An Example of Sampling Distributions: A Comparison of Simple Random Sampling, Stratified Random Sampling, Systematic Sampling, and Cluster Sampling (SRS, STS, SYS, and CLS)

Consider again the Population (N=6): 2 6 8 10 10 12

[Since we know the entire population, we know and]

I Simple Random Sampling (SRS)

The 15 possible simple random samples (SRSs) of size 2 with their means () are as follows:

Sample: 2,6 / 2,8 / 2,10 / 2,10 / 2,12 / 6,8 / 6,10 / 6,10 / 6,12 / 8,10 / 8,10/ 8,12 / 10,10 /10,12 / 10,12

Mean (): 4 / 5 / 6 / 6 / 7 / 7 / 8 / 8 / 9 / 9 / 9 / 10 / 10 / 11 / 11

Thus, the SRS sampling distribution for n = 2 is as follows: Frq [Prob]

4 1 [1/15]

5 1 [1/15]

6 2 [2/15]

7 2 [2/15]

8 2 [2/15]

9 3 [3/15]

10 2 [2/15]

11 2 [2/15]

The estimator (statistic) is unbiased since E() = = 8, and the sampling variance is Var() = = 64/15 [and so the (exact) standard error of the mean is = 2.07].

II Stratified Random Sampling (STS)

Suppose the Population were divided into the two strata: Stratum I: 2 6 8

Stratum II: 10 10 12

(Continued)

An Example of Sampling Distributions: STS (Continuation) Page Two

Now for a sample of size 2, choose one at random from each stratum. [Note: If we were to choose both from the same stratum, then such would illustrate *selection bias*.] Thus, the 9 possible STSs of size 2 with their means are as follows:

Sample: 2,10 / 2,10 / 2,12 / 6,10 / 6, 10 / 6,12 / 8,10 / 8,10 / 8,12

Mean (): 6 / 6 / 7 / 8 / 8 / 9 / 9 / 9 / 10

Thus, this STS sampling distribution is as follows: Frq

6 2 [2/9]

7 1 [1/9]

8 2 [2/9]

9 3 [3/9]

10 1 [1/9]

Since E() = 8, the estimator is unbiased, and the sampling variance can be found to be Var() =

16/9 [and so the (exact) standard error of the mean is = 4/3 1.33].

III Systematic Sampling (SYS)

There are N/n = 6/2 = 3 possible systematic samples of size n=2 here. Using the Population as ordered [2-6-8-10-10-12], they are first-fourth (2,10), second-fifth (6,10), and third-sixth (8,12).

Thus, these three samples and their means are as follows:

Sample: 2,10 / 6,10 / 8, 12

Mean (: 6 8 10

Thus, the SYS sampling distribution is as follows: Frq

6 1 [1/3]

8 1 [1/3]

10 1 [1/3]

(Continued)

An Example of Sampling Distributions: SYS (Continuation) Page Three

Again the estimator is unbiased, and its sampling variance is Var() = 8/3 [and so the (exact) standard error of the mean is 1.63].

IV Cluster Sampling (CLS)

Suppose the Population is made up of three clusters of two elements each with such made as ordered, making the clustering as follows:

Cluster I: 2 6

Cluster II: 8 10

Cluster III: 10 12

Here the sampling unit is a cluster, and so to choose a sample of size 2, we select one cluster at random. Thus, the three possible such samples of size 2 here with their means are as follows:

Sample: 2,6 / 8,10 / 10,12

Mean (): 4 / 9 / 11

Thus, the CLS sampling distribution is as follows: Frq

4 1 [1/3]

9 1 [1/3]

11 1 [1/3]

Again the estimator is unbiased, and its sampling variance is Var() = 26/3 [and so the (exact) standard error of the mean is 2.94].

Now let’s compare the four sampling methods with one another.

(Continued)

An Example of Sampling Distributions (Continuation) Page Four

Summary and Comparison

We summarize the results here:

Method Sample Size Standard Error Range

SRS 2 2.07 7

STS 2 1.33 4

SYS 2 1.63 4

CLS 2 2.94 7

As usually is the case among these four sampling schemes done separately on the same population, first stratifying the population and then sampling within each stratum will be most efficient in estimating the overall population mean, especially when the stratum averages are as different as possible while the variances within each stratum are as small as possible.

On the other hand, if we do likewise in forming clusters and sample entire clusters, then such will turn out least efficient, as above (with largest standard error of the mean among the four methods) But as clusters are typically formed according to the propinquity of members of a population, such often leads to clusters each of which do not represent the overall population well, and, as above, such will be frequently less efficient than simple random sampling (notice 2.94>>2.07 above).

Nevertheless, there is something we could do about such a state of affairs, and that is to use a *second stage* of cluster sampling, that is, we should *subsample*. Let’s illustrate this with the clustering above, which was

Cluster I: 2 6

Cluster II: 8 10

Cluster III: 10 12

Now to take a sample of size 2, first select two clusters randomly, which would be the *first stage* of sampling. For the clustering above, there are three possibilities, as follows:

I / II // I / III // II / III or 2, 6 / 8, 10 // 2, 6 / 10, 12 // 8, 10 / 10, 12

(Continued)

An Example of Sampling Distributions (Continuation) Page Five

Now the *second stage* of such sampling would be to subsample each cluster, which here would mean to select one element at random from each of the two previously selected clusters, thus giving us our desired sample size of 2. Thus, such would give the following, along with the corresponding sample means:

Selected Clusters Possible Stage- Corresponding

At Stage One Elements Two Samples Sample Means

I and II 2, 6 // 8, 10 2,8/2,10/6,8/6,10 5 / 6 / 7 / 8

I and III 2, 6 // 10, 12 2,10/2,12/6,10/6,12 6 / 7 / 8 / 9

II and III 8, 10 // 10, 12 8,10/8,12/10,10/10,12 9 / 10 / 10/ 11

Thus, the Two-Stage CLS sampling distribution is as follows: Frq

5 1 [1/12]

6 2 [2/12]

7 2 [2/12]

8 2 [2/12]

9 2 [2/12]

10 2 [2/12]

11 1 [1/12]

The estimator is unbiased [E() = 8 = ], and its sampling variance is now Var() = 38/12 = 19/6, and so its standard error is = 1.78, compared with the 2.94 from the poor one-stage cluster sampling above. Thus, two-stage sampling was much more efficient than single-stage sampling here, as it sort of “undid” the poor clustering design. In addition, it also was more efficient than simple random sampling (1.78 < 2.07).

An Example of Sampling Distributions (Continuation) Page Six

Problems for Practice

1. Consider this stratification: Stratum I: 2 6 12

Stratum II: 8 10 10

Suppose we again want an STS of size 2 from it. How would such compare with the previous

SRS, STS, SYS, and CLS we illustrated above?

2. Suppose the Population were ordered as follows: 2 8 10 6 10 12

Again we want an SYS of size 2 from it. How would such compare with the previous SRS, STS,

SYS, and CLS?

3. Consider this clustering: Cluster I: 2 6 8

Cluster II: 10 10 12

Again we want a CLS of size 2. If we select one cluster at random and then subsample two

elements from it, then would such be an improvement over the previous single-stage cluster

sampling scheme?

Problems to Ponder

4. For single-stage cluster sampling of two of the six Population elements, construct a clustering

with three equal-sized clusters that is more efficient than the SRS of size n=2. Also compare your

clustering with the STS and SYS.

5. Referring to the Example, show that systematic sampling (SYS) can be considered a special case

of single-stage cluster sampling (CLS). [*Hint*: Look back at Practice Problem 2.]

6. Could stratified sampling also be considered a special case of cluster sampling? What about SRS?