Calculating Fertility and Childhood Mortality Rates from Survey Data Using the `DHS.rates` R package

Extended Abstract

For the last 35 years, the Demographic and Health Surveys (DHS) has conducted more than 300 national surveys in more than 90 countries in Africa, Asia and South America. The DHS surveys are based on nationally representative samples that allow for national and regional estimates. Any standard DHS survey is designed to provide information about fertility, family planning, maternal and child health and childhood mortality levels. Most of the DHS surveys follow a two-stage sampling design, where census Enumeration Areas (EAs) are selected in the first stage as Primary Sampling Units (PSUs) from which samples of households are selected in the second stage. From the selected households, all women age 15-49 years who slept in the household the night before the survey are eligible for the individual interview. In addition to women, men age 15-49 or 15-59 might be eligible for a separate interview as well.

Among many other questions, in DHS surveys every woman is asked about her birth history, where information are captured about every birth given by the interviewed women including the deceased births. Information from the birth history, especially the date of birth, and the age of woman at the time of the survey are used to calculate key fertility indicators produced by the DHS surveys, such as the Total Fertility Rate (TFR), General Fertility Rate (GFR) and Age-Specific Fertility Rates (ASFR). Moreover, information about the deceased births of each woman are used to calculate the following childhood mortality indicators: the Neonatal Mortality Rate (NMR), the Post-Neonatal Mortality Indicator (PNMR), the Infant Mortality Indicator (IMR), the Child Mortality Indicator (CMR) and the Under-5 Mortality Indicator (U5MR).

Although these indicators are well-defined in the literature, using survey data to calculate these indicators has never been an easy task for data users. To produce these indicators, and all the DHS tables, the DHS program uses the Census and Survey Processing System (CSPro), a data processing software developed by the U.S. Census Bureau. Although the CSPro is capable enough to calculate all the indicators produced by the DHS surveys, using the CSPro as a statistical

package by data users has never been an option. Perhaps this is due to the complicated tabulation syntax the CSPro uses.

Developed as an implementation of the "S" programming language, the free software environment for statistical computing and graphics "R" has been extensively used among data analysts during the last two decades. In addition to the basic functions in R, developers are able to develop and submit their own packages and its documentation to the Comprehensive R Archive Network (CRAN), where data users can download the packages and use it in their analysis. These packages extended R capability to accommodate specialized statistical techniques, graphical presentation, import/export capabilities, data processing capabilities, and reporting tools. The `DHS.rates` R package was developed to calculate demographic indicators such as fertility and childhood mortality indicators. The same package can be used to calculate the same indicators from other surveys such as the Multiple Indicator Cluster Survey (MICS). This paper is aiming at 1) introducing the readers to the `DHS.rates` R package, 2) explaining how the package functions can be used to calculate fertility and childhood mortality rates and their precision indicators such as standard error, design effect and confidence interval, and 3) illustrating how the package work through several examples using datasets from DHS surveys and MICS surveys.