

“Life Expectancy Tables”

Getting SAS® to Run the Hard Math

Anna Vincent, Suting Zheng
Center for Health Statistics,
Texas Department of State Health Services

Abstract

Life expectancy (LE) tables are statistical tools typically used to portray the expectation of life at various ages and are a fundamental tool in survival analysis. They also provide information on the number of individuals who survive to various ages, median age at death, age-specific death rates, and the probability of dying at certain ages. LE at birth is the most frequently cited LE table statistic. The Texas Department of State Health Services (DSHS) creates LE for publication in the Vital Statistics Annual Report every year. Previously, these tables were constructed in a multi-step process using detailed Python and Java script. In this paper we provide a general framework for building LE tables using SAS.

Introduction

The LE table has become one of the most popular statistics in community health assessment. Like Infant Mortality Rates, they help to take a ‘snapshot’ of the health of a community or population. LE tables are typically used to portray the expectation of life in specific areas or populations at various ages. At the same time, LE tables provide information on survivals to various ages, the median age at death, age-specific death rates, and the probability of dying at certain ages. Life expectancy at birth simply constitutes our best estimate of how long a person might live if medical practices remain the same.¹ The Texas DSHS often get requests from our clients to “personalize” the LE table to their specific geographic areas, like Texas counties or by health regions. The time and resources involved in making these tables are currently prohibitive².

For the past twenty years, the Texas DSHS, Center for Health Statistics, have used several computer programs (a combination of Survival 5.0, Java, Python, and Microsoft Excel) to generate LE tables.³ To replace this detailed process, we searched the existing SAS papers for scripts to build LE tables but could not find one that suited our needs.⁴ Neither PROC LIFETEST nor PROC LIFEREG produced the output we desired. We therefore developed a SAS syntax which constructs LE tables in a few simple steps.

In the sections below, we describe the calculation method and the SAS syntax which generates a LE table. Two techniques which are used here are the LAG() function and the RETAIN statement.

Getting Started: The Mathematics¹

There are several columns which make up the life tables. Getting familiar with them will help us to understand how the life table is built. These parameters are:

- 1) Age interval (x to $x+n$): the period of life between two exact ages. Here x indicates the starting point for an age interval, and n is the interval length.
- 2) Proportion dying (q_x): the proportion of persons alive at the beginning of each age interval who die before reaching the end of the age interval.
 - a. For infants, q_0 is calculated as:

$$q_0 = \frac{\text{Infant Deaths}}{\text{Unweighted Live Births}}$$

- b. For age group of 75+, we set the $q_{75} = 1$ because this cohort will all die.
- c. For all other age cohorts, q_x is calculated as:

$$q_x = \frac{2 * n * d_r}{(2 + n * d_r)}$$

Where n is the number of years within an age interval and d_r is the age-specific death rate.

Proportion surviving (p_x): the proportion of persons alive at the beginning of each age interval who survive over the age interval.

$$p_x = 1 - q_x$$

- 3) Number surviving (l_x): number of people living at the beginning of each age interval. Life table usually represents a hypothetical cohort of 100,000 persons born at the same instant, therefore $l_0 = 100,000$. For the other age cohorts,

$$l_x = l_{x-n} * p_{x-n}$$

Number dying (d_x): number of persons dying during the age interval.

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

- 4) The number of person-years lived in each age interval:

$$L_x = n * l_{x+n} + d_x * a_x$$

Here a_x is the average amount of time lived in interval x to $x+n$ by those dying in the interval.

- 5) The number of person-years lived in each age interval and all subsequent age intervals:

$$T_x = L_x + L_{x+n} + \dots + L_w$$

or

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

Average remaining life time (e_x) or the expectation of life at any given age (the average years remaining to be lived by those surviving to that age).

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

Translating from Formulas to SAS Syntax

We will use the following data to generate a related life table. The numbers are from Texas Vital Statistics Annual Report, Table 24 for the reporting year of 2014⁵.

Table 1. Dataset used to generate Life Table for Texas Residents, 2014

SexRace	years	agegroup	population	deaths	births	ax
Texas Total	1	0	406,481	2,270	403,439	0.15
Texas Total	4	1	1,581,709	423	403,439	1.65
Texas Total	5	5	1,972,064	260	403,439	2.25
Texas Total	5	10	2,048,295	281	403,439	3.05
Texas Total	5	15	2,040,335	929	403,439	2.75
Texas Total	5	20	2,006,970	1,733	403,439	2.55
Texas Total	5	25	1,940,844	1,912	403,439	2.5
Texas Total	5	30	1,976,615	2,336	403,439	2.55
Texas Total	5	35	1,865,823	2,677	403,439	2.65
Texas Total	5	40	1,830,349	3,469	403,439	2.7
Texas Total	5	45	1,742,003	5,233	403,439	2.7
Texas Total	5	50	1,765,609	8,712	403,439	2.65
Texas Total	5	55	1,673,253	12,418	403,439	2.65
Texas Total	5	60	1,403,713	15,151	403,439	2.65
Texas Total	5	65	1,143,547	17,579	403,439	2.6
Texas Total	5	70	798,713	18,986	403,439	2.6
Texas Total	25	75	1,272,791	94,788	403,439	11.5

- 1) We imported the data into SAS® 9.4 and named the dataset 'test'.
- 2) We then calculated three new variables:
 - a. q_x , the proportion of persons alive at the beginning of each age interval who die before reaching the end of the age interval;
 - b. d_r , age specific death rate; and
 - c. p_x , the probability of surviving over the age interval.

For age group 0, we used the actual number of births in the calculation of

$q_0: q_0 = \frac{\text{Infant Deaths}}{\text{Unweighted Live Births}}$. For age group 75+, we set q_{75} to 1.0 since the cohort will eventually all die. For all other age groups, q_x was calculated as $q_x = \frac{2 \cdot n \cdot d_r}{(2 + n \cdot d_r)}$, where n is the column 'years' and d_r the age-specific death rate.

```

Data test1;
  Set test;
  Length qx dr px 8;
  If agegroup=0 then Do;
    dr=deaths/births;
    qx=deaths/births;
  End;
  Else Do;
    dr=deaths/population;
    If agegroup=1 then qx=2*4*dr/(2+4*dr);
    Else If agegroup=75 then qx=1.000000;
    Else qx=2*5*dr/(2+5*dr);
  End;
  px=1-qx;
  Format qx dr px 8.5;
Run;

```

- 3) In the next step, we calculated the number surviving (l_x), which denotes the number of people living at the beginning of each age interval. The first number of this column, l_0 , is an arbitrary

number which is usually set to 100,000, meaning 100,000 live birth happening at the same instant. Each successive number represents the number of survivors at the age(x). To calculate the probability of surviving at age(x), we create two new variables, x and xx. x is initially used to hold the value of p_0 . In each of the following age group, the value of x is passed to xx using a retain statement, and x itself is used to calculate the survival probability of the following age group. The surviving population at the beginning of each age interval is calculated as $I_x = 100,000 * xx$.

```

Data test2;
  Set test1;
  Length x xx Ix 8;
  If years=1 then Do;
    x=1;
    x=x*px;
    xx=1;
    Ix=100000;
  End;
  Else do;
    Retain x;
    xx=x;
    Ix=100000*xx;
    x=x*px;
  End;
Format Ix comma8.0;
run;

```

- 4) In order to calculate the expected number of deaths in each age group (d_x) and number of person-years lived (L_x), we first inverted the dataset so that we can use the LAG() function to capture the previous I_x , which is the number of people surviving at the beginning of the next age group. The expected number of deaths in each age group, d_x , was calculated as $d_x = I_x - Lag(I_x)$. The variable LagIx is set to 0 for the last age group, 75 years and older, which means that the entire population will die in the end.

Each of the d_x people who die during the interval x and $x+n$ has lived x complete years plus some fraction of the years n . The average of these fractions, denoted by a_x , is usually set as a constant and we have these values pre-determined in the table.

Since each member of the cohort who survives the year interval x to $x+n$ contributes “ n ” number of years to L_x . While those members who die during this period of time contributes, on the average, a fraction a_x of the “ n ” number of years, so that $L_x = n * LagI_x + d_x * a_x$.

```

Proc Sort Data=test2;
  by descending Agegroup;
Run;

Data test3;
  Set test2;
  Length LagIx dx Lx 8;
  LagIx=lag(Ix);
  if Agegroup=75 then LagIx=0;
  dx = Ix -LagIx;
  Lx=years*LagIx + dx*ax;
  Format LagIx dx Lx comma8.0;
Run;

```

- 5) Now we were very close to getting the life expectancies, but there was still one more step. We calculated the total number of years lived beyond age(x), T_x . This is equal to the sum of the number of years lived in each age interval beginning with age x, and can be calculated as $T_x = L_x + L_{x+n} + \dots + L_{75}$, or $T_x = L_x + T_{x+n}$. After we get T_x , the number of years yet to be lived by a person now at age x is calculated as $e_x = \frac{T_x}{I_x}$.

Now that we have completed all calculations, the table was inverted to the original order from the youngest age to the oldest.

```
Data Test4;
  Tx = 0;
  Do i = 1 to 17;
    Set Test3;
    Tx + Lx;
    Ex=Tx/Ix;
    output;
  End;
  Drop Tx i;
Format Tx 8.0 Ex 8.2;
Run;
Proc Sort Data=Test4;
  By Agegroup;
Run;
```

Creating the Report:

To create the final life table output, we used a template from a previous SAS paper⁴ to create our life table, and then used the ODS PDF output to create the output table. We formatted the variables using Proc Format (not shown here).

```
proc template;
  define style self.border;
    parent=styles.SansPrinter;
    style Table /
      rules = groups
      frame=hsides
      cellpadding = 3pt
      cellspacing = 0pt
      borderwidth = 2pt;
    style header /
      font_weight=bold
      background=white
      font_size=3;
  end;
run;

ods listing close;
ods pdf file="c:\temp\test.pdf" style=self.border;

Proc report data=final headline headskip nowd spacing=2 split='- ' center ;
format sexRace $sexf. ;
columns agegroup years deaths population qx Ix dx ax Lx ex;
define agegroup /display f=$agef. "-Age-Group" width=5 center;
define years /display "Years" width=5 center;
define deaths /display f=comma12.0 "Number-of-Deaths" width=9 center;
```

```

define population /display f=comma12.0 "-Estimated-Population" width=10
center;
define qx /display f=12.5 "(qx)" width=11 center;
define Ix /display f=comma12.0 "(Ix)" width=11 center;
define dx /display f=comma12.0 "(dx)" width=14 center;
define ax /display f=8.2 "(ax)" width=6 center;
define Lx /display f=comma12.0 "(Lx)" width=14 center;
define ex /display f=12.2 "(ex)" width=18 center;

title ="Abridged Life Tables for Texas Residents, 2014";
footnote =" Life expectancy at birth";
run;

ods pdf close;
ods listing;

```

Table 2: Output of the Proc Report: Abridged Life Table for Texas Residents, 2014

Abridged Life Tables for AR Table 24 by All Texas for 2014***

1

Age Group	Years	Number of Deaths	Estimated Population	(qx)	(Ix)	(dx)	(ax)	(Lx)	(ex)
< 1	1	2,320	401,044	0.00581	100,000	581	0.15	99,506	78.24
1-4	4	373	1,563,818	0.00095	99,419	95	1.65	397,454	77.70
5-9	5	262	1,963,917	0.00067	99,324	66	2.25	496,440	73.77
10-14	5	302	2,027,727	0.00074	99,258	74	3.05	496,147	68.82
15-19	5	906	1,989,542	0.00227	99,184	226	2.75	495,414	63.87
20-24	5	1,654	1,973,591	0.00418	98,959	414	2.55	493,780	59.01
25-29	5	1,874	1,910,287	0.00489	98,545	482	2.50	491,519	54.24
30-34	5	2,125	1,951,334	0.00543	98,063	533	2.55	489,009	49.50
35=39	5	2,487	1,813,357	0.00683	97,530	667	2.65	486,085	44.76
40-44	5	3,569	1,825,703	0.00973	96,864	942	2.70	482,152	40.05
45-49	5	5,356	1,709,853	0.01554	95,922	1,491	2.70	476,179	35.41
50-54	5	8,818	1,764,935	0.02467	94,431	2,330	2.65	466,679	30.93
55-59	5	12,291	1,630,397	0.03700	92,101	3,407	2.65	452,498	26.64
60-64	5	14,806	1,349,362	0.05340	88,694	4,736	2.65	432,338	22.57
65-69	5	16,343	1,083,217	0.07270	83,958	6,103	2.60	405,140	18.69
70-74	5	17,956	760,459	0.11148	77,854	8,679	2.60	368,441	14.95
75+	25	91,855	1,238,416	1.00000	69,175	69,175	11.50	795,513	11.50

In this way, we used SAS to produce a LE table that calculated all values as desired and only required minor cleanup (including adding subscripts and table notes) prior to publication.

Conclusion

We often receive requests to “personalize” Life Expectancy table to their specific geographic areas, like Texas counties or by health regions. The time and resources

involved in making these tables used to be prohibitive. The syntax created here greatly cuts down the time and effort to generate the life expectancy table, not only for creating the Texas Vital Statistics Annual Report but for all our requests.

Contact Information

Comments or Questions:

Anna Vincent
Center for Health Statistics, MC-1898
Texas Department of State Health
Services

PO Box 149317

Austin, TX 78714-9347

512-776-2724 work

512-458-3255 fax

Emails:

Anna.vincent@dshs.texas.gov

VSTAT@dshs.texas.gov

Website:

<http://www.dshs.texas.gov/chs/chs/default.htm>

Suting Zheng

Center for Health Statistics, MC-1898

Texas Department of State Health

Services

PO Box 149317

Austin, TX 78714-9347

512-776-6974 work

512-458-3255 fax

Emails:

Suting.Zheng@dshs.texas.gov

VSTAT@dshs.texas.gov

Website:

<http://www.dshs.texas.gov/chs/chs/default.htm>



Sources:

- 1) *Life Table Construction*, 2014 Texas Vital Statistical Annual Report, Technical Appendix
- 2) *The Methods and Materials of Demography (Condensed Edition)*, Henry S. Shryock and Jacob S. Siegel, Academic Press, NY, 1976.
- 3) Survival 5.0 a program written by David Smith at the University of Texas Health Science Center at Houston.
- 4) *SAS Macros for Generating Abridged and Cause-Eliminated Life Tables*, Zhao Yang and Xuezheng Sun, SAS Users Group International (SUGI) #31.
- 5) <http://dshs.texas.gov/chs/vstat/vs14/t24.aspx>; "Table 24 Life Tables by Race/Ethnicity and Sex, Texas, 2014 Total Texas Population"

The authors of the paper/presentation have prepared these works in the scope of their employment with the Texas Department of State Health Services and the copyrights to these works are held by the Center for Health Statistics.

Therefore, Anna Vincent hereby grants to SCSUG Inc a non-exclusive right in the copyright of the work to the SCSUG Inc to publish the work in the Publication, in all media, effective if and when the work is accepted for publication by SCSUG Inc.

This the 1st day of September, 2017.