Journal of Statistics Education

ISSN: (Print) 1069-1898 (Online) Journal homepage: https://www.tandfonline.com/loi/ujse20

Archaeological Sampling Strategies

Mary Richardson & Byron Gajewski

To cite this article: Mary Richardson & Byron Gajewski (2002) Archaeological Sampling Strategies, Journal of Statistics Education, 10:3, , DOI: 10.1080/10691898.2002.11910690

To link to this article: https://doi.org/10.1080/10691898.2002.11910690

Copyright 2002 Mary Richardson and Byron Gajewski

Published online: 01 Dec 2017.

Submit your article to this journal

Article views: 4071

View related articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ujse21

Journal of Statistics Education

Archaeological Sampling Strategies

Mary Richardson
Grand Valley State University

Byron Gajewski
The University of Kansas Medical Center

Journal of Statistics Education Volume 10, Number 3 (2002),
ww2.amstat.org/publications/jse/v10n3/richardson.html

Copyright © 2002 by Mary Richardson and Byron Gajewski, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Active learning; Advanced Placement statistics; Archaeology; Introductory statistics; Statistical sampling; Statistical simulation.

Abstract

This paper describes an interactive project developed to use for teaching statistical sampling methods in an introductory undergraduate statistics course, an AP statistics course, or, with adaptation, in a statistical sampling course or a statistical simulation course. The project allows students to compare the performance of simple random sampling, stratified random sampling, systematic random sampling, and cluster random sampling in an archaeological setting.

1. Introduction

In this paper, we will discuss an interactive project that we use to illustrate properties of statistical sampling techniques in an introductory level statistics course or, with adaptation, in higher-level statistics courses. The project allows students to compare the performance of different statistical sampling techniques in an archaeological context.

The project described here was initially developed for use in an undergraduate general education introductory statistics course. The students in this course have a limited mathematical background, with most having previously taken only basic algebra. We believe that some statistical concepts are generally difficult for the introductory students to learn in a lecture setting. Some statisticians even argue that lectures are no longer effective in an age where fast paced technology dominates our culture. Cobb (1992) states: “Shorn of all subtlety and led naked out of the protective fold of educational research literature, there comes a sheepish little fact: lectures don’t work nearly as well as many of us would like to think.” and in describing the Activities-Based Statistics Project, Scheaffer (1996) states: “Their fast-paced world of action movies, rapid-fire TV commercials and video games does not prepare today’s

students to sit and absorb a lecture, especially on a supposedly dull subject like statistics. To capture the interest of these students, teaching must move away from lecture-and-listen to innovative activities that engage students in the learning process."

In the introductory course, we use many hands-on interactive projects to illustrate key statistical concepts. The projects are completed collaboratively by groups of students during one-hour class periods. The use of hands-on explorations of statistical concepts helps to get the students more involved in the learning process and has been a very effective, fun way for the students to learn introductory statistics.

In addition to discussing the use of the project in the introductory course, we will discuss extensions of the project that can be used in an applied probability and simulation course and/or a statistical sampling course. We will also attempt to provide the reader with some realistic examples of statistical sampling applied in archaeological settings.

2. The Introductory Course Project

2.1 Background

As a motivation for our project, consider the following scenario. Three years ago, funding for a new campus library was awarded to a large public institution of higher learning. However, construction was delayed during the initial digging when archaeological artifacts were discovered. Specifically, human bones from an old grave were unearthed. This discovery required the university to perform an exhaustive, costly archaeological study on the excavation site of the library.

There are two possible ways the university could have avoided this costly, and somewhat embarrassing, controversy. The first would be to carefully study the legal documents, newspaper archives and all other information regarding the excavation site. This option is thorough, but time consuming, and it can take years to complete. Therefore, it is not a practical option. The second option would be to perform a kind of strategic digging, specifically to take a representative sample of the excavation site and, from this, determine if there are, in fact, archaeological "finds."

According to Orton (2000), the term site has many meanings. The example given above is of a development site -- an area of land subject to some form of proposed commercial, agricultural or infrastructural development. Orton (2000) 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 (perhaps newly discovered in a regional survey), 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.

The methods used to excavate a site may vary according to whether the site is largely invisible on the ground surface, or whether it has extensive visible remains (Orton (2000)). Orton (2000) notes that for some sites located in arid and semi-arid areas, such as the south-west USA, the visibility of archaeological remains on the surface is good. For these types of sites, fieldwalking, or pedestrian

survey can be employed and large areas can be covered at a reasonable cost in terms of time and labor. In other parts of the world (or of the USA), conditions at sites are completely different. The land surface may be covered in grassland, arable crops, or forest, so that even the ground itself, let alone any archaeological remains, may not be visible.

According to Orton (2000), the idea of using probabilistic sampling methods explicitly in archaeological survey is usually attributed to Binford (1964). Binford stressed the idea of the region (a collection of sites) as a natural unit of archaeological study, and, admitting the impossibility of total coverage at this scale, advocated probabilistic sampling techniques as the way of achieving representative and reliable data with less than total coverage. Before we begin our discussion of the introductory course project, we discuss some relevant definitions and terminology for applying statistical sampling in an archaeological context (extracted from Orton (2000)).

Prior to excavation, a site must be divided into sampling units (excavation units). The units should cover the entire site, and they should not overlap. Frequently, the choice of units is not obvious. In excavating a site, it would seem logical to proceed in stages, with the extensive use of non-invasive methods (such as fieldwalking) being followed by intensive sampling of "interesting" areas by invasive methods (such as trial excavation in trenches or test-pits). However, this approach is fairly uncommon in practice, perhaps because it fits badly within the time constraints usually involved in this sort of work. 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. If purposive sampling is used, the shape and size of the excavation units (trenches or test-pits) are likely to be determined by the nature of the visible evidence that points to their location. 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, although sizes up to 16 m square are sometimes used. Trenches are flexible in design, cheap per area stripped and good at detecting features, but are destructive and have a poor recovery rate for archaeological finds. Test-pits are good at detecting archaeological finds and are relatively non-destructive, but are labor-intensive, and therefore expensive.

In order to use an archaeological setting to demonstrate the use of statistical sampling, we will assume that a site will be sampled probabilistically. Further, we assume that the excavation units are test-pits.

2.2 Procedure

Prior to completing this project, students have been exposed to basic definitions and terminology related to statistical sampling and they have seen examples of simple random sampling, stratified random sampling, systematic random sampling, and cluster random sampling. For completeness, we define each of the four sampling techniques and comment on the use of each technique in the sampling of archaeological sites.

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. Orton (2000) notes that in an archaeological setting, some practitioners worry that a true simple random sample has the appearance of "bunching" and seek to avoid it. Systematic sampling is one way of doing so.

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 kth ordered unit. Orton (2000) notes that one disadvantage with systematic sampling is that the sampling interval may, by misfortune, relate to some regularity in the site. For example, when sampling grid squares on a map, systematic sampling may generate diagonal lines of sampled squares, which in turn

may relate to natural topographical or geological features. But, there are situations in which systematic sampling performs much better than simple random sampling. Archaeologists often seem to prefer systematic samples, partly because they appear to give "better coverage" or an "even spread" of a site, and partly because they are easier and quicker to select. These points must be weighed against the possibility of the sampling interval matching some regularity in the data.

Stratified random sampling is simply forming subgroups of the population units and selecting a simple random sample of units from within each subgroup. For example, stratification might be warranted if the archaeologist mandates that a representative sample be taken from the west end of a site as well as the east end of a site. Orton (2000) states that the possibility of stratification of a site should always be considered. The definition of strata could be based on any property, or combination of properties, such as geology or elevation, or on any aspect that is thought likely to affect the parameters under study, for example, the density of artifacts located in a site.

Cluster random sampling also requires the sampling units to be placed into subgroups. However the subgroups are typically obtained from units close in location. A simple random sample of the subgroups is then taken, and every unit within the selected subgroup is a part of the sample. For example, clustering by location of excavation units might be performed if the proximity of the units will provide efficient use of a backhoe when digging in the excavation units.

Part 1

The introductory project is completed in two parts. We use Part 1 and Part 2 in sequence, however, Part 1 could be used without using Part 2. Part 1 gives students an opportunity to practice the mechanical aspects of performing the four sampling techniques. The estimated interactive completion time for Part 1 is 30 minutes. Part 2 expands on Part 1 and helps students take mechanical knowledge of the sampling techniques one step further by requiring them to investigate the performance of the sampling techniques in differing scenarios. The estimated interactive completion time for Part 2 is 50 minutes. After completing both parts of the project, students will be able to perform the four types of sampling and have an understanding of which of the sampling procedures should be applied in populations with differing characteristics.

To begin Part 1, students are divided into groups of between two and four. Each student is given a copy of the Project Background sheet (see Appendix A.1), the Part 1 Worksheet (see Appendix A.2), and a random number table. The problem (which was taken from Lizee and Plunkett (1994)) is formalized as follows. The Project Background sheet contains an initial map of an archaeological site. The site is an area of approximately 6,400 square meters to be impacted by construction of a housing development. The area consists of a mature second growth forest of maple, ash, and oak trees. Since it is both time and labor intensive to excavate the entire site, a sampling strategy must be developed. Working within budgetary and time constraints, archaeologists can only excavate part of the site to determine the presence of buried artifacts. The site contains100 8 8 meter excavation units (test-pits) and there is only enough time to dig in 20 of the test-pits. The map is shown in Figure 1 below (with an X representing a test-pit that contains artifacts or "finds"). On the initial map, we randomly assigned finds to 20 of the test-pits.

Each classroom group is to use each of the four sampling strategies to select a sample of test-pits from the site. 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). Ten test-pits are selected from each stratum. Systematic random sampling is performed by rows, using the top row as ordered test-pit numbers 1 through 10 with the leftmost test-pit being 1 and the rightmost test-pit being 10, …, so that the bottom row represents

ordered test-pit numbers 91 through 100 with the leftmost test-pit being 91 and the rightmost test-pit being 100. Cluster random sampling is performed using the rows of test-pits for clusters. We ask the groups to use uniform starting locations on a random number table (the same seeds on a calculator) in order to have a classroom discussion of the results of selecting the different samples.

The goal is to use each sample of 20 test-pits to estimate the total number of test-pits containing finds (which, for brevity, we will refer to as the "total number of finds"), denoted by Y. After students have selected their simple random sample from Site 1, we ask them to explain how to use the sample number of test-pits containing finds to estimate the total number of finds at this site. For each sampling technique, 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, number of finds in the sample.

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, denoted by 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. Orton (2000) notes that in some cases, any archaeological remains may be deemed "significant," while in other cases the remains may have to occupy a specified total area, or a specified proportion of a site, before they can be called "significant."

Part 2

To begin Part 2, students are divided into groups of between two and four. Each student is given a copy of the Part 2 Worksheet (see Appendix A.3) and two sticky notes. The problem is formalized as follows. Through artificial examples, students are to explore scenarios where one sampling technique yields relatively more precise estimates of the total number of finds at a site than does another sampling technique. Students compare the performance of the four sampling techniques by examining three different layouts of archaeological sites. It is assumed that each site contains100 total test-pits and that 20 of the test-pits contain artifacts.

For Comparison #1, we 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 stratified random sampling of 20 test-pits would most likely produce a less variable (more accurate) estimate of the total number of artifact finds at the site than would repeated simple random sampling of 20 test-pits. Once again, we use column 1 through column 5 of test-pits for stratum I and column 6 through column 10 of test-pits for stratum II. The layout for this site is shown in Figure 2.

For this comparison, we do not use equal sample sizes from the two strata. Our 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. Redman (1987) discusses an excavation of a site at Shoofly Village in Arizona that was completed in three stages. In the first stage, stratification was used, with an area inside of an enclosure wall being sampled from at four times the intensity of an area immediately outside.

With this case study in mind, we instruct students to select sixteen test-pits from stratum I and four test-pits from stratum II and we ask them to explain how to use the sample number of test-pits containing

finds to estimate the total number of finds at the site. For the stratified samples, 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, number of finds in the sample. Note that for the simple random samples, it is still the case that, number of finds in the sample.

We begin our discussion of this 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. It has been our experience that this is a difficult question for students to answer. They are usually quite baffled until we quote a question from Aliaga and Gunderson (1999, p. 78): "When you form the strata, how should the variability of the units within each stratum compare to the variability between stratum?" After remembering that the variability within each stratum should be small and the variability between stratum should be large, many students see that by forcing one of the stratum to have all of the finds, we are forming very homogenous strata and we are giving stratified random sampling a better chance to produce consistent estimates of the total number of finds in repeated sampling.

One student wrote "I believe that stratified random sampling will produce less variable estimates than simple random sampling. I believe this will happen because all of the finds are so closely placed that doing just simple random it is likely that very few will be hit, while in stratified you know that sixteen attempts will be on the side where all the finds are and they will be hit more often and more uniformly from trial to trial." Another student wrote "Stratified would be less variable because if you use simple random sampling it would be possible to choose more test-pits in the right columns than in the left columns and find nothing." Another wrote "A stratified random sample would be the better choice because there would be a high number of finds in stratum one but zero in stratum two so the correct number of finds is more likely. In a simple random sample there is a higher chance of getting test-pits scattered about the entire site so there is more of a chance of not getting enough finds or too many."

Next, we ask each student to pick their own starting position on a random number table and select both a simple random sample and a stratified random sample of 20 test-pits from Site 2. Students record their estimated total number of finds on sticky notes, and place the sticky notes in the appropriate positions on frequency plots on the white board. In Figure 3, we have included example class results for the estimated total number of finds using each sampling technique.

Once everyone has selected their samples and placed their results on the white board, we begin a discussion of the class results. For each sampling technique, students calculate descriptive statistics for the class estimated total numbers of finds (mean, standard deviation, and quartiles) and construct comparative boxplots. Figure 4 shows descriptive statistics and comparative boxplots for the example class results.

We then 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. Note that for the example class results, the standard deviation of the stratified random sample estimated totals is 5.67 compared to 9.68 for the simple random sample estimated totals. And, clearly, from the comparative boxplots, we see that the simple random sample estimates vary considerably more than the stratified estimates. Note that another valuable aspect of collecting and analyzing the class data is that it enables the instructor to introduce the concept of unbiasedness. Students can see that the distributions of estimated totals (for both sampling techniques) are centered at approximately 20 finds.

Finally, we ask students to state whether they think that the class' repeated stratified random sampling of

test-pits from Site 2 produced 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.

For Comparison #2, we 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 (assuming that the systematic random sampling is performed by rows) 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. The layout for this comparison is shown in Figure 5 below.

We ask students to state whether they think that repeated systematic random sampling of test-pits from Site 3 will be likely to produce less variable estimates of the total number of finds. After having discussed Comparison #1 and working with a concrete example of repeated sampling from Site 2, most students, when given the layout for Site 3, are able to correctly identify 1-in-5 systematic random sampling of this site as producing the least variable estimated totals.

One student wrote "The 1-in-5 systematic sample is going to be more accurate than a simple random sample in this case because every time you do this you are going to get an estimated total of 20. The first four digs are going to hit artifacts no matter where you start. You won't be able to have as accurate of results with simple random sampling." Another wrote "A 1-in-5 systematic sample because of this particular layout whichever number you start with will always get you four finds. If you use a simple random sample then the number of finds will be less accurate. This is because each time the sample is taken the results could be different." Another wrote "With the one in five sample you will always get 20 estimated total no matter where you start. With repeated random sampling the estimated total should eventually even out around 20. However, with systematic, you always get 20."

For Comparison #3, we 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 (again assuming that the cluster sampling is performed by rows) would most likely produce a less variable estimate of the total number of finds than would repeated simple random sampling of 20 test-pits. Here, we 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). A correct answer for this comparison is shown in Figure 6 below.

For Comparison #3, assessment of student answers is straightforward. Following the hint, students will place two finds in each row of the site and cluster random sampling will provide exact estimates of the total number of finds. Simple random sampling will provide deviations from the true total number of finds.

3. A Case Study Between Stratified Random Sampling and Simple Random Sampling

In this section, we discuss an alternative approach to Comparison #1 for Part 2 of the introductory project.

Under the same layout for Site 2, students can be asked to compare the performance of repeated stratified random sampling to repeated simple random sampling (of 20 total test-pits) when ten test-pits are sampled from stratum I ( ) and ten test-pits are sampled from stratum II ( ). It has been our experience that, in a typical class, with only 25 to 30 replications, it is difficult to see a difference in the performance of these two sampling techniques for Site 2. However, to generate a larger-scale

simulation, each student could be asked to perform each of the sampling techniques four times so that the class as a whole provides roughly 100 simulated samples.

To explore the relationship between sample size from stratum I and the performance of stratified random sampling for Site 2, students could repeat their simulations, but for a different sample size in stratum I. It seems logical that more information should be taken from stratum I than from stratum II, since there are more artifacts in stratum I. Figure 7 displays the relationship between sample size from stratum I and the performance of stratified random sampling, for sample sizes from stratum I ranging from all the way to

The simulation results presented in Figure 7 show that the higher values of provide overall better estimates of the total number of finds at the site. Essentially this result demonstrates that all of our sampling resources need to be placed in the stratum with all of the information. The first stratum has all of the artifacts, therefore, optimally, the sample size allocated to the first stratum is and the sample size allocated to the second stratum is The reason this is optimal is due to the variation within each stratum. Consider placing a "1" in a test-pit with an artifact and a "0" in a test-pit without an artifact. The standard deviation for stratum II is 0 and the standard deviation for stratum I is 0.4949 (for Site 2). For equal sized strata, the optimum sample size allocated to each stratum is proportional to the standard deviation within that stratum. This is called "optimal allocation" or more generally "Neyman allocation." For more on this topic, for multiple strata and unequal sized strata, see Cochran (1977) or Lohr (1999).

4. The Simulation Course Project

For the simulation course project, we introduce the four sampling techniques and we assign Parts 1 and 2 of the introductory project as homework. For Part 2 of the introductory project, rather than have students select samples in class, we include an additional sheet that contains simulation results for estimated totals obtained by drawing repeated stratified and simple random samples from Site 2 (see Appendix A.4). After collecting this homework assignment, we give a follow-up assignment that requires students to write simulation programs to compare the performance of all four sampling techniques for each of Sites 1 through 4.

We ask students to base comparisons of the performance of the four sampling techniques on an examination of simulated values of Mean Squared Error (MSE) (in estimating the total number of finds). In general, if is an estimator of a parameter, we say that is an unbiased estimator of if Unbiased estimators of the same parameter can be compared in terms of the size of their variances: since is unbiased. If is not an unbiased estimator of a given parameter one way to compare to other estimators is on the basis of the Mean Squared Error:

Simple random sampling and stratified random sampling, for our problem, can be coded in MINITAB using the hypergeometric distribution. The commands to perform simple random sampling for any site layout are:

```
MTB > Random 10000 c1;
SUBC> Hypergeometric 100 20 20.
MTB > Let c2=(c1/20)*100
```

The "SUBC" command samples from a population of 100 test-pits with 20 successes (finds), using a sample size of 20. The second column (c2) contains the estimated totals for the 10,000 replications.

To generate stratified random sampling, under Site 2 (all of the artifacts in the first strata), one generates

two hypergeometric columns in MINITAB and combines them (via a weighted formula) to obtain the estimated total. For example, if we wish to sample and , we use the following commands:

```
MTB > Random 10000 c3;
SUBC> Hypergeometric 50 20 14.
MTB > Random 10000 c4;
SUBC> Hypergeometric 50 0 6.
MTB > Let c5=((c3/14*.5+c4/6*.5))*100
```

It has been our experience that our simulation students (often engineers, computer scientists and statisticians) prefer to program in C++, SAS, or MATLAB. Thus programming results can vary depending on the type of language used. The second author has created code to perform all of the sampling techniques in this paper using MATLAB. This MATLAB code can be obtained by contacting the second author.

Table 1 below gives example results for simulations of 10,000 samples of size 20 drawn using each of the four sampling techniques for each of Sites 1 through 4. The simulated MSE was calculated from the 10,000 values by taking the estimated totals, , and calculating:

The simulated MSE's provide very good approximations to the theoretical MSE's. The theoretical MSE is calculated by using probability arguments that take into account the structure of the site and the way that sampling was performed. The theoretical MSE is given by: , where denotes the probability of and the summation is over all unique Note that the concept of Mean Squared Error is easily illustrated from the systematic random sampling approach. By stacking the first five columns on the last five columns, the number of finds in each column times five would be the estimated total, where each total has a probability of selection of 1/5. For example, the five estimated totals for Site 1 are 10, 5, 30, 25 and 30. Thus, the theoretical MSE is: , corresponding almost exactly to the simulated value given in Table 1. Similarly, for systematic random sampling, the remaining three sites have theoretical MSE values of 280.0, 0.0, and 1600.0. Extending this idea to derive the theoretical MSE values for the other sampling designs can provide the foundation for a discussion of general variance formulas presented in a typical upper-division sampling course.

References

Deborah J. Rumsey
Department of Mathematics
The Ohio State University
Columbus, OH 43210
USA
rumsey@math.ohio-state.edu