

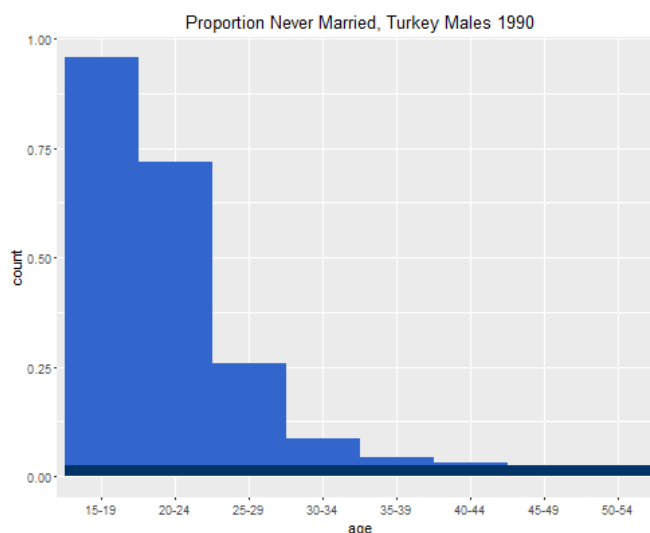
Current Status Life Tables

Singulate Mean Age at Marriage

Hajnal (1953) proposed a method for estimating the distribution of age at marriage from proportions single by age. The idea is that in a closed population with no mortality the proportion of women who remain single at age x is a direct estimate of $l(x)/l(0)$. Summing proportions single over age we obtain an estimate of time lived in the single state. From that we can obtain mean age at marriage by analogy with expectation of life. The only slight complication is that not everyone marries. The solution is quite simple, we rescale the survival function so it applies only to those who will marry, using

$$l^*(x) = \frac{l(x) - l(\infty)}{1 - l(\infty)}$$

In the website I go through the exercise on page 90 of the textbook, working with proportions single among Turkish men in 1990. I start at age 15 because no-one marries before then. The time lived single between ages 15 and 50 is just 5 times the sum of proportions single: 10.59 years. The proportion that remains single by age 50 is estimated at 2.33% by averaging the last two age groups. The mean age at marriage of those who marry by age 50 is then estimated as



$$15 + \frac{10.59 - 35 \times 0.0233}{1 - 0.0233} = 25.004$$

We start with 10.59 and subtract the time spent single by those who don't marry by age 50, namely $35 \times 0.0233 = 0.816$ years, scale up dividing by the proportion who marry by age 50, and add the 15 years everyone spends single from birth to age 15. In this example the adjustment makes little difference because almost every one marries by age 50.

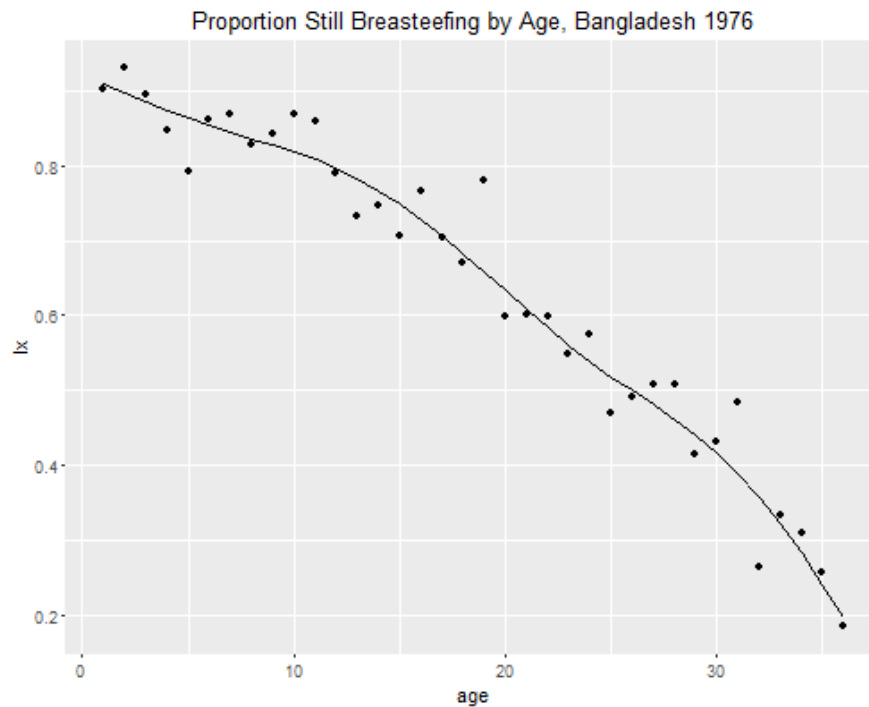
The resulting estimate is called the *singulate mean age at marriage* or SMAM. It does not represent the conditional mean age at marriage of a real cohort unless nuptiality has remained constant over the last 35 years or so. Even without that interpretation, however, it is a useful summary of period nuptiality, as the unscaled version represents the total time lived in the single state in a given period. For example in 1990 Turkish men spent 10.59 of the 35 years from ages 15 to 50 single. If people delay or forego marriage in a calendar period this indicator would be expected to increase.

Duration of Breastfeeding

Current status life tables are particularly useful in the study of duration of breastfeeding. One could, of course, build a cohort life table using retrospective reports of breastfeeding duration, but these are notoriously unreliable. We illustrate these ideas using data from the World Fertility Survey (WFS) of Bangladesh, conducted in 1976.

The WFS asked questions on breastfeeding for the last and next to last births only, and coded the answers in terms of the last closed and the open birth intervals. A life table based on retrospective reports of breastfeeding duration would combine both pieces of information, restricting the analysis to births in the last three years or so to ensure a representative sample of births. As we have seen earlier, the raw data show very substantial heaping on multiples of 12, and to a lesser extent some multiples of six. This could represent a real tendency for women to wean their children after achieving a milestone such as age two, but it could also represent bad data with substantial rounding to whole years.

An alternative approach is to tabulate whether the child is still being breastfed or not by current age. This is a direct estimate of $l(x)/l(0)$, with the estimate at very young ages representing the proportion who are ever breastfed. The figure below shows a current status life table by single months of age, based on all births in the 36 months before the survey. If the birth happened to be the last one I got breastfeeding status from the open interval data, otherwise I assumed it was weaned.



The first thing to note is that there is no precipitous drop in the proportion of children being breastfed around age two, as you would expect if the increased risk of weaning at this milestone was real. The other problem is that with single-month data we have small sample sizes and the proportions still breastfeeding are erratic, and fail to decline monotonically as any decent survival function should.

There is an algorithm called *pool adjacent violators* that obtains a monotone estimate by averaging successive entries that violate the monotonicity constraint. An alternative solution is to smooth the estimates, as I have done in the figure by using a regression spline with internal knots at 12 and 24 months. Note that the proportion still breastfeeding at age x completed months is attributed to exact age $x + 1/2$.

We can also compute mean duration of breastfeeding for all births by simply summing the proportions still breastfed by age, which is equivalent to computing time lived in the breastfeeding state. In our example the area under the curve is 23.0 months using the raw data and 22.8 months using the spline. If one wanted to compute the mean only for children who are breastfed one could adjust the survival curve dividing by the proportion ever breastfed. Assuming 95% breastfeed the mean would be 24.0 months.

Incidence-Prevalence

If all one wanted is the mean there is an even simpler estimate called the incidence-prevalence estimator. This is in common use in epidemiology, where one can approximate the duration of a disease in months dividing the number of new cases per month (incidence) by the number of existing cases at a given time (prevalence). Treating breastfeeding as a “disease” we would divide the number of births per month by the number of children being breastfed at the time of the survey. If the number of births is relatively constant over time this is exactly the same as summing the proportions ever breastfed by age. We show online that the incidence-prevalence estimate, computed as the overall proportion still breastfeeding over $0.95/36$, an estimate of “new cases” per month, is 23.8 months.

Right and Left Censoring

Note in closing that with current status data *all* observations are censored at their current age:

- children still breastfeeding are *right* censored; all we know is that they will breastfeed longer than their current age, and
- children weaned are *left* censored; all we know is that they were breastfeed less than their current age.

Yet we can still estimate a survival curve!