

Square GPS Sampling Algorithm

27 December 2016

Patrick D. Emond

Algorithm

Once a town has been selected for sampling the following algorithm is performed on that town:

1. determine width of squares based on user input “number of squares per side”*
2. select random point in town and identify square, S , the point falls into
3. make a set of all unselected buildings, B , which fall in square S
4. if B is empty then go back to step 2 (if this happens 1000 times then quit with an error)
5. select a random building, b , in set B
6. add all individuals** in all households in building b to the sample
7. if number of individuals in the sample is less than the sample size return to step 2

* “number of squares per side” is a parameter defined by the user when running the sampler and corresponds to how many squares there are along any one axis for all towns (eg: 10 would mean that every town is divided into $10 \times 10 = 100$ squares total since there are always exactly the same number of squares along the X and Y axes)

** selected individuals may be restricted to children only, or one individual per household based on input parameters defined by the user

Sampling Weights

Probability of selecting town i is:

$$P_i = P(\text{select town } i) = \frac{n_i}{N}$$

where n_i = population of town i , and N = total population. Strictly, it's number of buildings or households, but we ignore that, and assume the number of people per household is constant.

Probability of selecting square j is:

$$P_{j \square} \bigvee_i = P(\text{select square } j \text{ within town } i) = \frac{n_s}{S}$$

where n_s = number of squares chosen within the town, and S = number of squares in the town. The number of squares chosen will vary depending on the size of squares and population density.

Probability of selecting household k in square j is:

$$P_{k \square} \bigvee_j \square = P(\text{select household } k \text{ within square } j) = \frac{1}{n_{ij \square}}$$

where $n_{ij \square}$ = number of households in square j .

Therefore the probability of sampling household k is:

$$P_{ijk \square} = P(\text{select household } ijk) = P_i \times P_{j \square} \bigvee_i \times P_{k \square} \bigvee_j \square$$

Making the weight for household $i\overset{\square}{j}k$ equal to:

$$\begin{aligned}
w_{i\overset{\square}{j}k} &= \left[P_{i\overset{\square}{j}k} \right]^{-1} \\
&= \left[P_i \times P_{\overset{\square}{j}} \times P_k \right]^{-1} \\
&= \frac{n_{i\overset{\square}{j}} N S}{n_i n_s}
\end{aligned}$$

where $n_{i\overset{\square}{j}}$ = number of households in square $\overset{\square}{j}$, n_i = population of town i, n_s = number of squares chosen within the town, S = number of squares in the town and N = total population.