Comparing Period TFR for 2 groups

Carl Schmertmann

9 Nov 2019"

TFR for one group

Suppose that there are A possible single-year reproductive ages $a = 1 \ldots A$ in which individuals may have children. At some of these ages fertility rates may be zero for some groups (for example, for women at $a > 50$).

Suppose also that there are a total of B births in a given year to some group of interest (such as "all women", "all men", "Asian-American men", "residents of Montana" ...). Call π_x the fraction of the B births that happened to group members who were age x. That is, births disaggregated by age are

$$B_x = B\,\pi_x \quad x = 1, 2, \ldots, A$$

If there were N_x group members at age x at mid-year, then period TFR for this group is

$$TFR = \sum_x \frac{B_x}{N_x} = B \sum_x \pi_x \left(\frac{1}{N_x} \right) = B\,E_\pi \left(N_x^{-1} \right) \tag{1}$$

where the expectation in (1) is taken over the observed distribution of parents' ages for this year's births.

Comparing TFR for 2 groups

If there are two groups then the difference between their period TFRs is

$$TFR_1 - TFR_2 = B_1 E_{\pi_1} \left(N_{1x}^{-1} \right) - B_2 E_{\pi_2} \left(N_{2x}^{-1} \right) \tag{2}$$

In some cases (for example when the groups are "women" and "men") the total births are the same for both groups. For those cases:

$$TFR_1 - TFR_2 = B \left[E_{\pi_1} \left(N_{1x}^{-1} \right) - E_{\pi_2} \left(N_{2x}^{-1} \right) \right] \tag{3}$$

Using the very simple-but-useful results derived in the Appendix below, the difference can be approximated well by

$$TFR_1 - TFR_2 \approx B \left[\frac{1}{N_{1,m_1}} - \frac{1}{N_{2,m_2}} \right]$$

where m_1 and m_2 are the mean ages of parents in the two groups for this year's observed births.

Example: US TFRs in 1975

Start by loading ASFR and population data for 1975. Calculate male and female TFRs.

```
# USA male fertility
# from Dudel, C. and S. Klüsener. Male fertility data for high-income
# countries [unpublished data]. Submitted to the HFC by C. Dudel # on 15.05.2019.
#
# downloaded from Human Fertility Collection 2 Nov 2019
# https://www.fertilitydata.org/data/RAW_DATA/m_USA_51.zip


# Male 1975 rates from HFC, as described above
asfr_male = read.table(file='m_ASFR_USA.txt', header=TRUE) %>%
            filter(Year=='1975') %>%
            pull(ASFR)

# male data pulled is for ages 15-59. Add zero rates for 12-14
asfr_male = c( rep(0,3), asfr_male)
TFR_male  = sum(asfr_male)

# Female 1975 rates from HFD
asfr_female = read.table(file='USAasfrTR.txt', skip=2, header=TRUE) %>%
       mutate(Age = 11 + as.numeric(Age)) %>%
       group_by(Year,Age) %>%
       summarize(ASFR = mean(ASFR)) %>%
       filter(Year=='1975') %>%
       pull(ASFR)

# female data pulled is for ages 12-54. Add zero rates for 56-59
asfr_female = c( asfr_female, rep(0,4))
TFR_female  = sum(asfr_female)
```

For the US in 1975, female TFR was **1.77** and male TFR was **1.94**.

Now illustrate 1975 fertility rates, births, and populations, by age and sex.
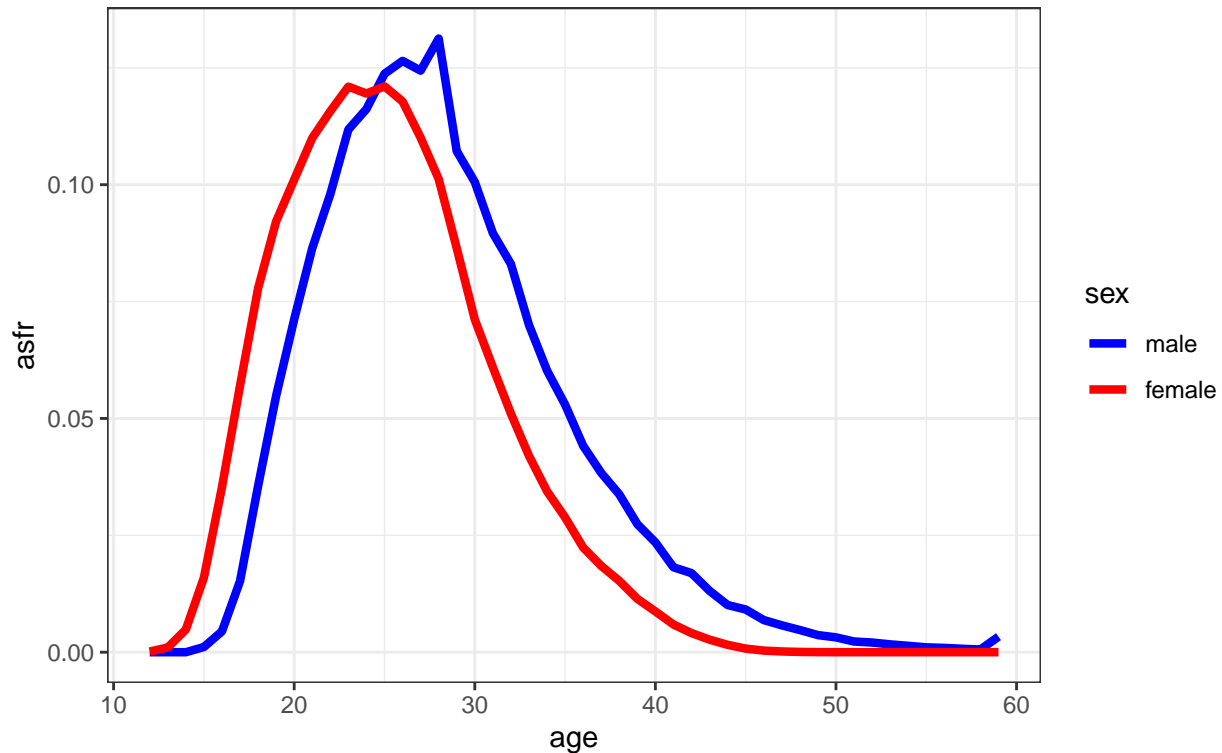
```
tmp = expand.grid(age=12:59, sex=c('male','female')) %>%
        add_column( asfr=c(asfr_male,asfr_female))

ggplot(data=tmp) +
  aes(x=age, y=asfr, color=sex) +
  geom_line(lwd=1.5) +
  scale_color_manual(values=c('blue','red')) +
  labs(title='US Age-Specific Fertility Rates 1975',
       subtitle=paste('Female TFR=',round(TFR_female,2),
                      '  Male TFR=',  round(TFR_male,2))) +
  theme_bw()
```

US Age–Specific Fertility Rates 1975
Female TFR= 1.77   Male TFR= 1.94


```r
# 1975 US populations

pop = read.table('Population.txt',skip=2, header=TRUE) %>%
        filter(Year=='1975', Age %in% 12:59) %>%
        select(Age,Female,Male)

B_male     = pop$Male   * asfr_male
B_female   = pop$Female * asfr_female

# these are slightly different. Use the avg to calculate the 1975 total
B = mean(sum(B_male), sum(B_female))

m_male   = weighted.mean( 12:59+.50, B_male)
m_female = weighted.mean( 12:59+.50, B_female)

msq_male   = weighted.mean( (12:59+.50)^2, B_male)
msq_female = weighted.mean( (12:59+.50)^2, B_female)

v_male   = msq_male   - m_male^2
v_female = msq_female - m_female^2

Nm_male   = approx(x=12:59+.50, y=pop$Male,   xout= m_male)$y
Nm_female = approx(x=12:59+.50, y=pop$Female, xout= m_female)$y

bump_y   = approx(x=12:59+.50, y=pop$Male,   xout= m_male + c(-0.5,+0.5))$y
Nm_prime_male = diff(bump_y)
```

```
bump_y     = approx(x=12:59+.50, y=pop$Female,   xout= m_female + c(-0.5,+0.5))$y
Nm_prime_female = diff(bump_y)

tmp = expand.grid(age=12:59, sex=c('male','female')) %>%
        add_column( births=c(B_male,B_female))

ggplot(data=tmp) +
  aes(x=age, y=births, color=sex) +
  geom_line(lwd=1.5) +
  geom_vline(xintercept = c(m_female,m_male), color=c('red','blue'),
             lwd=1.5,lty='dotted') +
  scale_color_manual(values=c('blue','red')) +
  labs(title='US Age-Specific Births 1975',
       subtitle=paste('Female Mean Age =',round(m_female,2),
                      '  Male Mean Age =',round(m_male,2))) +
  theme_bw()
```

US Age–Specific Births 1975
Female Mean Age = 24.94   Male Mean Age = 27.71
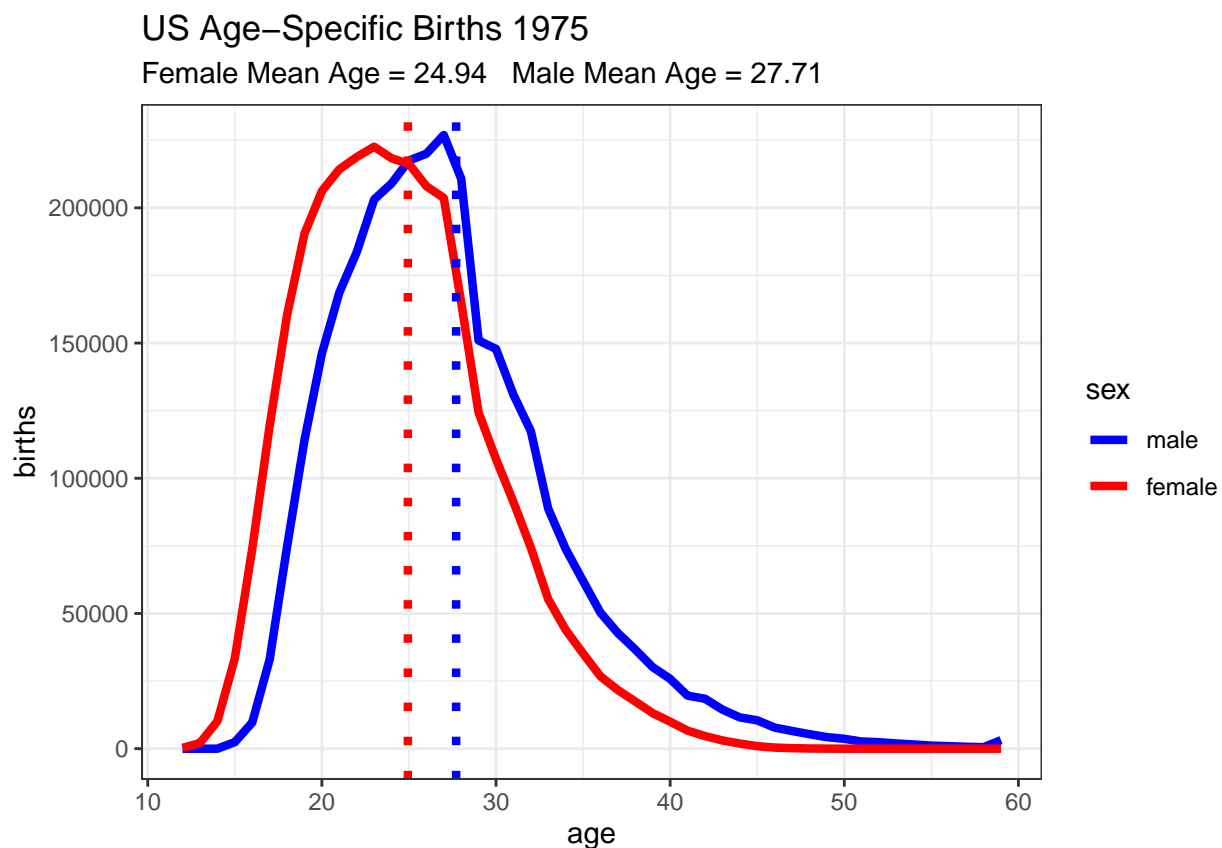

```
## population
tmp = expand.grid(age=12:59, sex=c('male','female')) %>%
        add_column( pop=c(pop$Male,pop$Female))

ggplot(data=tmp) +
  aes(x=age, y=pop, color=sex) +
  geom_line(lwd=1.5) +
  geom_vline(xintercept = c(m_female,m_male), color=c('red','blue'),
             lwd=1.5,lty='dotted') +
```

```
scale_color_manual(values=c('blue','red')) +
labs(title='US Age-Specific Populations 1975',
     subtitle=paste('Female Mean Age =',round(m_female,2),
                    '  Male Mean Age =',round(m_male,2))) +
theme_bw()
```

US Age–Specific Populations 1975
Female Mean Age = 24.94   Male Mean Age = 27.71
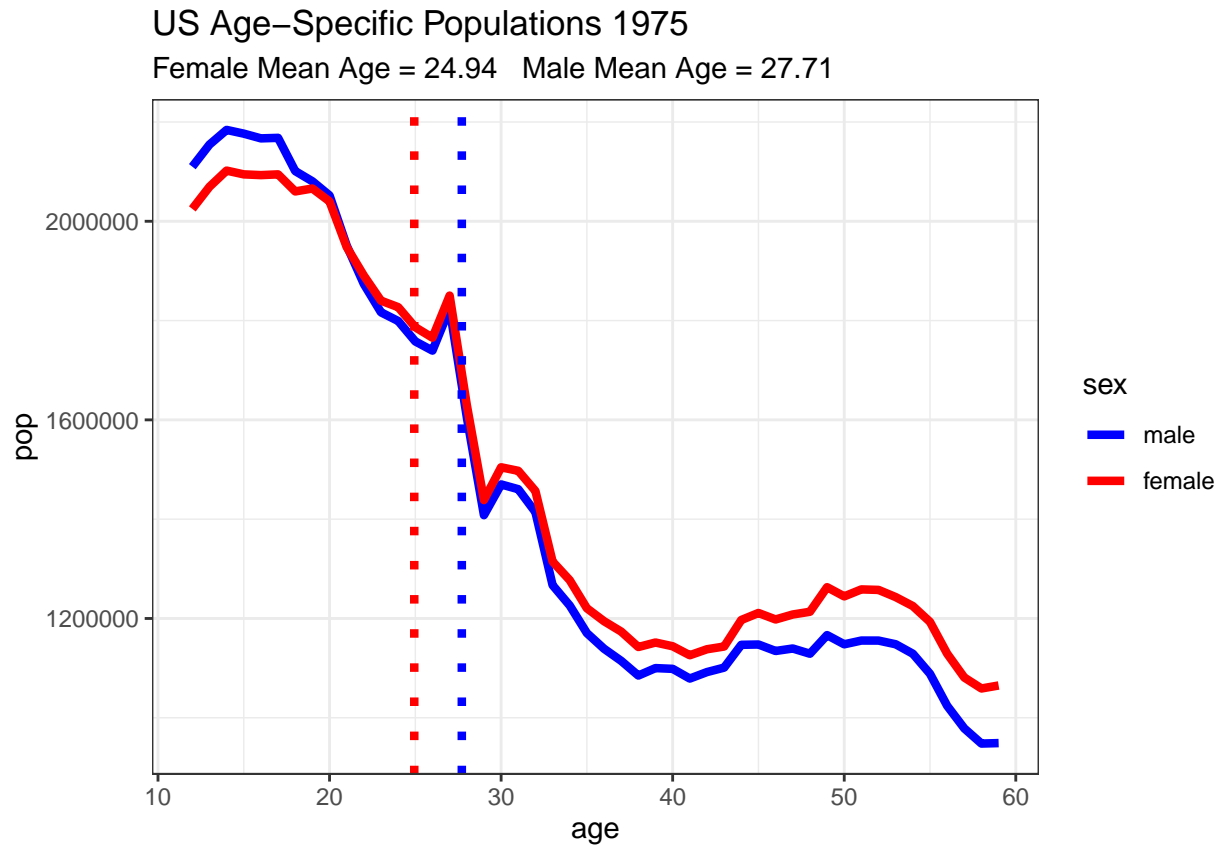
Now we will check the approximations based on

- mean parental ages for 1975 births: **24.94** for females and **27.71** for males
- variance of parental ages for 1975 births: **28.45** for females and **39.85** for males
- 1975 populations at mean ages: 1.809286×10^6 for females and 1.778486×10^6 for males

```
TFR_female
```

```
## [1] 1.76838
```

```
B/Nm_female
```

```
## [1] 1.710047
```

```
B * (1/Nm_female) * (1 + v_female/Nm_female * Nm_prime_female/Nm_female)
```

```
## [1] 1.710046
```

```
TFR_male
```

```
## [1] 1.93709
```

```
B/Nm_male
```

```
## [1] 1.739661
```

```
B * (1/Nm_male) * (1 + v_male/Nm_male * Nm_prime_male/Nm_male)
```

```
## [1] 1.739658
```

Appendix: Approximation

We can approximate the TFRs in (1) by considering a Taylor series approximation for N_x^{-1}. For the purposes of this derivation, temporarily ignore the discrete nature of ages in observed data, and momentarily consider N_x as a continuous function of exact age x.

Define the (period) mean age of parents in our group as m. A discrete approximation to m is $\sum_x \pi_x \left(x + \frac{1}{2}\right)$. The second-order Taylor approximation to the N_x^{-1} is

$$N_x^{-1} \approx N_m^{-1} + (x - m)\left[-N_m^{-2}N_m'\right] + \tfrac{1}{2}(x - m)^2\left[2N_m^{-3}N_m' - N_m^{-2}N_m''\right]$$

Thus the expectation in (1) is

$$E(N_x^{-1}) \approx N_m^{-1} + \tfrac{\sigma_x^2}{2}\left[2N_m^{-3}N_m' - N_m^{-2}N_m''\right]$$

where σ_x^2 is the observed variance of parental ages. Factoring out common multipliers and rearranging yields

$$E(N_x^{-1}) \approx N_m^{-1}\left[1 + \tfrac{\sigma_x^2}{2N_m}\left(2\tfrac{N_m'}{N_m} - N_m''\right)\right]$$

We can further simplify the approximation by noting that in most real populations the second derivative of exposure with age is likely very close to zero (very little local curvature in exposure by age). Thus $N_m'' \approx 0$ and

$$E(N_x^{-1}) \approx N_m^{-1}\left[1 + \tfrac{\sigma_x^2}{N_m}\left(\tfrac{N_m'}{N_m}\right)\right]$$

Further, variance in parental ages at childbearing is typically in the 25-50 range $\sigma_x^2 \in [25, 50]$, and the derivative of log exposure by age would typically be close to zero – perhaps $\tfrac{N_m'}{N_m} \in [-.05, +.05]$. Thus with a large population (high N_m) the second term in square brackets would likely be small, so that a simple approximation

$$E(N_x^{-1}) \approx N_m^{-1}$$

will often be quite useful.