SM 625: Week 6 Sampling Project Notes

Recall that the client and the data collection organization estimated that the data collection would cost $3,000 per primary stage cluster (school), and $50 per completed questionnaire within a cluster. We will now use this information for optimum subsample size calculations. Recall that the total budget for data collection will be $500,000.

Given this cost information and your estimates of roh for the three different variables of primary interest from last week, compute the **optimum subsample size** (and the corresponding optimal number of first stage clusters, given the total budget above) for each of the variables. How will you decide on a single overall optimum subsample size to use in your design? Think about a comparison of alternative cluster sample designs: under a fixed cost constraint, how would we decide which design would be best? What will be your overall sample size (n) under this new optimum subsample size? As you make progress in writing up what you have done so far, provide some discussion of the rationale for your choices in this regard.

Next, given this optimum subsample size and treating the values of roh as **portable**, compute the new expected DEFF for each estimate given the new design (this can be specific to each variable / estimate, given the different optimum subsample sizes). In addition, compute a new expected SRS variance for each variable under the new design, using the new "optimum" overall sample size (remember that you can treat the element variances for each variable estimated last week as **portable**). Finally, compute the new expected sampling variance for each estimate under this new cluster sample design. Are you still meeting the client's precision requirements?

The client has also provided other new information: the estimated size of the target population is N = 830,138. Given this population size and your overall sample size (n) under the new optimum subsample size computed above, what is your overall working sampling fraction (f)? Does it seem like finite population corrections will be necessary in your sampling variances if you choose to perform SRSWOR at some point?

The tables that you are developing and the text that accompanies them should carefully reflect the answers to all of the questions above.