Demographic Methods

Monica Alexander

Week 2: Mortality

Overview

- Life tables
- Mortality models
  - Skeleton rmds in git repo for Bayesian LC and Gompertz

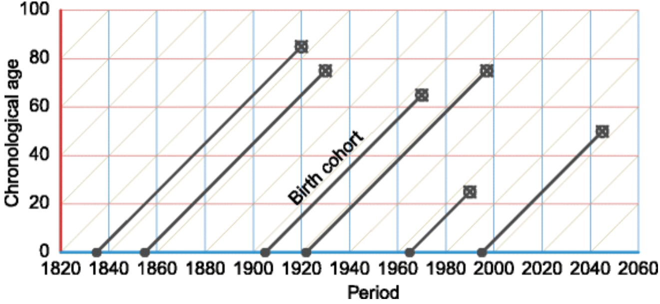
Life tables

What is a life table?

A life table describes the **survivorship by age** for a certain population.

There are different ways of describing survivorship (e.g. probability still alive, probability of dying, years of life left to live. . .), so a life table has many different columns.

Summarizing survivorship by age



(Riffe et al. 2017)

The life table

$x =$ age

$n =$ length of age interval

x	n	l_x	${}_nd_x$	${}_nq_x$	${}_np_x$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1								
1	4								
5	5								
10	5								
...									
100	∞								

The life table

- how we calculate life expectancy
- used a lot by actuaries
- summary measures to compare populations
- tell us something about the implied stationary population (more later)

History

17. In the next place, whereas many persons live in great fear and apprehension of some of the more formidable and notorious diseases following; I shall only set down how many died of each: that the respective numbers, being compared with the total 229,250, those persons may the better understand the hazard they are in.

Table of notorious diseases		Table of casualties	
Apoplexy	1,306	*Bleeding*	69
Cut of the Stone	38	*Burnt*, and *Scalded*	125
Falling Sickness	74	*Drowned*	829
Dead in the streets	243	*Excessive drinking*	2
Gowt	134	*Frighted*	22
Head-Ache	51	*Grief*	279
Jaundice	998	*Hanged themselves*	222
Lethargy	67	*Killed by several accidents*	1,021
Leprosy	6	*Murdered*	86
Lunatick	158	*Poisoned*	14
Overlaid, and *Starved*	529	*Smothered*	26
Palsy	423	*Shot*	7
Rupture	201	*Starved*	51
Stone and *Strangury*,	863	*Vomiting*	136
Sciatica	5		
Sodainly	454		

Graunt, 1662 (from Smith and Keyfitz)

History

9. Whereas we have found that of 100 quick conceptions about 36 of them die before they be six years old, and that perhaps but one surviveth 76, we, having seven decades between six and 76, we sought six mean proportional numbers between 64, the remainder living at six years, and the one which survives 76, and find that the numbers following are practically near enough to the truth; for men do not die in exact proportions, nor in fractions: from whence arises this Table following:

Viz. of 100 there dies within the first six years	36	The fourth	6
The next ten years, or decade	24	The next	4
The second decade	15	The next	3
The third decade	9	The next	2
		The next	1

Graunt, 1662 (from Smith and Keyfitz)

History

10. From whence it follows, that of the said 100 conceived there remains alive at six years end 64.

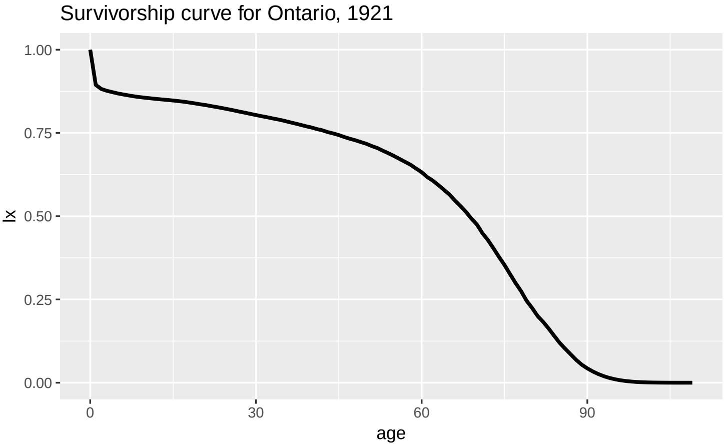
At sixteen years end	40	At fifty-six	6
At twenty-six	25	At sixty-six	3
At thirty-six	16	At seventy-six	1
At forty-six	10	At eighty	0

Graunt, 1662 (from Smith and Keyfitz)

The survivorship function l_x

$l_x =$ the number of survivors at age x

l_0 is the 'radix'. If $l_0 = 1$, then l_x is a probability of survival.



x	n	$\mathbf{l_x}$	${}_nd_x$	${}_nq_x$	${}_np_x$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1	1							
1	4	0.89							
5	5	0.86							
10	5	0.85							
...	...	...							
100	∞	0.0014							

Start by thinking of survival of a cohort of people moving through time.

Deaths ${}_nd_x$

${}_nd_x$ is the number of deaths between ages x and $x + n$.

$$ {}_nd_x = l_x - l_{x+n}. $$

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	${}_nq_x$	${}_np_x$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1	1	0.11						
1	4	0.89	0.03						
5	5	0.86	0.01						
10	5	0.85	...						
...	...	...	...						
100	∞	0.0014	0.0014						

Note that everyone who survived to the last age group must die (*memento mori*)

Probability of death ${}_nq_x$

${}_nq_x$ is the probability of dying between ages x and $x + n$.

$${}_nq_x = \frac{{}_nd_x}{l_x}$$

Note that this is conditional on having survived to age x.

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	${}_np_x$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1	1	0.11	0.11					
1	4	0.89	0.03	0.034					
5	5	0.86	0.01	0.011					
10	5	0.85	...	...					
...	...	...	...	...					
100	∞	0.0014	0.0014	1					

Again, note last interval!

Probability of surviving ${}_np_x$

${}_nq_x$ is the probability of surviving between ages x and $x + n$.

$$ {}_np_x = 1 -_n q_x. $$

- Note again: conditional probability.
- How else could we calculate the probability of surviving?

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_np_x}$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1	1	0.11	0.11	0.89				
1	4	0.89	0.03	0.034	0.97				
5	5	0.86	0.01	0.011	0.989				
10	5	0.85	…	…	…				
…	…	…	…	…	…				
100	∞	0.0014	0.0014	1	0				

Average years lived ${}_n a_x$

${}_n a_x$ is the number of years lived by those who died between ages x and $x + n$.

(pretend for now we observe the lifelines of all people so would have this info)

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_np_x}$	$\mathbf{{}_na_x}$	${}_nL_x$	T_x	e_x
0	1	1	0.11	0.11	0.89	0.3			
1	4	0.89	0.03	0.034	0.97	1.5			
5	5	0.86	0.01	0.011	0.989	2.5			
10	5	0.85	...	...	...	...			
...	...	...	...	...	...	...			
100	∞	0.0014	0.0014	1	0	1.5			

Person-years lived ${}_nL_x$

${}_nL_x$ is the number of person-years lived between ages x and $x + n$.

Total PYL = PYL by those who survived + PYL by those who died.

$$ {}_nL_x = n \cdot l_{x+n} +_n a_x \cdot_n d_x. $$

Note: last age interval, no survivors, so we just have ${}_na_x \cdot {}_nd_x$

x	n	$\mathbf{l_x}$	$\mathbf{{}_n d_x}$	$\mathbf{{}_n q_x}$	$\mathbf{{}_n p_x}$	$\mathbf{{}_n a_x}$	$\mathbf{{}_n L_x}$	T_x	e_x
0	1	1	0.11	0.11	0.89	0.3	0.92		
1	4	0.89	0.03	0.034	0.97	1.5	3.485		
5	5	0.86	0.01	0.011	0.989	2.5	4.275		
10	5	0.85	...	...	...	...	...		
...	...	...	...	...	...	...	...		
100	∞	0.0014	0.0014	1	0	1.5	0.0021		

- What's the max that ${}_nL_x$ can be?
- How does ${}_nL_x$ relate to the survival curve?
- So what's the continuous version (L_x)?

Person-years lived above age x T_x

T_x is the number of person-years lived above age x.

It is the sum of all the ${}_nL_x$'s at above age x, i.e.

$$T_x = \sum_{x}^{\infty} {}_nL_x$$

x	n	$\mathbf{l_x}$	$\mathbf{{}_n d_x}$	$\mathbf{{}_n q_x}$	$\mathbf{{}_n p_x}$	$\mathbf{{}_n a_x}$	$\mathbf{{}_n L_x}$	$\mathbf{T_x}$	e_x
0	1	1	0.11	0.11	0.89	0.3	0.92	58.01	
1	4	0.89	0.03	0.034	0.97	1.5	3.485	57.09	
5	5	0.86	0.01	0.011	0.989	2.5	4.275	53.61	
10	5	0.85	...	...	...	...	...	...	
...	...	...	...	...	...	...	...	...	
100	∞	0.0014	0.0014	1	0	1.5	0.0021	0.0021	

- How does T_x relate to the survival curve?
- So what's the continuous version of T_x?

Life expectancy e_x

e_x is the average number of remaining years of life for those who reach age x.

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

- you may be familiar with e_0, life expectancy at birth

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_np_x}$	$\mathbf{{}_na_x}$	$\mathbf{{}_nL_x}$	$\mathbf{T_x}$	$\mathbf{e_x}$
0	1	1	0.11	0.11	0.89	0.3	0.92	58.01	58.01
1	4	0.89	0.03	0.034	0.97	1.5	3.485	57.09	64.1
5	5	0.86	0.01	0.011	0.989	2.5	4.275	53.61	62.4
10	5	0.85	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
100	∞	0.0014	0.0014	1	0	1.5	0.0021	0.0021	1.5

- Does life expectancy have to monotonically decrease over age?

More on survival and probabilities

Example:

- For Kenyan children born in 2005, ${}_1q_0 = 0.053$ and ${}_4q_1 = 0.26$. What is ${}_5q_0$?

Survival probabilities multiply

Converting between 1- and 5-year

- Life table above was (mostly 5-year age groups)
- What if we wanted to convert these to 1-year age groups?
- E.g. For Taiwan in 1978, ${}_5q_{25} = 0.0247$. What is ${}_1q_{27}$?

Assume probability of surviving is constant in the absence of any other information.

Period life tables

Creating a synthetic cohort

- So far, we have assumed we are dealing with cohort data
- But cohort data are not available in a timely manner
- Create a synthetic cohort using mortality rates observed in a period
- Construct life table the same way
- Interpretation is different: e.g. life expectancy at birth is the number of years a newborn could expect to life if all age-specific mortality rates stayed the same in future

${}_nq_x$ conversion

- We observe period mortality rates ${}_nM_x$
- Recall from last week, these are number of deaths / person years lived. So in life table notation:

$$ {}_nM_x = \frac{{}_ndx}{{}_nLx} $$

- We want ${}_nq_x$ (then calculate a whole life table)
- Use conversion formula:

$$ {}_nq_x = \frac{n \cdot {}_nM_x}{1 + (n - {}_na_x) \cdot {}_nM_x} $$

then all other columns can be derived as before, expect. . . .

Average years lived ${}_na_x$

${}_na_x$ is the number of years lived by those who died between ages x and $x + n$.

- If we have individual lifelines, we can work this out
- We don't for period data. **For most age groups** using

$$ {}_na_x = n/2 $$

is a fine approximation.

For which age groups would it not be fine?

Average years lived ${}_na_x$

Somewhat rough approximations:

- First age group:

$$ {}_1a_0 = 0.07 + 1.7\,{}_1M_0 $$

- Second age group:

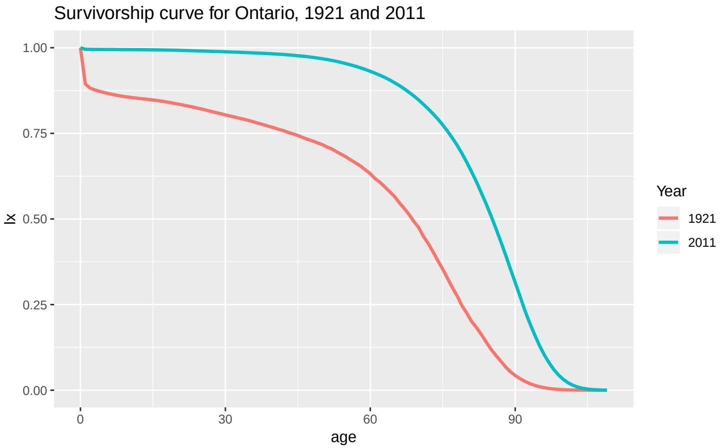
$$ {}_4a_1 = 1.5 $$

- Last age group:

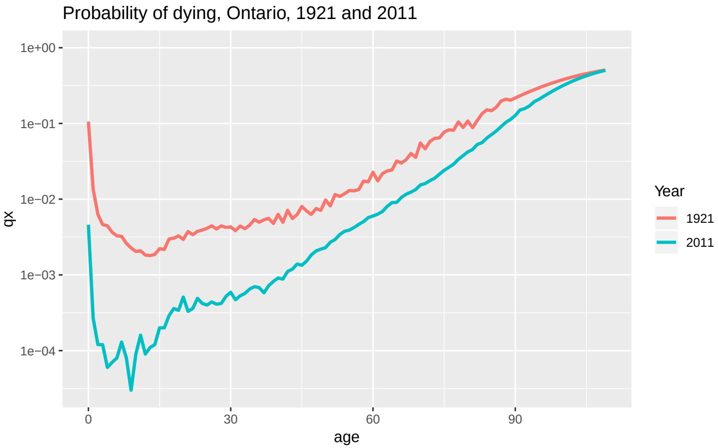
$$ {}_\infty a_\omega = 1/{}_\infty M_\omega $$

where ${}_\infty M_\omega$ is the age-specific mortality rate for the last age interval (check units).

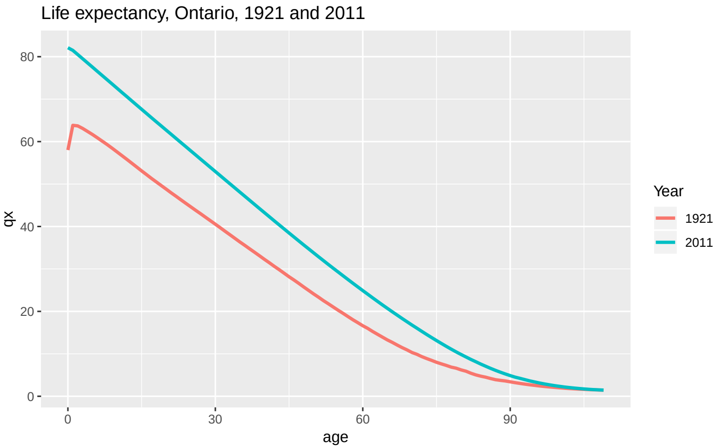
Characteristic shapes



Characteristic shapes



Characteristic shapes



Multiple decrements and cause-deleted life tables

Single versus multi-decrements

- So far we have only considered decrements over age due to all deaths
- But people die of different things, and age patterns of causes of death are important
- Can extend the single-decrement life table into a series of **multiple decrement life tables**

Not just mortality! Life table approach could be used to marriage (divorce, widowhood), contraception use, etc.

Multiple decrements

Say we observe mortality rates by cause.

${}_nM_x^i$ is the mortality due to cause i.

${}_nM_x^{-i}$ is the mortality due to all causes apart from i.

Note the conversion formula, thus the dependence of probabilities of dying:

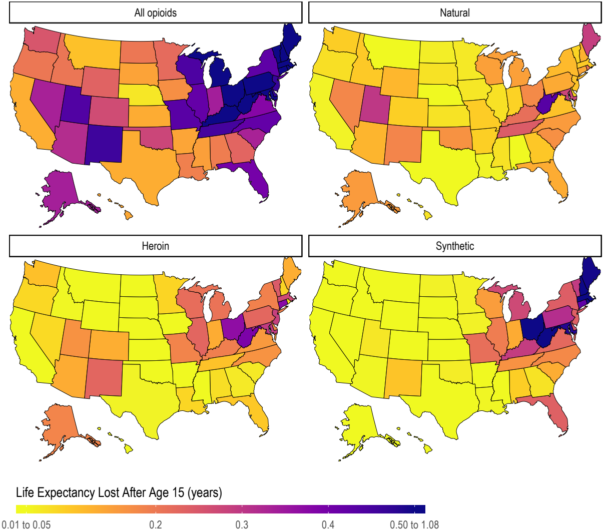
$${}_nq_x^i = \frac{n \cdot {}_nM_x^i}{1 + (n - {}_na_x) \cdot {}_nM_x}$$

Also note we don't have to use this:

$${}_nq_x^i = {}_nq_x \cdot \frac{{}_nMx^i}{{}_nMx} = {}_nq_x \cdot \frac{{}_nDx^i}{{}_nDx}$$

Use this technique to create **cause-deleted life tables**, looking at the hypothetical situation where we omit all deaths due to one cause but everything else stays the same.

Life expectancy lost due to opioids



More here: https://sanjaybasu.shinyapps.io/opioid_geographic/

Lifespan disparity

- Life expectancy, the most common mortality measure, gives us a measure of survival on average
- Also important is the variation in lifespan, to study survival disparities within populations
- E.g. can look at standard deviation in age at death:

$$\sqrt{\sum_{0}^{\omega} \frac{(x - e_0)^2 \, d_x}{\sum_0^{\omega} d_x}} = \sqrt{\sum_{0}^{\omega} \frac{(x - e_0)^2 \, d_x}{l_0}}$$

- In general, as life expectancy goes up, disparity goes down
- Potentially useful to compare populations/time points
  - e.g. for a given level of life expectancy in two populations, what's the disparity?

Mortality models

Continuous mortality

The continuous version of mortality rates are called **hazard rates** $h(x)$ (or $\lambda(x)$) or, if you're vintage demography, the **force of mortality** $\mu(x)$.

Let's take the limit of ${}_nM_x$ as $n \rightarrow 0$

$$
\begin{aligned}
h(x) &= \lim_{n \rightarrow 0} \frac{{}_nd_x}{{}_nLx} \\
&= \lim_{n \rightarrow 0} \frac{l_x - l_{x+n}}{n \cdot l_x} \\
&= -\frac{dln(l_x)}{dx}
\end{aligned}
$$

Hazard rates

$$h(x) = -\frac{dln(l_x)}{dx}$$

Implies

$$l(x+n) = l(x)e^{-\int_x^{x+n} h(x)dx}$$

Note the similarities between this and exponential growth from last week.

If h is constant in the interval then $l(x+n) = l(x)e^{-hn}$.

Continuous mortality

Given instantaneous mortality, have the continuous version of ${}_nd_x$, which is called the density of deaths, $d(x)$. Importantly,

$$d(x) = h(x)l(x)$$

Note that the total number of deaths has to equal $l(0)$, so if the radix is one then

$$\int d(x) = 1$$

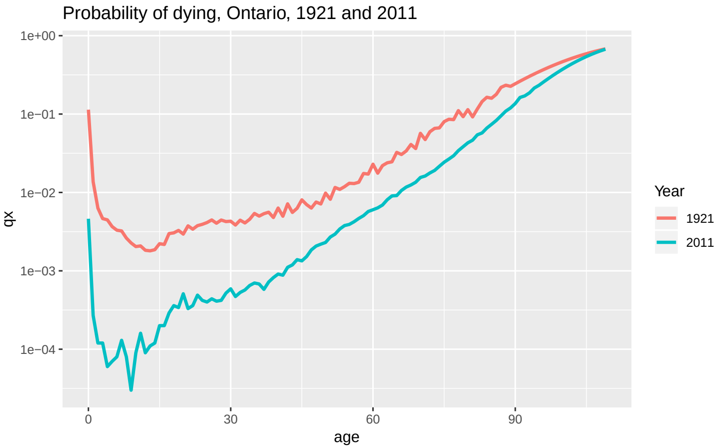
so $d(x)$ is a pdf.

Can also calculate ${}_nd_x$ in continuous form, i.e.

$${}_nd_x = \int_x^{x+n} l(a)h(a)da$$

Modeling mortality rates over age

Modeling mortality rates over age



Gompertz model

Gompertz (1825). Hazards are log linear:

$$h(x) = \alpha e^{\beta x}$$

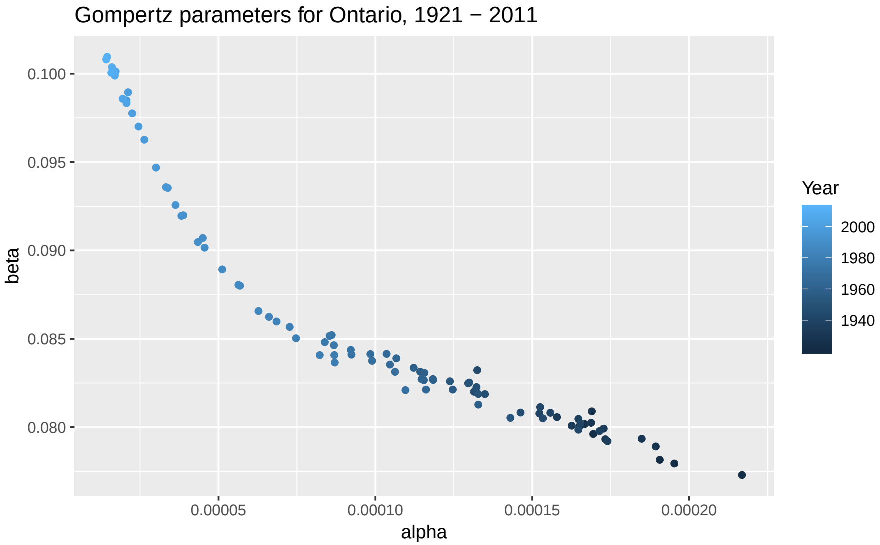
or

$$\log h(x) = \log(\alpha) + \beta x$$

So we can take the log of mortality rates over age and fit a simple regression to get estimates of α and β.

Only fit to adult ages!

Gompertz parameters



Gompertz distribution

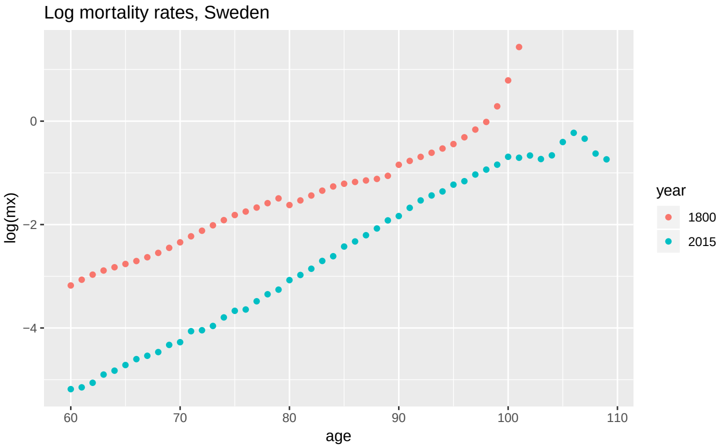
Note that because we know the form of Gompertz hazards $h(x)$, we can get closed form expressions for survivorship l_x etc. For example, the Gompertz death distribution is

$$d(x) = \alpha \exp\left(\beta x - \frac{\alpha}{\beta}\left(e^{\beta x - 1}\right)\right)$$

This is a PDF because $\int d(x) = 1$ i.e. everyone dies.

Other parametric models

Gompertz is unrealistic at young ages, but may also be misleading at old ages



Other parametric models

e.g., the log quadratic model

$$h(x) = e^{\alpha + \beta x + \gamma x^2}$$

allows for deceleration at older ages.

There's a good discussion of parametric mortality models in Feehan (2018).

The ongoing debate

Is there a human mortality plateau at older ages?

The plateau of human mortality: Demography of longevity pioneers

Elisabetta Barbi[1,*], Francesco Lagona[2], Marco Marsili[3], James W. Vaupel[4,5,6,7], Kenneth W. Wachter[8]
+ See all authors and affiliations

Science 29 Jun 2018:
Vol. 360, Issue 6396, pp. 1459-1461
DOI: 10.1126/science.aat3119

All they do is fit a Gompertz model, with proportional effects for cohort and gender

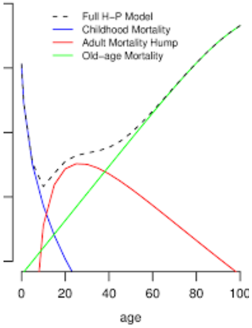
$$h(x) = \alpha e^{\beta x} e^{\beta_1 C + \beta_2 M}$$

Parametric models get tricky across all ages

Most well-known parametric model for the whole mortality curve is Heligman-Pollard (1980):

$$ {}_1q_x = A^{(x+B)^C} + De^{-E(ln(x)-ln(F))^2} + GH^x $$

Eight parameters! Hard to fit, even with good data.



Relational models

- Mortality across age is very non-linear
- But the general 'shape' is quite regular across populations
- Use information from one population as the basis for a model for another population.
- Add in parameters to shift and twist shape of mortality curve

Brass relational logit

Define a mortality standard as Y_x. Then the model is

$$\text{logit } lx = \alpha + \beta Y_x$$

- α is a level parameter, β shifts the balance between young and old age mortality.
- Why is this set-up useful? Can get sensible mortality curves for populations where we have limited data.
- DIY standard: choose your favorite l_x, take the logit.

Lee-Carter mortality forecasting model

- forecasting mortality rates is important (insurance, social security)
- life expectancy is a non-linear function of age-specific mortality rates; trends are driven by trends at different ages
- want to take into consideration age patterns but reduce the dimensionality of the model

Lee-Carter model

$$\log m_{x,t} = a_x + b_x \cdot k_t + \varepsilon_{x,t}$$

- $m_{x,t}$ are age-specific mortality rates for time t
- a_x is some baseline mortality schedule
- b_x is the contribution of age group x to mortality change over time
- k_t is a time index, telling us how much mortality is changing

So if you have a_x and b_x, then you can forecast k_t to obtain forecasts of mortality.

Obtaining values of the parameters

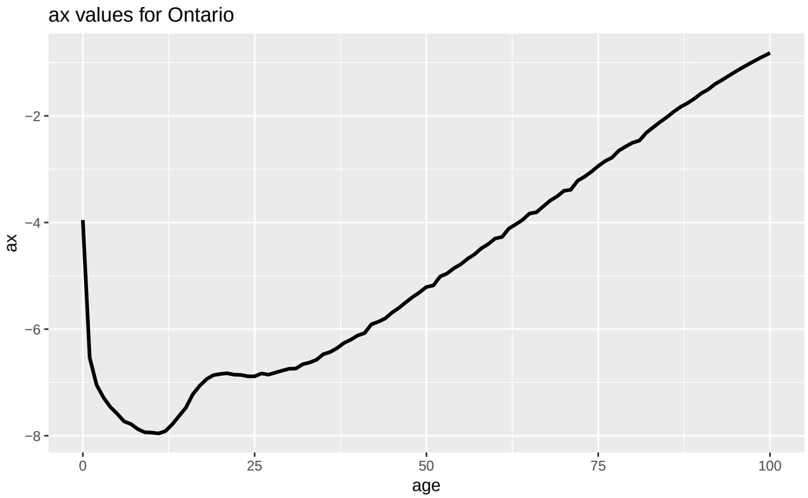
Do Singular Value Decomposition (SVD) on matrix of demeaned, logged age-specific rates over time. Call this matrix X. Then the SVD is

$$X = UDV'$$

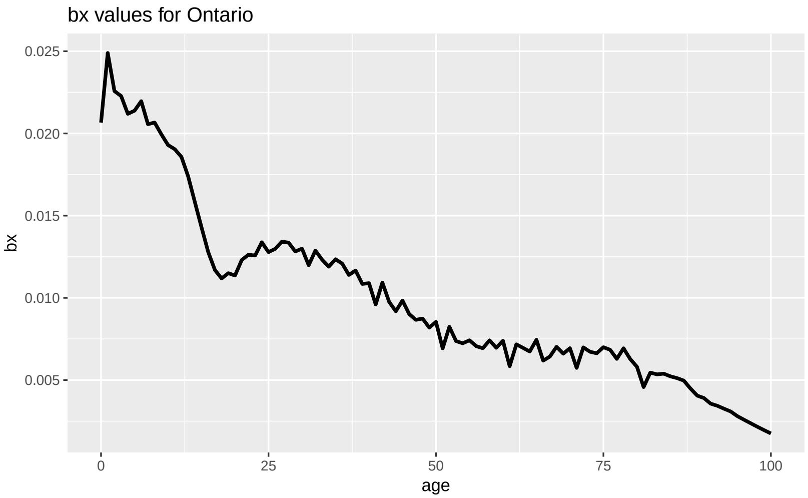
- a_x is just the mean age schedule
- b_x is the first right singular vector, v_1, normalized to sum to 1.
- k_t is the first left single vector, multiplied by the first singular value $u_1 \cdot d_{1,1}$

Once you have these, you can forecast k_t using ARIMA etc

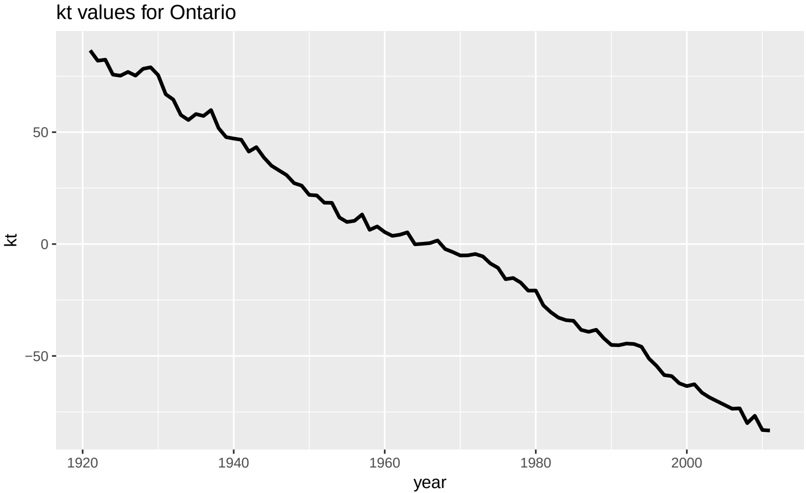
Lee-Carter for Ontario



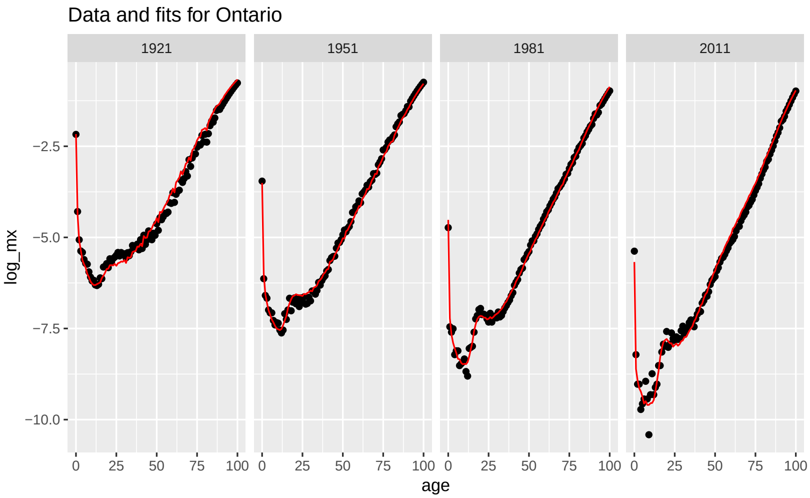
Lee-Carter for Ontario



Lee-Carter for Ontario



Lee-Carter for Ontario



Mortality as a more general regression framework

- So far, have considered mortality rates or probabilities with no exposure adjustment
- But often want to incorporate information about the size of the population (especially when estimating in smaller populations)
- Can consider observed deaths as an outcome of a Poisson process (or Binomial if dealing with smaller populations)

e.g. We could use Brass in a hierarchical model!

Consider deaths / mortality by population group g

$$
\begin{aligned}
D_{xgt} &\sim \text{Poisson}(P_{xgt} \cdot h_{xgt}) \\
l_{xgt} &= e^{-\sum h_{xgt}} \\
\text{logit} ł_{xgt} &= \alpha_{gt} + \beta_{gt} Y_x \\
\alpha_{gt} &\sim N(\mu_{sex[g]} + \mu_{race[g]}, \sigma_\alpha^2) \\
\beta_{gt} &\sim N(\beta_{g,t-1}, \sigma_\beta^2)
\end{aligned}
$$

Dataset ideas

Data sources

- UN WPP
- IPUMS
  - ACS
  - Censuses
  - DHS
- Mortality
  - StatCan
  - HMD
  - Canada HMD
  - US HMD
  - CDC Wonder
  - Linked births and deaths
  - Multiple cause of death data
  - Opioids / suicides / drugs ...