EXERCISE 4. ANALYZING CHANGE OVER TIME AND DISPARITIES BETWEEN GROUPS

IPUMS-DHS Training, Nairobi

STEP ONE. Getting the data onto your computer (after making your data extract)

Step 1. Download the data

- While you are logged in (with either your own DHS account or the workshop account)
- Go to the Home page and click "My Data Extracts"
- Right click on the STATA link under Formatted Data, next to the extract you created (the one with your name on it if you are using the workshop account)
- Choose "Save Link As …" (or "Save Target As …")
- Save into "Documents" (which should pop up as the default location)

Step 2. Decompress the data

- Find the "Documents" folder under the Start menu
- Right click on the ".dta.gz" file
- Use decompression software to unzip the compressed file (if using 7-zip, choose Extract to "idhs_0001.dta\")

Step 3. Read in the Data

- Open the "\idhs_0001.dta" folder
- Double click on the file to open it in Stata

The exercise

We're going to first look at change over time in measures of educational attainment in Kenya

Enter the following commands to produce tables of 3 measures of educational attainment for Kenya, by year:

```
tab educlvl  sample  [aw=perweight]  if educlvl  < 7, col

tab edachiever sample [aw=perweight]  if edachiever < 7, col

tab litbrig sample [aw=perweight] if litbrig <30, col
```

"tab" is short for table. When "tab" is followed by two variables it will create a crosstab of them. The "aw=perweight" command applies the proper analytic weights to the data. (Note that if you used the original DHS files, you would have to adjust the weight before you could use it. That is not necessary with IPUMS-DHS). "col" means to calculate percentages by columns (rather than rows).

Look at the documentation to find out which responses "if fccirc < 7" and "if litbrig <30" exclude. Explain: _____

_____

Now review your results. What do you learn about the spread of women's education over time in Kenya by looking at these 3 variables? _____

_____

_____

_____

Sometimes we want to look at differentials or disparities between groups, rather than changed over time. For example, perhaps women living in urban areas, in wealthier households, or in younger age groups have more access to education. Let's test that by looking at educational disparities for one year, focusing on the 2008 Kenyan DHS. Pick one of the educational variables above, and make tables to look at disparities at one point in time. For example,

```
tab edachiever urban [aw=perweight] if edachiever < 7 & year == 2008, col

tab edachiever wealthq  [aw=perweight] if edachiever <7 & year == 2008, col

tab edachiever age5year [aw=perweight] if edachiever < 7 & year == 2008, col
```

To what extent do women living in rural areas, in the 2 lowest wealth quintiles, and in the oldest age groups lag behind, as of 2008, in their educational attainment?

_____

_____

_____

_____

_____

Perhaps you want to see when a disadvantaged group of women made significant gains in educational attainment. Then you could make a table for each of the Kenyan samples and compare them to study changes in disparities over time. For example,

```
tab litbrig wealthq [aw=perweight] if litbrig <30 & year == 1989, col

tab litbrig wealthq [aw=perweight] if litbrig <30 & year == 1993, col

tab litbrig wealthq [aw=perweight] if litbrig < 30 &  year == 1998, col

tab litbrig wealthq [aw=perweight] if litbrig < 30 & year == 2003, col

tab litbrig wealthq [aw=perweight] if litbrig <30 & year == 2008, col
```

Looking at subcategories' educational attainment over time (for example, the poorest wealth quartile), what do you find?

_____