|  |
| --- |
| QMSS2 Immigration and Population DynamicsSummer School on: PROJECTION METHODS FOR ETHNICITY AND IMMIGRATION STATUSSchool of Geography, University of Leeds, UK , 2 – 9 July 2009 |

Practical life table construction in R Pia Wohland

Friday 3rd of July 2009, 0900-1030,

**The first aim of this practical is to generate 1998 UK period life tables for men and women**

Use yesterdays lecture notes to calculate: mortality rates, mortality probabilities, survival probabilities, number surviving, number dying, life years lived, total life years lived, life expectancy and survivorship and non survivorship probabilities.

**Then construct cohort life tables**

Project mortality rates into the future for women and compute cohort life expectancy for those born in 2008. To do this you need to assume a rate of decline in mortality.

For 1998 to 2008, use a 2% per annum decline.

From 2008 to 2108 make one of the following assumptions, based on your knowledge of mortality trends and philosophical position:

*Things will get even better:* 2.5% decline (male improvement over past 25 years)

*Business as usual:* 2% decline (average of male and female improvement over past 25 years)

*Slowdown in mortality improvement:* 1% (average improvement over the past 100 years)

*The obesity epidemic will stop improvement:* 0% decline (constant mortality)

*The obesity epidemic will reduce mortality:* -1% decline.

When you have extrapolated mortality for 110 years, extract the experience of the cohort born this year, in 2008. This is a set of cells along a diagonal, starting from age 0 in 2008 to age 100+ in 2108.

**Finally answer these questions:**

What are the 1998 life expectancies for ages 0, 60, 65, 70, 75, 80, 85 for men and women

In 1998 what was the median life expectancy for men and women

What are the life expectancies for ages 0, 65, 75, 85 for men and women for the 2008 cohort life table

What is your predicted median life expectancy for men and women for 2008 cohort life table

**Data you need**

The file PD\_UK1998.csv file contains deaths and population data by single year of age from birth to the age of 100+ for men and women, UK 1998. **If you are familiar with R, just get going.**

The Column names in the file denote for:

"age" -age (single year of age 0-100+)

"men.Dx" - men, number of deaths at age

 "men.Px" - women, population at age

"wom.Dx" -women, number of deaths at age

"wom.Px" -women, population at age

**#If you are New to R, this practical will talk you through some of the basics**

**#by constructing the required life table step by step and Visualising your results**

##The # is used in R to add comments to your script, lines starting with # are ignored by R if you ##run your script. R commands/functions are highlighted in **red.** There are explanations for most of ##them in this script. For further information or to understand the once that are used without ##explanation, use the **help()** functions. When you type the command, do not type the > as well. ##It is the command prompt sign.

##Save the file (PD\_UK1998.csv) in your working directory and start R from the desktop icon.

##To see in which directory R is currently working in use

**>getwd()**

##if you want to change your working directory, you can do so with

**>setwd()**

##for example:

>**setwd(**"C:\\Documents and Settings\\geopnwo\\Desktop\\SummerSchoolPracticals")

##or

**>setwd(**"C:/Documents and Settings\\geopnwo/Desktop/SummerSchoolPracticals")

##Check it worked

> **getwd()**

##In my case this will show:

[1] "C:/Documents and Settings/geopnwo/Desktop/SummerSchoolPracticals"

##Explore your directory:

**list.files() ## shows the files in the directory**

**ls() ## lists the R object in your folder**

##(There should not be any R objects in the directory at that point in time)

##To work in R you need your data accessible in R. In R you work R objects to. You can create ##them by typing in your data but a probably faster way is, import them.

##Create objects in R

##To create an object in R, you use <- or alternatively =

> x<-1 ## creates an object of the name x with the value 1

> x ## type x to see the value of x

[1] 1 ##not surprising, it is one

##But be careful, if you now assign x the value of 2, R all forgets about the 1.

> x<-2

> x

[1] 2

##This might sound very trivial right now, but it is well worth keeping in mind!

## Creating an R object by reading in a csv file produces a data frame (

##The command for reading in a csv file is

**##read.csv()** ##Note: R has more options to read in data, from variety of file formats

PD\_UK1998 **<- read.csv**("PD\_UK1998.csv")

##Explore the data object



#Try: **fix()** to look at the spreadsheet. If you wanted to change ##single values in the cells, you could do that here. Beware, you ##need to close the window before any other command will be ##executed

**fix(**PD\_UK1998**)**

**##names()** to see the header names of each column.

**names(**PD\_UK1998**)**

"age" -age (single year of age 0-100+

"men.Dx" - men, number of deaths at age

 "men.Px" - women, population at age

"wom.Dx" -women, number of deaths at age

"wom.Px" -women, population at age

##Some examples for extracting data using [ ] or $

##If you use square brackets to extract data from a matrix or a dataframe: [rows, columns]

##look at the first five rows of the data

> PD\_UK1998 [1:5,]

 age men.Dx men.Px wom.Dx wom.Px

1 0 2327 364800 1752 347900

2 1 190 375100 150 357100

3 2 117 369000 100 349800

4 3 86 373600 61 354800

5 4 70 384000 36 365300

##look at the second column

> PD\_UK1998 [,2]

##or

> PD\_UK1998$men.Dx

##will both show the same data. You also see the single elements are numbered. [13] means the ##second row starts with element 13

 [1] 2327 190 117 86 70 67 53 63 52 48 57 65

 [13] 60 75 86 113 163 178 290 314 276 271 317 337

 [25] 338 367 406 436 448 463 468 462 559 546 529 564

 [37] 582 600 588 633 653 736 748 867 931 1011 1040 1204

 [49] 1308 1433 1655 1978 1898 2015 2283 2394 2383 2455 2962 3152

 [61] 3429 3794 4071 4399 4739 5127 5731 6311 6693 7143 7835 8326

 [73] 9074 9411 10067 10305 11227 11910 12293 8946 8360 9091 9627 10070

 [85] 9639 9000 8098 7263 6409 5677 4781 3793 3025 2289 1728 1245

 [97] 832 562 370 215 319

##So

PD\_UK1998$men.Dx [1]

##Will give the first element of the column

[1] 2327

## And you will get the same result with:

> PD\_UK1998 [1,2]

[1] 2327

> PD\_UK1998$men.Dx[-1] ## shows the same column without the first element.

 [1] 190 117 86 70 67 53 63 52 48 57 65 60

 [13] 75 86 113 163 178 290 314 276 271 317 337 338

 [25] 367 406 436 448 463 468 462 559 546 529 564 582

 [37] 600 588 633 653 736 748 867 931 1011 1040 1204 1308

 [49] 1433 1655 1978 1898 2015 2283 2394 2383 2455 2962 3152 3429

 [61] 3794 4071 4399 4739 5127 5731 6311 6693 7143 7835 8326 9074

 [73] 9411 10067 10305 11227 11910 12293 8946 8360 9091 9627 10070 9639

 [85] 9000 8098 7263 6409 5677 4781 3793 3025 2289 1728 1245 832

 [97] 562 370 215 319

#########################################################

**#We now start to calculate the period life table for men, UK 1998#**

#########################################################

##From this point onwards, best write your code in the R editor, because you might want to reuse it ##later on. To open the editor go to File and then New Script.

##Have a look at the file containing your data again.

**names(**PD\_UK1998**)**

# The first step to calculate a period life table is to calculate the mortality (mx) rate in the given year from # the population and deaths data.
#This is easily calculated as:

mx <- PD\_UK1998$ men.Dx / PD\_UK1998$ men.Px

# If the two vectors are the same size, R will divide one by the other element wise.

mx ## to look at what you

##If you use PD\_UK1998 for various calculations, you can avoid using the $ sign to refer to columns, by “attaching” the object with the function, **attach()**

**attach**(PD\_UK1998)

then you can do the same calculation by just typing:

mx <- men.Dx /men.Px

rad<-100000 ##radix (define the radix for your l, to easily vary it

ax <- c(0.3,rep(0.5,99),1/mx[101]) ##fraction lived.

 ##This is a good example to show,

 ##how function can be combined in R

 ##the **rep()** function repeats a value a given time

` ##**rep** (value, times)

 ##The **c()** function c for concatenate

 ##In our case, we build a vector of length 101,

 ##the first element is 0.3, the next 99 elements

 ##are 0.5 and the final is 1 divided by the last element

 ##of the mortality rate.

qx <- mx / (1+(1-ax)\*mx) ##Probability of dying (qx) between the exact ages x and ##x+1

qx[101] <- 1 ##The last element of qx is 1.

px<-1-qx ##probability of surviving (px) between the exact ages x and x+1

lx <- c(1\*rad,(**cumprod**(1-qx[-101]))\*rad) ##numbers surviving (lx).

dx <- lx\*qx ## numbers dying (dx)

Lx <- lx - (1-ax)\*dx ##number alive between ages

Lx[101] <- ax[101]\*dx[101] ##Again, the last element is treated differently and you ##replace it with the appropriate formula

Tx <- **rev**(**cumsum**(**rev**(Lx))) ##look up in help what cumsum() and rev() are doing.

ex <- Tx/lx ##life expectancy (ex)

sx1<-Lx[1]/lx[1], ##survivorship probability (sx)

sx<-Lx[2:100]/Lx[1:99]

sxlast<-Lx[101]/(Lx[101]+Lx[100]))

sx<-c(sx1, sx , sxlast)

## or all in one sx<-c(Lx[1]/lx[1],Lx[2:100]/Lx[1:99],Lx[101]/(Lx[101]+Lx[100]))##

ux<-1-sx ##non survivorship probability (ux)

##Now all data are calculated, to create a datframe which contains all:

age<-0:100

ltable\_m <-**cbind**(age,**round**(ax,2), **round**(**cbind**(mx,qx,px),5), **round**(**cbind**(lx,dx,Lx,Tx),0), (round(ex,2)), round(cbind(sx,ux),5))

**colnames**(ltable\_m) <- c("age","ax","mx","qx","px","lx","dx","Lx","Tx","ex","sx","ux")

ltable\_m<-**data.frame**(ltable\_m)

##Well done! The first life table is constructed

##Now, generate the life table for UK women, 1998.

##In the following, I will assume the resulting life table for women is called ltable\_w

##Once finished, reveres attach by detaching the object

**detach**(PD\_UK1998)

**##Now explore your data with some graphics**

##if you want to plot age versus mortality rate, simply type

##plot men mortality rate versus age

**plot**(ltable\_m$age,ltable\_m$mx)

## the same, but your own labels

**plot**(ltable\_m$age,ltable\_m$mx, xlab="Age",ylab="Mortality rate" )

##the same, but the logarithmic mortality rate. Another example how you easily combine ##functions in R

**plot**(ltable\_m$age,(**log**(ltable\_m$mx,base = exp(1))), xlab="Age",ylab="Mortality rate")

##plot men cohort survivors, a different style of graph

**plot**(ltable\_m$age,ltable\_m$lx/1000,type = "h", col="grey", xlab="Age",ylab="Cohort survivors in thousend")

##plot life expectancy for men (blue) and women (red). Note you use data from two different ##dataframes and combine the in one graph.

**plot**(ltable\_m$age,ltable\_m$ex, xlab="Age",ylab="Life expectancy",col="blue", ylim=c(0,80))

**points**(ltable\_f$age,ltable\_f$ex, xlab="" , ylab="", col="red", ylim=c(0,80))

**Now construct the cohort life tables**

##Before constructing the cohort life table, have a look at this short example for a **for** loop in R

##Fibonacci numbers: a sequence of numbers, in which the first two numbers are 0 and 1.

##Each remaining number is the sum of the two previous ones.

##The first 12 numbers:

Fibonacci <- numeric(12) ##create an empty vector of length 12

Fibonacci[1]<-0 ##define the first two elements

Fibonacci[2]<-1 ##

**for** (i in 3:12) Fibonacci[i]<- Fibonacci[i-1]+Fibonacci[i-2] ##calculates the remaining elements

#Now construct the cohort life table for men

##Here you predict future

##

dec1<-2 ##defines the percentage decline

dec2<-2.5 ##

cohortm<-**matrix**(nrow=101,ncol=111) ##creates an empty matrix with

cohortm[,1]<-ltable\_m$mx ##the

**colnames**(cohortm)<-**seq**(1998,2108, by=1)

for (i in 2:10) cohortm[,i]<- cohortm[,i-1]\*(1-dec1/100)

for (i in 11:111) cohortm[,i]<- cohortm[,i-1]\*(1-dec2/100)

cohort<-cohortm[,11:111] ##creates a matrix containing data ##from 2008-2108

cohortm<-**diag**(cohort) ##gives you the diagonal of the matrix

##Now do the same for the women, change your decline rates if you want to.

##Once done that, create your 2008 cohort life tables for both, men and women.

##Now you have all the data you need to answer the questions from page one!

################################################################################

##Well done! Play tic tac toe with your neighbour to relax now

##adapted from (http://www.stats.uwo.ca/faculty/murdoch/2864/tictactoe.R)

plot(c(0.5,3.5), c(0.5, 3.5), type="n", axes=FALSE)

segments(x0 = c(1.5, 2.5, 0.5, 0.5), y0=c(0.5, 0.5, 1.5, 2.5),

 x1 = c(1.5, 2.5, 3.5, 3.5), y1=c(3.5, 3.5, 1.5, 2.5))

## place X or O like this

text(1, 1, "X", cex=3)