Problem Set 1, Question 1(b)
February 12, 2015

1. The general issue of weighted averages.

Suppose we have a dataset with N people, each denoted i. We observe wages, w_i. People further belong to group j, there are J groups in total, and each group has N_j people.
We can write the average wage as
$$\frac{\sum_{i=1}^{N} w_i}{N}$$
We can write the average for any group j as

$$\frac{\sum_{k=1}^{N_j} w_k}{N_j}$$

Any group j is a share of the total population, and we can write that share as N_j/N.
Given this information, we can write a weighted total of the group averages as

$$\left(\frac{N_1}{N} \frac{\sum_{k=1}^{N_1} w_k}{N_1} \right) + \left(\frac{N_2}{N} \frac{\sum_{k=1}^{N_2} w_k}{N_2} \right) + \ldots + \left(\frac{N_J}{N} \frac{\sum_{k=1}^{N_J} w_k}{N_J} \right)$$

Note that this equation simplifies to the first equation for average wage.

2. What if we start with a weighted average?
Now further suppose that we observe a weight for each person in the dataset, r_i.
Note that each person's share of the weight is $\frac{r_i}{\sum_{i=1}^{N} r_i}$. Let $s_i = N \frac{r_i}{\sum_{i=1}^{N} r_i}$.
We can therefore write the weighted average for the full sample in two equivalent ways

$$\frac{\sum_{i=1}^{N} s_i w_i}{N} = \sum_{i}^{N} \left(\frac{r_i}{\sum_{i=1}^{N} r_i} \right) w_i$$

Just as we did before, we can write the overall weighted average as a weighed sum of weighted averages.

$$\left(\frac{N_1}{N} \frac{\sum_{k=1}^{N_1} s_k w_k}{N_1} \right) + \left(\frac{N_2}{N} \frac{\sum_{k=1}^{N_2} s_k w_k}{N_2} \right) + \ldots + \left(\frac{N_J}{N} \frac{\sum_{k=1}^{N_J} s_k w_k}{N_J} \right)$$

As before, this simplifies to the overall weighted average.
This formula also points out why you had trouble doing this problem in Stata. In the weighted average equation above, the sampling weights for each group are the weights that we would use **in the full sample**. In other words, the denominator in calculating s_k is $\sum_{i=1}^{N} r_i$, not $\sum_{k=1}^{N_j} r_k$. Stata does the latter when you use the automatic weighting.
If this explanation doesn't make sense, please look at the excel sheet I posted with an

example. It uses two groups and shows how to apply weights.

3. How to implement in Stata

To get to the right answer in Stata, I made group averages using weights as above, rather than allowing Stata to weight. I then weighted the group averages by each group's share of the population (n_j/N). See details in the do file posted for this problem set.