Module 1: Life Tables

Marcus Milazzo and Jack Conway

"There must be some mistake. According to our actuary tables I'm going to live to 83."

1.1 Life Table Components

Life tables, also known as **mortality tables**, are a way to display the conditional probability that a person will die in a certain year given that they are a certain age. They often take into account many factors including gender, occupation, socioeconomic status, lifestyle habits, and more in order to predict how much longer a life is likely to last. Life tables are useful in a variety of fields, but a couple examples related to actuarial work are insurance companies using them to help price products, and government employees using them to predict financial expenses for Social Security. Typical life tables include similar summary information, and often appear like the one below.

Figure 1.1. Example of Life Table

	Male			Female		
Exact age	Death probability [a]	Number of lives [b]	Life expectancy	Death probability [a]	Number of lives [b]	Life expectancy
0	0.006680	100,000	76.10	0.005562	100,000	80.94
1	0.000436	99,332	75.62	0.000396	99,444	80.39
2	0.000304	99,289	74.65	0.000214	99,404	79.43
3	0.000232	99,259	73.67	0.000162	99,383	78.44
4	0.000172	99,235	72.69	0.000132	99,367	77.46
5	0.000155	99,218	71.70	0.000117	99,354	76.47
6	0.000143	99,203	70.71	0.000106	99,342	75.47
7	0.000131	99,189	69.72	0.000099	99,332	74.48
8	0.000115	99,176	68.73	0.000093	99,322	73.49
9	0.000096	99,164	67.74	0.000090	99,313	72.50
10	0.000082	99,155	66.74	0.000090	99,304	71.50
11	0.000086	99,147	65.75	0.000096	99,295	70.51
12	0.000125	99,138	64.76	0.000111	99,285	69.52
13	0.000205	99,126	63.76	0.000137	99,274	68.52
14	0.000319	99,106	62.78	0.000170	99,261	67.53

To introduce the fundamentals of mortality tables, we will begin with a highly simplified version.

Table 1.1. Simplified Life Table

```
 x   lx  dx
 0 1000 100
 1  900 100
 2  800 100
 3  700 100
 4  600 100
 5  500 100
 6  400 100
 7  300 100
 8  200 100
 9  100 100
10    0   0
```

The x column represents the age of the person at the beginning of the year. The lx column represents the expected number of people alive at the beginning of year x, which is mathematically expressed as l_x. The dx

column represents the number of people that died during year x, which is expressed as d_x.

Before, we mentioned that a life table should display the conditional probability of survival given you are age x. While *Table 1.1* does not display that yet, we can calculate those probabilities easily.

Let's begin by focusing on a person that is 0 years old, or a newborn. We denote a person of age x by putting the age inside parentheses. So, in this example, we denote a life of age 0 with (0). Suppose we want to find the probability that (0) survives their first year of life. It's as simple as making sure we understand what the information in the life table means, and then taking advantage of the definitions of l_x and d_x.

Row 1, when $x = 0$, represents the beginning of the first year where we know from the life table that we start with 1000 people. Row 2, when $x = 1$, tells us that $l_1 = 900$ which means we expect 900 people to be alive at the beginning of year two. This necessarily means that we expect 100 people to die during the first year which is reflected by d_0. Therefore, we can represent the probability as follows:

$$Pr((0)\ survives\ the\ first\ year) = \frac{900}{1000} = .9$$

Now, lets calculate the probability that (1) survives to age two. Like the previous example, we will take the expected number of survivors at the beginning of year two and divide it by the expected number of survivors at the end of year two.

$$Pr((1)\ survives\ the\ year) = \frac{800}{900} \approx .889$$

The probabilities that we have calculated above are denoted with p_x where x is the age of the person of interest. So, we write the probability that (0) survives the year as p_0, and the probability that (1) survives the year as p_1.

We then have natural notation for the probability that a person dies within a certain year, denoted with q_x, where

$$q_x = 1 - p_x$$

.

In *Table 1.2* we have added columns p_x and q_x to our previous life table.

Table 1.2.

x	lx	dx	px	qx
0	1000	100	0.900	0.100
1	900	100	0.889	0.111
2	800	100	0.875	0.125
3	700	100	0.857	0.143
4	600	100	0.833	0.167
5	500	100	0.800	0.200
6	400	100	0.750	0.250
7	300	100	0.667	0.333
8	200	100	0.500	0.500
9	100	100	0.000	1.000
10	0	0	NA	NA

With *Table 1.2*, we can now calculate more conditional probabilities. In the above examples, notice that we calculated the probabilities of surviving year x, but what if we want to know the probability of surviving for periods longer than that? For example, suppose we want to know the probability that a newborn survives the year they were born, *and* the year after that. We can represent that as follows:

$$
\begin{aligned}
& Pr((0)\ survives\ the\ first\ year\ and\ the\ second\ year) \\
&= Pr((0)\ survives\ the\ first\ year) Pr((0)\ survives\ the\ second\ year \mid (0)\ survived\ the\ first\ year) \\
&= Pr((0)\ survives\ the\ first\ year) Pr((1)\ survives\ the\ second\ year) \\
&= p_0 p_1 = (\frac{900}{1000})(\frac{800}{900}) = .8
\end{aligned}
$$

Thankfully, we have notation for this type of problem. We will represent the probability that (x) survives to year $x+t$ with ${}_tp_x$. So, in the previous example, we would write ${}_2p_0$ to represent the probability that a newborn reaches age 2. In general, we can calculate the probability that a person age x reaches age $x+t$ with

$$
{}_tp_x = p_x p_{x+1} p_{x+2} \cdots p_{t-1}
$$

.

Another intuitive calculation for ${}_tp_x$ is

$$
{}_tp_x = \frac{l_{x+t}}{l_x}
$$

This makes sense because it is the proportion of people that we expect to still be alive at year $x+t$ given that we started with l_x.

One more important calculation that can come from a life table is the probability that (x) will die in some age range $u+t$. In other words, the probability that a person age x will live to age t and then die within the next u years. We will represent this probability with

$$
\begin{aligned}
{}_{t|u}q_x &= {}_{t+u}q_x - {}_tq_x \\
&= {}_tp_x - {}_{t+u}q_x
\end{aligned}
$$

It can also be shown that

$$
{}_{t|u}q_x = {}_tp_x \ {}_uq_{x+t}.
$$

We will use this to calculate $2 \mid 3q_4$, the probability that a person of age 5 lives to age 7 and then dies before age 9.

$$
\begin{aligned}
{}_{2|3}q_4 &= {}_2p_4 \ {}_3q_{4+2} \\
&= p_4 p_5 (1-p_6)(1-p_7)(1-p_8) \\
&= (.833)(.800)(1-.750)(1-.333)(1-.500) \\
&= .0556
\end{aligned}
$$

If $u = 1$ we just write ${}_{t|}q_x$.

The future years completed by (x) prior to death is called the *curtate-future-lifetime* of (x). The expected value of this quantity, also known as the *curtate-expectation-of-life", in life tables is commonly denoted as e$_x$, the remaining life expectancy at age x.

2.1 Life Table Construction

Now that you have a basic understanding of the information contained in life tables, it would be a good time to get an idea of how these life tables are constructed.

There is an immense body of knowledge out there for constructing life tables, and it would take an entire book to dive into all of the details. Here, we will opt to provide a background of the different factors that are relevant to their building.

The most straightforward way to construct a complete life table is by observing a population at the time of birth and following them throughout the years, recording the number of deaths each year until there are no longer any survivors. The group that is being followed is called a **cohort**; so these are called **cohort life tables**. In the United States and many other countries, the numbers of births and deaths are recorded each year. For the U.S., there is data for births and deaths dating all the way back to 1933. If we know exactly how many people are born at 1933, we can treat them as a cohort and record deaths at each age to calculate death rates like in Section 2.1.

The major drawback of a cohort life table is that you only have data in the table up until the current point in time, and can only make rough predictions for future probabilities. For example, if we began examining a cohort that was born in 1985, we would only have data in 2016 that reflects a life table through age 31. Trying to predict the death rates for people aged 40, 50, and 60, could put you at risk of large error by extrapolation.

This leads us to another common type of life table, called a **static**, or **period**, **life table**. This method involves looking at an entire population over a short period of time; most commonly 1-3 years. This is often done around a census period, where one can utilize information about total number of people alive in a country or region and what age they are. By then looking at the number of deaths and age of death over the short time period, you can estimate the death rates at each age.

The downside to period life tables is that this snapshot of death rates may not be reflective of the death rates that younger people will experience when they're older. For example, the probability of death for an 85 year old in the current period life table is not likely to be the probability of death when I actually turn 85 years from now due to advances in medicine or other changes in the environment.

Another common approach to constructing a life table is by statistical modeling. Like you've seen in your statistics classes, it may be the case that fitting a model will provide deeper insight into your data, and a more accurate way to estimate or forecast. There are several popular methods for modeling death rates, and they all try to compete to be the best one by including different parameters or making different assumptions. A detailed comparison of some of the most popular methods can be found via a link in the appendix.

In practice, it may depend what your goals are with the life table when choosing which method, or combination of methods, to employ. Like any other model, the one you build is not 100% right, but you're hoping it's useful.

Problems

1.

Referring to table below, evaluate the probability that someone that survives to age 2 will
a. live to year 4.
b. die before year 5.
c. die during 5th year of life.

Life Table for the Total Population: United States, 1979-81:

Period of Life between Two Ages x to $x + t$	Proportion of Persons Alive at Beginning of Age Interval Dying during Interval ${}_tq_x$	Number Living at Beginning of Age Interval L_x	Number Dying during Age Interval ${}_td_x$
0-1	0.01260	100000	1260
1-2	0.00093	98740	92
2-3	0.00065	98648	64
3-4	0.00050	98584	49
4-5	0.00040	98535	40
5-6	0.00037	98495	36

2.

(Saving the soa.08 data as a data frame)
Using the soa.08 data set, recreate the following columns in R.
a. (p_x) (hint: take next l_x / current l_x)
b. (e_x) (hint: sum expectation during remaining years / current l_x)

Use the following code to subset the data:

```
new_soa.08 = soa.08[,1:2]
```

Solutions

Problem 1:

1a.

$$\frac{s(4)}{s(2)} = \frac{l_4}{l_4} = \frac{98535/100000}{98648/100000} = \frac{98535}{98648} = 0.99885$$

1b.

$$\frac{s(2)-s(5)}{s(2)} = 1 - \frac{l_5}{l_2} = 1 - \frac{98495}{98648} = 0.001551$$

1c.

$$\frac{s(4)-s(5)}{s(2)} = \frac{(l_4 - l_5)}{s(2)} = \frac{(98535 - 98495)}{98648} = 0.0004055$$

Problem 2:

2a.

There are many ways to code this, one option for (p_x) would be:

```
for (i in 1:length(row.names(new_soa.08))){
       new_soa.08$px[i] =  new_soa.08$lx[(i+1)]/ new_soa.08$lx[i]
}

new_soa.08$px[length(row.names(new_soa.08))] = 0
```

2b.

One option for creating (e_x):

```
for (i in 2:length(row.names(new_soa.08))){
     new_soa.08$ex[i] =
         sum(new_soa.08$lx[-c(1:(i-1))] * new_soa.08$px[-c(1:(i-1))])
         / new_soa.08$lx[i]

}

new_soa.08$ex[1] = sum(new_soa.08$lx * new_soa.08$px)
                 / new_soa.08$lx[1]
```