

SOC 756: Problem Set 1

Mateo Frumholtz

September 9, 2025

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 .

```
# nNx = midyear population
# nDx = deaths between ages x and x + n
# nmx ~ nDx/nNx

lt <- ps1 |>
  arrange(x) |>
  mutate(
    time_diff = lead(x) - x,
    nmx = nDx / nNx,
    nqx = (time_diff * nmx) / (1 + ((time_diff - nax) * nmx)),
    nqx = case_when(
      is.na(nqx) ~ 1,
      TRUE ~ nqx
    ),
    npx = 1 - nqx,
    lx = accumulate(npx, `*`, .init = 100000)[-1],
    lx = lag(lx, default = 100000),
    ndx1 = lx - lead(lx),
    # Other option for ndx
    # ndx1 = case_when(
    #   is.na(ndx1) ~ lx,
    #   TRUE ~ ndx1
    # ),
```

```

ndx = nqx * lx,
nLx = (time_diff * lead(lx)) + (nax * ndx),
nLx = case_when(
  is.na(nLx) ~ lx / nm,
  TRUE ~ nLx
),
Tx = rev(cumsum(rev(nLx))),
ex = Tx / lx
)

```

Life Table for French Males in 1985

x	nqx	lx	ndx	nLx	nm	Tx	ex
0	0.00976	100,000.0	975.7	99,109.1	0.00985	7,131,028.1	71.3
1	0.00197	99,024.3	195.3	395,608.8	0.00049	7,031,918.9	71.0
5	0.00140	98,829.0	138.5	493,798.4	0.00028	6,636,310.1	67.1
10	0.00156	98,690.4	153.6	493,139.6	0.00031	6,142,511.7	62.2
15	0.00489	98,536.8	482.2	491,608.2	0.00098	5,649,372.1	57.3
20	0.00792	98,054.6	776.2	488,390.1	0.00159	5,157,763.9	52.6
25	0.00778	97,278.5	756.9	484,509.2	0.00156	4,669,373.8	48.0
30	0.00848	96,521.6	818.3	480,632.5	0.00170	4,184,864.6	43.4
35	0.01138	95,703.3	1,089.4	475,963.9	0.00229	3,704,232.1	38.7
40	0.01723	94,613.9	1,630.1	469,315.3	0.00347	3,228,268.2	34.1
45	0.02849	92,983.8	2,649.5	458,811.8	0.00577	2,758,952.9	29.7
50	0.04566	90,334.3	4,125.0	442,031.5	0.00933	2,300,141.1	25.5
55	0.06824	86,209.3	5,883.2	417,074.2	0.01411	1,858,109.6	21.6
60	0.09473	80,326.2	7,609.4	383,375.8	0.01985	1,441,035.4	17.9
65	0.13127	72,716.7	9,545.3	340,818.3	0.02801	1,057,659.6	14.5
70	0.20498	63,171.5	12,948.6	284,754.8	0.04547	716,841.3	11.3
75	0.31172	50,222.8	15,655.5	212,570.3	0.07365	432,086.5	8.6
80	0.45848	34,567.3	15,848.4	132,676.7	0.11945	219,516.2	6.4
85	1.00000	18,718.9	18,718.9	86,839.4	0.21556	86,839.4	4.6

b. Graph the following life table functions using either `plot()` or `ggplot()`: `lx`; `ndx`; and `nm`. What do you observe?

Figure 1 below shows the three following life table functions for the French male population in 1985: `lx` (the number of people still living), `ndx` (number of people dying between each age interval), and `nm` (the death rate for each cohort). We can see that these three functions are largely complementary, as expected. As age advances, `lx` declines at a steady rate as `ndx` starts to increase at a steady rate. Towards later life, we see a lower `lx` and higher `ndx` and

nm_x , signaling the higher mortality rate of older cohorts and the fewer number of them left alive. On an interesting note related to these functions, we observe some major demographic mortality trends very clearly: the early life increased mortality rate, the young adult mortality bump, and the later life steady mortality increase.

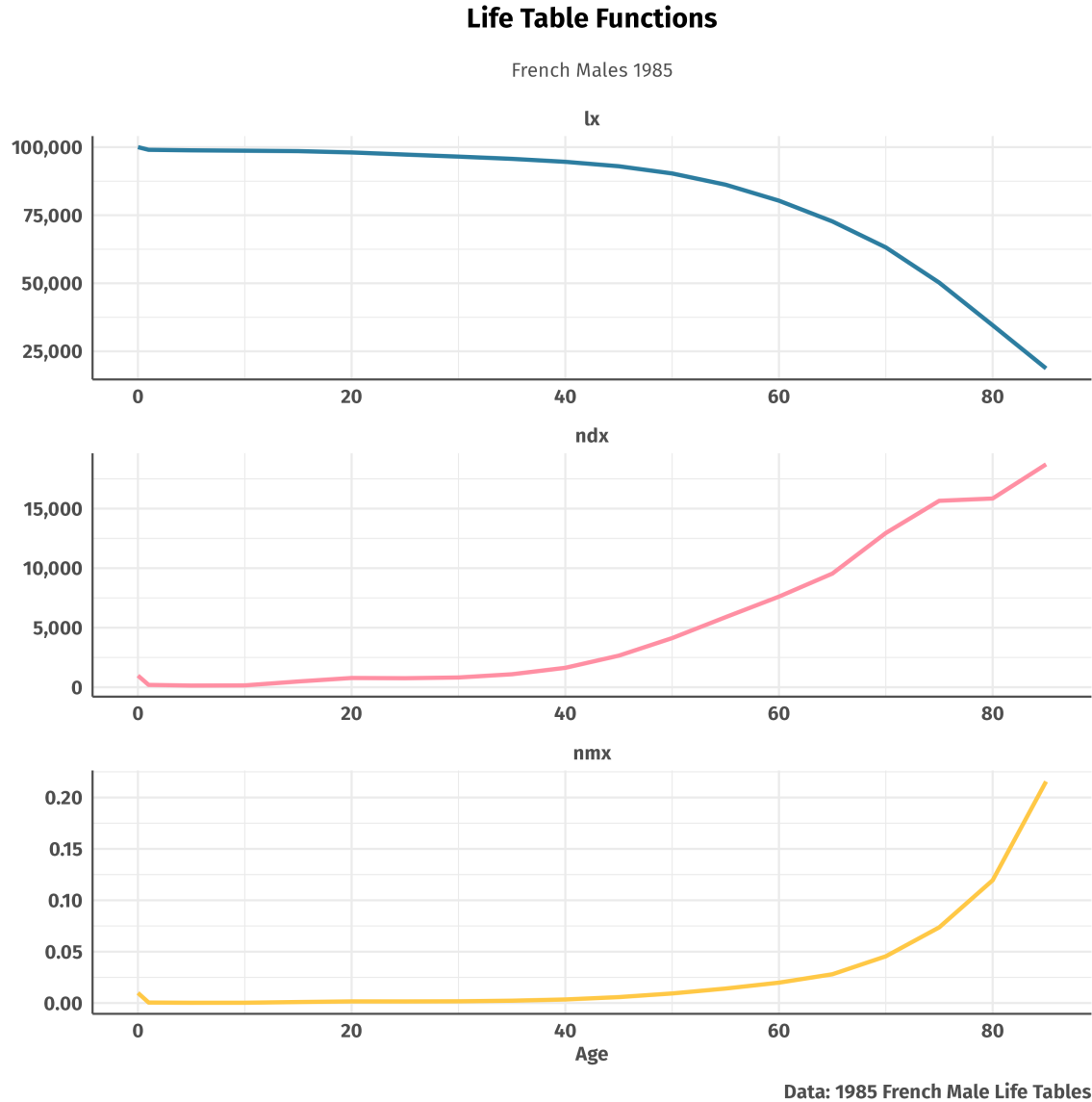


Figure 1: Life table functions for French male population in 1985. The top panel shows l_x , the middle panel shows nd_x , and the bottom panel shows nm_x .

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

In 1985, the life expectancy for French males at age 40 was 34.12 more years. This means that on average, French males that were 40 could expect to live to 74.12 years of age.

Life Expectancy at Age 40

x	ex
40	34.12045

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

The probability of surviving from birth to age 30 was 0.9652159

```
lt$lx[lt$x == 30] / lt$lx[lt$x == 0]
```

```
[1] 0.9652159
```

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

The probability of surviving to age 65 for those that survived to age 30 was 0.7533727.

```
lt$lx[lt$x == 65] / lt$lx[lt$x == 30]
```

```
[1] 0.7533727
```

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

The probability that a newborn will die between 50 and 55 is 0.04124974.

```
(lt$lx[lt$x == 50] - lt$lx[lt$x == 55]) / lt$lx[lt$x == 0]
```

```
[1] 0.04124974
```

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

A newborn can expect to live 45.91713 person-years in the 15-65 year interval.

```
(lt$Tx[lt$x == 15] - lt$Tx[lt$x == 65]) / lt$lx[lt$x == 0]
```

```
[1] 45.91713
```

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?)

With only the `nax` column, we might be able to compare populations' mortality. Crudely, we could compare where in the range they fall, with those falling closer to 0 representing higher likely mortality rates for that age group. Specifically, the first `nax` value where $x = 0$ can tell us about child mortality rates and compare it to other populations.

i. If the French population were stationary, what would be the crude death rate?

The CDR would be 14.023 deaths per 1,000 people.

```
(1 / lt$ex[lt$x == 0]) * 1000
```

```
[1] 14.02322
```

j. Extra credit part 1: push your code to your Github page and list the URL in your submitted answers.

The github link for this class can be found [here](#).

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`.

We can use the [demCore](#) package available on GitHub to check some of our estimates. Overall, as Table 1 shows, our estimates for `lx` and `qx` appear to be consistent.

```
dtr <- demCore::gen_lx_from_qx(
  data.table::as.data.table(tr),
  id_cols = c("age_start", "age_end")
)
```

Comparing Manual Life Table Functions with `demCore`

x	package_nqx	manual_nqx	nqx_diff	package_lx	manual_lx	lx_diff
0	0.010	0.010	0	100000.00	100000.00	0
1	0.002	0.002	0	99024.26	99024.26	0
5	0.001	0.001	0	98828.95	98828.95	0
10	0.002	0.002	0	98690.42	98690.42	0
15	0.005	0.005	0	98536.79	98536.79	0
20	0.008	0.008	0	98054.61	98054.61	0

25	0.008	0.008	0	97278.46	97278.46	0
30	0.008	0.008	0	96521.59	96521.59	0
35	0.011	0.011	0	95703.25	95703.25	0
40	0.017	0.017	0	94613.89	94613.89	0
45	0.028	0.028	0	92983.77	92983.77	0
50	0.046	0.046	0	90334.31	90334.31	0
55	0.068	0.068	0	86209.34	86209.34	0
60	0.095	0.095	0	80326.18	80326.18	0
65	0.131	0.131	0	72716.74	72716.74	0
70	0.205	0.205	0	63171.49	63171.49	0
75	0.312	0.312	0	50222.85	50222.85	0
80	0.458	0.458	0	34567.33	34567.33	0
85	1.000	1.000	0	18718.90	18718.90	0

Table 1: Generated

2. Think about the social phenomena / processes that most interest you. Might any of these processes be measured in the form of a lifetable? If yes:

Maybe we could use life tables to understand the demographic profile of evictions.

a. What events would constitute “births” and “deaths”?

Births would still be births, but deaths would actually be evictions.

b. What could you learn from using a lifetable?

We don't have that much information on the demographics of tenants that are evicted, so this would be able to tell us sex and age specific eviction rates.

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?

This is the hard part. I know that Nick Graetz has done some work on this effort using the Federal Statistical Research Data Centers, so that might be the best source of information, linked with eviction data from Princeton's Eviction lab.

d. What issues might limit how the information produced in your lifetable can be interpreted?

I think all that we could say about this would be limited age-specific eviction rates, with data source-specific limitations. I'm also not sure how different this would be from a cox-regression model?

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