Subject: Pooling data & DV weights
Posted by 12345Ap on Thu, 27 Jun 2019 20:31:17 GMT
View Forum Message <> Reply to Message

Hi,

I am analyzing domestic violence experiences among women in Kenya, Tanzania, and Uganda. These three countries were chosen for practical and theoretical reasons. The goal is to profile DV experiences among women in this region using latent class analysis. I will be running the latent class analysis for all three countries combined, differential item function will assess the potential differences in item responses for each of the DV items by country and if necessary paths will be fixed, and then the latent class model will be run for each of the countries separately to examine if the same or close to same profile of DV emerges in all four scenarios. If this is the case and the initial latent class model is replicated across countries, I will also be adding covariates in the merged country model to 1) assess whether the same profile of DV continues to emerge even with the addition of covariates (socio-demographics, characteristics known to be related to DV items, and country of residence to account for further differences by country), and 2) to assess whether covariates are related deferentially to the different DV types or classes. I have some questions about weights that are somewhat unclear.

1) For basic descriptive statistics and crosstabs - I should be de-normalizing the weight to account for sample size differences correct? Would it ever make sense not to do this and just to use the original DV weight recognizing that the larger country is going to pull the frequencies/descriptive stats?

2) When running the latent class model if I use the regular DV weight and then include country as a covariate, doesn't this essentially account for potential biases based on the country (and different N sizes), and negate the need to use a de-normalized DV weight?

Thanks, I'm struggling to grasp some of the issues around weights.

Subject: Re: Pooling data & DV weights
Posted by Bridgette-DHS on Mon, 01 Jul 2019 18:09:50 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

There have been many postings on pooling and weights. The term "denormalize the weights" means different things to different people, and I avoid the term. My preference would be to give equal weight to each survey. Say the sum of the weights in each of the three surveys is T1, T2, and T3. You can re-scale the weights in survey 1 by multiplying the original weights by (1/T1). Re-scale the weights in survey 2 by multiplying the original weights by (1/T2). Similarly for survey 3. If you are using Stata and pweights you shouldn't have to do anything else. You could then multiply all the weights by 1000000 but that won't change any results that use pweight.

If you don't adjust the weights this way then the results will be pulled to the largest survey, not the

largest country. To adjust to the population size, you would have to go through the preceding step and then multiply the weights for country 1 by P1, where P1 is the population (of the relevant type of respondent) in country 1, multiply the weights for country 2 by P2, where P2 is the population (of the relevant type of respondent) in country 2, and similarly for country 3. It may not be easy to determine P1, P2, and P3 at the time of the survey. When doing this, you may find that the results are too dominated by the largest country and hardly depend at all on the smallest country.

Leaving the weight alone and using country as a covariate will only adjust the intercept. It is likely that all of the associations are different across the three countries. I suggest that you try this; I believe you will see that all the results are still sensitive to changing or not changing the weights.