Hi Lucas,

Hope you had a good weekend and a nice visit with your mother today.

While the attached is not complete, I wanted to send you what I have so far for UNRWA.  In the attached you have all the sampling information (note the large sample frame file contains frames for each of the 5 sites separately as well as combined: the sites’ samples were drawn independently, so you don’t need the combined one), the samples drawn for each site, the log-in forms for each site, datasets for 4 of the 5 sites plus the final questionnaire.

I’m still pulling together the last dataset and Laura promised she would get to the annotated questionnaire file for you today.  (She is on leave from tomorrow and promised to do it before going on leave.)

One thing I realized I forgot to mention to you, since you are going back to the original sampling frames, is that we habitually remove schools with enrollments under 40 from the sampling frames prior to selecting the samples.  Thus, if you are going to use the original sampling frames to calculate the probability of selection with PPS, you would need to edit these frames in the same way.   (I wouldn’t code this, as there are rare exceptions, but just remove the small schools from the frame used as input to the R code.) You can see the edited frames in each of the final sample files for each site, but I know those versions of the frames only have the total enrollment, not the enrollment by grade and sex which you need for post-adj.

Two things to note in particular to UNRWA:

1. All the sites’ samples were stratified by type of school – you’ll see on the “sample distribution” tab in the large sample frame file how this was done.  It’s usually 3 strata (“male”, “female” and “coed”), but UNRWA Gaza has just male and female schools.
2. This site was a “double draw” for GSHS plus another survey on tobacco (Global Youth Tobacco Survey – GYTS).  PCSample allows for you to do this and you can choose to pick double the number of classes per school or double the number of schools.  Usually (in all the cases I have seen), we go with the former.  This is particularly important for the weighting as the **school interval** reflects this double draw and is therefore **half**of what it really would be for just the GSHS.  To adjust for this double draw, **the school interval needs to be doubled** when doing the GSHS weighting calculation.  I would not suggest coding this, but simply making the adjustment in the input file.

I think we discussed talking around 11 tomorrow – does that still work for you?

Have a good evening,

Melanie