Week 2: Mortality I

SOC6708 ADA

Monica Alexander

Read in data

```
library(tidyverse)
library(here)
library(readxl)
library(janitor)
```

For this example we are using mortality in Canada by cause, 2012-2016. Available from StatCan.

```
d <- read_csv(here("data", "CAN_age_cod.csv"))
head(d)
```

```
# A tibble: 6 x 5
  start_age cause                                  year number  rate
      <dbl> <chr>                                 <dbl>  <dbl> <dbl>
1         0 Total, all causes of death  [A00-Y89]  2012   1818   483
2         0 Total, all causes of death  [A00-Y89]  2013   1884   492
3         0 Total, all causes of death  [A00-Y89]  2014   1794   471
4         0 Total, all causes of death  [A00-Y89]  2015   1737   454
5         0 Total, all causes of death  [A00-Y89]  2016   1741   450
6         0 Salmonella infections  [A01-A02]       2012      0     0
```

Create a life table

Let's create a life table using the mortality rates from all causes of death in 2016. Get the data we need:

```r
dl <- d %>%
  filter(year==2016, cause=="Total, all causes of death  [A00-Y89]") %>%
  mutate(age = start_age, Mx = rate/100000) %>%
  select(age, Mx)

head(dl)
```

```
# A tibble: 6 x 2
    age      Mx
  <dbl>   <dbl>
1     0 0.0045
2     1 0.00019
3     5 0.00008
4    10 0.00013
5    15 0.00033
6    20 0.00053
```

We need to create columns:

- n
- ${}_na_x$
- ${}_nq_x$
- ${}_nd_x$
- ${}_nL_x$
- T_x
- e_x

```r
d_lt <- dl |>
     mutate(n = case_when(
       age==0 ~ 1,
       age==1 ~ 4,
       TRUE ~ 5
     ),
       ax = case_when(
       age==0 ~ 0.07 + 1.7*Mx,
       age==1 ~ 1.5,
       age==90 ~ 1/Mx,
       TRUE ~ 2.5
     ),
      qx = n * Mx / (1 + (n - ax)* Mx),
      px = 1 - qx,
```

```
      lx = lag(cumprod(px), default = 1),
      dx = lx - lead(lx, default = 0),
      Lx = n * lead(lx, default = 0) + (ax* dx),
      Tx = rev(cumsum(rev(Lx))),
      ex = Tx / lx
      )
```

What's the life expectancy at age 10?

Exercise

Calculate the lifespan disparity for Canada in 2016.

Calculate cause-deleted life expectancy

Now calculate life expectancy if all intentional injuries were deleted. Get the data we need:

```
dls <- d %>%
  filter(year==2016, cause=="Total, all causes of death  [A00-Y89]"|cause=="Intentional self-
  mutate(age = start_age, Mx = rate/100000) %>%
  select(age, cause, Mx) %>%
  mutate(cause = ifelse(cause=="Intentional self-harm (suicide)  [X60-X84, Y87.0]", "suicide"
  spread(cause, Mx) %>%
  rename(Mx_i = suicide,
         Mx = total)

head(dls)
```

```
# A tibble: 6 x 3
    age    Mx_i      Mx
  <dbl>   <dbl>   <dbl>
1     0 0       0.0045
2     1 0       0.00019
3     5 0       0.00008
4    10 0.00003 0.00013
5    15 0.00009 0.00033
6    20 0.00012 0.00053
```

Exercise: Cause-deleted life expectancy

You need to create the same columns as above, but with the cause-deleted versions $(-i)$. Do this by first creating the ratio $R_x^{-i} = \frac{M_x^{-i}}{M_x}$, use this to get ${}_nq_x^{-i}$, and the rest is the same.

What's the cause-deleted life expectancy at age 10? What's the implied life lost due to suicide?

Decomposition

Let's read in WPP data from last week and calculate the age-specific mortality rates:

```
d_male <- read_xlsx(here("data/WPP2024_POP_F01_2_POPULATION_SINGLE_AGE_MALE.xlsx"), skip = 1
d_male$sex <- "Male"
d_male <- d_male |> drop_na(Year)
d_female <- read_xlsx(here("data/WPP2024_POP_F01_3_POPULATION_SINGLE_AGE_FEMALE.xlsx"), skip
d_female$sex <- "Female"
d_female <- d_female |> drop_na(Year)

d <- rbind(d_male, d_female)
rm(d_male, d_female)

d <- d  |>
  clean_names() |>
  select(region_subregion_country_or_area, iso3_alpha_code, year, x0:sex) |>
  rename(region = region_subregion_country_or_area) |>
  mutate(across(x0:x100, as.numeric))

d_male <- read_xlsx(here("data/WPP2024_MORT_F01_2_DEATHS_SINGLE_AGE_MALE.xlsx"), skip = 16)
d_male$sex <- "Male"
d_male <- d_male |> drop_na(Year)
d_female <- read_xlsx(here("data/WPP2024_MORT_F01_3_DEATHS_SINGLE_AGE_FEMALE.xlsx"), skip =
d_female$sex <- "Female"
d_female <- d_female |> drop_na(Year)

dm <- rbind(d_male, d_female)
rm(d_male, d_female)

dm <- dm  |>
  clean_names() |>
  select(region_subregion_country_or_area, iso3_alpha_code, year, x0:sex) |>
```

```r
  rename(region = region_subregion_country_or_area) |>
  mutate(across(x0:x100, as.numeric))

d_long <- d  |>
  pivot_longer(x0:x100, names_to = "age", values_to = "pop") |>
  mutate(age = as.numeric(str_remove(age, "x")))

dm_long <- dm  |>
  pivot_longer(x0:x100, names_to = "age", values_to = "deaths") |>
  mutate(age = as.numeric(str_remove(age, "x")))

# join these two tibbles and calculate rates

asmr <- d_long |>
  left_join(dm_long) |>
  mutate(mx = deaths/pop)
```

Do the decomposition of the difference between Kenya and Canada:

```r
asmr |>
  filter(region == "Kenya",year==2023)  |>
  select(sex, age, pop, mx) |>
  rename(pop_kenya = pop, mx_kenya = mx) |>
  left_join(asmr |>
              filter(region == "Canada",year==2023)  |>
              select(sex, age, pop, mx) |>
              rename(pop_can = pop, mx_can = mx) ) |>
  mutate(prop_kenya = pop_kenya/sum(pop_kenya),
         prop_can = pop_can/sum(pop_can)) |>
  mutate(rate_diff = mx_kenya - mx_can,
         prop_diff = prop_kenya - prop_can) |>
  mutate(ave_rate = (mx_kenya+mx_can)/2,
         ave_prop = (prop_kenya+prop_can)/2) |>
  mutate(age_contr = prop_diff*ave_rate,
         rate_contr = rate_diff*ave_prop) |>
  summarize(age_total_contr = sum(age_contr),
            rate_total_contr = sum(rate_contr)) |>
  mutate(total_diff = age_total_contr+rate_total_contr)
```

```
# A tibble: 1 x 3
  age_total_contr rate_total_contr total_diff
```

```
             <dbl>            <dbl>      <dbl>
1          -0.0107          0.00998  -0.000724
```

Check that the difference is actually the difference between the two CDRs

```r
asmr |>
  filter(region == "Kenya"|region=="Canada",year==2023) |>
  group_by(region) |>
  summarize(cdr = sum(mx*pop)/sum(pop)) |>
  summarise(diff = cdr[region=="Kenya"] - cdr[region=="Canada"])
```

```
# A tibble: 1 x 1
       diff
      <dbl>
1 -0.000724
```

Exercise

Decompose the difference in CDRs between USA and Japan in the year 2023. Is the majority of the difference due to age structure or mortality?