

- **Average years lived in last year of life** of those dying having reached age x, a_x: this entails the average time lived at age x by those individuals in the fictitious cohort who die at the aforementioned age.

- **Seasonal population** aged x, L_x: this corresponds to the total lived (measured in years) by individuals from the fictitious generation having reached the age of x. As each person surviving until the age of x contributes a year to that time and, on average, those dying at the aforementioned age contribute a_x years each, the aforementioned function is traditionally estimated using the expression $L_x = l_{x+1} + a_x \cdot d_x$.

-Total of lifetime since X age, Tx, it is defined as the total time to live the generation has, from X age to its disappearance. It is represented as:

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

- Specific mortality rate at age x**,** m_x: is defined as the number of individuals in the fictitious cohort who die aged x by time exposed to risk of death by individuals in the aforementioned generation. In other words, it involves the quotient between the number of deaths of individuals having reached the age of x and the total time (measured in years) lived by individuals in the cohort of the aforementioned age, in other words, $m_x = \frac{d_x}{L_x}$. The specific mortality rate at each age thus measures the incidence or relative intensity of the phenomenon at each age.

- Likelihood or risk of death having reached age x, q_x: is defined as the likelihood of an individual belonging to the initial fictitious cohort surviving until reaching age x years dying at the aforementioned age. It is thus defined as the quotient between the number of occurrences of the phenomenon (theoretical deaths aged x, d_x), and the total possible cases or population subject to risk thereof (survivors aged x, l_x), in other words, $q_x = \frac{d_x}{l_x}$.

Conversely, taking the estimated relationship between seasonal population and survival function and that of the actual definition of the specific mortality rate at each age, a classic approximate is derived between risk of death and mortality rate at each age x:

$$q_x = \frac{m_x}{1 + (1 - a_x) \cdot m_x}$$

- Life expectancy at age x, e_x: represents the average number of years that an individual aged x belonging to the initial fictitious cohort would have left to live. Its value is the result of the quotient between the total time (measured in years) left to live after reaching age x by individuals from the fictitious generation until their complete expiry and the number of survivors of the same age x. That is, $e_x = \frac{\sum_{y \geq x} L_y}{l_x}$.

So then, the annual mortality tables presented therefore maintain the purpose of describing the short-term behaviour of mortality of the resident population, for each sex and both sexes, in Spain, its Autonomous Communities and Provinces. Therefore a *fictitious cohort* of 100,000 individuals is submitted to the defined age, basically, by specific mortality rates observed regarding the population studied in the reference year and other *biometric functions* from the mortality table will be derived regarding this.

Moreover, for the aforementioned purpose, we use all of the statistical-demographic information available on the subject. Specifically, the mortality tables presented are calculated with the results of the deaths occurring in Spain each year, provided by the Vital Statistics, and of the figures of the population resident at 1 January of each year that the INE uses as a reference in all of its statistical production, constituted by Intercensal Population Estimates until 2011 and by Population Figures since 2012 onwards.

The estimate of specific mortality rates observed regarding the population studied in the reference year and other biometric functions from the mortality table will be carried out in accordance with the methodology described below.

Lastly, we must warn that the mortality tables on a national level offer results that are broken down by simple age, whereas the results of the Autonomous and provincial mortality tables are provided broken down into five-year age groups, except for the ages of 0 and 1.

2 The life tables for Spain

The life table for Spain measures the incidence of mortality regarding the population resident in the country during the reference year simulating its behaviour regarding a *cohort* or *fictitious generation* of individuals subject to a mortality pattern by identical age to that observed regarding the population studied during the observation period. Specifically, the aforementioned simulation consists of applying the incidence of mortality at each given age to a fictitious generation of individuals, basically, by specific rates observed regarding the population resident in Spain during the reference year and deriving, taken from these, the other functions comprising their mortality table.

The **specific mortality rate at** age x observed regarding the population, m_x, is estimated, under the hypothesis of the uniform distribution of birthdays of all of the individuals of the population of a certain age who do not die throughout the year, and also a uniform distribution throughout the year of the day of arrival of the individuals who join the population under study during said year, and of the day of departure of the individuals who leave said population during the year of study, from the expression:

$$m_x = \frac{D(t,x,s)}{\frac{(P(t,x,s)-D_2(t,x,s))}{2} + \sum_{i=1}^{D_2(t,x,s)} b_2(t,x,s,i) + \frac{P(t+1,x,s)}{2} + \sum_{i=1}^{D_1(t,x,s)} b_1(t,x,s,i)},$$

$x = 0,1,...,99.$

In which:

t, the observation year or period.

x, the age reached, with $x = 0,1,2,...,99$.

s, the sex, which may take attributes of male, female or both sexes.

$P(t,x,s)$ is the resident population stock as of 1 January of year t aged x and of sex s.

$D(t,x,s)$ is the number deaths in year t aged x and with sex s.

$D_1(t,x,s)$ is the number of deaths in year t aged x and with sex s, who reach age x during the year t.

$D_2(t,x,s)$ is the number of deaths in year t aged x and with sex s, who reached age x during the year t-1.

$b_1(t,x,s,i)$ is defined as the difference (in number of years) between the date of the death and the birthday (during the year t) of each individual i with sex s dead during the year t aged x and reaching age x during the year t. It should be taken into account that this quantity is the time lived (in number of years) with age x by each individual dead with this age in the year t of the generation which reach the age x along such year.

$b_2(t,x,s,i)$ is defined as the difference (in number of years) between the date of the death and the 1st of January of the year t for each individual I with sex s dead during the year t with age x and who reached the age x during the year t-1. It should be taken into account that this quantity is the time lived (in number of years) with age x by each individual dead with this age in the year t of the generation which reached the age x along the year t-1.

For the open age group (100 years old and over), the aforementioned indicator is estimated taking:

$$m_{100+} = \frac{D(t,100+,s)}{\frac{(P(t,100+,s)-D_2(t,100+,s))}{2} + \sum_{i=1}^{D_2(t,100+,s)} b_2(t,100+,s) + \frac{P(t+1,100+,s)}{2} + \sum_{i=1}^{D_1(t,100,s)} b_1(t,100,s,i)}$$

where:

t, the observation year or period.

s, the sex, which may take attributes of male, female or both sexes.

$D(t,100+,s) = \sum_{x \geq 100} D(t,x,s)$ is the total for deaths occurring during year t of individuals aged 100 years old or more and of sex s.

$D_1(t,100,s)$ is the number of deaths in year t aged 100 and with sex s, who reach age 100 during the year t.

$D_2(t,100+,s)$ is the number of deaths in year t aged 100 or more and with sex s, who reached age 100 during the year t-1.

$P(t,100+,s) = \sum_{x \geq 100} P(t,x,s)$ is the resident population stock as of 1 January of year t aged 100 years old or more and of sex s.

$b_1(t,x,s,i)$ is defined as the difference (in number of years) between the date of the death and the birthday (during the year t) of each individual i with sex s dead during the year t aged 100 and reaching age 100 during the year t. It should be taken into account that this quantity is the time lived (in number of years) with age 100 by each individual dead with this age in the year t of the generation which reach the age x along such year.

$b_2(t,x,s,i)$ is defined as the difference (in number of years) between the date of the death and the 1st of January of the year t for each individual I with sex s dead during the year t with age 100 or more and who reached the age 100 during the year t-1. It should be taken into account that this quantity is the time lived (in number of years) with age 100 or more by each individual dead with this age in the year t of the generation which reached the age 100 along the year t-1.

Assimilating the estimated values of the specific mortality rates of the population studied with those corresponding to the specific mortality rates at every age x of a fictitious cohort of 100,000 individuals, the likelihood or risk of death aged x, q_x, of the aforementioned cohort of individuals, which presents the same incidence of mortality at each age as the population observed in the reference year, is estimated using the expression:

$$q_x = \frac{m_x}{1+(1-a_x)\cdot m_x}, x = 0,1,...,99.$$

where a_x is the average of years lived in the last year of life by those individuals in the fictitious cohort who die having reached age x.

The aforementioned function, a_x, is also estimated reflecting the short-term incidence of mortality regarding the population studied in the reference year, taking the average time lived at age x by individuals of the aforementioned population who die at said age during that year, that is:

$$a_x = \frac{\sum_{i=1}^{D(t,x,s)} a(t,x,s,i)}{D(t,x,s)}, \text{ x=0,1,...,99.}$$

$$a_{100+} = \frac{\sum_{i=1}^{D(t,100+,s)} [a(t,x,s,i)+(x-100)]}{D(t,100+,s)}$$

where $a(t,x,s,i)$ is the time lived by individual i of the population under study, of sex s, who died at age x during reference year t. For the individuals of t-x generation, it is consistent with b_1. For the individuals of t-x-1 generation it is calculated like $1+b_1$

For the open age group (100 years old and over) considered, for whom death is a certainty, there is:

$$q_{100+} = 1$$

$$a_{100+} = \frac{1}{m_{100+}}$$

The functions of survivors, l_x, and of theoretical deaths, d_x, from the table are obtained recurrently:

$$l(0,s) = 100000$$

$d(x,s) = l(x,s) \cdot q(x,s)$ and $l(x+1,s) = l(x,s) - d(x,s)$, for $x = 0,1,2\ldots,99,100+$.

Moreover, the total lived (measured in years) by individuals from the fictitious generation, of sex s, having reached age x or seasonal population of the table, is derived from the expression:

$L_x = l_{x+1} + a_x \cdot d_x$, for $x = 0,1,\ldots,100+$

Lastly, the function of life expectancy at age x for sex s results from:

$$e_x = \frac{\sum_{y \geq x} L_y}{l_x}$$ for x=0, 1, ... , 99, 100+.

INE National Statistics Institute

Changes made in the mortality tables since December 2015

In December 2015, the following changes are made:

1) Data update for period 1991-2013

Throughout 2014 series of intercensal population for the periods 1981-1991 and 1991-2001 were calculated in order to obtain a continuous and unified set of population figures beginning on January 1st of 1981 and it is linked to the last period of the figures published population statistics.

In December 2015 mortality tables calculated on the basis of the new population intercensal will be published.

Methodologial chages:

Until 2015, closing mortality tables was calculated using the following formulas:

$$a_w = \frac{1}{m_w}$$

$$q_w = 1$$

$$e_w = a_w$$

Since December 2015, the closure results are obtained in accordance with the following formulas:

$$a_{w+} = \frac{\sum_{i=1}^{D(t,w+,s)} [a(t,x,s,i) + (x-w)]}{D(t,w+,s)}$$

$$q_w = 1$$

$$e_w = a_w$$

That is, before, with m_w calculated using the observed data, a_w was inferred to q_w be equal to 1. In turn, a_w is obtained from the data so that life expectancy at age closing deduct also observed values rather than obtained from the closure assumption.

3) The segment b1 of the denominator of the rate estimate until 2015 is estimated to begin on January 1st of the year t. Since December 2015 it is estimated to start on 12 December t-1, in order to have the January 1st in the risk exposure of the deceased persons.

3 Life Tables of Autonomous Communities and Provinces

The life table for an Autonomous Community or Province measures the incidence of mortality regarding the population resident therein during the reference year by simulating its behaviour regarding a *cohort* or *fictitious generation* of individuals subject to a mortality pattern by identical age to that observed regarding the population studied during the observation period. Specifically, the aforementioned simulation consists of applying the incidence of mortality at each given age to a fictitious generation of individuals, basically, by specific rates observed regarding the population resident in the Autonomous Community or Province considered during the reference year and deriving, using these, the other functions comprising their mortality table.

So then, maintaining the objective of providing a measurement of the short-term incidence of the phenomenon during the reference year, a results aggregation procedure is adopted from a complete mortality table by simple age, in five-year age groups, in order to avoid unwanted distortions regarding the results provided which might impede their interpretation as a direct result of the randomness of the information pertaining to smaller populations.

In this way, taking the function of survivors, l_x, and of the seasonal population, L_x, from a complete mortality table calculated with a similar methodology to the one used for the national total, the values are determined from the function of survivors, theoretical deaths, and seasonal population from the Autonomous Community or Provincial mortality table with results aggregated five-year age groups ($[0,1),[1,5),[5,9),\ldots,[95,\infty)$)[1]:

$l_0 = 100.000$

l_x, for $x = 1,5,\ldots,95$

$d_{0,1} = l_0 - l_1$

$d_{1,5} = l_1 - l_5$

$d_{x,x+n} = l_x - l_{x+n}$, for $x = 5,\ldots,95$ and $n = 5$

where $d_{x,x+n}$ are those individuals from the fictitious generation or cohort who died at the age reached belonging to the group $[x, x+n)$.

$d_{95+} = l_{95}$

where

[1] In the case of there not being any deaths observed in the Autonomous Community or Province considered during the reference year in the highest age group, the results will be offered considering as the open final age group that resulting from aggregating the latter to the immediately preceding one. In any case, the results for Ceuta y Melilla present as the last age group that aged 90 years old and over.

$$L_{0,1} = L_0$$

$$L_{1,5} = \sum_{1 \le y < 5} L_y$$

$$L_{x,x+n} = \sum_{x \le y < x+n} L_y \text{ , for } x = 5,\ldots,90 \text{ and } n = 5$$

where $L_{x,x+n}$ is the total lived (measured in years) by individuals from the fictitious cohort, between ages x and $x+n$.

The season population in the open group aged 95 years old and over takes value L_{95+} from the complete start table.

Moreover, the function of likelihood or risk of death aged x then results from the expression:

$$q_{0,1} = \frac{d_{0,1}}{l_0}$$

$$q_{1,5} = \frac{d_{1,5}}{l_1}$$

$$q_{x,x+n} = \frac{d_{x,x+n}}{l_x} \text{ , for } x = 5,\ldots,95 \text{ and } n = 5$$ [2]

where $q_{x,x+n}$ is the likelihood or risk of individuals of the fictitious generation or cohort surviving to the age of x dying before reaching the age of $x+n$ years.

$$q_{95+} = 1$$

Therefore, life expectancy at age x is calculated as:

$$e_x = \frac{\sum_{y \ge x} L_y}{l_x} \quad \text{for } x = 1,5,\ldots,95$$

Lastly, both the average of years lived in the last year of life by those individuals in the fictitious cohort who die belonging to group $[x, x+n)$ and the specific mortality rate at age $[x, x+n)$, are consistently estimated, by means of the expressions:

[2]In the extreme case of deaths observed in the Autonomous Community or Province considered during the reference year being inconsistent with the population estimates used in the calculation resulting in an impossible value of likelihood of death in a specific age group older than 1, the aforementioned value would be accounted for by 1 and the results would be provided with the final age group aggregated taking the latter.

$$a_{0,1} = 1 - \frac{l_0 - L_{0,1}}{d_{0,1}}$$

$$a_{1,5} = 1 - \frac{4 \cdot l_1 - L_{1,5}}{4 \cdot d_{1,5}}$$

$$a_{x,x+n} = 1 - \frac{n \cdot l_x - L_{x,x+n}}{n \cdot d_{x,x+n}}, \quad \text{for } x = 1,5,\ldots,95 \text{ and } n = 5$$

where $a_{x,x+n}$ is the average years lived in the last year of life by those individuals in the fictitious cohort who die belonging to the group $[x, x+n)$.

$$a_{95+} = \frac{1}{m_{95+}}$$

where a_{95+} is the average years lived by those surviving to the age of 95 years from the fictitious cohort from the aforementioned age.

$$m_{0,1} = \frac{d_{0,1}}{L_{01,}}$$

$$m_{1,5} = \frac{d_{1,5}}{L_{1,5}}$$

$$m_{x,x+n} = \frac{d_{x,x+n}}{L_{x,x+n}}, \text{ for } x = 1,5,\ldots,95 \text{ and } n = 5$$

where $m_{x,x+n}$ is the specific mortality rate in age group $[x, x+n)$ of the fictitious cohort.

$$m_{95+} = \frac{d_{95+}}{L_{95+}}$$

where m_{95+} is the mortality rate for age 95 years old or over from the fictitious cohort.