

Sampling in Archaeology

Mary Richardson
Grand Valley State University
richamar@gvsu.edu



Published: March 2012

Overview of Lesson

This activity allows students to practice taking simple random samples, stratified random samples, systematic random samples, and cluster random samples in an archaeological setting. Additionally, students compare the performance of simple random sampling and stratified random sampling within the context of a specific archaeological problem.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. The four components are: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level C activity.

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.
6. Attend to precision.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

Common Core State Standards Grade Level Content (High School)

S-ID. 1. Represent data with plots on the real number line (dot plots, histograms, and box plots).

S-ID. 2. Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

S-ID. 3. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

S-IC. 1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

S-IC. 3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.

S-IC. 4. Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

S-IC. 5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability Standards for Grades 9-12

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them:

- understand the differences among various kinds of studies and which types of inferences can legitimately be drawn from each;
- understand histograms and parallel box plots and use them to display data;
- compute basic statistics and understand the distinction between a statistic and a parameter.

Select and use appropriate statistical methods to analyze data:

- for univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics;
- display and discuss bivariate data where at least one variable is categorical.

Develop and evaluate inferences and predictions that are based on data:

- use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions;
- understand how sample statistics reflect the values of population parameters and use sampling distributions as the basis for informal inference.

Understand and apply basic concepts of probability:

- use simulations to construct empirical probability distributions.

Prerequisites

Prior to completing this activity students have been exposed to definitions and terms related to sampling and they have seen basic examples of simple random sampling, stratified random sampling, systematic random sampling, and cluster random sampling.

Learning Targets

Students will be able to select a simple random sample, stratified random sample, systematic random sample, and cluster random sample. Students will be able to calculate numerical summaries and boxplots and use them to compare two data distributions.

Time Required

1 to 2 class periods.

Materials Required

A copy of the Activity Sheet; a graphing calculator; two sticky notes per student.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

I. Formulate Question(s)

To start the activity the teacher may wish to provide students with background on the use of sampling in an archaeological setting. According to Orton (2000), the term *site* has many meanings. Orton states that for a *development site*, the goal is to detect the presence and extent of any significant archaeological remains, and to either record them before damage or destruction, or to mitigate the damage by redesign of the proposed development. For an *archaeological site*, the goal may be to determine the extent and character of a site, or there may be a more site-specific research design.

According to Lizee and Plunkett (1994), one of the challenges that an archaeologist faces after the discovery of an excavation site is how to determine the locations within the excavation site that will be dug in order to uncover artifacts. Obviously, digging everywhere within a site would be the maximal way to locate artifacts, but usually, time and resources do not allow for the total excavation of a site. Archaeologists must develop cost and time-efficient strategies for digging.

Prior to excavation, a site must be divided into sampling units (excavation units). Typically, a site is either sampled in a *purposive* way, in that the digging is targeted on possible features (which have already been identified), or in a *probabilistic* way, if little is known in advance about the site. For probabilistic sampling, the choice of excavation units is usually either 2 m-wide machine-dug trenches, often 30 m long, or hand-dug test-pits, usually 1 m or 2 m square.

Explain to students that in order to use an archaeological setting to demonstrate the use of statistical sampling, it will be assumed that a site will be sampled probabilistically. Further, it will be assumed that the excavation units are test-pits.

II. Design and Implement a Plan to Collect the Data

Prior to data collection the teacher may wish to review the definitions of the four sampling techniques that will be used in this activity. **Simple random sampling** is the foundation for all of the sampling techniques. Simple random sampling is such that each possible sample of size n units has an equal chance of being selected. **Systematic random sampling** requires the user to order the population units in some fashion, randomly select one unit from among the first k ordered units, and then select subsequent units by taking every k^{th} ordered unit. **Stratified random sampling** is simply forming subgroups of the population units and selecting a simple random sample of units from within each subgroup. **Cluster random sampling** also requires the sampling units to be placed into subgroups. A simple random sample of the *subgroups* is then taken, and *every* unit within the selected subgroup is a part of the sample.

After a review of the sampling techniques each student is given a copy of the Activity Sheet. The problem is formalized as follows. Since it is both time and labor intensive to excavate an entire site, a sampling strategy must be developed. A site contains 100 8×8 meter excavation units (test-pits) and there is only enough time to dig in 20 of the test-pits. A map of the site is shown in Figure 1 below (with each square representing an excavation unit and an X representing a test-pit that *contains* artifacts or “finds”). On this map, finds were randomly assigned to 20 of the test-pits.

SITE 1

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Figure 1. Initial map of an archaeological site.

Each student is to use each of the four sampling strategies to select a sample of $n = 20$ test-pits from Site 1. Explain to students that the goal is to use the sample of 20 test-pits to estimate the *total* number of test-pits containing finds. After students have selected their samples they are asked to explain how to use the sample number of test-pits containing finds to estimate the total number of finds at the site. Since 1/5 of the site’s test-pits are being sampled, five times the number of finds out of the 20 sampled test-pits serves as an estimate of the total number of finds.

The motivation behind estimating the total number of finds at an archaeological site is that, if the estimated total number of finds at a site is above some predetermined threshold value, then spending more time and money to dig in more than 20 test-pits at the site might be justified. For each of the sampling techniques students are asked to use a uniform SEED for random number generation on the TI 84 calculator in order to have a classroom discussion of the results of selecting the different samples. Simple random sampling is performed using a SEED of 2000. To perform stratified random sampling, the site is divided into two equally sized strata containing 50 test-pits each (using column 1 through column 5 of test-pits for stratum I and column 6 through column 10 of test-pits for stratum II) and a SEED of 1981 is used. Ten test-pits are selected from each stratum. In order to obtain a sample of 20 test-pits, 1-in-5 systematic

sampling is used with a SEED of 2003. Cluster sampling is performed using the rows of test-pits for clusters and a SEED of 2004.

After students have had a chance to perform each of the four sampling techniques, have a summary discussion. Remind the students about the specific details of the various sampling techniques and compare and contrast the techniques.

Now give each student two sticky notes. Additionally; provide a new map of an archaeological site (see Figure 2 below).

SITE 2

Stratum I					Stratum II				
1	2	3	4	5	6	7	8	9	10
11	12	13X	14	15	16	17	18	19	20
21	22X	23X	24X	25	26	27	28	29	30
31	32X	33X	34X	35	36	37	38	39	40
41	42X	43X	44X	45	46	47	48	49	50
51	52X	53X	54X	55	56	57	58	59	60
61	62X	63X	64X	65	66	67	68	69	70
71	72X	73X	74X	75	76	77	78	79	80
81	82	83X	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Figure 2. Site layout for comparing simple random sampling to stratified random sampling.

On Site 2, an X has been placed in the appropriate test-pits in order to illustrate the layout of a site for which *repeated* stratified random sampling of 20 test-pits would most likely produce a less variable estimate of the total number of artifact finds at the site than would *repeated* simple random sampling of 20 test-pits. Once again, column 1 through column 5 of test-pits make up stratum I and column 6 through column 10 of test-pits make up stratum II.

For stratified sampling from Site 2, *do not use equal sample sizes from the two strata*. The motivation for sampling from the strata at different rates is based on an attempt to realistically illustrate the use of stratification in archaeological sampling. Orton (2000) discusses a case study for which an urban site contains clearly visible structures and notes that many urban sites fall into this category, especially if they have been deserted and not re-occupied or built over. Orton (2000) states that 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. With this in mind, instruct students to select sixteen test-pits from stratum I and four test-pits from stratum II and ask them to explain how to use the sample number of test-pits containing finds to estimate the total number of finds at Site 2. Since 16/50 of the test-pits are being sampled from stratum I (and stratum II contains no finds), for each selected sample, 50/16 times the number of finds out of the sixteen sampled test-pits in stratum I serves as an estimate of the total number of finds at the site. For the simple random samples, it is still the case that the estimated total is 5 times the number of finds in the sample.

Begin the discussion of the comparison by asking students to examine Site 2 and state whether they think that repeated stratified random sampling of test-pits from this site will be likely to produce less variable estimates of the total number of finds at the site. This may be a difficult question for students to answer. To provide a hint, the teacher may wish to point out a ‘worst case scenario’ for simple random sampling in which all 20 sampled test pits come from the side of the site that does not contain finds.

In order to simulate the performance of simple random samples versus stratified samples for Site 2, have each student select her own SEED and select both a simple random sample and a stratified random sample of 20 test-pits from the site. Have students record their estimated total number of finds on sticky notes, and place the sticky notes in appropriate positions on frequency plots on the white board.

III. Analyze the Data

For sampling from Site 1, for the simple random sample, once the SEED is set, random numbers are generated between 1 and 100. The test pits with the corresponding numbers are included in the sample. For a SEED of 2000, the selected test pits are: 13, 81, 72, 46, 39, 85, 82, 44, 31, 66, 92, 28, 6, 27, 18, 63, 54, 70, 56, 90. Test pits 13, 44, 66, and 54 contain artifacts. Since 1/5 of the test pits at the site were sampled, we can multiply the number of artifacts found by 5 to obtain an estimate of the total number of finds at the site: $(5)(4) = 20$.

For the stratified sample, once the SEED is set, random numbers are generated between 1 and 100. After each number is generated, the corresponding test pit must be located (either in Stratum I or Stratum II) and selected. For a SEED of 1981 the selected test pits from Stratum I are: 54, 44, 55, 34, 24, 61, 2, 5, 51, 21. Pits 54, 44, 55, and 51 contain artifacts. The selected

test pits from Stratum II are: 96, 100, 28, 60, 89, 70, 49, 87, 29, 38. Only pit 100 contains an artifact. Thus, the estimated total number of finds at the site is $(5)(5) = 25$.

For the systematic sample since there are a total of 100 test pits at the site, if we wish to obtain 20 test-pits in our sample, we need to take a 1-in-5 systematic sample. Using a SEED of 2003, and generating an integer at random between 1 and 5, the starting test pit would be test pit #2. After test pit 2, every 5th test pit is selected: 2, 7, 12, 17, 22, 27, 32, 37, 42, 47, 52, 57, 62, 67, 72, 77, 82, 87, 92, 97. Only test pit 47 contains artifacts. Thus, the estimated total number of finds at the site would be: $(5)(1) = 5$.

For cluster sampling, using a SEED of 2004 and generating integers at random between 1 and 10, the selected clusters would be row 2 and row 3. So the sampled test pits are: 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30. Pits 13, 19, 20, and 25 contain artifacts so the estimated total number of finds at the site is: $(5)(4) = 20$

For the simple random and stratified samples selected from Site 2 once everyone has selected their samples and placed their results on the white board, the class results can be analyzed. Here is some example class data for the comparison of sampling strategies for Site 2:

stratified random sampling estimated totals:

15.625, 18.750, 28.125, 15.625, 18.750, 21.875, 31.250, 9.375, 28.125, 21.875, 18.750, 18.750, 21.875, 25.000, 15.625, 25.000, 25.000, 12.500, 15.625, 25.000, 31.250, 21.875, 31.250, 25.000, 15.625, 15.625, 18.750, 25.000, 18.750

simple random sampling estimated totals:

10, 25, 5, 40, 25, 35, 20, 5, 20, 25, 35, 5, 40, 25, 20, 25, 20, 15, 10, 20, 25, 15, 30, 20, 25, 10, 10, 15, 25

For each sampling technique students calculate descriptive statistics for the class estimated total numbers of finds. Figures 3 and 4 show descriptive statistics and comparative dotplots for the example class results (the frequency plots made from sticky notes on the white board will resemble the dotplots). Additionally, comparative boxplots are constructed from the class estimated totals. Figure 5 shows comparative boxplots for the example class results.

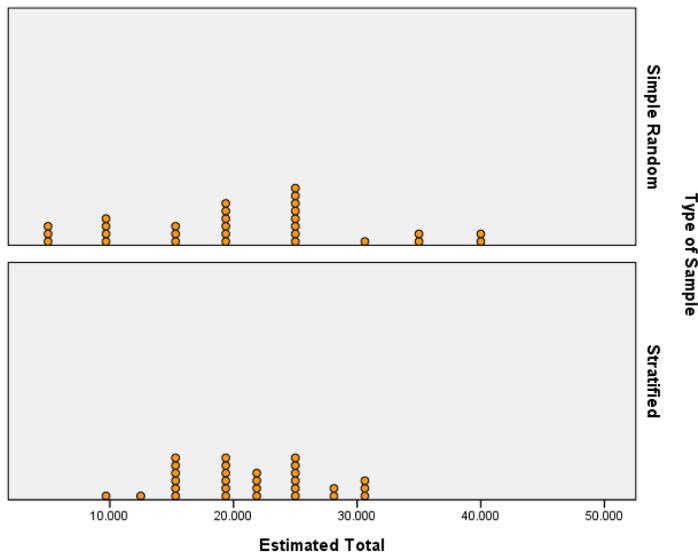


Figure 3. Dotplot of example class results for comparing simple random sampling to stratified random sampling.

Stratified Random Sampling

mean = 21.23
 standard deviation = 5.75
 first quartile = 15.625
 median = 21.875
 third quartile = 25.00

Simple Random Sampling

mean = 20.69
 standard deviation = 9.79
 first quartile = 12.50
 median = 20.00
 third quartile = 25.00

Figure 4. Descriptive statistics of class results for comparing simple random sampling to stratified random sampling.

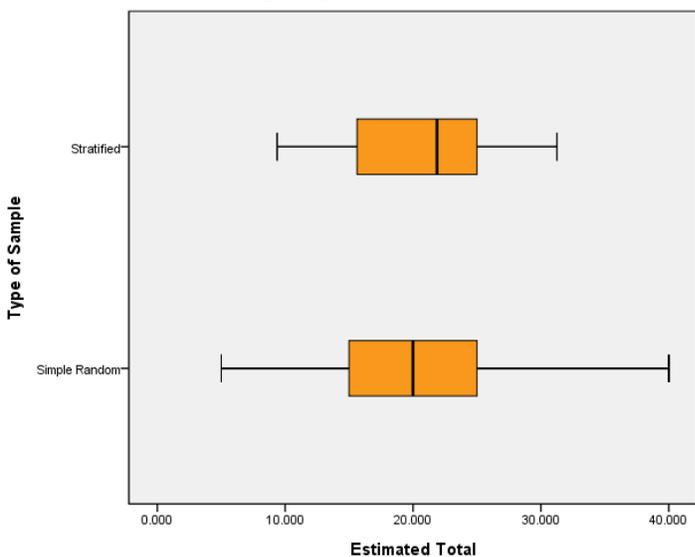


Figure 5. Comparative boxplots of simple random sampling and stratified random sampling estimated totals based upon example class data.

IV. Interpret the Results

Ask students to use results from the data analysis to discuss whether they think that repeated stratified random sampling of test-pits from Site 2 would produce less variable estimates of the total number of finds at the site. Students must justify their answers by using the numerical descriptive statistics and the graphs produced from the class estimated totals.

Students should discuss how the numerical calculations and graphs of the class estimated totals support the fact that, for Site 2, repeated stratified random sampling is more likely to produce less variable estimates than repeated simple random sampling. For the example class results, the standard deviation of the stratified random sample estimated totals is 5.75 compared to 9.79 for the simple random sample estimated totals. From the comparative boxplots it can be seen that the simple random sample estimates vary more than the stratified estimates with a larger overall range and interquartile range.

Note that another valuable aspect of collecting and analyzing class data is that it enables the teacher to introduce the concept of unbiasedness. Students can see that the distributions of estimated totals for both sampling techniques are centered on approximately 20 finds.

Assessment

1. A farmer has four orchards of apple trees which are located at different locations on his farm. Each orchard has 200 apple trees. He wishes to find out whether the apple trees are infested with a certain type of insect. If this is so, he would hire a crew to spray his trees. Instead of examining all 800 trees, he decides to select a sample of trees and just examine these. There are three proposed sampling plans described below:

Plan 1: Randomly select 100 trees from the 800 trees.

Plan 2: Randomly select 20 trees from each of the 4 orchards.

Plan 3: Randomly select 1 orchard from the 4 orchards, and then select all trees from the selected orchard.

For each of the above plans, identify the type of sampling method being proposed (simple random sample, stratified random sample, cluster sample, systematic sample).

Plan 1: _____

Plan 2: _____

Plan 3: _____

2. There are 16 first class passengers scheduled on a flight. In addition to the usual security screening, 4 of the passengers will be subjected to a more complete search. Here is the first class passenger list, denoted by which Section that the passenger is seated in.

Section 1	Section 2	Section 3	Section 4
1. Bergman	5. DeLara	9. Frongillo	13. Swafford
2. Cox	6. Forrester	10. Roufaiel	14. Clancy
3. Fontana	7. Rabkin	11. Castillo	15. Febo
4. Perl	8. Burkhauser	12. Dugan	16. LePage

(a) Select a simple random sample of 4 passengers. Use a **SEED** of **845**.
selected passengers:

(b) Select a cluster sample of 4 passengers. Use the Sections of passengers as clusters. Use a **SEED** of **332**.
selected passengers:

(c) Select a 1-in-4 systematic sample of 4 passengers. Use a **SEED** of **75**.
selected passengers:

3. A population consists of 12 people. The two columns divide up the population into 2 Strata, labeled I and II. The population is also divided into 3 Clusters by row.

POPULATION

	STRATUM I	STRATUM II
CLUSTER 1	Ann Betty	George John
CLUSTER 2	Carrie Donna	Bob Steve
CLUSTER 3	Ellen Fran	Paul Tom

Thus, Cluster 1 consists of: Ann Betty George John
 Cluster 2 consists of: Carrie Donna Bob Steve
 Cluster 3 consists of: Ellen Fran Paul Tom

Stratum I consists of: Ann Betty Carrie Donna Ellen Fran
 Stratum II consists of: George John Bob Steve Paul Tom

A sample of 4 people was obtained. Listed below are three different samples. Consider the following sampling methods: (i) simple random sampling, (ii) stratified random sampling with equal sample sizes from each stratum, (iii) cluster sampling by rows.

For each sample **determine which sampling method(s) could have generated that sample**, by circling yes or no for each. Hint: more than one method is possible.

SAMPLING METHOD

SAMPLE	(i) Simple Random?	(ii) Stratified?	(iii) Cluster?
(a) Carrie, Donna, Bob, Steve	Yes No	Yes No	Yes No
(b) Ann, Fran, Carrie, Betty	Yes No	Yes No	Yes No
(c) Carrie, Donna, George, Tom	Yes No	Yes No	Yes No

Answers

1.

Plan 1 describes simple random sampling.

Plan 2 describes stratified random sampling.

Plan 3 describes cluster sampling.

2.

(a) 16, 4, 7, 8

(b) Section 3: 9, 10, 11, 12

(c) 4, 8, 12, 16

3.

SAMPLE	SAMPLING METHOD					
	(i) Simple Random?		(ii) Stratified?		(iii) Cluster?	
(a) Carrie, Donna, Bob, Steve	<u>Yes</u>	No	<u>Yes</u>	No	<u>Yes</u>	No
(b) Ann, Fran, Carrie, Betty	<u>Yes</u>	No	Yes	<u>No</u>	Yes	<u>No</u>
(c) Carrie, Donna, George, Tom	<u>Yes</u>	No	Yes	<u>No</u>	Yes	<u>No</u>

Possible Extensions

1. Ask students to place an X in the appropriate test-pits on a blank grid in order to illustrate the layout of an archaeological site for which *repeated* (1-in-5) systematic random sampling of 20 test-pits would most likely produce a less variable estimate of the total number of artifact finds at the site than would *repeated* simple random sampling of 20 test-pits. Have students choose their own SEED and perform 1-in-5 systematic sampling and random sampling on the site created. Have students interpret the class results.

2. Ask students to place X's in the appropriate test-pits on a blank grid in order to illustrate the layout of an archaeological site for which *repeated* cluster random sampling of 20 test-pits would most likely produce a less variable estimate of the total number of finds than would *repeated* simple random sampling of 20 test-pits. Give a hint that challenges students to create a layout that will produce exactly four finds in every possible cluster sample of 20 test-pits (two rows). Have students choose their own SEED and perform cluster sampling and random sampling on the site created. Have students interpret the class results.

References

1. *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) Report, ASA, Franklin et al., ASA, 2007 <http://www.amstat.org/education/gaise/>.
2. Activity adapted from: Richardson, M. and Gajewski, B. (2003). Archaeological Sampling Strategies. *Journal of Statistics Education*, Volume 11, Number 1.
3. Orton, C. (2000), *Sampling in Archaeology*, Cambridge: Cambridge University Press.
4. Lizee, J., and Plunkett, T. (1994), "Archaeological Sampling Strategies," http://www.tiwanakuarcho.net/corso_arqpreh/rec/3/BAK/lizee_sampling/04-Anth292-2004.pdf
5. Assessment questions taken from: Aliaga, M. and Gunderson, B. (1999), *Interactive Statistics*, New Jersey: Prentice Hall.

Sampling in Archaeology Activity Sheet

Background: One question often asked of archaeologists is, “How do you know where to dig?” When archaeologists are working in areas that have not been previously explored, they must decide how to determine if the area contains any artifacts. Usually time and resources do not allow for the total excavation of a site, so archaeologists must develop a cost-effective strategy to allow for the maximum coverage of a site.

Problem: Suppose that the Map of an Archaeological Site represents an area that contains 100 8×8 meter test-pits (excavation units). In the map shown below, each square represents a test-pit and an X represents a test-pit that *contains* artifacts or “finds.” Twenty finds were randomly assigned to the test-pits on this map. There will only be enough time and resources allotted to dig in approximately 20 of the test-pits at the site. However, if a large enough number of the 20 selected test-pits contain artifacts, then more resources may be allocated to dig in more pits at the site.

Map of an Archaeological Site

1	2	3	4	5	6	7	8	9	10
							X		
11	12	13	14	15	16	17	18	19	20
		X						X	X
21	22	23	24	25	26	27	28	29	30
				X					
31	32	33	34	35	36	37	38	39	40
				X					
41	42	43	44	45	46	47	48	49	50
			X			X	X		
51	52	53	54	55	56	57	58	59	60
X		X	X	X			X		
61	62	63	64	65	66	67	68	69	70
					X				
71	72	73	74	75	76	77	78	79	80
			X	X				X	
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
							X		X

Taking Samples

NOTE:

To enter a Seed Value into the TI-84:

Choose any positive whole number. For example 1967.

1967 STO MATH PRB 1:RAND ENTER ENTER

The calculator will show your seed value. Only do this one time for your calculator.

To generate Random Integers on the TI-84: For example, to generate a list of random integers between 1 and 100:

MATH PRB 5:RANDINT(1, 100) ENTER

Continue to push ENTER to get more integers.

Strategy #1:

Select a **simple random sample** of 20 of the test-pits at the site. **Use a SEED of 2000.** Sample without replacement.

How many of the selected test-pits contain artifacts?

How can the number of finds out of 20 sampled test-pits be used to estimate the total number of finds at the site?

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

Strategy #2:

Select a **stratified random sample** of 20 of the test-pits at the site. Divide the site into 2 equally sized strata (use columns 1 through 5 of test-pits for Stratum I and columns 6 through 10 of test-pits for Stratum II). Sample 10 test-pits from each stratum. Sample without replacement. **Use a SEED of 1981.**

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

Strategy #3:

Select a **systematic random sample** of 20 test-pits at the site. **Use a SEED of 2003.**

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

Strategy #4:

Select a **cluster random sample** of 20 test-pits at the site. Use the rows of test-pits for clusters. Randomly select 2 clusters (use the top row as Row 1 and use the bottom row as Row 10). **Use a SEED of 2004.**

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

Comparison of Sampling Strategies

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. Thus, archaeologists might wish to sample the area that contains the visible structure at a higher intensity than the remainder of the site.

Stratum I					Stratum II				
1	2	3	4	5	6	7	8	9	10
11	12	13 X	14	15	16	17	18	19	20
21	22 X	23 X	24 X	25	26	27	28	29	30
31	32 X	33 X	34 X	35	36	37	38	39	40
41	42 X	43 X	44 X	45	46	47	48	49	50
51	52 X	53 X	54 X	55	56	57	58	59	60
61	62 X	63 X	64 X	65	66	67	68	69	70
71	72 X	73 X	74 X	75	76	77	78	79	80
81	82	83 X	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

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 SEED** 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 SEED**, 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 =

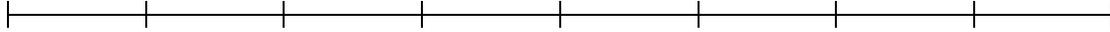
first quartile =

median =

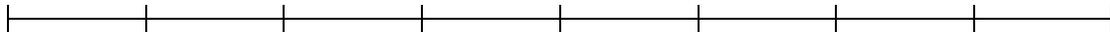
third quartile =

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 less variable estimates of the total number of artifact finds at the site than would *repeated* simple random sampling of test-pits? Why? Or, why not?