DemTech2_Problem Set 1

Sara Peters

2023-09-23

Problem Set 1

Note to Jenna- I use ChatGPT to learn how to write code that I'm unfamiliar with given that some weeks I often have to figure out the assignment in small chunks of time because of life commitments. I'm not sure how to cite this in the homework but I'm happy to print out the chat history for this assignment so you can see the iterative process I use to figure out how to write code. I still don't think I've got the code right for all the values but I ran out of time to follow up with someone to figure it out so I would appreciate feedback.

Use R to program the answers to question 1 and graph the lifetable functions. Submit your program stapled to your answers. 1. Table 1 (next page) contains deaths by age for French males in 1985. These data also include mid-year population estimates and a set of nax values for French males for 1985. Table 1 is on our webpage in .csv format. a. Use these data to construct a life table for the male population. Do this by performing operations on the vectors. You will need to calculate and fill in the following life table columns: nqx; lx; ndx; nLx; nmx; Tx; and ex. b. Graph the following life table functions using either plot() or ggplot(): lx; ndx; and nmx . What do you observe? c. What was life expectancy at age 40? How would you interpret this number? d. What was the probability of surviving from birth to age 30? e. What was the probability of surviving to age 65 for those who survived to age 30? f. What was the probability that a newborn would die between 50 and 55? g. How many years could a newborn expect to live in the interval 15-65? h. If you only had the fourth column of Table 1, would you be able to distinguish this population as one with high mortality or low mortality? (What nax value in particular might help distinguish between the two?) i. If the French population were stationary, what would be the crude death rate? j. Extra credit part 1: push your code to your Github page and list the URL in your submitted answers. k. Extra credit part 2: install the Lifetables package in R. With nmx in hand, use lt.mx() to populate the other functions. Check your work in 1(a), noting discrepancies if you set nax=NULL.

###Load Data

```
#Set working directory
setwd("C:/Users/saraa/OneDrive - UW-Madison/SOC 756- Demography Techniques II/Problem Sets/Probl
em Set 1")

#load libraries
# install.packages(dplyr)
# install.packages(tidyverse)
# install.packages(ggplot2)

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ forcats   1.0.0     ✔ readr     2.1.4
## ✔ ggplot2   3.4.2     ✔ stringr   1.5.0
## ✔ lubridate 1.9.2     ✔ tibble    3.2.1
## ✔ purrr     1.0.1     ✔ tidyr     1.3.0
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
library(ggplot2)

#Load data
base_data<- read.csv("ps1_data_F2023.csv")


#transform base data into numeric values
# Define the column(s)  to process
columns_to_process <- c("nNx", "nDx", "nax")

# Remove commas and convert to numeric for selected columns
for (col_name in columns_to_process) {
  base_data[, col_name] <- as.numeric(gsub(",", "", base_data[, col_name]))
}

str(base_data)
```

```
## 'data.frame':    19 obs. of  4 variables:
##  $ x  : int  0 1 5 10 15 20 25 30 35 40 ...
##  $ nNx: num  379985 1559722 1896295 2160190 2179837 ...
##  $ nDx: num  3741 770 532 673 2138 ...
##  $ nax: num  0.087 1.5 2.5 2.966 2.769 ...
```

```r
# Set the 'scipen' option to a large value to prevent scientific notation
options(scipen = 999)

# Set the 'digits' option to control the number of decimal places
options(digits = 6)  # Change the number to the desired decimal places

#nNx= mid-year population estimates
#nDx = age-specific mortality rates
#nax = average person years lived in the interval by this dying in the interval
```

##Construct life table

```
life_table<-base_data


#####create vector of age interval lengths (n)
life_table$n <- numeric(length(life_table$nNx))


#calculate rows
rows<-nrow(life_table)- 2

#create n vector
n_values<- c(1, 4, rep(5, rows))

#add values to life table
life_table$n<-n_values

######Construct nmx - death rate of cohort between ages x and x + n
# Initialize an empty vector to store nmx values
life_table$nmx <- numeric(length(life_table$nNx))

# Calculate nMx values using a for loop
for (i in 1:nrow(life_table)) {
  life_table$nmx[i] <- life_table$nDx[i] / life_table$nNx[i]
}


####Construct nqx- probability of dying x to x + n
# Initialize an empty vector to store nqx values
life_table$nqx <- numeric(length(life_table$nNx))

# Calculate nqx values using a for loop
for (i in 1:(nrow(life_table) - 1)) {  # Exclude the last row
  n <- life_table$n[i]
  nmx <- life_table$nmx[i]
  nax <- life_table$nax[i]

  life_table$nqx[i] <- (n * nmx) / (1 + (n - nax) * nmx)
}

# Set the last value of nqx to 1
life_table$nqx[nrow(life_table)] <- 1

##### construct lx- n of survivors exact age x
# Initialize an empty vector to store lx values
life_table$lx <- numeric(length(life_table$nNx))

# Set the initial population (l0)
l0 <- 100000
life_table$lx[1] <- l0

# Calculate lx values using a for loop
for (i in 2:nrow(life_table)) {
```

```
  nqx <- life_table$nqx[i-1]

  life_table$lx[i] <- life_table$lx[i - 1] * (1 - nqx)
}


#####construct ndx- number of deaths x to x + 1
# Initialize an empty vector to store ndx values
life_table$ndx <- numeric(nrow(life_table))


# Calculate ndx values using a for loop
for (i in 1:nrow(life_table)) {
  lx <- life_table$lx[i]
  nqx <- life_table$nqx[i]

  life_table$ndx[i] <- lx * nqx
}

####Construct nLx- number of person-years lived x to x+ n
# Initialize an empty vector to store nLx values
life_table$nLx <- numeric(nrow(life_table))

# Calculate nLx values using a for loop
# Calculate nLx values using a for loop
for (i in 1:nrow(life_table)) {
  n <- life_table$n[i]
  lx <- life_table$lx[i]
  nax <- life_table$nax[i]
  ndx <- life_table$ndx[i]

  life_table$nLx[i] <- (n * lx) + (nax * ndx)
}

# Calculate the last value of nLx as lx/nmx
life_table$nLx[nrow(life_table)] <- life_table$lx[nrow(life_table)] / life_table$nmx[nrow(life_t
able)]


#####construct Tx- person-years lived above age x

#Initialize an empty vector to store Tx values

# Initialize an empty vector for Tx
life_table$Tx <- numeric(length(life_table$nLx))

# Calculate the cumulative sum using a for loop
for (i in 1:length(life_table$nLx)) {
  if (i == 1) {
    # For the first value, Tx = sum of all values in nLx
    life_table$Tx[i] <- sum(life_table$nLx)
  } else {
```

```
    # For other values, Tx = sum of all values in nLx except values above the current row in Tx
    life_table$Tx[i] <- sum(life_table$nLx[-(1:i)])
  }
}


# Set the last value of Tx to the last value of nLx
life_table$Tx[nrow(life_table)] <- life_table$nLx[nrow(life_table)]


#####Construct ex- expectation of life at age x
life_table$ex <- numeric (length(life_table$Tx))


#calculate ex using a for loop
for (i in 1:nrow(life_table)) {
  life_table$ex[i] <- life_table$Tx[i] / life_table$lx[i]
}
```

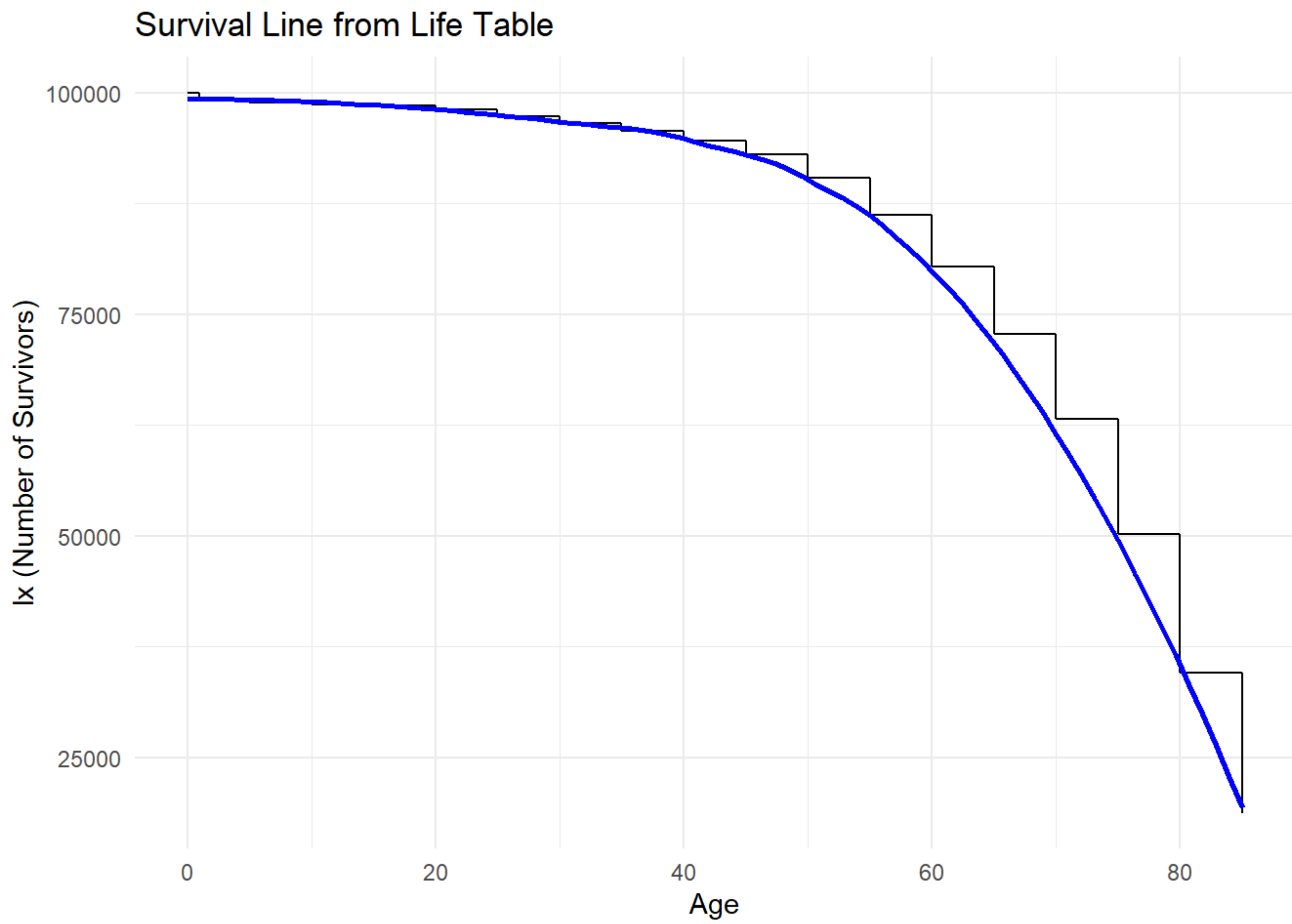
Graph the following life table functions using either plot() or ggplot(): lx; ndx; and nmx .

What do you observe?

```
#Plot lx
# Create the survival line graph
ggplot(life_table, aes(x = x, y = lx, group = 1)) +
  geom_step(direction = "hv") +
  geom_smooth(method = "loess", se = FALSE, color = "blue", size = 1) +
  labs(
    title = "Survival Line from Life Table",
    x = "Age",
    y = "lx (Number of Survivors)"
  ) +
  theme_minimal()
```

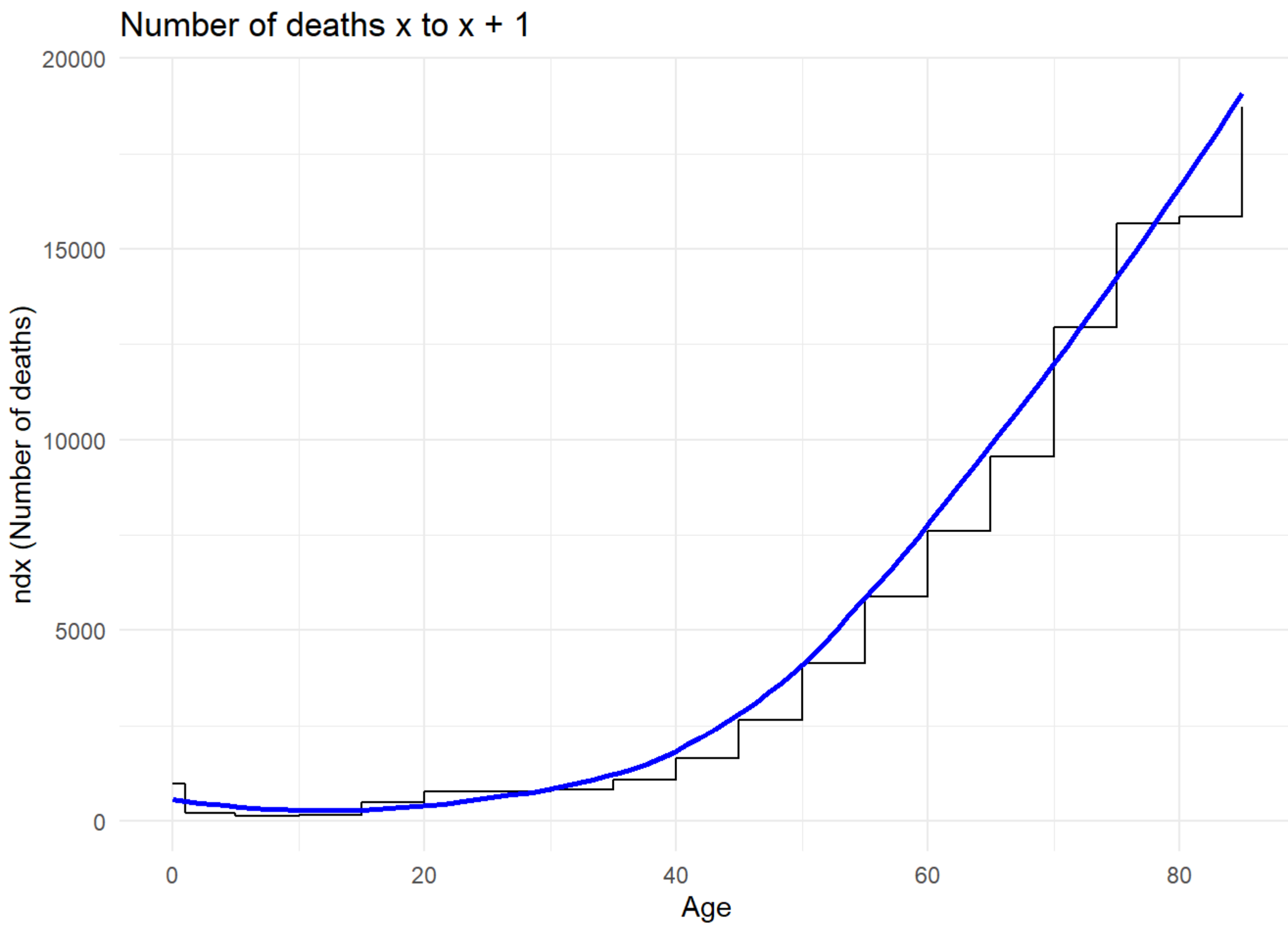
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
#Plot ndx
ggplot(life_table, aes(x = x, y = ndx, group = 1)) +
  geom_step(direction = "hv") +
  geom_smooth(method = "loess", se = FALSE, color = "blue", size = 1) +
  labs(
    title = "Number of deaths x to x + 1",
    x = "Age",
    y = "ndx (Number of deaths)"
  ) +
  theme_minimal()
```

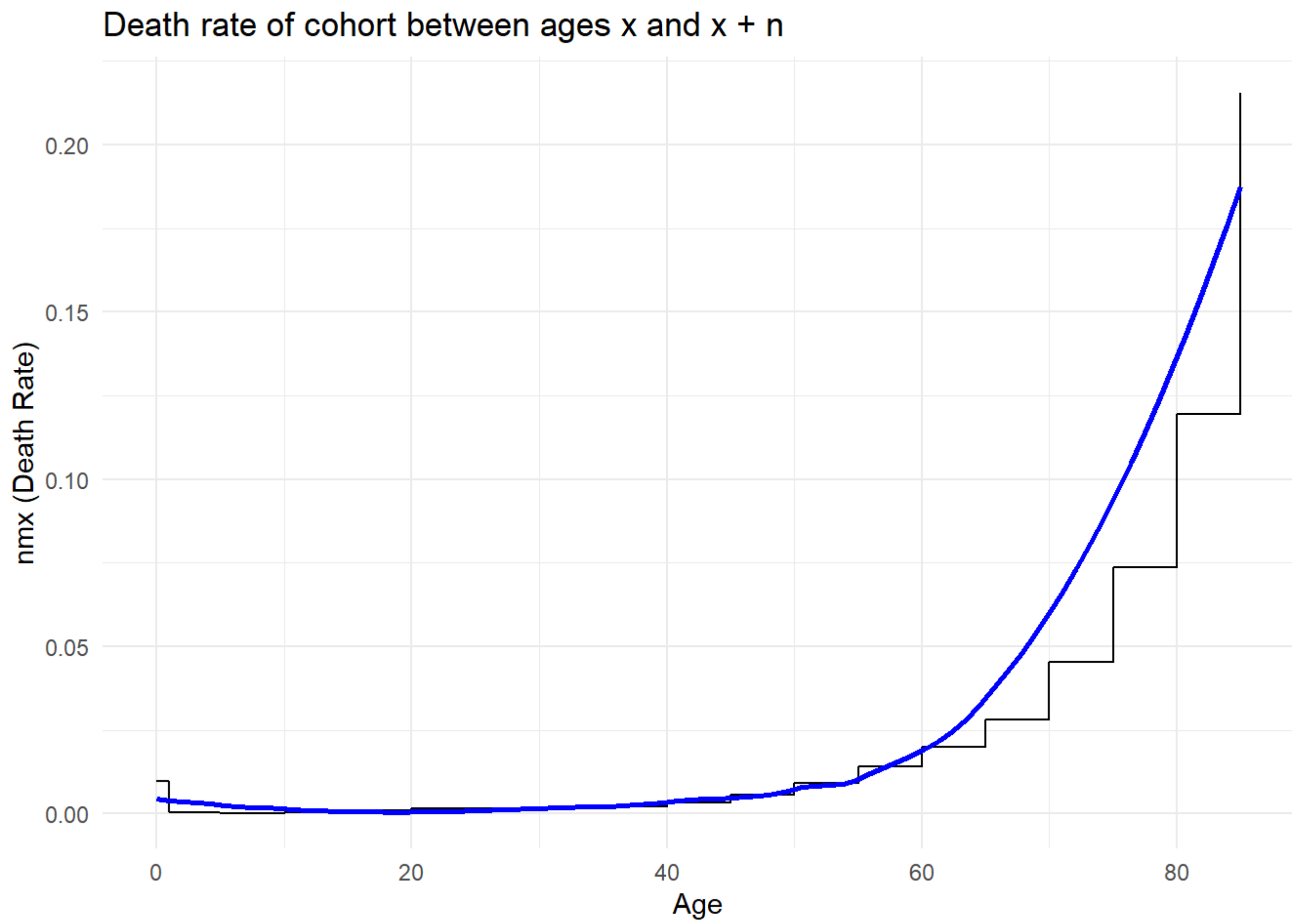
```
## `geom_smooth()` using formula = 'y ~ x'
```




```
#plot nmx

ggplot(life_table, aes(x = x, y = nmx, group = 1)) +
  geom_step(direction = "hv") +
  geom_smooth(method = "loess", se = FALSE, color = "blue", size = 1) +
  labs(
    title = "Death rate of cohort between ages x and x + n",
    x = "Age",
    y = "nmx (Death Rate)"
  ) +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



lx plot I observe a population with lower early life mortality and an increase in mortality at middle ages (around x=40.)

ndx plot Understandably the curve that represents the number of deaths in the population is the inverse of the survival curve. Low number of deaths exact ages 5 to exact ages 40. Higher number of deaths 0-4 and above age 40. In this curve the higher mortality rate at the very younge ages is more obvious to me.

nmx plot Unlike the survival curve (lx) and the curve plotting the number of deaths (ndx) the deathe rate for the cohort in this plot doesn't start an upward curve until closer to exact age 50 with a dramatic increase after 60. I think this is indicative of a cohort's experience of mortality rather than the mortality trends I see in the period life table- which are in the previous two plots.

```
print(life_table)
```

```
##     x     nNx   nDx   nax n         nmx        nqx       lx       ndx      nLx
## 1   0  379985  3741 0.087 1 0.009845125 0.00975742 100000.0   975.742 100084.9
## 2   1 1559722   770 1.500 4 0.000493678 0.00197228  99024.3   195.303 396390.0
## 3   5 1896295   532 2.500 5 0.000280547 0.00140175  98829.0   138.534 494491.1
## 4  10 2160190   673 2.966 5 0.000311547 0.00155675  98690.4   153.636 493907.8
## 5  15 2179837  2138 2.769 5 0.000980807 0.00489333  98536.8   482.173 494019.1
## 6  20 2159556  3432 2.574 5 0.001589216 0.00791556  98054.6   776.157 492270.9
## 7  25 2106750  3291 2.512 5 0.001562122 0.00778037  97278.5   756.862 488293.5
## 8  30 2147845  3657 2.586 5 0.001702637 0.00847834  96521.6   818.343 484724.2
## 9  35 2165387  4956 2.657 5 0.002288736 0.01138264  95703.3  1089.356 481410.7
## 10 40 1516952  5269 2.697 5 0.003473412 0.01722924  94613.9  1630.126 477465.9
## 11 45 1498630  8654 2.695 5 0.005774607 0.02849377  92983.8  2649.458 472059.1
## 12 50 1552746 14490 2.663 5 0.009331855 0.04566342  90334.3  4124.974 462656.4
## 13 55 1476770 20831 2.625 5 0.014105785 0.06824271  86209.3  5883.159 446490.0
## 14 60 1350479 26805 2.601 5 0.019848513 0.09473176  80326.2  7609.440 421423.0
## 15 65  722430 20233 2.615 5 0.028006866 0.13126622  72716.7  9545.251 388544.5
## 16 70  842589 38315 2.598 5 0.045472941 0.20497600  63171.5 12948.638 349498.0
## 17 75  636848 46903 2.538 5 0.073648657 0.31172109  50222.8 15655.521 290848.0
## 18 80  372059 44443 2.466 5 0.119451485 0.45848006  34567.3 15848.430 211918.9
## 19 85  175169 37759 4.639 5 0.215557547 1.00000000  18718.9 18718.897  86839.4
##           Tx       ex
## 1  7533335.3 75.33335
## 2  7036860.4 71.06199
## 3  6542369.3 66.19891
## 4  6048461.5 61.28722
## 5  5554442.5 56.36923
## 6  5062171.6 51.62604
## 7  4573878.1 47.01841
## 8  4089153.9 42.36517
## 9  3607743.2 37.69719
## 10 3130277.3 33.08475
## 11 2658218.1 28.58798
## 12 2195561.8 24.30485
## 13 1749071.8 20.28866
## 14 1327648.8 16.52822
## 15  939104.3 12.91455
## 16  589606.3  9.33342
## 17  298758.3  5.94865
## 18   86839.4  2.51218
## 19   86839.4  4.63913
```

##Observations:

Additional Questions c. What was life expectancy at age 40? How would you interpret this number?

According to my life table calculations the life expectancy of french males at exact age 40 is 73 (40+33). I would interpret this number as representative of the history that the birth cohort currently aged 40 has experienced. Since the period life expectancy of e0 is 75, those who are 40 years old in this period are expecting a higher mortality rate.

d. What was the probability of surviving from birth to age 30?

The chance of survival to a certain age from birth is calculated from lx+n/l0. So to calculate the survival from birth to age 30 would be l30/l0. My life table puts these values at 96522/100000 = 0.96522

e. What was the probability of surviving to age 65 for those who survived to age 30?

The chance of survival from age a to age b is calculated from lb/la. So to calculate survival from age 30 to age 65 would be l65/l30. My life table puts these values at 72717/96522= 0.75337

f. What was the probability that a newborn would die between 50 and 55?

The number expected to die between two exact ages would be la-lb and to find out how many from a different exact age would die during that interval you would divide la-lb by the number alive at exact age x (i.e. lx). In this case the formula is l50-l55/l0 which in my life table is 90334-86209/100000 =0.04125.

g. How many years could a newborn expect to live in the interval 15-65?

The average future lifetime of persons aged x = Tx/lx.

Number of persons aged l0= 100000 Years lived between 15 and 65 is T15-T65 = 5554443-939104=4615339 Interval size 65-15= 50 years lived those who reach age 65 is l65= 72716

= 4615339-50 *72716 = 979,539

h. If you only had the fourth column of Table 1, would you be able to distinguish this population as one with high mortality or low mortality? (What nax value in particular might help distinguish between the two?)

I think we could distinguish the motrality trends given just the nax value. The nax value for l0 in the information provided is .08, which is higher than a theoretical value of .06 that I would normally use for the first age interval in a life table which indicates to me that this population might be experiencing higher mortality.

##Extra credit part 2: Install the Lifetables package in R. With nmx in hand, use lt.mx() to populate the other functions. Check your work in 1(a), noting discrepancies if you set nax=NULL.

```
# install.packages("LifeTables")
library ("LifeTables")
```

```
## Loading required package: mclust
```

```
## Package 'mclust' version 6.0.0
## Type 'citation("mclust")' for citing this R package in publications.
```

```
##
## Attaching package: 'mclust'
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
life_tableEC<-mod.lt(0.087, child.mort=1, e0.target= 70, sex= "male")

LT_EC<-lt.mx(nmx, sex="male", age = c(0, 1, seq(5)), nax = NULL)
```

I tried this but I don't really understand the output

2. Think about the social phenomena / processes that most interest you. Might any of these processes be measured in the form of a lifetable? I think that the experience of poverty or food insecurity might be an interesting events to see in a life table and that it would be an appropriate use of these calculations.

If yes: a. What events would constitute "births" and "deaths"? For poverty "births" would be births like a traditional life table the hypothetical size of the population (the radix) and "deaths"would be the first experience of poverty. Similar to poverty, to measure food insecurity, birth would be "birth" and the first experience of food insecurity would be "death".

b. What could you learn from using a lifetable? By using a life table I could understand how poverty and food insecurity effect a population over the life course. For example, if poverty or food insecurity are higher at younger ages it would provide an indication of what mechanisms to study which might be impacting poverty or food insecurity for younger populations.

c. Where might you start looking for data to identify the size of the population at risk and the age-specific "death" rates or probabilities? There are various Census and USDA data sets that track poverty and food insecurity rates. Rather than using a predetermined measure of poverty I could also calculate this basaed on income. Data would likely need to be aggregated unless I had access to a restricted dataset.

d. What issues might limit how the information produced in your lifetable can be interpreted? I think a life table as calculated here would only allow me to look at the first experience of either poverty or food insecurity and I would not be able to track how people's experience of poverty or food insecurity change over time. Most individuals go through periods of experiencing poverty, some only once, some multiple times throughout their life and a number of experiences of poverty would have different implications than one experience of poverty. The same is true of food insecurity.

If no, describe the issues that would make the lifetable an inappropriate analytical tool for the social processes that you study. N/A