QUESTION 1: What is the difference between planned and unplanned domains?

Assume that I want to evaluate a given phenomenon in Tuscany. I must select a sample size (say 2000 units for the whole region) and then I can compute the parameters of interest for the whole region by using standard efficient direct estimators. Tuscany is the planned domain. In other words, survey is planned only to produce estimators of a good level of precision for the whole region. In this case an unplanned domain is represented by a municipality, for instance Prato. For the unplanned domain the direct estimators may be affected by a high sample variability and SAE methods should be used to obtain “statistically sound estimates”.

QUESTION 2: HT estimator can be applied when the domain is planned for the estimation at the stage of design of survey, and if not planned, SAE is more often used. In reality, in what situation the institution (official, private) need the estimation of the unplanned domain?  Is there any example that the institution suddenly needed to focus on specific domain?

EUSilc survey is an example: it is planned in each Member State to provide reliable estimates at NUTS2 level (planned domains) and not provide surely credible estimates at NUTS3 level. This could be possible (oversampling, planning the domain in advance) but it is not operational because it is too costly. This is why the Eurostat is not planning it in advance.

*QUESTION 3: When using HT estimation, is it necessary to know the target population size Nd in advance? At first sight I thought it is necessary for calculation of πk (and weight), but seeing the example of stratification in Happy land, πk can be calculated by πk= πih· πk|ih, so I am a bit confused.*

No, it is not necessary in general.

It is necessary when the sampling design *p(s)* is simple random sampling without replacement from the domain. In this case you need *Nd* to compute *πk=nd/Nd* as it is equal to the sampling fraction.

Under other sampling designs it may not be needed. For instance, you can select units by probability proportional to a Measure Of Size (MoS) of the unit. In Agricultural Surveys you can select farms in a domain (k=1…Nd) using probability proportional to the quota of their cultivated area (xk) to the total cultivated surface area in the domain (Xd). In this case *πk=xk/Xd* . You do not need to know *Nd*.

Coming back to the Happy land example, the domain is equal to the stratum. In that case you select a simple random sample of villages (1st stage) and then a simple random sample of households (2nd stage). To compute *πk= πih· πk|ih,* you need to know the number of villages in each stratum (domain) and then the number of households in each village. Indeed, you know *Nd*also in this case: the stratum size (domain size) is *Nd=Nh=50x200=10000.*

*QUESTION 4:* Concerning the answer to the first question provided in the previous word document, I encountered some difficulties to understand how to index the two different probabilities of inclusion. As far as I know, the slides supposed that i represents villages, k represents households, and h is stratum. Therefore, this lead me to an ambiguity to understand how the probability of inclusion at the second stage is indexed πk|ih=nih/Nih. πk|ih is without any confusion, but nih and Nih are indexed with i, while, they are presenting the numbers of households.

In the notation nih and Nih, the pedix *ih* indicates the village in the h-th stratum. So nih is the number of selected households in the i-th village of the h-th stratum