

The Editor's Corner



Angela Walmsley

It is with great pleasure that I write my first Editor's Corner for the Statistics Teacher Network. I have served as an associate editor for a number of years, and I'm thrilled to have this opportunity! I would first like to say thank you to **Rebecca Pierce**, who has been the senior editor for the past three years and who has graciously agreed to serve as an associate editor so we can call on her expertise!

You'll find some diverse and interesting articles in Issue 85. The first article is Part III of the LOCUS project (parts I and II are in previous issues of STN). "Lessons from the LOCUS Assessments" was written by **Steven Foti**, a doctoral fellow, and **Tim Jacobbe**, an associate professor, from the University of Florida. In this article, they discuss the importance of having a true understanding of boxplots and the five-number summary. The LOCUS project has focused on developing assessments within the GAISE framework.

The second article is from an elementary school teacher, **Heather Ristau**, who teaches third grade, but has taught

various grades at the elementary level. She introduces us to the concept of "what are the chances" according to third graders' understanding of statistics. It is always great to have articles written by elementary teachers, since they can show us all how statistical thinking develops in young minds.

Our last article is from our new associate editor, **Doug Rush** from Saint Louis University. It focuses on the important and often confusing concepts of power and effect size. Doug explains hypothesis testing, type I and type II errors, and effect size in a straightforward manner. He also offers a short and more in-depth look at effect size for those who are interested.

I encourage you to read all three!

Please send any articles or ideas you have for consideration to pastpresident@pdkwa.org.

Regards,

Angela Walmsley, Editor,
Northeastern University

Articles

Lessons from the LOCUS Assessments (Part 3: Boxplots).....	2
What Are the Chances of That?.....	6

Power and Effect Size.....	7
Sampling in Archaeology	11
Announcements	20

Lessons from the LOCUS Assessments (Part 3): Boxplots

Steven Foti – University of Florida
Tim Jacobbe – University of Florida

Since its publication, the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework* (Franklin et al., 2007) has been influential in the field of statistics education. The developmental levels—A, B, and C—through which students are hypothesized to progress provide convenient touchstones for curriculum and lesson design.

Despite the impressive contributions to statistics education in terms of instructional recommendations and the aforementioned developmental progression, the GAISE framework says less about what types of assessments are recommended or should be considered as a model. The NSF-funded Levels of Conceptual Understanding in Statistics (LOCUS) project focused on developing statistical assessments in the spirit of the GAISE framework. These assessments emphasize conceptual (rather than procedural) understanding and can be used to classify students as having understanding at level A, B, or C.

The assessments—now available at <http://locus.statisticseducation.org>—consist of four forms: a pre- and post-test targeting the A and B levels and a pre- and post-test targeting the B and C levels. The A/B assessment was designed for students in grades 6–9, and the B/C assessment was designed for students in grades 10–12. Two versions of these are available: one with 23 multiple-choice items and five free-response items and another with 30 multiple choice items only.

The items from which these forms were constructed were piloted in spring of 2013 with a total of 2,075 students for the A/B assessment and 1,249 students for the B/C assessment. (Although not every item was piloted with every student, each item was piloted with several hundred students.) While the pilot administration was large and included students of many backgrounds and ability levels, it was not selected to be a representative sample of students in the United States. We do report some overall performance indicators, but we caution these should not be over-interpreted. Rather, the indicators are included to paint a more complete picture of the students and item.

Student work can be a valuable resource for teachers. The size and scope of the LOCUS pilot assessments yielded considerable variation in student responses. While there were some ‘textbook’ correct answers, students also were able to demonstrate correct statistical reasoning in imaginative ways. Incorrect answers often illustrate specific misunderstandings and, if identified as such, can suggest areas for more attention.

The LOCUS item being examined here is shown in Figure 1. This item addresses the following Common Core State Standards (CCSS):

- 6.SP.4 “Display numerical data in plots”

- 7.SP.3 “Draw informal comparative inferences about two populations”
- S-ID.1 “Represent data with plots on the real number line”
- S-ID.2 “Compare center and spread of two or more data sets”
- S-ID.3 “Interpret differences in shape, center, and spread in context”

The “Analyze Data” and “Interpret Data” components of the GAISE framework at Level B also are addressed by this item.

Student Responses

The free-response item shown was piloted with a total of 539 students. Of those, 524 were students in grades 6–8, and the remaining 15 were students in grades 9–12 who took both the A/B and B/C forms of the assessment. Free-response items were scored out of 4 points, with a 4 indicating a “complete” response (allowing for mistakes such as minor computational errors that are not indicative of a misunderstanding). For each item, a small team of graders established a rubric and conducted initial item scoring verbally as a group. Once every grader felt comfortable with applying the rubric, scoring continued individually. Any discrepancies or questions were brought to the group’s attention. The distribution of students’ scores for the item shown is given in Table 1: 2.0% earned a 4, 7.4% earned a 3, 23.2% earned a 2, 13.4% earned a 1, and the remaining 54.0% earned either a 0 or did not provide a response. The responses given provide useful illustrations of students’ varying conceptions.

Table 1: Distribution of Student Scores for the Item Shown

Score	Percent of Students
4	2.0%
3	7.4%
2	23.2%
1	13.4%
0 or no response	54.0%

Questions related to free response item in Figure 1:

- (a) Based on these summaries, construct boxplots that will help you compare the scores between the two classes.

The first part of this item has the students construct boxplots from tabulated numerical summaries of mathematics test scores from two hypothetical classes. Common Core (6.SP.4) first mentions boxplots in grade 6 and only requires students to construct the display. In fact, many students from the pilot administration were able to accurately construct the boxplots from the given data. Students who did not correctly construct

Figure 1. A mathematics test is given to two classes. The scores are summarized in the table.

	Class I	Class II
Minimum	42	57
First Quartile	53	66
Median	72	72
Third Quartile	89	78
Maximum	98	84
Range	56	27
IQR (interquartile range)	36	12

(a) Based on these summaries, construct boxplots that will help you compare the scores between the two classes.

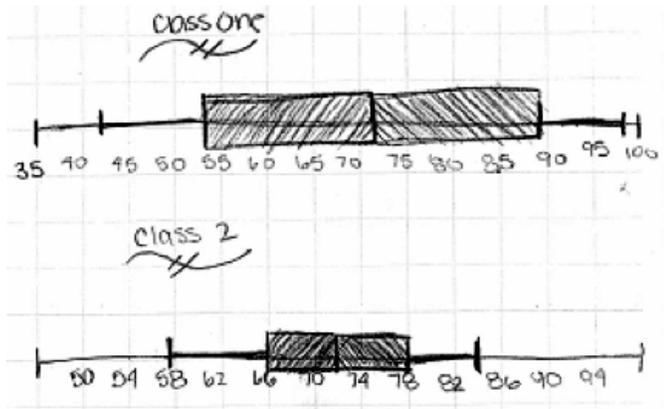


(b) Describe in context the important similarities between the scores for the two classes. Support your observations with references to the graphical display and/or numerical summaries. Make sure to address the amount of overlap between the two groups.

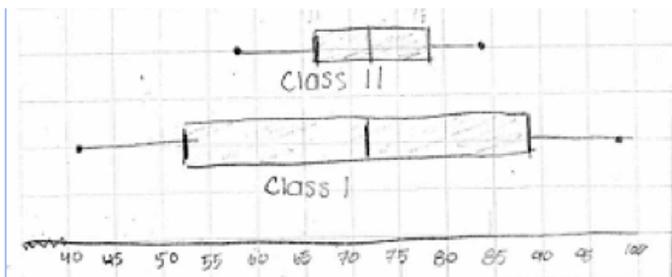
(c) Describe in context the important differences between the scores for the two classes. Support your observations with references to the graphical display and/or numerical summaries.

the boxplots showed evidence that they simply did not know the necessary steps to create the display from the given data.

The more informative piece of evidence of the students' understanding of boxplots is revealed in the piece of the question that requires students to construct the boxplots that will help you compare the scores between the two classes. One of the main reasons boxplots were created was to be able to visually compare groups. Several students, however, constructed the boxplots on different scales, which would result in visual comparisons that are misleading or incorrect.



To make informal comparisons between the two groups, students needed to construct both boxplots on the same scale.



Questions related to free response item in Figure 1

- (b) Describe in context the important similarities between the scores for the two classes. Support your observations with references to the graphical display and/or numerical summaries. Make sure to address the amount of overlap between the two groups.
- (c) Describe in context the important differences between the scores for the two classes. Support your observations with references to the graphical display and/or numerical summaries.

These two parts of the item require the students to use the boxplots to make informal comparisons by analyzing similarities and differences between the two classes. There is a consensus in the research on student understanding of boxplots that, while

they are fairly easy to construct, interpreting them provides challenges for many students (Bakker et al., 2004; Watson, 2012). Due to the complexity of interpreting boxplots, some researchers recommend waiting until high school to introduce them (Bakker et al., 2004). The GAISE framework recommends introducing boxplots when students reach Level B of their statistical understanding development (Franklin et al., 2007). Common Core (7.SP.3, S-ID.1, S-ID.2, S-ID.3) is generally in line with these recommendations, with the exception of making informal comparisons between two populations in Grade 7 using displays such as boxplots.

The responses from parts (b) and (c) of the item revealed that only a few students are able to effectively use the graphical display to make meaningful comparisons between the two groups. The GAISE framework emphasizes that boxplots are useful to compare groups by evaluating differences in center, spread, shape, and visual overlap (Franklin et al., 2007, p.46–48, 55, 76–77). However, many students resorted to using the numerical information available in the table when discussing similarities and differences. This is evident in responses that only make numerical comparisons such as, “the median for both classes is 72 ... while the range is quite different, the maximums for both classes are similar ...” Most student who scored poorly on this item were able to accurately construct the boxplots, but were not able to use them to make any general comparisons between the two groups.

A common response among students who scored higher on the item showed some evidence of understanding of boxplots; however, they misused some common terms. The interquartile range (IQR), for example, is a measure of spread; however, many students have difficulty understanding it as one (Bakkar et al., 2004). More commonly, students refer to the actual “box” in the display as the IQR, which is shown in responses such as, “class two was all within the IQR of class one” or “the entire range of class 2 is contained inside the IQR of class one.” These responses show promising evidence that students begin to understand the concept of overlapping boxplots; however, their use of measures of spread to describe the picture is not accurate.

There were some responses that received full credit for this item. They showed an advanced level of understanding of terms used to describe and compare boxplots. One response says “both of the classes had a median of 72. The entire class two scored between class one’s third and first quartile.” This response uses more accurate language to describe the overlap between the two boxplots because the quartiles are actual locations on the display, where the IQR, for example, is merely a number.

Discussion

While some students were able to score well on the item, it is evident that most students do not have a solid understanding of using boxplots to compare two distributions. Similarities and differences the students discussed were mostly derived from the table of numerical values and often failed to lead them to draw meaningful conclusions about the data. Although many students

fell far short of a perfect score, as defined by the rubric, a strong majority of them displayed their ability to construct boxplots from the Five-Number Summary. Constructing boxplots has been in the mathematics curriculum for quite some time; however, their importance as a data analysis tool is becoming more relevant in the curriculum.

This item focuses on the “Analyze Data” and “Interpret Data” components of the GAISE framework at level B and targets Common Core standards in Grade 6, Grade 7, and high school. GAISE and the Common Core standards clearly emphasize that while it is certainly important for students to be able to quickly construct boxplots, it is more important to be able to use them as a tool to summarize and compare distributions.

For teachers who would like to focus their instruction on the content covered in this item, there are multiple resources available. Various lesson plans are available on the STEW website (www.amstat.org/education/stew) that address the use of boxplots to compare two groups. For example, the activity “Armspans” has students compare the arm span lengths of the boys and girls within their class, using boxplots, to determine if there is any difference. Other examples—such as “Don’t Spill the Beans!”—has students use comparative boxplots as a tool for analyzing data from an experiment. The video series “Learning Math: Data Analysis, Statistics, and Probability” (WGBH, 2001) is intended to introduce K–8 teachers to statistical concepts

through classroom case studies. In particular, video 4, “The Five-Number Summary,” discusses boxplots in depth and concludes with an activity that uses boxplots comparatively. All these lessons could be implemented in the classroom and would provide students with examples that explore the usefulness and importance of boxplots as summative and comparative tools.

Further Reading

Bakker, A., R. Biehler, and C. Konold. 2004. Should young students learn about box plots? *Curricular Development in Statistics Education: International Association for Statistical Education*, 163–173.

Franklin, C. A., G. Kader, D. Mewborn, J. Moreno, R., Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*. Retrieved from www.amstat.org/education/gaise

National Governors Association Center for Best Practices and Council of Chief State School Officers. 2010. *Common core state standards for mathematics*.

Watson, J. M. 2012. Box plots in the Australian curriculum. *The Australian Mathematics Teacher* 68(3):3.

WGBH Educational Foundation (Producer). 2001. *Learning math: Data analysis, statistics, and probability*. Available from www.learner.org/resources/series158.htm



K–12 Teachers:

JOIN THE ASA

A special offer tailored for K–12 educators!

The American Statistical Association wants to help you enhance your students' statistical education.

Visit www.amstat.org/membership/k12teachers for details.



Free Trial Membership

Sign up today for your **FREE** 3-month trial membership and gain access to:

- *Amstat News*, the ASA's monthly membership magazine, and *Significance*, an ASA and RSS partnership magazine aimed at international outreach to enhance both organizations, the statistics profession, and statisticians.
- The ASA's top journals and resources—including online access to *CHANCE* magazine, the *Journal of Statistics Education*, and *The American Statistician*—in addition to discounts on all ASA meetings and products.
- Teaching resources, including webinars, the *Statistics Teacher Network*, *GAISE: A Pre-K–12 Curriculum Framework*, and the *Statistical Significance* series.
- A new online community for ASA K–12 Teacher Members that allows participation in online discussions and sharing of resources with other members.

What Are the Chances of That?

Heather Ristau, Stormonth Elementary School

This is a question I have heard asked many times in my third-grade classroom this year. It got me thinking that maybe we should talk about what that expression means and what the chances of something really are.

I have always taught my students about data and data collection. Children love surveying other people and finding out if more like chocolate or vanilla ice cream or which sport is most popular. We even vote each week on the debate of the week for Scholastic News and look at the collected data from all over the country. Looking back to when I taught first grade and remembering graphs and projects I have seen my kindergarten teacher friends do, it is evident that kids get excited about collecting, organizing, and analyzing data from early on.

Now that I teach slightly older students, I thought I might take it a step further and delve into the age-old question, “What are the chances of that?” So the next time I heard that question from a particular student who is fond of asking it, I stopped and repeated it. “What are the chances of that?” I asked the students if they knew what it meant.

I started with vocabulary that went along with this question. We read a book titled *That’s a Possibility!* by Bruce Goldstone. From this book, I pulled vocabulary words such as probable, possible, impossible, most likely, and least likely. Then, I challenged the students to start using these words and thinking about the world around them like mathematicians do. We read through the situations in the book and the students discussed the questions posed—sometimes as a group and sometimes with just the people around them—to figure out if something was possible, probable, or impossible.

We then moved to discussing the actual chance of something, which was introduced in the book. I introduced fraction vocabulary to the class so they could talk about the chances, as in there is a 1 in 4 chance. After seeing the fraction that went with the particular situation (six marbles, one is white: $1/6$), the students discussed in groups what each of the numbers meant.



“I introduced fraction vocabulary to the class so they could talk about the chances, as in there is a 1 in 4 chance. After seeing the fraction that went with the particular situation (six marbles, one is white: $1/6$), the students discussed in groups what each of the numbers meant.”

There were many interesting ideas, and I was proud of how the students were thinking about the numbers. Some of the students figured out what the numbers meant and shared with the class. The best explanation was something similar to this: The bottom number tells the number of marbles available. The top number tells how many are white. They were so excited to be able to say there was a 1 in 6 chance and to know what that actually meant. The students made the connection between this one example and what the numbers in the fraction meant in a more general sense. They asked for example after example so they could figure out the chances. The students and I had a great deal of fun being statisticians!

So, the next time someone asks about the chances of something, take some time to figure it out. Or ask a kid—they are smarter than some people give them credit for!

2015 Poster and Project Competitions



Introduce your K–12 students to statistics through the annual poster and project competitions directed by the ASA/NCTM Joint Committee on Curriculum in Statistics and Probability. The competitions offer opportunities for students to formulate questions and collect, analyze, and draw conclusions from data. Winners will be recognized with plaques, cash prizes, certificates, and calculators. Also, their names will be published in *Amstat News*. Posters (grades K–12) are due every year on April 1. Projects (grades 7–12) are due on June 1. For more information, visit www.amstat.org/education/posterprojects.

Note: There is an updated rubric for the project competition and new additional guidance under the project competition rules link posted at www.amstat.org/education/posterprojects.

Power and Effect Size

Douglas K. Rush, Saint Louis University

A biology teacher gives her students a field assignment. She asks them to go to the small pond behind the school and determine whether there is life in the pond and, if life is present, to bring samples to the classroom.

The students observe the pond. They see no fish, tadpoles, frogs, turtles, water bugs, or other forms of life. They return and report that there is no life in the pond.

The teacher then asks the students to go back to the pond and bring a dropper full of water to the classroom. On their return, she places a small sample of water on a slide and puts it under a microscope. Observation of the water through the microscope reveals flagellates, amoebae, ciliates, and other protozoan life.

I frequently use this example to stimulate discussion of statistical power and effect size. In the first example, the teacher failed to designate the size of life (the effect size) she wanted her students to try to detect. When they looked into the pond, their eyes did not have the ability (power) to detect the small life actually present. As a result, the students reached the incorrect conclusion that there was no life in the pond.

A Refresher on Making Inferences and Testing Hypotheses

There are techniques for making inferences by which researchers try to gain an understanding of a population characteristic (called a parameter) by studying data from a sample randomly drawn from that population. Common techniques for making inferences include the following:

- Techniques that attempt to find differences between groups (e.g., Do ninth grade girls have a different mean, or average, algebra test score when compared to ninth grade boys?). Common techniques include t-tests and analysis of variance (ANOVA).
- Techniques that attempt to find relationships between a dependent variable and one or more independent variables (e.g., Can we predict first-year college grade point averages based on ACT/SAT scores, high-school grade point averages, social economic status, and the education of the students' parents?). Common techniques to detect relationships include correlation and multiple regression.

Researchers making inferences typically pose a research question and then create an appropriate null hypothesis (H_0) and alternate hypothesis (H_a). These two hypotheses form alternate possibilities of reality, only one of which can be true. In biomedical and science research, the alternate hypothesis is called the research hypothesis.

Using our algebra test score example, the null hypothesis would be: “There is no difference in ninth-grade algebra test scores when comparing the mean score of girls to the mean score of boys” [H_0 : mean score of girls = mean score of boys].

The alternate hypothesis would be: “There is a difference in ninth-grade algebra test scores when comparing the mean score of girls to the mean score of boys” [H_a : mean score of girls \neq mean score of boys].

In this example, researchers would draw a random sample of ninth-grade student algebra scores, compute the sample mean algebra score for boys and the sample mean algebra score for girls, and then use an independent sample t -test to make an inference about whether there is a difference in the population mean algebra test scores when comparing all ninth-grade boys and girls in the population.

Type I and Type II Errors

Researchers may draw erroneous conclusions when they use statistical techniques to make inferences about a population parameter (e.g., mean algebra test scores) based on collecting data from a sample selected from the population. No matter how carefully selected, the sample will not exactly mirror the population from which it is selected. One sample mean score selected from a population will differ from other sample mean scores selected from the same population, as well as from the true population mean score. This is referred to as sampling error. As a result, researchers know there will be error in the estimate of a population parameter when based on a sample taken from the population.

There are two possible types of error when testing a hypothesis. A Type I (α) error is erroneously rejecting a true null hypothesis (H_0) in favor of a false alternate hypothesis (H_a). Conversely, a Type II (β) error is failing to reject a false null hypothesis (H_0) in favor of a true alternate hypothesis (H_a).

Consider Table 1, which has been aptly named a “confusion matrix.”

Table 1: Confusion Matrix

		Reality	
		H_0 True	H_a True
Researcher's Statistical Decision	H_0 True	Correct Decision Probability* $1 - \alpha = .95$	Type II β Error Probability** $\beta = .20$
	H_a True	Type I α Error Probability* $\alpha = .05$	Correct Decision Probability** $1 - \beta = .80$

*Assuming researcher uses a significance level of .05 for statistical decision

**Assuming researcher selects sample size to yield power = .80

The reality in any hypothesis test is that either the null hypothesis (H_0) is true or the alternate hypothesis (H_a) is true. The researcher uses an inferential technique to reach a statistical decision to either retain or reject the null hypothesis. However, the researcher never knows the reality of whether the null or the alternate hypothesis is true. The researcher's decision is based on how strongly the sample data supports the null hypothesis.

As noted above, in most social science research, a researcher conducting an inferential statistical technique will not reject the null hypothesis in favor of an alternate hypothesis unless the (α) significance level is .05 or below. As we can see in Table 1, if the researcher conducts an inferential test using an alpha (α) .05 decision level, there is a $1 - \alpha = 1 - .05 = .95$ probability of making a "correct" decision to retain the null hypothesis where the null hypothesis is, in reality, true [See Researcher's Statistical Decision: H_0 True and Reality: H_0 True]. However, the researcher also has a .05 probability of making a Type I (α) error that is the incorrect decision to reject a true null hypothesis in favor of a false alternate hypothesis [See Researcher's Statistical Decision: H_a True and Reality H_0 : True]. This is also known as a "false positive" in medical research.

Conversely, a Type II (β) error is failing to reject a false null hypothesis in favor of a true alternate hypothesis [See Researcher's Statistical Decision H_0 : True and Reality H_a : True]. A Type II error is also known as a "false negative" in medical research.

Statistical Power Analysis

Power is the probability of not making a Type II error. Power is defined as $(1 - \beta)$, where β is the probability of making a Type II error. A power of .80 would provide an 80% probability of rejecting a false null hypothesis in favor of a true alternate hypothesis. Unfortunately, a power of .80 also leaves a probability of .20 of making a Type II error, which is failing to reject a false null hypothesis in favor of a true alternate hypothesis.

Statistical power is a function of three factors:

- 1 *Effect size*
- 2 *Sample size*
- 3 *Significance level alpha (α)*

In preparing a study, a researcher selects a desired power level (usually .80 in social science research) and alpha (α) level (typically .05) and designates the effect size the researcher wishes to detect. The researcher can then calculate the sample size necessary to conduct the study. Power is increased by designating larger effect sizes, choosing larger sample sizes, or increasing significance levels (α).

Effect Size

Once the research question is developed and the hypotheses stated, a researcher needs to determine what magnitude of difference or strength of relationship (effect size) is important or of interest for the purpose of the study. By convention, we use an alpha (α) level of .05 and a power of .80 in social science research. Therefore, we can calculate the required number of

subjects in our study (sample size) if we identify the effect size of interest in our study. What is the effect size the researcher seeks to detect using sample data if the effect is truly present in the population? Think back to our algebra example. What magnitude of difference in algebra scores (effect size) was the school district interested in detecting when comparing the mean algebra score for boys to the mean algebra score for girls? Let's assume for the purpose of this example that the algebra test is not being given to all ninth graders in the district.

This is the point at which the calculation becomes a little difficult. Achievement tests are usually reported in raw or scaled scores. Let's assume our hypothetical algebra test is reported in raw scores that are normally distributed, have a range from 0 to 100, and have a mean score of 70 and standard deviation of 10. One common measure of effect size for mean differences is known as Cohen's d . Cohen's d is a measure of difference in standard deviation units, rather than raw score units.

Researchers commonly use effect sizes based on standard deviation units because it allows them to compare scores that are not based on the same scale. Consider how a researcher would compare ACT test scores, which typically have a mean of 24 and an approximate standard deviation of 4, with SAT test scores, which have a mean of 500 and an approximate standard deviation of 100? The researcher would transform each score to a standardized z score, which is based on standard deviation units, and then compare the ACT z score to the SAT z score.

Returning to our algebra test example, what would be an important difference (effect size) in algebra scores when comparing boys' mean test scores with girls' mean test scores? One answer may be how large the difference in mean algebra test scores needs to be to result in the school district undertaking corrective action. This becomes a policy question. It is doubtful a school district would want to develop and fund a special remedial program if it found there was a one-point difference in algebra test scores when comparing boys with girls.

But what if there was a five-point difference? If a district found a five-point difference in algebra test scores when comparing boys' scores with girls' scores, would it undertake a program to try to identify why such a difference existed and expend the necessary funds to implement corrective or remedial programs? Let us suppose the district decided prior to the study that if a five-point difference exists, corrective action would be taken.

Sample Size

In our opening biology example, the students did not ask the teacher what size of life (effect size) they were to try to detect in the pond. As a result, their eyes did not have the power to see the life that was actually present in the pond. The students made a Type I error. They failed to reject a false null hypothesis (no life in the pond) in favor of a true alternate hypothesis (there is life in the pond).

Similarly, students frequently make errors in their proposed statistical research because they do not consider the magnitude or size of the relationships or differences (effect size) they are

trying to detect and, as a result, fail to properly calculate the number of subjects needed in their study to detect effects that are actually present. Students will ask how many subjects are enough in their study. My response is, “What effect size are you trying to detect?”

Now let’s examine our algebra example. How many students, randomly selected with equal numbers of boys and girls, do we need to give the algebra test to detect (infer) whether a five-point score difference exists in our school district? We first need to convert our raw score difference to standard deviation units. Recall that our hypothetical algebra test had a standard deviation of 10.

The formula for calculating an effect size difference in standard deviation units is:

$$\text{Cohen's } d = \frac{\mu_1 - \mu_2}{\sigma}$$

Where μ_1 = girls algebra mean score; μ_2 = boys mean algebra score; and σ = population standard deviation (assuming equal σ for boys and girls).

A five-point score difference becomes:

$$\text{Cohen's } d = \frac{5}{10} = .50$$

Suppose the district decided a two-point difference would be enough to develop and implement a corrective action plan. How many students, selected at random with equal numbers of boys and girls, need to take the algebra test to detect such a two-point difference if the difference really exists in our school district? Again, we convert the raw or scaled score to standard deviation difference units.

$$\text{Cohen's } d = \frac{2}{10} = .20$$

Now we have all the information necessary to calculate the number of students we need to test to answer the research question. We will use a power of .80, an alpha of .05, and a Cohen’s d (effect size) of either .50 or .20, depending on what difference the school district determines will call for corrective action.

There are several free online statistical power calculators available to complete this sample size calculation. However, Cohen has come to our rescue and created a simple power chart that can be used for most common .80 power calculations. An abbreviated version of his table is shown as Table 2.

Cohen (1992) has suggested that an effect size of .20 standard deviations should be considered a small effect, an effect size of .50 standard deviations should be considered a medium

Further Information on Effect Size

I suggest reading Cohen’s (1992) article, which is cited in the Further Reading section, for a better understanding of the use of Table 2. I also would point out two cautions in the table’s use. First, note that Cohen uses different effect size measures that are appropriate to different tests. For example, the “Sig r^* ” row uses effect sizes based on Pearson correlation coefficients (r) rather than Cohen’s d standard deviation units as the appropriate effect size.

Second, note the use of the asterisk (*) for Sig r^* and Multiple Regression (Multi R^*) in the “Test” column in Table 2. Those sample sizes refer to the total sample size needed. The rows in Table 2 that represent tests not marked with an asterisk refer to the total sample size needed “per group.”

I would like to briefly discuss two other common effect sizes you may see referred to in social science research: the Pearson product-moment correlation coefficient and eta squared (η^2).

The Pearson product-moment correlation coefficient (r) represents the linear relationship, or correlation, between two continuous numeric variables. For example, what is the relationship or correlation between age (X variable) and price (Y variable) of used cars? Pearson’s r ranges from -1.0 to $+1.0$. The negative (-) and positive (+) signs represent the direction of the relationship, and the numerical values represent the strength of the relationship. A negative relationship or correlation would mean that as the value of the X variable (age) increased, the value of the Y variable (price) decreased. A positive relationship or correlation would mean that as the value of one variable increased, the value of the other variable increased. The strength of the relationship or correlation increases as it moves from 0 to either $+1.0$ or from 0 to -1.0 .

Cohen (1988) proposes that “small” effect sizes are $r = \pm .10$ to $.29$, “medium” effect sizes are $r = \pm .30$ to $.49$, and “large” effect sizes are $r = \pm .50$ to 1.0 .

It is also interesting to note that r^2 is the coefficient of determination. It represents the variance in the Y variable, which is explained by the X variable. For example, assume the Pearson r between the age of used cars and the price of used cars is $-.70$. Squaring $r = -.70$ results in a coefficient of determination of $.49$, which informs us that 49% of the price of a used car is explained by its age.

Eta squared (η^2) is a measure of the proportion of the variance in the numeric continuous dependent variable, which is explained by the difference in group membership. For example, what is the proportion of variance in algebra test scores explained by being a boy as opposed to being a girl? Eta squared effect sizes can be used for both t-tests and analysis of variance (ANOVA). Cohen (1988) proposes an eta squared of $.01$ represents a small effect size, an eta squared of $.06$ represents a medium effect, and an eta squared of $.14$ represents a large effect.

effect, and an effect size of .80 standard deviations should be considered a large effect.

For our algebra study, we have two effect sizes of interest: .20 and .50. The independent sample *t*-test is an inferential technique used to determine whether a difference exists in two population means when comparing two sample means. Referring to Table 2, the first row “Means” displays the required sample sizes for *t*-tests to detect small, medium, and large differences (effect sizes) at a .80 power level.

Reading across the “Means” row until we get to the alpha (α) .05 columns, we find we would need to have a sample of algebra

test scores from 393 girls and 393 boys if we were interested in detecting the “small” .20 Cohen’s *d* standard deviation difference in algebra test scores. However, if we wanted to detect the “medium” .50 Cohen’s *d* standard deviation difference, we would only need 64 boys’ scores and 64 girls’ scores.

As you can see, the “magic” number of subjects needed in a study really is kind of magical. It depends on the effect size, the sample size, and the alpha (which is related to beta). All are interrelated and important. Cohen’s *d* is one measure of effect size (and a common one); however, there are other ways to measure effect size and sample size depending on the type of study involved (see Further Information on Effect Size).

Table 2: Sample Sizes Required to Detect Small, Medium, and Large Effects for Selected Alpha (α) Levels for Power = .80

α									
Test	.01			.05			.10		
	Small	Medium	Large	Small	Medium	Large	Small	Medium	Large
Means	586	95	38	393	64	26	310	50	20
Sig r*	1,163	125	41	783	85	28	617	68	22
P dif	584	93	36	392	63	25	309	49	19
χ^2 1df	1,168	130	38	785	87	26	618	69	25
χ^2 2df	1,388	154	56	964	107	39	771	86	31
χ^2 3df	1,546	172	62	1,090	121	44	880	98	35
χ^2 4df	1,675	186	67	1,194	133	48	968	108	39
χ^2 5df	1,787	199	71	1,293	143	51	1,045	116	42
χ^2 6df	1,887	210	75	1,362	151	54	1,113	124	45
ANOVA 2 groups	586	95	38	393	64	26	310	50	20
ANOVA 3 groups	464	76	30	322	52	21	258	41	17
ANOVA 4 groups	388	63	25	274	45	18	221	36	15
ANOVA 5 groups	336	55	22	240	39	16	193	32	13
ANOVA 6 groups	299	49	20	215	35	14	174	28	12
Mult R* 2 var.	698	97	45	481	67	30			
Mult R* 3 var.	780	108	50	547	76	34			
Mult R* 4 var.	841	118	55	599	84	38			
Mult R* 5 var.	901	126	59	645	91	42			
Mult R* 6 var.	953	134	63	686	97	45			
Mult R* 7 var.	998	141	66	726	102	48			
Mult R* 8 var.	1,039	147	69	757	107	50			

Note: Means = *t* test for two independent sample means; Sig r = correlation; P dif = population proportion for two independent populations; Mult R = multiple regression.

*Sample size for Mult R and Sig r is total sample size. All other sample sizes are per group or cell.

Adapted with permission from “A Power Primer,” by J. Cohen, 1992, *Psychological Bulletin*, 112, p. 158.

Further Reading

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
 Cohen, J. 1992. A power primer. *Psychological Bulletin* 112(1):155-159. doi: 10.1037/0033-2909.112.1.155

Lenth, R. V. 2001. Some practical guidelines for effective sample size determination. *The American Statistician* 55(3):187-193.
 Weiss, N. A. 2008. *Elementary statistics* (7th ed.). Boston, MA: Pearson/Addison Wesley.

Additional resources accompanying this lesson also are posted.

Sampling in Archaeology

Mary Richardson, Grand Valley State University

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 can 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 boxplots).

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 boxplots 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 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 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

1–2 class periods

Materials Required

A copy of the activity sheet, a graphing calculator, and 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 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 an archaeologist faces after the discovery of an excavation site is how to determine the locations within the excavation site that will be dug to uncover artifacts. Obviously, digging everywhere within a site would be the maximal way to locate artifacts, but time and resources usually 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 to possible features (which have 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 to use an archaeological setting to

demonstrate the use of statistical sampling, it will be assumed a site will be sampled probabilistically. Further, it will be assumed 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 used in this activity. Simple random sampling is the foundation for all 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 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 8x8-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 (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 X	9	10
11	12	13X	14	15	16	17	18	19X	20X
21	22	23	24	25X	26	27	28	29	30
31	32	33	34	35X	36	37	38	39	40
41	42	43	44X	45	46	47X	48X	49	50
51X	52	53X	54X	55X	56	57	58X	59	60
61	62	63	64	65	66X	67	68	69	70
71	72	73	74X	75X	76	77	78	79X	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98X	99	100X

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 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 so they can have a classroom discussion about the results of selecting the different samples. Simple random sampling is performed using a SEED of 2,000. 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 1,981 is used. Ten test-pits are selected from each stratum. To obtain a sample of 20 test-pits, 1-in-5 systematic sampling is used with a SEED of 2,003. 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 them.

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

On Site 2, an X has been placed in the appropriate test-pits 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)

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	22	23	24	25	26	27	28	29	30
	X	X	X						
31	32	33	34	35	36	37	38	39	40
	X	X	X						
41	42	43	44X	45	46	47	48	49	50
	X	X							
51	52	53X	54X	55	56	57	58	59	60
	X								
61	62	63	64	65	66	67	68	69	70
	X	X	X						
71	72	73	74X	75	76	77	78	79	80
	X	X							
81	82	83	84	85	86	87	88	89	90
		X							
91	92	93	94	95	96	97	98	99	100

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

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 16 test-pits from Stratum I and four test-pits from Stratum II and then 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 16 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 five 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 repeated stratified random sampling of test-pits from this site would 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.

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 2,000, 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 1,981 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 2,003 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 2,004 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

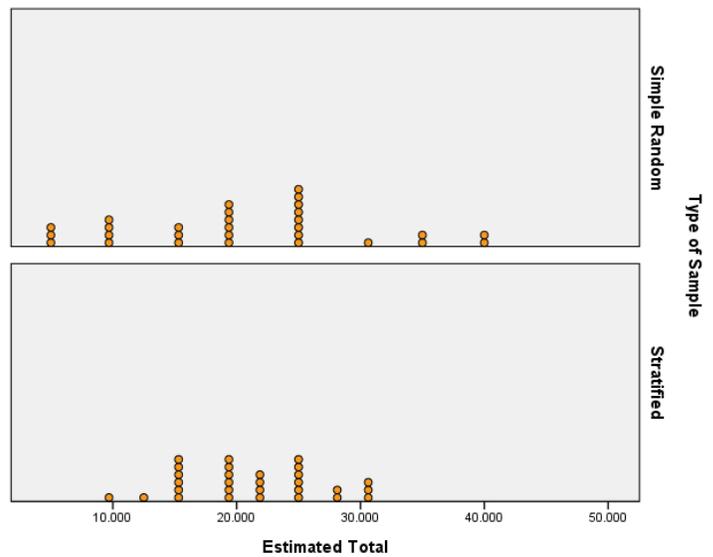


Figure 3. Dotplot of example class results for comparing simple random sampling to stratified random sampling

Stratified Random Sampling	Simple Random Sampling
mean = 21.23	mean = 20.69
standard deviation = 5.75	standard deviation = 9.79
first quartile = 15.625	first quartile = 12.50
median = 21.875	median = 20.00
third quartile = 25.00	third quartile = 25.00

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

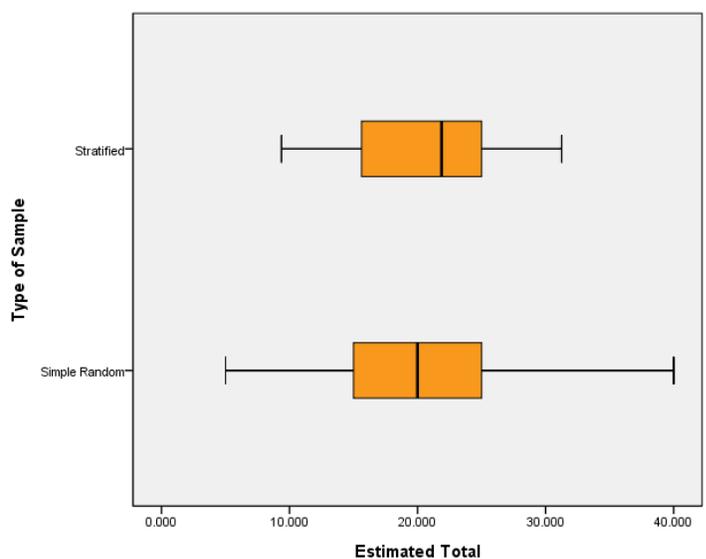


Figure 5. Comparative boxplots of simple random sampling and stratified random sampling estimated totals based on example class data

analyzed. Here are 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, 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.

IV. Interpret the Results

Ask students to use results from the data analysis to discuss whether they think 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 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 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 were 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:

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

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

Plan 3: Randomly select one orchard from the four 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: _____

Answer

Plan 1 describes simple random sampling.
Plan 2 describes stratified random sampling.
Plan 3 describes cluster sampling.

2. There are 16 first-class passengers scheduled on a flight. In addition to the usual security screening, four of the passengers will be subjected to a more complete search. Here is the first-class passenger list, denoted by which section 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 four passengers. Use a **SEED** of **845**.
 Selected passengers:

(b) Select a cluster sample of four passengers. Use the sections of passengers as clusters. Use a **SEED** of **332**.
 Selected passengers:

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

Answer:

- (a) 16, 4, 7, 8
- (b) Section 3: 9, 10, 11, 12
- (c) 4, 8, 12, 16

3. A population consists of 12 people. The two columns divide the population into two strata, labeled I and II. The population also is divided into three 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 four people was obtained. Listed below are three 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

Answer:

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

Possible Extensions

1. Ask students to place an X in the appropriate test-pits on a blank grid 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 Xs in the appropriate test-pits on a blank grid 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.

Further Reading

- Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K–12 Curriculum Framework*. Alexandria, VA: American Statistical Association. www.amstat.org/education/gaise.
- Richardson, M., and B. Gajewski. 2003. Archaeological sampling strategies. *Journal of Statistics Education* 11(1).
- Orton, C. 2000. *Sampling in archaeology*. Cambridge: Cambridge University Press.
- Lizee, J., and T. Plunkett. 1994. Archaeological sampling strategies. <http://archnet.asu.edu>.
- Aliaga, M., and B. Gunderson. 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 the Map of an Archaeological Site represents an area that contains 100 8 x 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 contains 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
11	12	13X	14	15	16	17	18	19X	20X
21	22	23	24	25X	26	27	28	29	30
31	32	33	34	35X	36	37	38	39	40
41	42	43	44X	45	46	47X	48X	49	50
51X	52	53X	54X	55X	56	57	58X	59	60
61	62	63	64	65	66X	67	68	69	70
71	72	73	74X	75X	76	77	78	79X	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98X	99	100X

Taking Samples

NOTE: To enter a seed value into the TI-84:

Choose any positive whole number (e.g., 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 2,000.** 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 two strata of equal size (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 1,981.**

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 2,003.**

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 two clusters (use the top row as Row 1 and use the bottom row as Row 10). **Use a SEED of 2,004.**

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

Comparison of Sampling Strategies

Xs 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 two strata (using columns 1 through 5 of test-pits for Stratum I and columns 6 through 10 of test-pits for Stratum II). Assume 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	13X	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
	X	X	X						
31	32	33	34	35	36	37	38	39	40
	X	X	X						
41	42	43	44X	45	46	47	48	49	50
	X	X							
51	52	53X	54X	55	56	57	58	59	60
	X								
61	62	63	64	65	66	67	68	69	70
	X	X	X						
71	72	73	74X	75	76	77	78	79	80
	X	X							
81	82	83	84	85	86	87	88	89	90
		X							
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 four 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 *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 four 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 =

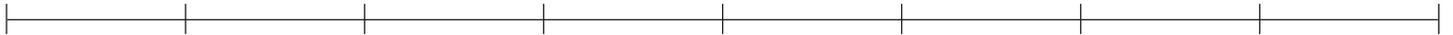
third quartile =

6. Construct side-by-side boxplots for the class estimated totals.

Stratified random sampling



Simple random sampling



7. Based on the above calculations, do you think *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?

ANNOUNCEMENTS

Significance Opens Archives

Significance magazine has opened its 10-year archives for access by the public. The magazine's volumes 1 through 10 are available to read, free of charge, at www.statslife.org.uk/significance/back-issues. Further, all magazine content will be made freely available one year after its initial publication. Editor Brian Tarran believes open access will demonstrate the importance of statistics and the contributions it makes in all areas of life. Royal Statistical Society and ASA members and subscribers will continue to enjoy exclusive access to the latest magazine content.



Useful Websites for Statistics Teachers

The ASA hosts a listing of websites useful for statistics teachers. The list was updated recently, though it is a work in progress. Visit the site at www.amstat.org/education/usefulsitesforteachers.cfm. If you have recommendations or additions, contact Rebecca Nichols at rebecca@amstat.org.

Episode 13 of STATS+STORIES Available

Episode 13 (“Reading, Writing, and Statistics? Data Analysis and Statistical Literacy for All”) of S+S is available at www.statsandstories.net. S+S guest Christine Franklin, Lothar Tresp Honoratus Honors Professor and undergraduate coordinator in statistics at the University of Georgia, joined the Stats+Stories regulars to talk about educating students to be statistically literate citizens. To listen, visit www.statsandstories.net or iTunes.

Census at School Reaches More Than 28,600 Students

The ASA's U.S. Census at School program (www.amstat.org/censusatschool) is a free international classroom project that engages students in grades 4–12 in statistical problem solving. The students complete an online survey, analyze their class census results, and compare their class with random samples of students in the United States and other participating countries. The project began in the United Kingdom in 2000 and now includes Australia, Canada, New Zealand, South Africa, Ireland, South Korea, and Japan. The ASA is seeking champions to expand the U.S. Census at School program nationally. For more information about how you can get involved, visit <http://magazine.amstat.org/blog/2012/02/01/censusatschool-2> or email Rebecca Nichols at rebecca@amstat.org.

Judges Sought for Statistics Project Competition

The ASA/NCTM Joint Committee on Curriculum in Statistics and Probability is seeking judges for the 2015 Statistics Project Competition (www.amstat.org/education/posterprojects). Judging takes place via email during the summer and requires about four hours of your time. If interested, email Daren Starnes, head judge, at dstarnes@lawrenceville.org.

ASA Statistics Education of Teachers Report Expected in April

In light of the Common Core State Standards, the Conference Board of the Mathematical Sciences (CBMS) released *The Mathematical Education of Teachers II (MET2)*, which focuses on the mathematics and statistics preparation of K–12 teachers. The ASA review of *MET2* was well received by CBMS, which encouraged the ASA to expand recommendations to a white paper. The ASA board recently funded this project to create a companion report on the statistics education of teachers, led by Christine Franklin and Tim Jacobbe, both members of the ASA/NCTM Joint Committee. For more information, see <http://magazine.amstat.org/blog/2014/03/01/education-of-teachers>. The report should be released online at www.amstat.org/education/SET in April.

Free Statistics Education Webinars

The ASA offers free webinars on K–12 statistics education topics at www.amstat.org/education/webinars. This series was developed as part of the follow-up activities for the Meeting Within a Meeting Statistics Workshop. The Consortium for the Advancement of Undergraduate Statistics Education also offers free webinars on undergraduate statistics education topics at www.causeweb.org.

PROJECT-SET

PROJECT-SET is an NSF-funded project to develop curricular materials that enhance the ability of high-school teachers to foster students' statistical learning regarding sampling variability and regression. All materials are geared toward helping high-school teachers implement the Common Core State Standards for statistics and are closely aligned with the learning goals outlined in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE): A Pre-K–12 Curriculum Framework*. For more information, visit <http://project-set.com>.

LOCUS

LOCUS (<http://locus.statisticseducation.org>) is an NSF-funded project focused on developing assessments of statistical understanding across levels of development as identified in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE)*. The intent of these assessments is to provide teachers, educational leaders,

ANNOUNCEMENTS

assessment specialists, and researchers with a valid and reliable assessment of conceptual understanding in statistics consistent with the Common Core State Standards.

World of Statistics Website and Resources

The free international statistics education resources created during the 2013 International Year of Statistics are available and ongoing through The World of Statistics website. Teachers everywhere can access a wealth of statistics instruction tools and resources from around the world at www.worldofstatistics.org.

STEW Lesson Plans

For free, peer-reviewed lessons, visit www.amstat.org/education/stew.

Explore Census at School Data with TuvaLabs

TuvaLabs provides free, real data sets, lessons, and visualization tools to enable teachers to teach statistics and quantitative reasoning in the context of real-world issues and topics. The ASA has provided TuvaLabs with a clean Census at School data set with 500 cases and 20 attributes that is now freely available in TuvaLabs for students and teachers to explore online with their visualization tool and Census at School–adapted lesson plans. Start exploring Census at School data with TuvaLabs at https://tuvalabs.com/datasets/census_at_school_clean_data/#/. Other TuvaLabs data sets and lessons are available at www.tuvalabs.com.

Meeting Within a Meeting (MWM) Statistics Workshop for Middle- and High-School Mathematics and Science Teachers – Seattle, Washington – August 11–12

MWM will take place in conjunction with the Joint Statistical Meetings this summer in Seattle, Washington. The workshop is meant to strengthen K–12 mathematics and science teachers' understanding of statistics and provide them with hands-on activities aligned with the Common Core State Standards that they can use in their own classrooms. The cost of the workshop is \$50. Online registration is available at www.amstat.org/education/mwm.

Beyond AP Statistics (BAPS) Workshop – Seattle, Washington – August 12

The ASA/NCTM Joint Committee is pleased to sponsor a Beyond AP Statistics workshop at the annual Joint Statistical Meetings. Organized by Roxy Peck, the BAPS workshop is offered for experienced AP Statistics teachers and consists of enrichment material just beyond the basic AP syllabus. The cost of the workshop is \$50. Online registration is available at www.amstat.org/education/baps.

UPCOMING CONFERENCES

National Council of Teachers of Mathematics (NCTM) Annual Meeting & Exposition
www.nctm.org/Conferences-and-Professional-Development/Annual-Meeting-and-Exposition
April 15–18, Boston, Massachusetts
Stop by the ASA booth in the exhibit hall for materials and resources.

U.S. Conference on Teaching Statistics (USCOTS)
www.causeweb.org/uscots
May 28–30, State College, Pennsylvania



Joint Statistical Meetings (JSM)
www.amstat.org/meetings/jsm/2015
August 8–13, Seattle, Washington



The Meeting Within a Meeting Statistics Workshop for Math and Science Teachers
www.amstat.org/education/mwm
August 11–12, Seattle, Washington (JSM 2015)



Beyond AP Statistics (BAPS) Workshop
www.amstat.org/education/baps
August 12, Seattle, Washington (JSM 2015)



HELP US RECRUIT THE **NEXT GENERATION** OF STATISTICIANS

The field of statistics is growing fast. Jobs are plentiful, opportunities are exciting, and salaries are high. So what's keeping more kids from entering the field?

Many just don't know about statistics. But the ASA is working to change that, and here's how you can help:

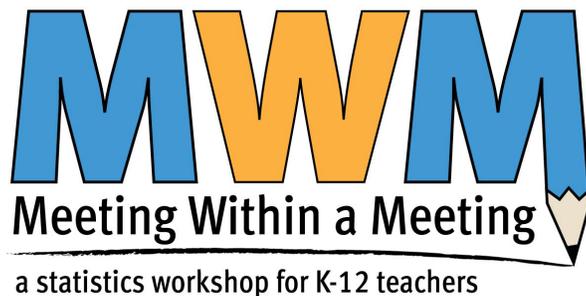
- Send your students to www.ThisIsStatistics.org and use its resources in your classroom. It's all about the profession of statistics.
- Download a handout for your students about careers in statistics at www.ThisIsStatistics.org/educators.



If you're on social media, connect with us at [www.Facebook.com/ThisIsStats](https://www.facebook.com/ThisIsStats) and www.Twitter.com/ThisIsStats. Encourage your students to connect with us, as well.

The site features include:

- Videos of young statisticians passionate about their work
- A myth-busting quiz about statistics
- Photos of cool careers in statistics, like a NASA biostatistician and a wildlife statistician
- Colorful graphics displaying salary and job growth data
- A blog about jobs in statistics and data science
- An interactive map of places that employ statisticians in the U.S.



MWM Statistics Workshop for Middle- & High-School Mathematics and Science Teachers

SPONSORED BY THE AMERICAN STATISTICAL ASSOCIATION (ASA)

www.amstat.org/education/mwm

Based on the Common Core State Standards for Mathematics (corestandards.org) and Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K–12 Curriculum Framework (www.amstat.org/education/gaise)

- Dates:** Tuesday, August 11, and Wednesday, August 12, 2015, 8:00 a.m. to 4:00 p.m.
- Place:** Seattle, Washington, Washington State Convention Center or nearby hotel (workshop meeting room TBD)
- Audience:** Middle- and high-school mathematics and science teachers. Multiple mathematics/science teachers from the same school are especially encouraged to attend.
- Objectives:** Enhance understanding and teaching of statistics within the mathematics/science curriculum through conceptual understanding, active learning, real-world data applications, and appropriate technology
- Content:** Teachers will explore problems that require them to formulate questions; collect, organize, analyze, and draw conclusions from data; and apply basic concepts of probability. The MWM program will include examining what students can be expected to do at the most basic level of understanding and what can be expected of them as their skills develop and their experience broadens. Content is consistent with Common Core standards, GAISE recommendations, and *NCTM Principles and Standards for School Mathematics*.
- Presenters:** GAISE report authors and prominent statistics educators
- Format:** Middle-school and high-school statistics sessions. Activity-based sessions, including lesson plan development
- Provided:** Refreshments. Handouts. Certificate of participation from the ASA certifying professional development hours. Optional graduate credit available
- Cost:** The fee for the two days is \$50. Attendees do not need to register for the Joint Statistical Meetings* to participate.
- Follow up:** Follow-up activities and webinars (www.amstat.org/education/k12webinars). Networking with statisticians and teachers to organize learning communities
- Registration:** More information and online registration is available at www.amstat.org/education/mwm. Space is limited. If interested in attending, please register as soon as possible.
- Contact:** Rebecca Nichols at rebecca@amstat.org or (703) 684-1221, Ext. 1877

*The Joint Statistical Meetings is the largest annual gathering of statisticians, where thousands from around the world meet to share advances in statistical knowledge. The JSM activities include statistics education sessions, poster sessions, and the exhibit hall.



A Workshop for Experienced Teachers

Sponsor: ASA-NCTM Joint Committee on Curriculum in Statistics and Probability

Wednesday, August 12, 2015 | 8:00 a.m. – 4:30 p.m. | Seattle, Washington

The ASA/NCTM Joint Committee is pleased to sponsor a Beyond AP Statistics (BAPS) Workshop at the annual Joint Statistical Meetings* in Seattle, Washington, August 12, 2015. Organized by Roxy Peck, the BAPS Workshop is offered for AP Statistics teachers and consists of enrichment material just beyond the basic AP syllabus. The course is divided into four sessions led by noted statisticians. Topics in recent years have included experimental design, topics in survey methodology, multiple regression, logistic regression, what to do when assumptions are not met, and randomization tests.

Cost

The fee for the full day is \$50. Attendees do not need to register for the Joint Statistical Meetings (JSM) to participate in this workshop, although there is discounted JSM registration for K–12 teachers available at www.amstat.org/meetings/jsm/2015.

Location

Seattle, Washington, Washington State Convention Center or nearby hotel (meeting room location to be announced)

Provided

- Refreshments (lunch on your own)
- Handouts
- Pass to attend the exhibit hall at the Joint Statistical Meetings
- Certificate of participation from the ASA certifying professional development hours
- Optional graduate credit available

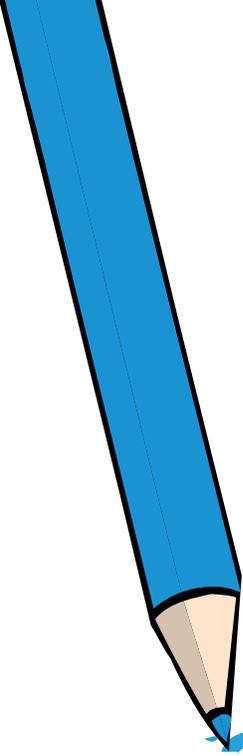
Registration

More information and online registration is available at www.amstat.org/education/baps. Registrations will be accepted until the course fills, but should arrive no later than July 21, 2015. Space is limited. If interested in attending, please register as soon as possible.

Questions

Contact Rebecca Nichols at rebecca@amstat.org or (703) 684-1221, Ext. 1877.

*The Joint Statistical Meetings is the largest annual gathering of statisticians, where thousands from around the world meet to share advances in statistical knowledge. JSM activities include statistics and statistics education sessions, poster sessions, and the exhibit hall.



Statistics Education Web

STEW

Lesson Plans Available on Statistics Education Web for K–12 Teachers

Statistics Education Web (STEW) is an online resource for peer-reviewed lesson plans for K–12 teachers. The lesson plans identify both the statistical concepts being developed and the age range appropriate for their use. The statistical concepts follow the recommendations of the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*, *Common Core State Standards for Mathematics*, and *NCTM Principles and Standards for School Mathematics*. The website resource is organized around the four elements in the GAISE framework: formulate a statistical question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the data in the context of the original question. Teachers can navigate the site by grade level and statistical topic. Lessons follow Common Core standards, GAISE recommendations, and NCTM Principles and Standards for School Mathematics.

Lesson Plans Wanted for Statistics Education Web

The editor of STEW is accepting submissions of lesson plans for an online bank of peer-reviewed lesson plans for K–12 teachers of mathematics and science. Lessons showcase the use of statistical methods and ideas in science and mathematics based on the framework and levels in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE)* and *Common Core State Standards*. Consider submitting several of your favorite lesson plans according to the STEW template to steweditor@amstat.org.

For more information, visit www.amstat.org/education/stew.





Teaching Statistics Through Data Investigations

New Online Course Starting March 9

www.mooc-ed.org/tsdi

Learn with colleagues near and far in this 8-week, online professional development course, designed for teachers of statistics in grades 6-12 and post-secondary contexts. This course can help you learn to teach statistics using investigations, with real data and real cool tools! The course is **FREE** and can lead to continuing education credits.

Offered through the Friday Institute for Educational Innovation at NC State University Lead instructor: **Dr. Hollylynn Lee**, Professor of Mathematics Education





FREE international classroom project to engage students in statistical problem solving

Teach statistical concepts, statistical problem solving, measurement, graphing, and data analysis using your students' own data and data from their peers in the United States and other countries.

Complete a brief online survey (classroom census)

13 questions common to international students, plus additional U.S. questions
15–20-minute computer session

Analyze your class results

Use teacher password to gain immediate access to class data.

Formulate questions of interest that can be answered with Census at School data.

Collect/select appropriate data

Analyze the data—including appropriate graphs and numerical summaries for the corresponding variables of interest

Interpret the results and make appropriate conclusions in context relating to the original questions.

Compare your class census with samples from the United States and other countries

Download a random sample of Census at School data from United States students.

Download a random sample of Census at School data from international students (Australia, Canada, New Zealand, South Africa, and the United Kingdom).

International lesson plans are available, along with instructional webinars and other free resources.

www.amstat.org/censusatschool

For more information about how you can get involved,
email Rebecca Nichols at rebecca@amstat.org.

Bridging the Gap

Between
Common Core State Standards
and Teaching Statistics



Twenty data analysis and probability investigations for
K–8 classrooms based on the four-step statistical process as defined by the
Guidelines for Assessment and Instruction in Statistics Education (GAISE)

www.amstat.org/education/btg



MAKING SENSE OF STATISTICAL STUDIES



Finally,
a statistics book
that makes sense!



You're talking
and I understand
everything!



Making Sense of Statistical Studies consists of student and teacher modules containing 15 hands-on investigations that provide students with valuable experience in designing and analyzing statistical studies. It is written for an upper middle-school or high-school audience having some background in exploratory data analysis and basic probability.

www.amstat.org/education/msss