GNSUM Meeting notes- 20-10-23

Step 5 Equation 23

Answer to 3 – total number of people in the frame

Divided by average degree of everyone in F

Calculated total degree of everyone in F divided by F (number in the sampling frame)

Probe alter groups – a better way is to use know group population size.

Two quantities in the denominator

The total number of nodes (people who know people)

Number of domestic workers in the UK

Difficult to estimate.

Can use average to estimate known size.

Equation 23/24 need

As noted, in appendix B4 in on-line appendix.

Denominator -average number of connections within the frame –

Take totals, or do the scale up within a known population

Line 343

First divide answer 4 by answer 1, then use that as the denominator – put brackets round this

Caveat: for this to be consistent and essentially unbiased, this number is an estimate of the sample frame,

This is an average, and can be calculated using an estimate of the known population.

For each person we known who has been through the NRM and

If subset -assumption that filipino respondents only know other filipino NRM

Requires us to know the probability of inclusion. Unless it’s a random sample

Perhaps used the wrong inclusion probabilities. So estimates created.

We don’t know the inclusion possibilities.

Work around. If we didn’t use the referrals, could we plausible claim we have a simple random sample. Consider the inclusion possibilities as equal

Stratified samples over-sample minorities or choose particular regions. Sampling design gives rise to the possibility that any one person is observed. If equal, OK, but if deliberate oversampling every person in the UK does not have an equal chance of being in the sample. The probability of being in the sample is different. Probability is unequal. Called survey weights, inclusion weights. We are missing this.

Social networks used rather than attempting probalistic random sampling.

This matters for uncertainty estimates. – however no uncertainty estimates for bootstrapping

Step 6

How many people are exploited in the overall population

Sample A 5,382 are being exploited in domestic work

Sample B 2,623 Filipinos are potentially exploitated.

What is required to believe these numbers is that the sample is representative of F.

Some flawed assumptions

Reasons why overseas domestic workers visas proves a representation of the overall population is a questionable assumption.[Need to describe this as a limitation of the estimate’

In relation to sub-sample, Homophily immigrants tend to cluster – form intra-national social groups.

For the scale up:

In the survey – sum of how many people do you know who have gone through the NRM.

(using sum of answers to question 83 5f15, divided by the total of NRM referrals which is 73,019)

Sensitivity Neighbourhood bootstrap

Resample referrals in secondary and tertiary waves

Package sent but Scott can do this manually. Take a sample with replacement until you get a sample the same size as your original. Repeat 1000 times and then take the variance

Scott needs a function that will calculate the point estimates each time. i.e. a function that will do steps 1..5

Scott will provide the data set and the question that will indicate exploitation (the hidden population) and the size of the known population size of the NRM.

Line 343

Line 348

Selim give back 2 numbers one from sample a and one from sample b

Bootstrapping takes the randomness of the sample, original sample created by chance. NB recreates a sample that we didn’t see. On the basis of this sample, repeated 10,000 times look at empirical distribution and gives us a 95% confidence interval. Ranges don’t have to be symmetrical. No

Selim will automate steps 1- 5

Scott will think about probe alters and think about NB

For sample b, assumption that filipino workers would only know other filipino workers

Selim clarified that Node 1 is the recruiter, Node 2 is who they recruited

All the blanks are seeds, the non-blanks are those that have been referred