

Appendix A.3 Part 2 (Comparison of Sampling Strategies)

This worksheet accompanies the article:

Richardson, M., and Gajewski, B. (2003) "Archaeological Sampling Strategies," *Journal of Statistics Education* [Online], 11(1).
<http://www.amstat.org/publications/jse/v11n1/richardson.html>

Comparison 1:

X's have been placed in test-pits below to illustrate the layout of finds in an archaeological site with 20 test-pits containing artifacts. [It is assumed that the site is divided into 2 strata (using columns 1 through 5 of test-pits for stratum I and columns 6 through 10 of test-pits for stratum II).] Assume that the site is an urban site that contains a clearly visible structure in stratum I and no visible structure in stratum II. According to Orton (2000), many urban sites contain partial structural remains, especially if they have been deserted and not re-occupied or built over. For urban sites, stratification may be more useful and more feasible than in other situations. A site may be divisible into zones (e.g. religious, industrial, domestic), which can be demarcated as statistical strata and sampled from at different rates according to the nature of the research questions. An example of such a site is a site found in Shoofly Village in Payson, Arizona. Part of this site contained an enclosure wall and part of this site contained no visible surface artifacts. Archaeologists sampled the area inside of the enclosure wall at four times the intensity of the area immediately outside of the enclosure wall (pictures of part of the Shoofly Village excavation site can be viewed at: <http://www.geocities.com/dlcobb/shoofly.html>).

We want to use class data to determine if, for an archaeological site with the above layout, ***repeated stratified random sampling*** of 20 test-pits (**16 test-pits from stratum I and 4 test-pits from stratum II**) will result in estimated total numbers of finds that are less variable than the estimated totals resulting from ***repeated simple random sampling*** (20 test-pits sampled).

1. Do you think that *repeated* stratified random sampling of test-pits from the above site will be likely to produce less variable estimates of the total number of finds at the site than will *repeated* simple random sampling of test-pits? Why? Or, why not?
2. Using ANY starting point on the random number table, select a **simple random sample** of 20 of the test-pits from the above site.

What is the estimated total number of finds at the site?

Write your estimated total number of finds for your simple random sample on one of your sticky notes and place your sticky note in the appropriate position on the frequency plot on the white board labeled “Simple Random Sample Estimated Totals.”

3. Using ANY starting point(s) on the random number table, select a **stratified random sample** of 20 of the test-pits from the above site (**sample 16 test-pits from stratum I and 4 test-pits from stratum II**).

How can the number of finds out of 20 sampled test-pits be used to estimate the total number of finds at the site? (**Hint:** You sampled 16/50 of the test-pits from stratum I.)

What is the estimated total number of finds at the site?

Write your estimated total number of finds for your stratified random sample on one of your sticky notes and place your sticky note in the appropriate position on the frequency plot on the white board labeled “Stratified Random Sample Estimated Totals.”

4. Record the class estimated totals for each of the sampling techniques below.

stratified random sampling estimated totals:

simple random sampling estimated totals:

5. Calculate descriptive statistics for the class estimated totals.

Stratified Random Sampling

mean =

standard deviation =

first quartile =

median =

third quartile =

Simple Random Sampling

mean =

standard deviation =

first quartile =

median =

third quartile =

6. Construct side-by-side box plots for the class estimated totals.

stratified random sampling



simple random sampling



7. Based on the above calculations, do you think that *repeated* stratified random sampling of test-pits from this site would most likely produce a more accurate estimate of the total number of artifact finds at the site than would repeated simple random sampling of test-pits? Why? Or, why not?

Comparison 2:

X's have been placed in test-pits below to illustrate the layout of finds in an archaeological site with 20 test-pits containing artifacts.

X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X

For the above site, do you think that *repeated 1-in-5 systematic random sampling* of test-pits would most likely produce a more accurate estimate of the total number of artifact finds at the site than would *repeated simple random sampling* of test-pits? Why? Or, Why not? [For the systematic sampling, assume that we are using the top row as ordered test-pit numbers 1 through 10, ..., so that the bottom row represents ordered test-pit numbers 91 through 100.]

Comparison 3:

Place X's in the appropriate excavation units below to illustrate the layout of an archaeological site with 20 test-pits containing artifacts for which *repeated cluster random sampling* of test-pits would most likely produce a more accurate estimate of the total number of artifact finds than would *repeated simple random sampling* of test-pits. [Use rows of test-pits as clusters and assume that two clusters will be randomly selected.]

Hint: With our definition of cluster sampling for this site, we know that two rows of test-pits will be selected. Think of a layout that would guarantee four finds using cluster sampling.
