Outline: Finding the Optimal Sample Allocation for the Mean of a Stratified Random Sample

1) Calculate the sample variance for each of our four strata – i.e., for each h in {1,2,3,4}, calculate s\_h^2 = (1/(n\_h – 1)\*sum((y\_i – y^bar\_h)^2, from i = 1 to n\_h), where n\_h here refers to the number of units from stratum h contained in the pilot sample (which we will replace with the “adjusted” n\_h)

2) Use s\_h^2 as an estimate for S\_h^2 in the algorithm

3) Let the initial allocation be (n\_11, n\_21, n\_31, n\_41) = (1, 1, 1, 1)

4) Calculate enough (perhaps 10 per stratum) “priority values” to ensure that we can complete the algorithm: (N\_1^2)\*(s\_1^2) / (1\*2\*N^2), (N\_1^2)\*(s\_1^2) / (2\*3\*N^2),…, (N\_1^2)\*(s\_1^2) / (10\*11\*N^2), (N\_2^2)\*(s\_2^2) / (1\*2\*N^2), (N\_2^2)\*(s\_2^2) / (2\*3\*N^2),…, (N\_4^2)\*(s\_4^2) / (10\*11\*N^2), where N denotes the size of the entire population (which we will replace with the “adjusted” population size); we should keep these values in squared form, so that we can simply subtract them from the updated variance as the algorithm proceeds

5) Create a vector of the priority values in decreasing order; let (1) denote the largest value, (2) denote the second largest value, etc.

6) Compute the initial variance: V(y^bar\_str | n\_11=1, n\_21=1, n\_31=1, n\_41=1) = (1 / N^2) \*sum((N\_h – 1)\*N\_h\*s\_h^2, from h = 1 to 4)

7) Take priority value (1) and increase the associated stratum’s sample size by 1 (e.g., if value (1) is (N\_2^2)\*(s\_2^2) / (1\*2\*N^2), then we would let n\_22 = 2 replace n\_21 = 1); compute the updated variance V(y^bar\_str | n12, n22, n32, n42) = V(y^bar\_str | n\_11, n\_21, n\_31, n\_41) – priority value (1)

8) If V(y^bar\_str | n12, n22, n32, n42) <= V\_0 (the desired variance), then the optimal allocation is (n\_12, n\_22, n\_32, n\_42); otherwise, go to step 9

9) Take priority value (2) and increase the associated stratum’s sample size by 1; compute the updated variance V(y^bar\_str | n13, n23, n33, n43) = V(y^bar\_str | n\_12, n\_22, n\_32, n\_42) – priority value (2)

10) If V(y^bar\_str | n13, n23, n33, n43) <= V\_0, then the optimal allocation is (n13, n23, n33, n43); otherwise, continue in a similar fashion (subtracting the largest of the unused priority values from the updated variance) until step “j,” where V\_0 is attained and the optimal allocation is (n1j, n2j, n3j, n4j)