Direct and Indirect Standardization

POP 502 / ECO 572/ SOC 532 • SPRING 2017

Rates as Weighted Averages

Any general rate can be written as a weighted average of group-specific rates, with weights given by the proportion of exposure time in each group. The basic result (see page 23 in the textbook) is

$$R = \sum c_i m_i$$

where c_i reflects the composition and m_i the rate for group i. A general rate can change over time (or differ across groups) because of changes in composition, even if the group-specific rates are constant. (Frequently the groups are defined by age.)

Example: The CDR in Kazakhstan is lower than in Sweden: 7.42 compared to 10.55 deaths per 1000 population. We suspect this is due to its younger age structure. Note: All calculations are shown in the website using R and Stata. The key results are summarized in the following table combining different rates and compositions:

Composition	Rates		
	Kazakhstan	Sweden	Average
Kazakhstan	7.42	4.20	5.81
Sweden	16.34	10.55	13.44
Average	11.88	7.37	

Direct Standardization

The purpose of standardization is to facilitate comparison of rates over time (or across groups) by removing the effect of composition. The direct standardized rate for a given population combines the population's group-specific rates m_i with the composition of a standard population c_i^S:

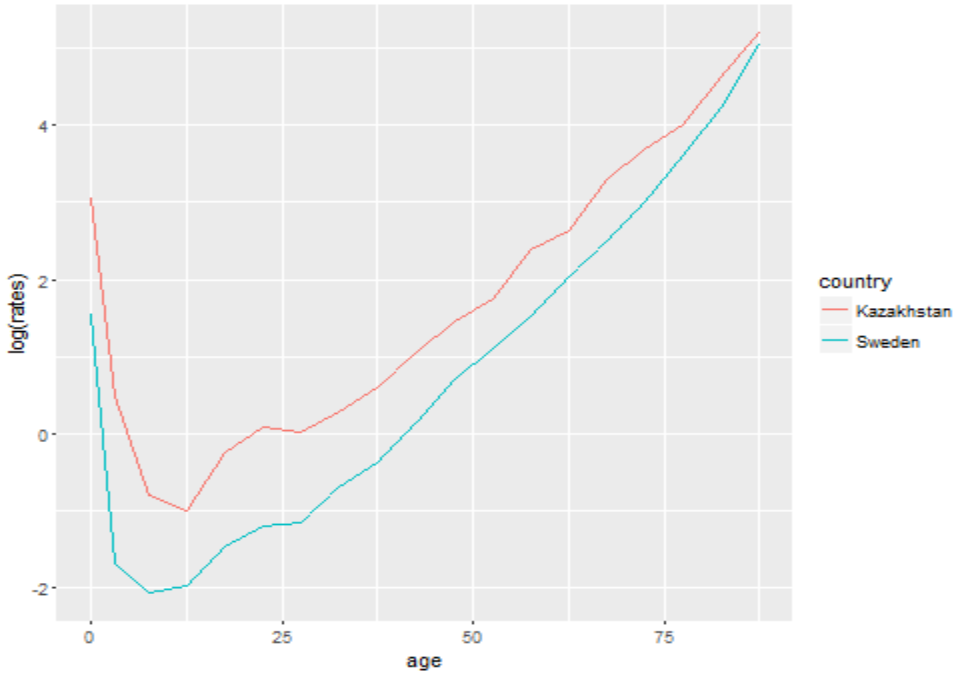
$$DSR = \sum c_i^S m_i$$

We interpret this rate as a *counterfactual*: the rate that would be observed if the population had the standard composition but its own group-specific rates. Using a uniform distribution as the standard composition yields the average rate. (A closely related measure is the total fertility rate, which has a synthetic cohort interpretation.) A good choice of standard to compare two populations is the average composition.

Example: If Kazakhstan had Sweden's age structure the CDR would be 16.34, which is 77% higher than the CDR in Sweden. Standardizing both countries using the average age structure gives rates of 11.88

and 7.37, or 61% higher mortality in Kazakhstan. The Swedish standard gives more weight to older ages where the relative difference is smaller, see the graph.

Figure 1. Age-Specific Mortality Rates



Decomposition of a Difference between Two Rates

The difference between two general rates can be due to differences in group-specific rates and differences in composition:

- The effect of the rates can be ascertained by comparing standardized rates that use a standard composition and the actual rates, say $\sum c^S m_i^A$ for population A and $\sum c^S m_i^B$ for population B, where c_i^S denotes the standard composition.
- The effect of the composition can be ascertaining by comparing summary measures that use the actual compositions and a set of standard rates, say $\sum c_i^A m_i^S$ and $\sum c_i^B m_i^S$, where m_i^S denotes the standard rates.

One difficulty with this approach is that these two components do not always add up to the original difference, which has led some authors to propose adding "interaction" terms. A much simpler approach is to use as standard the average of the two populations, taking

$$c_i^S = \frac{1}{2}(c_i^A + C_i^B) \text{ and } m_i^S = \frac{1}{2}(m_i^A + m_i^B).$$

The decomposition is then exact, and no additional terms are needed (see page 28 in the textbook). This works for proportions too!

Example: The CDR is 3.12 points *lower* in Kazakhstan than in Sweden. If both countries had the average age composition the CDR would be 4.51 *higher* in Kazakhstan. If both countries had the same rates the CDR would be 7.63 *lower* in Kazakhstan. In other words the difference of -3.12 between Kazakhstan and Sweden results from higher rates (+4.51) compensated by a younger age structure (-7.63) in Kazakhstan.

Standardized Mortality Ratio

Computation of a direct standardized rate requires knowledge of the group-specific rates. Can we adjust for compositional effects if these are not known? Yes, provided we know the composition. We can compute the rate that would be observed if the population had its own composition but a standard set of rates, and compare this to the observed rate (which results from the same composition but the actual rates).

The standardized mortality ratio is the ratio of observed to expected deaths (or death rates) under the standard:

$$SMR = \frac{R}{\sum c_i m_i^S}$$

We interpret this ratio as a *counterfactual*: proportionally how much higher (or lower) mortality would be if the group-specific rates were the same as in the standard. The same idea applies to other events. (The Princeton fertility index I_f, for example, has the same construction.)

Example: The SMR for Kazakhstan using Sweden as the standard is 7.42/4.20 = 1.77. Thus, the CDR in Kazakhstan is 77% higher than it would be if it had Sweden's age-specific mortality rates (but its own age structure).

Indirect Standardization

A closely-related approach is to approximate the direct standardized rate using a two-step procedure: we figure out the effect of composition using the observed and standard compositions with the standard rates, and then apply this as a correction to the observed rate. In symbols:

$$ISR = R\,\frac{R^S}{\sum c_i m_i^S}$$

where R^S is the general rate in the standard population.

Example: The indirect standardized mortality rate for Kazakhstan using Sweden as the standard is 7.42 x (10.55/4.20) = 18.64. (This is an approximation to the direct standardized rate of 16.34 obtained earlier.)