Comparing Period TFR for 2 groups

Carl Schmertmann

10 Nov 2019

TFR for one group

Suppose that there are A possible single-year reproductive ages $a = 1 \dots A$ 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

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

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 ratio of their period TFRs is

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

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

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

Using the results derived in the Appendix below, the ratio in (3) can be approximated well by

$$\frac{TFR_1}{TFR_2} \approx \frac{N_{2,m_2}}{N_{1,m_1}}$$

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 1970

Start by loading ASFR and population data for 1970. 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 Male fertility data for high-income countries [unpublished data].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.

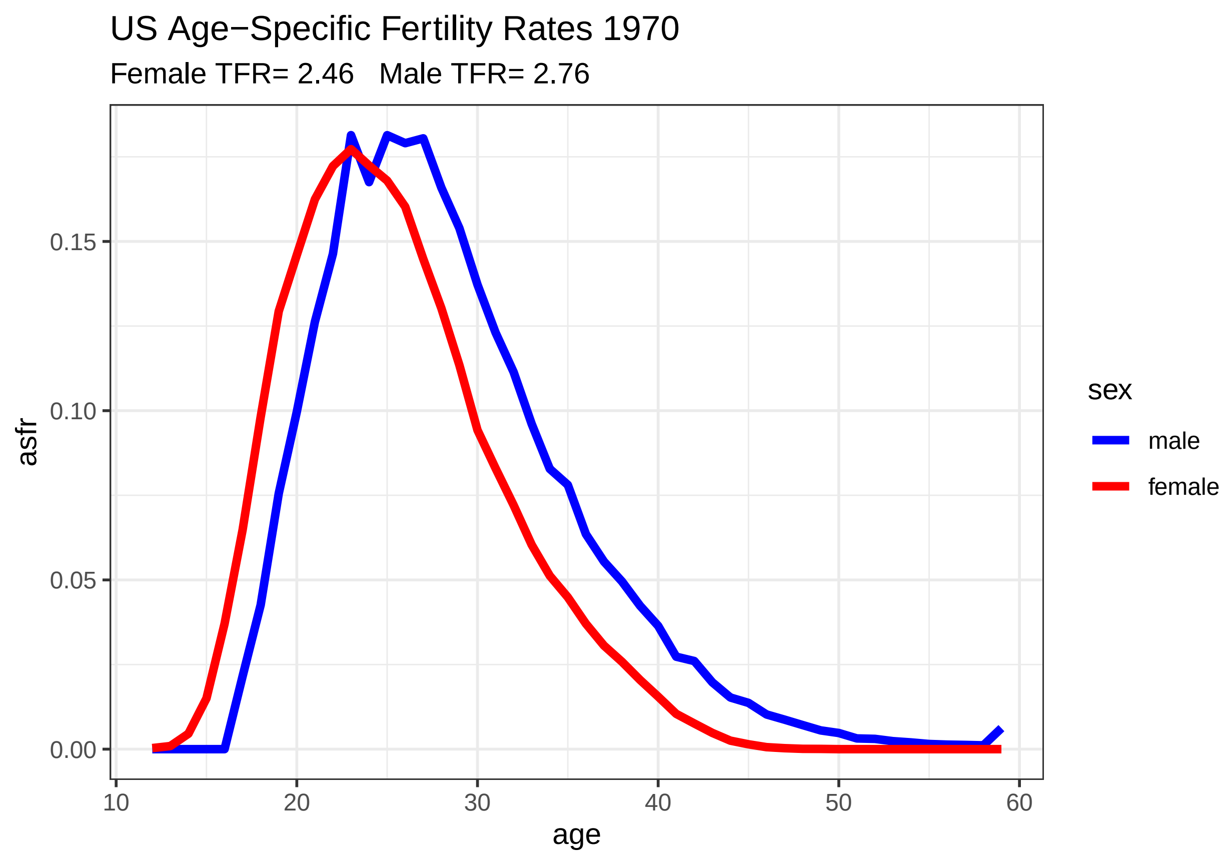
downloaded from Human Fertility Collection 2 Nov 2019 https://www.fertilitydata.org/data/RAW_DATA/m_USA_51.zip

```
# Male rates from HFC, as described above
asfr_male = read.table(file='m_ASFR_USA.txt', header=TRUE) %>%
              filter(Year== as.character(this_year)) %>%
              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 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==as.character(this_year)) %>%
       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 1970, female TFR was **2.46** and male TFR was **2.76**.

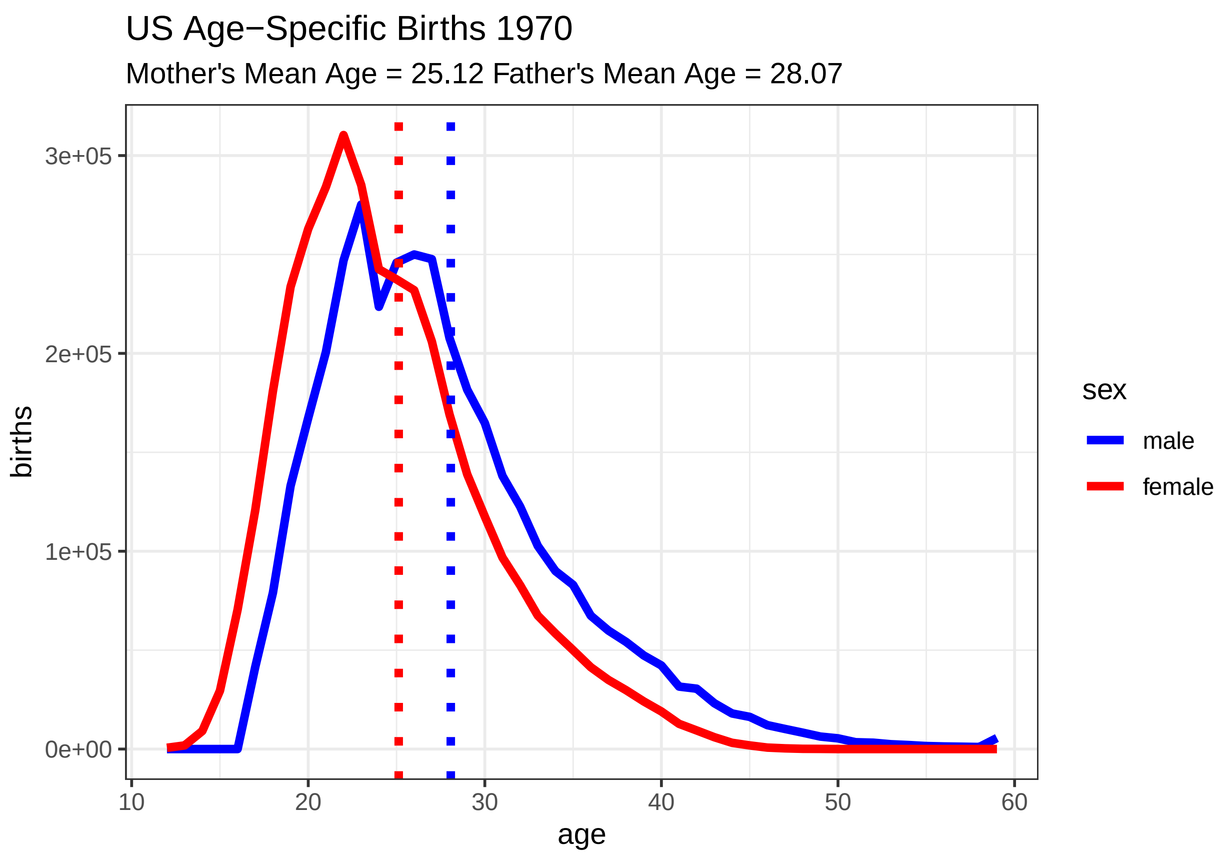
Now illustrate 1970 fertility rates, births, and populations, by age and sex.

```
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=paste('US Age-Specific Fertility Rates',this_year),
       subtitle=paste('Female TFR=',round(TFR_female,2),
                      '  Male TFR=',  round(TFR_male,2))) +
  theme_bw()
```




```
# US populations

pop = read.table('Population.txt',skip=2, header=TRUE) %>%
        filter(Year==as.character(this_year), 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 period 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

## function to approximate the pop at the mean age of births
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
```

```r
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=paste('US Age-Specific Births',this_year),
       subtitle=paste('Mother\'s Mean Age =',round(m_female,2),
                      'Father\'s Mean Age =',round(m_male,2))) +
  theme_bw()
```




```r
## 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=paste('US Age-Specific Populations',this_year)) +
  theme_bw()
```

Now we will check the approximations based on

- mean parental ages for 1970 births: **25.12** for females and **28.07** for males
- variance of parental ages for 1970 births: **30.97** for females and **45.09** for males
- 1970 populations at mean ages: 1.410823×10^6 for females and 1.303519×10^6 for males

```
TFR_female
```

```
## [1] 2.46046
```

```
B/Nm_female
```

```
## [1] 2.591371
```

```
TFR_male
```

```
## [1] 2.75801
```

```
B/Nm_male
```

```
## [1] 2.804689
```

Thus WORKING HERE...

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. Typical values for the variance term are in $[25, 50]$, so when exposure N is measured in thousands or millions, the second-order term is extremely small relative to the first-order term.

This leads to a very simple approximation for TFR that is fairly robust:

$$TFR \approx \frac{B}{N_m} = \frac{\text{total annual births}}{\text{mid-year population at mean parental age}}$$