

SOC 756: Problem Set 2

Mateo Frumholtz

September 23, 2025

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

In this problem set, you will be calculating gender differences in the expected duration of years lived in poverty for the United States in 2004. You will calculate variance estimates for these expectancies to determine if the estimates of differences between men and women are statistically meaningful, and at what ages.

1. Sullivan's method requires high quality lifetable data to construct total person years in each age group. Set up an account on the Human Mortality Database: <https://www.mortality.org/> and get the nL_x values for U.S. males and U.S. females in 2004 from the 5x1 abridged tables, e.g. for women: https://www.mortality.org/hmd/USA/STATS/ftper_5x1.txt Install the R package HMDHFDplus and get the lifetable values directly: <https://cran.r-project.org/web/packages/HMDHFDplus/HMDHFDplus.pdf> The commands ask you to supply your user name and password, you'll still need to sign up at HMD first.

We can download the HMD files directly from R using the HMDHFDplus R package (Riffe 2015).

```
# Downloaded with the following script -  
# commented out so I don't have to show my password  
  
# m_lt <- readHMDweb(  
#   CNTRY = "USA",  
#   item = "mltper_5x1",  
#   username = "frumh002@umn.edu"  
# ) |>  
#   filter(Year == 2004) |>  
#   rename(  
#     year = Year,  
#     age = Age  
#   ) |>  
#   select(-OpenInterval) |>
```

```

# mutate(
#   sex = "male"
# )

# fm_lt <- readHMDweb(
#   CNTRY = "USA",
#   item = "fltper_5x1",
#   username = "frumh002@umn.edu"
# ) |>
# filter(Year == 2004) |>
# rename(
#   year = Year,
#   age = Age
# ) |>
# select(-OpenInterval) |>
# mutate(
#   sex = "female"
# )

# blt <- rbind(m_lt, fm_lt) |>
# mutate(
#   age = as.numeric(age)
# )

```

2. I have used General Social Survey data to estimate the poverty prevalence by age separately for men and women. I have defined the poverty prevalence as the proportion living at or below the poverty line at the time of the survey. You will find these values, as well as the size of the sample used to compute them in a table on the class webpage. Because the GSS uses simple random sampling, we should be able to estimate standard errors using the simplified formula made available in Molla, Wagener, and Madans (2001), as opposed to correcting for the sampling scheme with a more complicated approach.

(a) Calculate the expected number of years lived in poverty above age x separately for men and women ages 0-100 (using the given intervals).

We can use the following equation from Imai and Soneji (2007) to estimate Sullivan's method for each sex.

$$\hat{e}_x^{POV} = \frac{1}{l_x} \sum_{i \in x} (1 - \hat{\pi}_i)_{ni} L_i$$

Where:

- \hat{e}_x^{POV} represents the average years lived in poverty above age x

- l_x represents the people that are still alive in the interval i
- ${}_{ni}\hat{\pi}_i$ represents the percent of people living in poverty (to estimate those living in poverty we just use the ${}_{ni}\hat{\pi}_i$ estimate)
- ${}_{ni}L_i$ represents the person-years lived in interval i

```
pov_lt <- df |>
  mutate(
    adj_Lx = Lx * pov
  ) |>
  group_by(sex) |>
  mutate(
    Tx = rev(cumsum(rev(adj_Lx))),
    adj_ex = (1 / lx) * Tx
  ) |>
  ungroup()
```

Estimated years spent in poverty by sex
United States, 2004

Age	ex	
	Male	Female
0-1	10.62	16.29
1-4	10.49	16.18
5-9	9.66	15.36
10-15	8.75	14.45
15-20	7.95	13.65
20-25	7.18	12.87
25-30	6.28	11.54
30-35	5.79	10.68
35-40	5.43	9.89
40-45	5.16	9.16
45-50	4.85	8.56
50-55	4.55	8.04
55-60	4.22	7.53
60-65	3.87	6.98
65-70	3.57	6.29
70-75	3.16	5.40
75-80	2.63	4.40
80-85	2.19	3.43
85-90	1.71	2.50
90-95	1.19	1.73
95-100	0.85	1.20
100-105	0.63	0.86

105-110	0.50	0.65
110+	0.63	0.55

Table 1: Estimated ex for male and females living in poverty in the US, 2004.

(b) Assess whether, at birth, the total expected number of years lived in poverty differs for men and women and whether this difference is statistically significant.

We can estimate the variance of the prevalence estimates using methods from Molla, Wagener, and Madans (2001) for each sex.

$$S^2({}_n\pi_x) = [{}_n\pi_x * (1 - {}_n\pi_x)] / {}_nN_x$$

Where:

- ${}_nN_x$ represents the number of people used to estimate the sample prevalence in age interval x to x+n
- ${}_n\pi_x$ represents the percent of people living in poverty

We can then use these variance estimates to get the overall variance of \hat{e}_x :

$$VAR(\hat{e}_x) = \frac{1}{l_x^2} \sum_{i=x}^w [{}_xL_i^2 * S^2(1 - {}_n\pi_x)]$$

Where:

- l_x represents the people that are still alive in the interval i
- ${}_niL_i$ represents the person-years lived in interval i
- $S^2(1 - {}_n\pi_x)$ is the variance of the prevalence estimates

```
pov_var <- pov_lt |>
  select(age, sex, lx, Lx, pov, n, adj_ex) |>
  mutate(
    s = (pov * (1 - pov)) / n,
    sx = ((Lx^2) * s)
  ) |>
  group_by(sex) |>
  arrange(age) |>
  mutate(
    adj_sx = rev(cumsum(rev(sx)))
  ) |>
  ungroup() |>
  mutate(
    var = (1 / (lx^2)) * adj_sx,
```

```

    se = sqrt(var),
    conf_low = adj_ex - se,
    conf_high = adj_ex + se
  ) |>
  select(age, sex, adj_ex, var, se, conf_low, conf_high) |>
  mutate(
    sex = str_to_title(sex)
  )

pov_stat <- pov_var |>
  select(age, sex, adj_ex, var, se) |>
  mutate(
    sex = tolower(sex)
  ) |>
  pivot_wider(names_from = sex, values_from = c(adj_ex:se)) |>
  mutate(
    num_z = adj_ex_female - adj_ex_male,
    denom_z = se_male + se_female,
    z = num_z / denom_z,
    p = 2 * pnorm(q = z, lower.tail = FALSE),
    pval = case_when(
      p < 0.001 ~ "< 0.001",
      p < 0.05 ~ "< 0.05",
      TRUE ~ "> 0.05"
    )
  ) |>
  select(-denom_z, -p)

```

Statistical test for sex-differences in years spent in poverty
United States, 2004

Age	ex		Diff.	Variance		SE		Z-value	p-value
	Female	Male		Female	Male	Female	Male		
0	16.3	10.6	5.7	0.049	0.040	0.221	0.199	13.48	< 0.001
1	16.2	10.5	5.7	0.049	0.040	0.222	0.200	13.49	< 0.001
5	15.4	9.7	5.7	0.047	0.037	0.217	0.193	13.90	< 0.001
10	14.5	8.8	5.7	0.045	0.035	0.212	0.188	14.26	< 0.001
15	13.7	8.0	5.7	0.043	0.033	0.208	0.182	14.61	< 0.001
20	12.9	7.2	5.7	0.041	0.031	0.203	0.177	14.98	< 0.001
25	11.5	6.3	5.3	0.038	0.029	0.195	0.170	14.39	< 0.001
30	10.7	5.8	4.9	0.037	0.028	0.191	0.167	13.66	< 0.001
35	9.9	5.4	4.5	0.035	0.027	0.187	0.164	12.72	< 0.001
40	9.2	5.2	4.0	0.033	0.026	0.182	0.162	11.61	< 0.001

45	8.6	4.9	3.7	0.032	0.026	0.178	0.160	10.95	< 0.001
50	8.0	4.5	3.5	0.030	0.025	0.174	0.158	10.53	< 0.001
55	7.5	4.2	3.3	0.029	0.024	0.170	0.156	10.19	< 0.001
60	7.0	3.9	3.1	0.027	0.024	0.166	0.154	9.72	< 0.001
65	6.3	3.6	2.7	0.026	0.024	0.161	0.155	8.58	< 0.001
70	5.4	3.2	2.2	0.025	0.025	0.158	0.157	7.09	< 0.001
75	4.4	2.6	1.8	0.024	0.027	0.155	0.163	5.56	< 0.001
80	3.4	2.2	1.2	0.024	0.029	0.155	0.170	3.82	< 0.001
85	2.5	1.7	0.8	0.022	0.029	0.147	0.170	2.48	< 0.05
90	1.7	1.2	0.5	0.014	0.018	0.117	0.134	2.15	< 0.05
95	1.2	0.8	0.4	0.008	0.011	0.091	0.105	1.82	> 0.05
100	0.9	0.6	0.2	0.005	0.007	0.070	0.082	1.53	> 0.05
105	0.6	0.5	0.2	0.003	0.004	0.055	0.067	1.25	> 0.05
110	0.5	0.6	-0.1	0.002	0.007	0.048	0.086	-0.59	> 0.05

Table 2: Statistical findings from expected differences in years spent living in poverty between male and females in the US, 2004.

At birth, males have an estimated 10.6 expected number of years lived in poverty while females have an estimated 16.3 years (Figure 1). These differences appear to be statistically significant at a $p < 0.001$ - level (Table 2).

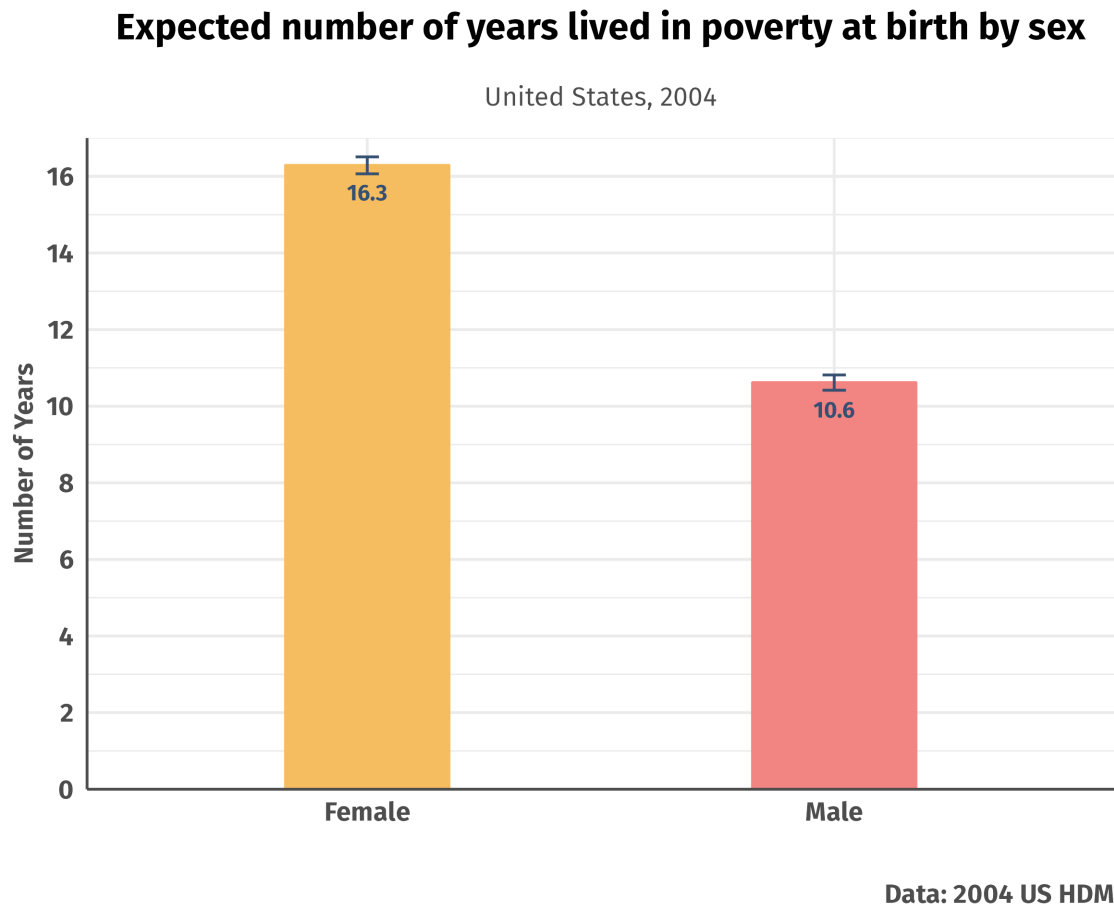


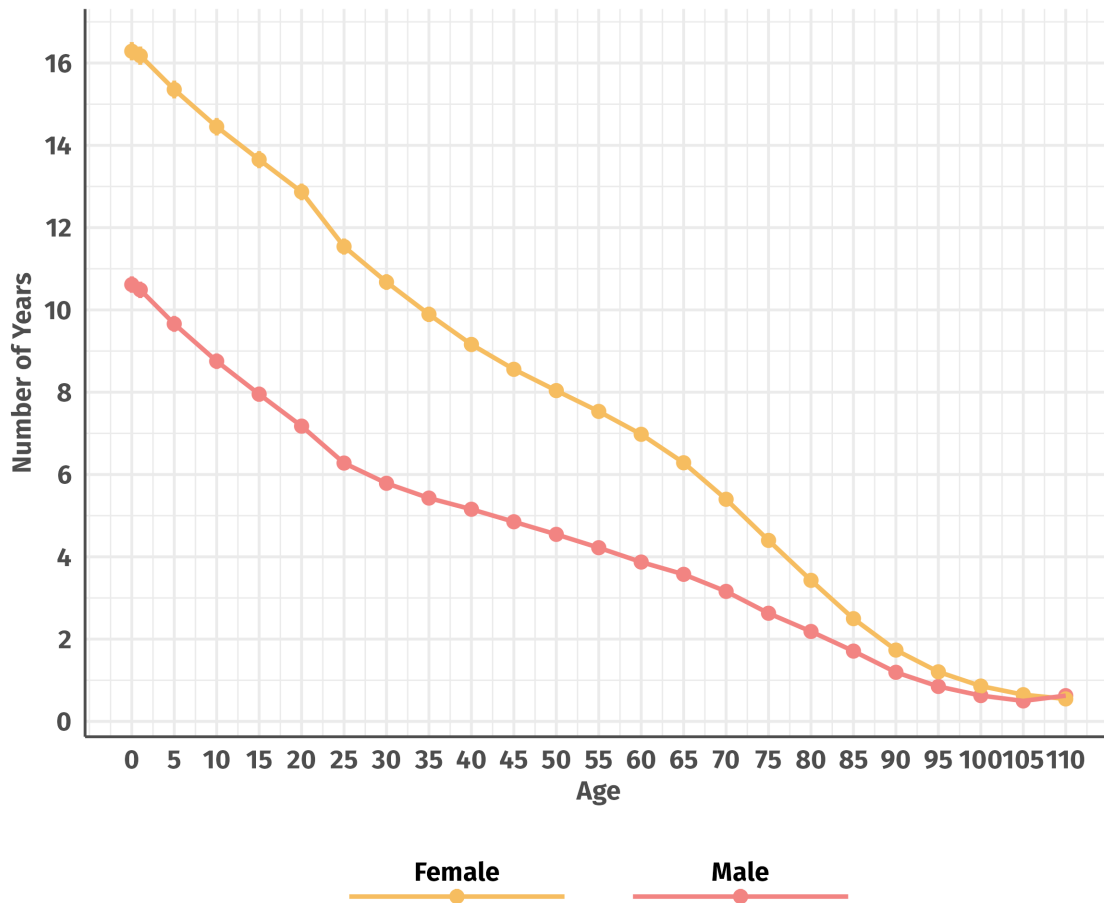
Figure 1: Expected number of years lived in poverty at birth by sex in the United States, 2004.

(c) Assess these differences at each age interval.

At each age interval, males have a lower expected number of years lived in poverty than females (Figure 2). These differences appear to be significant across all age intervals until age 80, at a $p < 0.001$ level, and until age 90, at a $p < 0.05$ level (Table 2).

Expected number of years lived in poverty at birth by sex

United States, 2004



Data: 2004 US HDM

Figure 2: Expected number of years lived in poverty by sex in the United States, 2004.

References

- Imai, Kosuke, and Samir Soneji. 2007. "On the Estimation of Disability-Free Life Expectancy: Sullivan's Method and Its Extension." *Journal of the American Statistical Association* 102 (480): 1199–1211. <https://doi.org/10.1198/016214507000000040>.
- Molla, Michael T., Diane K. Wagener, and Jennifer H. Madans. 2001. "Summary Measures of Population Health: Methods for Calculating Healthy Life Expectancy: (583762012-001)." <https://doi.org/10.1037/e583762012-001>.

Riffe, Timothy. 2015. "Reading Human Fertility Database and Human Mortality Database Data into r." Rostock: Max Planck Institute for Demographic Research. <https://doi.org/10.4054/MPIDR-TR-2015-004>.