SURVMETH / SURV 625
Applied Sampling
Winter / Spring 2025

Homework 5

1. The following is the distribution of occupied housing units and blocks for two counties in Michigan. A stratified PPeS sample of $n = 3000$ housing units in $a = 100$ blocks with minimum sufficient size b^* per block is to be selected across the counties. The counties and the subdivisions within each county of river and non-river blocks serve as four (4) strata.

   a) Allocate the sample blocks across the four strata <u>proportionate to occupied housing units</u>. How have you handled rounding of sample blocks?

   b) Give the overall sampling fraction f across the four strata.

   c) Maintaining *epsem*, what value of b^* should be used in each of the four strata?

State	County	Type	Occupied HUs	Blocks
Allegan				
26	005	All	37,265	3,593
26	005	River	8,521	940
26	005	River %	22.9%	26.2%
Kalamazoo				
26	077	All	94,523	4,494
26	077	River	17,224	1,318
26	077	River %	18.2%	29.3%
All				
26	All	All	131,788	8,087
26	All	River	25,745	2,258
26	All	River %	19.5%	27.9%

2. Download the Excel file provided with this homework (a geographically sorted list of blocks).

   a) Select a systematic PPeS sample of $a = 20$ blocks from the list of $A = 4{,}495$ in the Excel file. <u>The minimum sufficient size for a block is 50 occupied housing units</u>. Specify the sampling interval, and the random start. List selected blocks, the selection number (the random start plus interval), and the total number of occupied housing units in the sample blocks. Document how the random start was determined, in a way that would allow someone to replicate your selection process exactly.

   b) Do you have oversized units? If any, describe how you have handled oversized units.

c) Suppose the following are the first five blocks selected.

ID	STATE	COUNTY	TRACT	BLKGRP	BLOCK
142	26	077	000201	1	1064
349	26	077	000500	2	2002
452	26	077	000600	4	4001
655	26	077	001000	6	6013
789	26	077	001300	1	1016

i. Implement the 'linking after selection' procedure described in the lecture notes for each of these five selections, and describe the five resulting linked blocks that would be included in the sample.

ii. The overall sampling fraction is $f = 0.001$. What sampling rates should be used within each of the five sets of linked blocks to achieve *epsem* across the two stages (blocks, then occupied housing units)? (*Hint:* Remember that the housing units at the second stage will be randomly selected from *all* units in the set of linked blocks, not just the units in the five original selections above.)

iii. For the first "linked" selection, suppose that field canvassing procedures yield $B_1 = 90$. What two possible values might you see for the achieved subsample size from the first selection?