Digital Commons @ LMU and LLS

Digital Commons@
Loyola Marymount University
and Loyola Law School

Mathematics Faculty Works | Mathematics

1-1-2014

Simulation of the Sampling Distribution of the Mean Can Mislead

Ann E. Watkins
California State University, Northridge

Anna E. Bargagliotti
Loyola Marymount University, anna.bargagliotti@lmu.edu

Christine Franklin
University of Georgia

Repository Citation

Watkins, Ann E.; Bargagliotti, Anna E.; and Franklin, Christine, "Simulation of the Sampling Distribution of the Mean Can Mislead" (2014). *Mathematics Faculty Works*. 86.
http://digitalcommons.lmu.edu/math_fac/86

Recommended Citation

Watkins, Ann E., Anna Bargagliotti, and Christine Franklin. "Simulation of the Sampling Distribution of the Mean Can Mislead." Journal of Statistics Education 22.3 (2014), www.amstat.org/publications/jse/v22n3/watkins.pdf.

Citation / Publisher Attribution

Permission has been granted by the authors to supply this article for educational and research purposes. More info can be found about the Journal of Statistics Education at https://ww2.amstat.org/publications/jse/. Copyright © 2014 by Ann E. Watkins, Anna Bargagliotti, and Christine Franklin, all rights reserved.

Simulation of the Sampling Distribution of the Mean Can Mislead

Ann E. Watkins
California State University, Northridge

Anna Bargagliotti
Loyola Marymount University

Christine Franklin
University of Georgia

Journal of Statistics Education Volume 22, Number 3 (2014),
www.amstat.org/publications/jse/v22n3/watkins.pdf

Copyright © 2014 by Ann E. Watkins, Anna Bargagliotti, and Christine Franklin, 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: Simulated sampling distribution; Sampling variability; Variance of means; Variance of variances; Central Limit Theorem.

Abstract

Although the use of simulation to teach the sampling distribution of the mean is meant to provide students with sound conceptual understanding, it may lead them astray. We discuss a misunderstanding that can be introduced or reinforced when students who intuitively understand that "bigger samples are better" conduct a simulation to explore the effect of sample size on the properties of the sampling distribution of the mean. From observing the patterns in a typical series of simulated sampling distributions constructed with increasing sample sizes, students reasonably—but incorrectly—conclude that, as the sample size, n, increases, the mean of the (exact) sampling distribution tends to get closer to the population mean and its variance tends to get closer to σ^2/n, where σ^2 is the population variance. We show that the patterns students observe are a consequence of the fact that both the variability in the mean and the variability in the variance of simulated sampling distributions constructed from the means of N random samples are inversely related, not only to N, but also to the size of each sample, n. Further, asking students to increase the number of repetitions, N, in the simulation does not change the patterns.

1. The Sampling Distribution of the Mean in the Classroom

Sampling distributions are the bridge from summary and display of a random sample to inference about the population from which the sample was taken, so instructors in introductory statistics courses devote much time and effort to helping students understand them. Instruction usually focuses on the most important of sampling distributions, the sampling distribution of the mean, and its special case, the sampling distribution of the proportion. As preparation for statistical inference, students learn three properties of the sampling distribution of the mean:

> The sampling distribution of the mean (SDM), for random samples of size *n* selected from a population with mean μ and (finite) standard deviation σ, has
>
> 1. mean, $\mu_{\bar{X}_n}$, equal to the mean of the population: $\mu_{\bar{X}_n} = \mu$.
> 2. standard deviation, $\sigma_{\bar{X}_n}$, equal to the standard deviation of the population divided by the square root of the sample size: $\sigma_{\bar{X}_n} = \sigma/\sqrt{n}$.
> 3. (Central Limit Theorem) a shape that is normal if the population is normal; for other populations with finite mean and variance, the shape becomes more normal as *n* increases.

The first of these properties often is thought to be obvious to students, which perhaps it is for symmetric populations, so instruction centers on the second and third, which are of deeper theoretical interest. After all, "For means, it's centered at the population mean. What else would we expect?" (De Veaux, Velleman, and Bock 2012, p. 442). Similarly, in an online textbook (Lane 2014), all three properties are stated, but only the second and third are justified.

However, when the population is skewed, it certainly is not intuitively obvious to students that $\mu_{\bar{X}_n} = \mu$. For example, we presented 40 post-calculus students taking introductory statistics with the skewed distribution of the salaries of National Basketball Association players (ESPN.com 2013). After exploring the concept of a sampling distribution for various statistics, students were asked to predict whether the sampling distribution of the mean salary for $n = 10$ has a mean that is larger than, smaller than, sometimes larger and sometimes smaller than, or equal to the population mean of $4.5 million. Four students choose larger, 17 choose smaller (a sensible choice because most values in the skewed population are smaller than the mean), 8 choose sometimes larger and sometimes smaller, and only 11 of the 40 students correctly choose equal. Although they did not use this language, more than half of these post-calculus students thought that the mean of a random sample is a biased estimator of the mean of this population, at least for a sample size of 10. The choice of "smaller" by so many students is consistent with a prediction by Chance, delMas, and Garfield (2004), which they base on classroom observations, contributions of colleagues, and analysis of student performance.

We gave a similar problem to 41 students in an introductory class without a calculus prerequisite, this time after instruction about the SDM. The results were better, but hardly impressive, with only 23 of the 41 students saying that $\mu_{\bar{X}_n} = \mu$.

Results such as these may be a consequence of a specific misunderstanding about the mean of the SDM (and, sometimes, about the standard deviation). Many students incorrectly believe that the first property is this:

> The mean, $\mu_{\bar{X}_n}$, of the SDM more closely approximates the mean, μ, of the population *as the sample size increases.*

And, consistently, they also may incorrectly believe that the second property is this:

> The standard deviation, $\sigma_{\bar{X}_n}$, of the SDM more closely approximates $\sigma/\sqrt{n}$ *as the sample size increases.*

These misunderstandings may be invisible to the instructor because, in standard textbook exercises, the sample size is large enough for the Central Limit Theorem to come into play. With a large sample size, the student feels justified, not only in using a normal approximation to the SDM, but also in approximating $\mu_{\bar{X}_n}$ with μ and $\sigma_{\bar{X}_n}$ with $\sigma/\sqrt{n}$ in needed formulas.

How do these misunderstandings arise? One reason, the subject of this paper, is the use of simulation to demonstrate properties of the SDM. Other possible reasons will be discussed in Section 8.

2. Research About the Use of Simulation to Teach the SDM

Students find the concept of sampling distribution difficult to grasp. Through the use of simulation, instructors hope to demonstrate the properties of the SDM in a hands-on and intuitive manner that promotes conceptual understanding and appeals to students. They find wide support for using simulation to teach the SDM in numerous papers, textbooks, and online applets.

2.1. Sampling Distributions Are Difficult to Understand

Garfield and Ben-Zvi (2007) present an enlightening, and rather depressing, summary of the literature on how students learn statistics. For example, even students who complete an introductory college statistics course with high grades retain only a "disappointing" conception of the mean, standard deviation, and Central Limit Theorem.

Empirical research about specific conceptual difficulties among students is rare. In a thorough review of the literature on misconceptions about statistical inference published between 1990 and early 2006, Sotos, Vanhoof, Van den Noortgate, and Onghena (2007) found 500 references, but only 17 different research studies that provided empirical evidence (beyond personal observation) of misconceptions among university students. Those research studies confirm that students find the concept of sampling distributions and, specifically, the sampling distribution of the mean and the Central Limit Theorem, difficult to understand. Students tend to confuse sampling distribution, distribution of a (data) sample, and distribution of the population. Further, many students ignore the effect of sample size on the variability of sample means. (See also, for example, delMas, Garfield, and Chance 1999, Doerr and Jacob 2011, and Noll and Sharma 2014.)

2.2. Using Simulation to Teach the SDM Is Widely Recommended But Rarely Evaluated

Since at least 1960, a vast number of articles have been published that recommend using simulation to teach the SDM, particularly the Central Limit Theorem, to introductory statistics students. Early articles include Jowett and Davies (1960) and, in an ASA journal, Gentleman (1977). Undoubtedly the most influential was the May 1971 statement of the Committee on the Undergraduate Program in Mathematics (1972, p. 490) of the Mathematical Association of America:

> Since the Central Limit Theorem should only be stated and not proved, evidence of its operation will need to be given to the student. Some texts contain exact sampling distributions for different sample sizes or the results of sampling experiments. Printouts of computer runs simulating sampling distributions for different sample sizes can also be distributed to students and discussed. If a computer is not available on campus, printouts could be obtained from a computer located elsewhere.
>
> These approaches are helpful but, in our opinion, are not as effective as having students participate in sampling experiments. A simple experiment is to sample a rectangular distribution, either from a table of random numbers, by drawing chips from a bowl, or by computer. If a computer is used, it will also be easy to sample other kinds of populations. Sampling a moderately skew population may help convince students of the Central Limit Theorem in the absence of symmetry. Indeed, the use of several populations (e.g., rectangular, exponential) can demonstrate to the student that the rapidity with which the sampling distribution of $(\bar{x}-\mu)\div\left(\sigma/\sqrt{n}\right)$ approaches a normal distribution as n increases depends on the population from which the samples are selected.

Echoing these recommendations, articles describe how to simulate the SDM using a wide variety of physical objects, a graphing calculator, or a computer. The demonstrations tend to use skewed or bimodal populations, so that students are impressed with the counter-intuitive result. Invariably, the authors anticipate that "the student will observe that the center of the distribution remains about the same and the distribution becomes narrower. That is, as sample size gets larger the approximations to the mean do not get better, but the variability about the mean decreases." (Koehler 2006, pp. 264-265).

Rarely is the method evaluated or compared with a non-simulation approach, either with respect to student time needed or as to how well students understand sampling distributions (Chance et al. 2004). The formal research that has been conducted to compare student understanding of sampling distributions following instruction with and without simulation generally has found no difference or a modest difference in favor of simulation (Mills 2002; Meletiou-Mavrotheris 2003; Chance et al. 2004; Pfaff and Weinberg 2009).

2.3. Previous Warnings about Simulation and the SDM

While many researchers have discussed how misconceptions about sampling distributions can be challenged using simulation, we have found but two warnings about how conceptual difficulties can arise or be reinforced through the use of simulation. Hodgson and Burke (2000, p. 94) found that a computer simulation of the SDM resulted in 6 of their 18 students believing that "one *must* draw multiple samples in order to make valid statistical inferences." Hesterberg (1998) warns that simulations should have a large number of replications or else students "may have trouble distinguishing randomness due to random selection of data from randomness due to using small numbers of replications."

In Section 1, we described two misunderstandings we observed in our own students, that students believe that the mean of the SDM gets closer to the mean of the population as the sample size increases and the standard deviation gets closer to $\sigma/\sqrt{n}$ as the sample size increases. Lunsford, Rowell, and Goodson-Espy (2006) observed the second of these misunderstandings among their post-calculus introductory statistics students:

> In addition, we believed that some of our students confused the *limiting* result about the shape of the sampling distribution (i.e. as *n* increases the shape becomes approximately normal, via the CLT) with the fixed (i.e. nonlimiting) result about the magnitude of the variance of the sampling distribution …

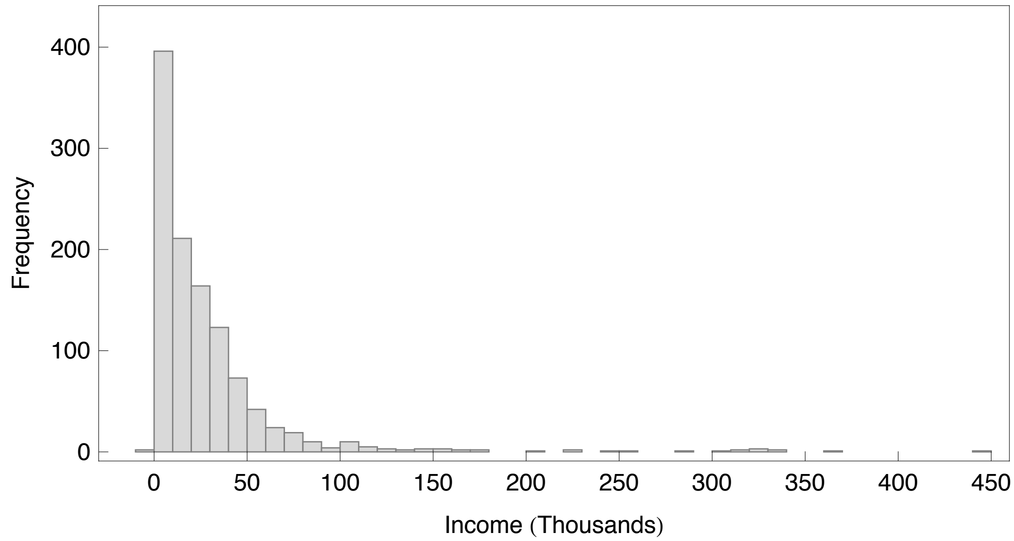
While Lunsford et al. note this misunderstanding, they do not connect its formation to the use of simulation by their students. In the following sections, we will show that the combination of student intuition that "larger samples are better" with the use of simulation turns out, not to be a marriage made in heaven for teaching the SDM, but rather a mismatch that leads some students, quite logically, into developing or reinforcing the misunderstandings described in Section 1 about the first and second properties of the SDM.

3. Results from the Classroom About the Estimated Mean of the SDM

Through an NSF-funded project, a professional development class was offered to a group of nine high school teachers, all of whom had some experience teaching statistics. The teachers spent five three-hour class meetings working on activities related to sampling distributions.

At the end of the fifth meeting, the teachers worked individually with a familiar population with known mean and standard deviation, the skewed incomes of the residents of "Mira Beach," shown in Figure 1. They were asked to construct three simulated SDMs, for sample sizes of 5, 15, and 30, using 100 random samples each, and to compute their means and standard deviations. The final task was, "Compare the three distributions that you constructed. What can you say about the shape of the distribution as the sample size, *n*, increases? What can you say about the mean? What can you say about the standard deviation?"

Figure 1. Incomes of the residents of "Mira Beach," with mean μ = 27,394 and standard deviation σ = 42,572.



When comparing the three simulated sampling distributions that they constructed, the teachers correctly were able to describe that, as the sample size increases, the variability of the simulated SDMs decreases and the shape becomes more approximately normal. However, when discussing the mean of the SDM, none of the teachers gave the description that we were expecting. Instead, most observed that the mean of the SDM *tends to get closer* to the mean of the population as the sample size increases. For example, the means of one teacher's three simulated sampling distributions are shown in Table 1. As the sample size increases, the means do, in fact, get closer to the population mean of 27,394. So the teacher wrote about the pattern in the three means of the SDMs,

As expected when the sample size increases the mean approaches the true mean.

Other teachers made similar statements, although several seemed surprised at the pattern they observed.

Table 1. Means of three simulated sampling distributions of the mean, each constructed using 100 random samples

Sample Size	Mean of Simulated SDM	Absolute Difference of the Mean of the Simulated SDM and Population Mean
5	23,472	3,922
15	25,704	1,690
30	27,601	207
Population mean	27,394	

In the next section we will show that the teachers were correct about the means of *simulated* SDMs—they *do* tend to get closer to the population mean as the sample size, n, increases.

4. Variability in the Mean of Simulated SDMs

While theory tells us that the mean of the SDM—a parameter—is equal to the population mean for all sample sizes, we do not expect the mean of a simulated SDM—an estimate of the parameter—to be exactly equal to the population mean. What is unexpected is that if $n_1 > n_2$, the mean of a simulated SDM constructed using N samples each of size n_1 tends to be closer to the population mean, μ, than the mean of a simulated SDM constructed using N samples each of size n_2. We will prove this result about the mean in this section and prove a similar result about the variance of simulated SDMs in Section 7 by analyzing the five related distributions that are summarized in Table 2.

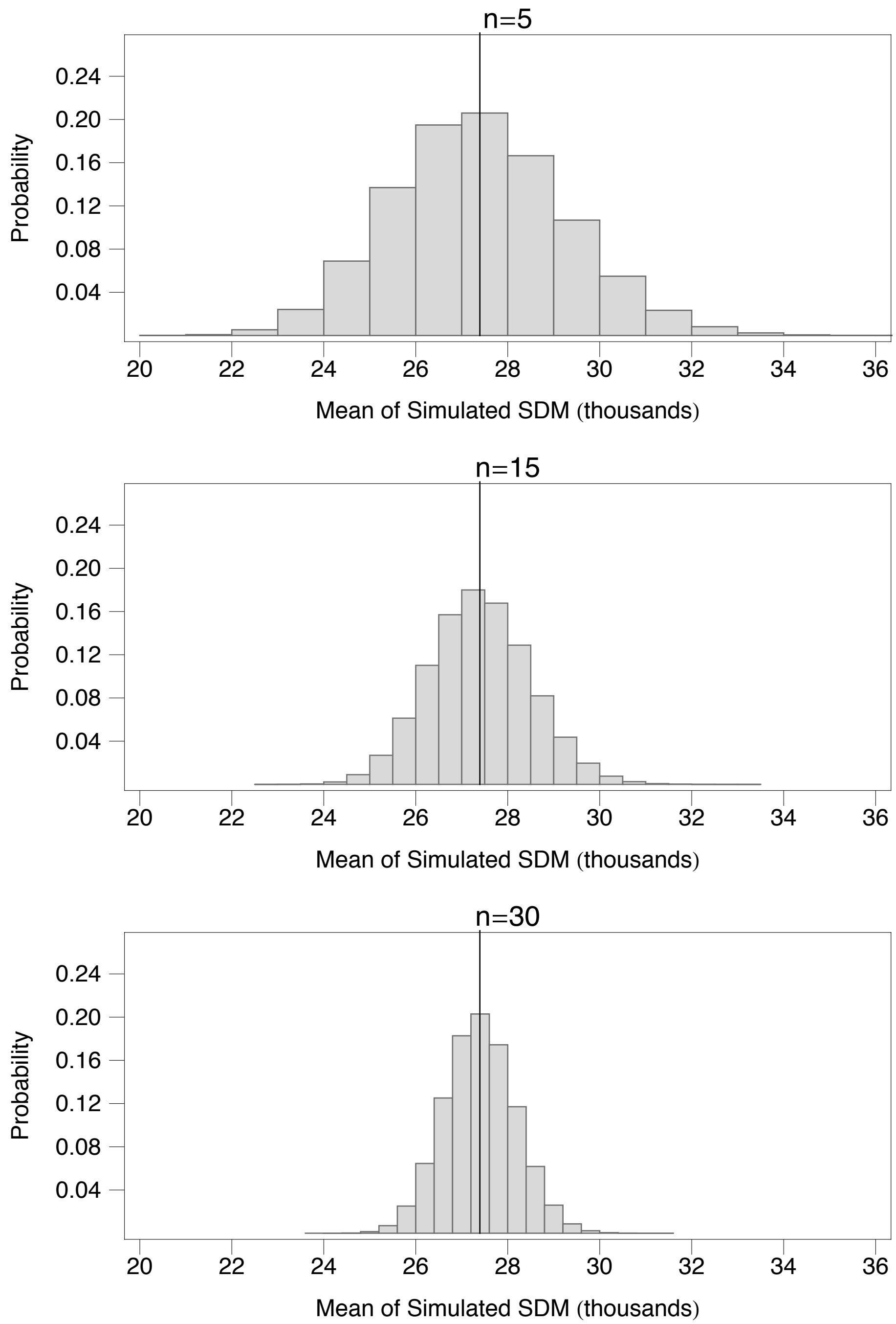
Table 2. The five distributions with mean and variance

Distribution	Notation	Mean	Variance
Population	X	μ	σ^2
SDM for samples of size n	$\overline{X}_n$	$\mu_{\overline{X}_n} = \mu$	$\sigma^2_{\overline{X}_n} = \sigma^2/n$
Simulated SDM from N samples each of size n	$\overline{X}_{n,N}$	$\overline{\overline{x}}_{n,N}$ (varies)	$s^2_{n,N}$ (varies)
Sampling distribution of the means of simulated SDMs	$\overline{\overline{X}}_{n,N}$	$E\left(\overline{\overline{X}}_{n,N}\right) = \mu$	$Var\left(\overline{\overline{X}}_{n,N}\right) = \sigma^2/nN$
Sampling distribution of the variances of simulated SDMs	$S^2_{n,N}$	$E\left(S^2_{n,N}\right) = \sigma^2/n$	$Var\left(S^2_{n,N}\right) \approx \dfrac{2\sigma^4}{n^2(N-1)}$ (if $\overline{X}_n$ is approx. normal)

So far, we have discussed two sampling distributions. The first is the (exact) SDM for samples of size n, denoted by $\overline{X}_n$, which has mean $\mu_{\overline{X}_n} = \mu$ and variance $\sigma^2_{\overline{X}_n} = \sigma^2/n$. The second is a simulated SDM, constructed by taking N random samples of size n from the population and computing the mean of each. Equivalently, a more useful way to describe the simulated SDM is that it consists of N values taken at random from $\overline{X}_n$.

Our third sampling distribution will be the (exact) sampling distribution, $\overline{\overline{X}}_{n,N}$, of the means of simulated SDMs. Figure 2 shows $\overline{\overline{X}}_{n,N}$ for $N = 100$ samples of size $n = 5$, $n = 15$, and $n = 30$ taken from the Mira Beach incomes. That is, each of the values in the histograms in Figure 2 is the mean of a simulated SDM constructed with $N = 100$ values randomly selected from $\overline{X}_n$. Note how much closer the means of the simulated SDMs tend to be to the population mean, $\mu_{\overline{X}_n} = \mu = 27{,}394$, as the sample size increases.

Figure 2. Three sampling distributions, $\overline{\overline{X}}_{n,N}$, of the means of simulated SDMs constructed using $N = 100$ samples taken from the Mira Beach incomes, for sample sizes of $n = 5$, 15, and 30. The vertical line is located at $\mu_{\bar{X}_n} = \mu = 27{,}394$.



More generally, because $\overline{\overline{X}}_{n,N}$ is a sampling distribution of a mean, composed of the means of random samples of N values taken from $\overline{X}_n$, we can apply the three properties in Section 1 to it. From the first property, $\overline{\overline{X}}_{n,N}$ has a mean equal to the mean of $\overline{X}_n$, which is μ. That is, $E\left(\overline{\overline{X}}_{n,N}\right)$ = μ. From the second property, $\overline{\overline{X}}_{n,N}$ has a variance equal to the variance of $\overline{X}_n$ divided by N,

$$Var\left(\overline{\overline{X}}_{n,N}\right) = \frac{\sigma^2/n}{N} = \frac{\sigma^2}{nN}$$

Thus, for fixed N, as the sample size, n, increases, the variance of $\overline{\overline{X}}_{n,N}$ decreases. Finally, from the Central Limit Theorem, $\overline{\overline{X}}_{n,N}$ will be normal or approximately so for a reasonably large number of samples, N, even if the sample size, n, is small.

It follows from the three properties of $\overline{\overline{X}}_{n,N}$ that the means of simulated SDMs do tend to get closer to the population mean, μ, as n increases, which is the pattern we see from the simulations summarized in Table 1.

To look at it in a different way, the mean of a simulated SDM, constructed from N random samples each of size n, can be found either by averaging the means of the N samples or by averaging the nN individual values. For example, when teachers computed the mean of a simulated SDM constructed from 100 random samples of size 5, in essence they were estimating $\mu_{\overline{X}_n} = \mu$ by averaging 500 randomly selected values. This observation makes it clear why $\overline{\overline{X}}_{n,N}$ is at least approximately normal and why nN is in the denominator of the formula for its variance.

5. The Simulation Cannot Be Fixed

To "fix" the results of a simulation gone wrong, an instructor's first impluse is to increase the number of repetitions in the simulation (see Hesterberg 1998, as quoted in Section 2.3). Certainly, increasing the number of samples, N, means that the simulated SDM should better approximate the exact SDM and the size of the difference, $\left|\overline{\overline{x}}_{n,N} - \mu\right|$, where $\overline{\overline{x}}_{n,N}$ is the mean of the simulated SDM, should get smaller. But unfortunately, increasing N cannot change the pattern students see that the difference tends to become smaller with increasing sample size.

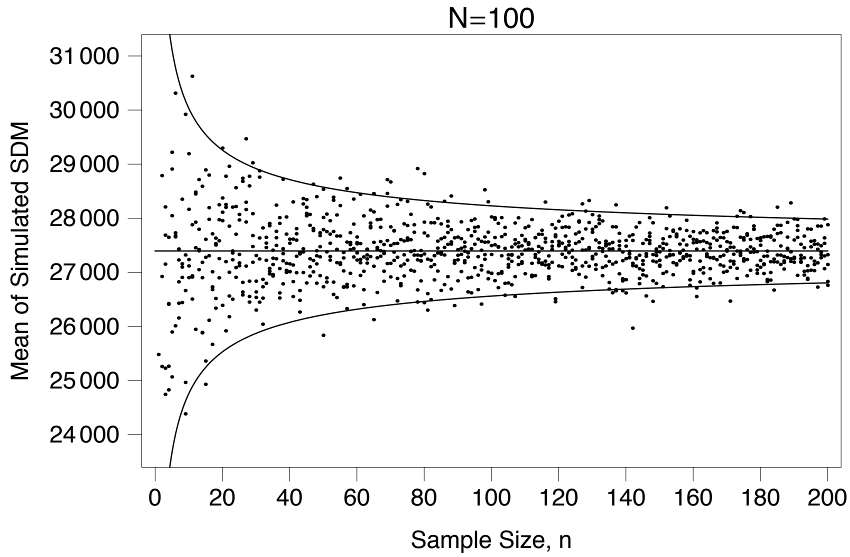
As we saw in Section 4, both the sample size, n, and the number of samples, N, contribute to the precision of the estimate of $\mu_{\overline{X}_n}$, so no matter how large N is, the estimate, $\overline{\overline{x}}_{n,N}$, of $\mu_{\overline{X}_n} = \mu$ from a simulation tends to be closer to $\mu_{\overline{X}_n} = \mu$ for larger n than for smaller n. An even stronger statement is true: For any two sample sizes, n_1 and n_2, the ratio of the variances of $\overline{\overline{X}}_{n_1,N}$ and $\overline{\overline{X}}_{n_2,N}$ does not depend on N,

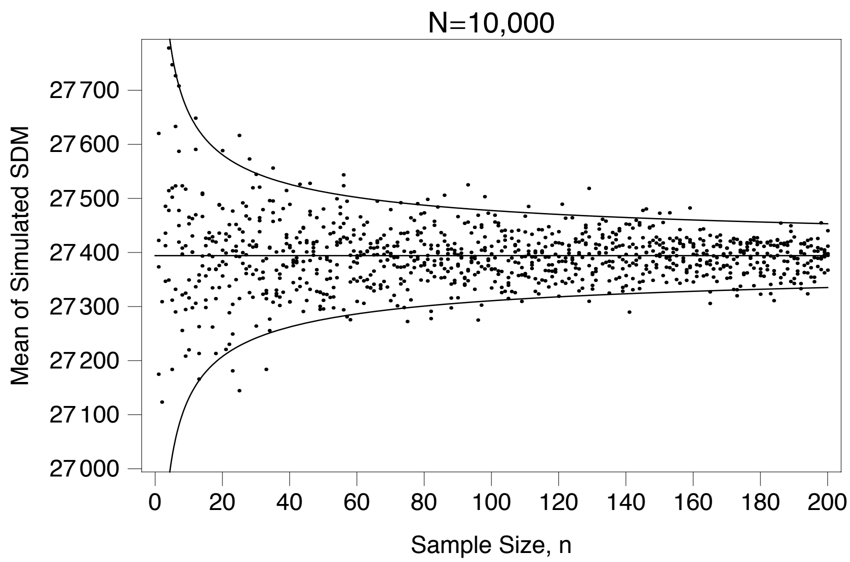
$$\frac{Var\left(\overline{\overline{X}}_{n_1,N}\right)}{Var\left(\overline{\overline{X}}_{n_2,N}\right)} = \frac{\sigma^2/n_1N}{\sigma^2/n_2N} = \frac{n_2}{n_1}$$

The implications of this fact are illustrated by the graphs in Figure 3 of means of simulated SDMs plotted against sample size, one for $N = 100$ and the other for $N = 10{,}000$. With the rescaling of the vertical axis, the graphs show exactly the same pattern. For example, while both $\left|\overline{\overline{x}}_{20,N} - \mu\right|$ and $\left|\overline{\overline{x}}_{100,N} - \mu\right|$ tend to be smaller for $N = 10{,}000$ than for $N = 100$, the probability that $\left|\overline{\overline{x}}_{20,N} - \mu\right| > \left|\overline{\overline{x}}_{100,N} - \mu\right|$ is the same for both $N = 100$ and $N = 10{,}000$.

Figure 3 also suggests that the misleading pattern will be especially persuasive to students if they plot their estimated means against the sample sizes, as is sometimes recommended in the literature. For example, the plots of estimated means versus sample sizes in Renolls and Massay (1991, p. 72) and Mulekar and Siegel (2009, p. 37 and 40) show a clear trend for the means of the simulated SDMs to get closer to the population mean as the sample size increases.

Figure 3. Means, $\bar{\bar{x}}_{n,N}$, of simulated SDMs plotted against sample size, *n*. Each point is a mean of the means of *N* random samples, each of size *n* taken from the Mira Beach incomes. The horizontal line is the population mean, $\mu = 27{,}394$. The curves are the graphs of $\bar{\bar{x}}_{n,N} = 27{,}394 \pm 1.96 \cdot \sigma / \sqrt{Nn}$, where σ = 42,572. The vertical axes are scaled to show that the pattern is similar for *N* = 100 and *N* =10,000.





6. Results from the Classroom About the Estimated Standard Deviation of the SDM

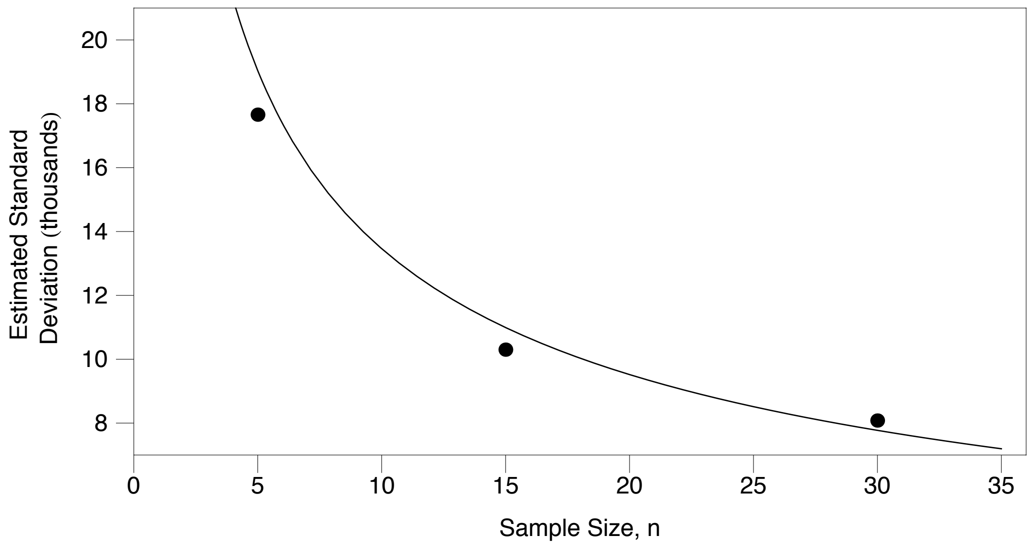
Similar to the behavior of the mean, the estimate, denoted $s_{n,N}$, from a simulated SDM of the standard deviation of the SDM tends to be closer to $\sigma_{\bar{X}_n} = \sigma/\sqrt{n}$ when the simulated SDM is constructed using *N* larger samples than when constructed using *N* smaller samples. For example, the estimated standard deviations in Table 3 come from the work of the teacher whose means are given in Table 1.

Table 3. Comparison of the estimated standard deviation from three simulated SDMs, each constructed using $N = 100$ samples, with the standard deviation of the SDM computed using $\sigma = 42{,}572$

Sample Size, n	Standard Deviation Estimated from Simulated SDM, $s_{n,N}$	Standard Deviation of the SDM, $\sigma/\sqrt{n}$	Absolute Difference $\left\|s_{n,N} - \sigma/\sqrt{n}\right\|$	Relative Difference $\dfrac{\left\|s_{n,N} - \sigma/\sqrt{n}\right\|}{\sigma/\sqrt{n}}$
5	17,665.3	19,038.7	1,373.4	.072
15	10,297.9	10,992.0	694.1	.063
30	8,077.0	7,772.5	304.5	.039

Similar to the case for the means in Table 1, the estimate from the simulated SDM of the standard deviation of the SDM gets closer to $\sigma/\sqrt{n}$ as the sample size increases. None of our teachers noticed this, however, because they were not prompted to compute the absolute or relative differences. Nor did we ask them to plot the estimates of the standard deviation from their simulated SDMs against the sample size *n* and compare to a graph of the exact standard deviation, as in Figure 4. Such graphs, however, are commonly recommended (see Renolls and Massay 1991 and Mulekar and Siegel 2009, for example), so it is quite possible that a perceptive introductory statistics student would observe that $s_{n,N}$ tends to get closer to $\sigma/\sqrt{n}$ as the sample size increases and make the incorrect generalization observed by Lunsford et al. (2006, as quoted in Section 2.3).

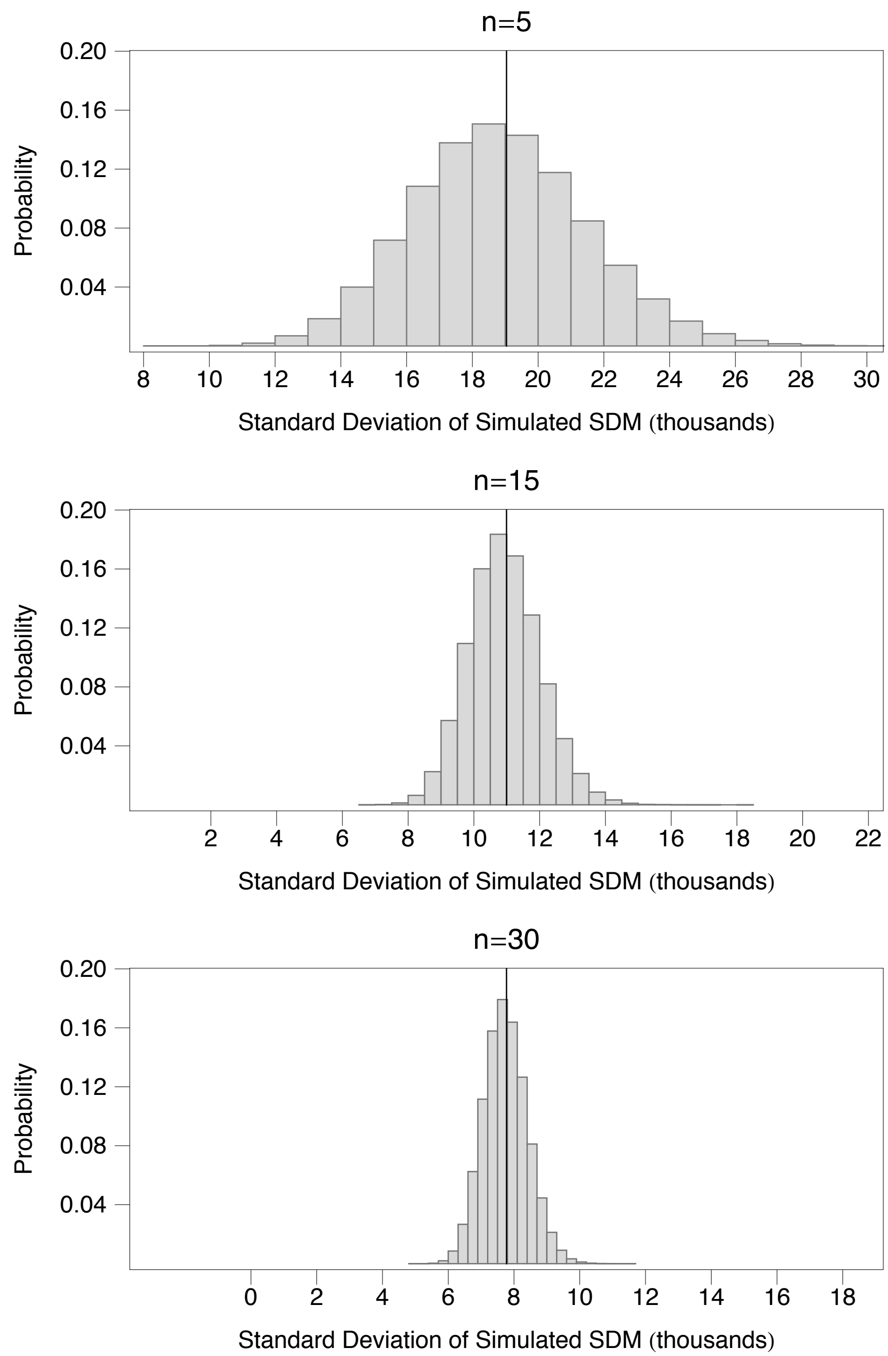
Figure 4. Estimated standard deviations from Table 3 plotted against sample size, n. As n increases, the estimates tend to get closer to $\sigma/\sqrt{n}$, shown by the curve.



7. Variability in the Standard Deviation of Simulated SDMs

Now, consider the (exact) sampling distribution, $S_{n,N}$, of the standard deviations, $s_{\bar{X}_{n,N}}$, of simulated SDMs. Figure 5 illustrates $S_{n,N}$ for $N = 100$ samples of size $n = 5$, $n = 15$, and $n = 30$ taken from the Mira Beach incomes. That is, each of the values in the histograms in Figure 5 is the standard deviation of a simulated SDM constructed with $N = 100$ values randomly selected from $\bar{X}_n$. A vertical line representing the standard deviation of the SDM, $\sigma/\sqrt{n}$, (from Table 3) is drawn on each plot. Note how, with increasing sample size, $s_{\bar{X}_{n,N}}$ tends to get closer to the standard deviation it is estimating.

Figure 5. Three sampling distributions of the standard deviations, $s_{n,N}$, of simulated SDMs constructed using $N = 100$ samples taken from the Mira Beach incomes, for samples of size 5, 15, and 30. The vertical line is located at $\sigma/\sqrt{n}$.



More generally, suppose that the population, with mean μ and standard deviation σ, is normal or the sample size *n* is large enough so that the SDM, $\bar{X}_n$, is approximately normal. Let $s^2_{n,N}$ be the variance of a simulated SDM, $\bar{X}_{n,N}$, composed of *N* values randomly selected from $\bar{X}_n$.

The (exact) sampling distribution, $S^2_{n,N}$, of the variances, $s^2_{n,N}$, is approximately normal for a large number of samples *N*, and has mean $E\left(S^2_{n,N}\right) = \sigma^2/n$ and variance $Var\left(S^2_{n,N}\right) = \dfrac{2\sigma^4}{n^2(N-1)}$.

These follow because the "population" $\bar{X}_n$ is $N\left(\mu,\ \sigma^2/n\right)$, so $\dfrac{(N-1)S^2_{n,N}}{\sigma^2/n}$ is χ^2 with $(N-1)$ degrees of freedom.

First, $S^2_{n,N}$ is approximately normal for any reasonably large number of samples, *N*, because a χ^2 distribution with large degrees of freedom is approximately normal. Second, a χ^2 distribution has a mean equal to its degrees of freedom, so,

$$E\left(\frac{(N-1)S^2_{n,N}}{\sigma^2/n}\right) = N-1 \quad \text{or} \quad E\left(S^2_{n,N}\right) = \sigma^2/n$$

(Alternatively, the mean comes from the fact that the sample variance is an unbiased estimator of the variance of its population, $\bar{X}_n$.) Third, a χ^2 distribution has a variance equal to twice its degrees of freedom, so,

$$Var\left(\frac{(N-1)S^2_{n,N}}{\sigma^2/n}\right) = 2(N-1) \quad \text{or} \quad Var\left(S^2_{n,N}\right) = \frac{2\sigma^4}{n^2(N-1)}$$

The formulas for $E\left(S^2_{n,N}\right)$ and $Var\left(S^2_{n,N}\right)$ show that, for fixed *N*, the larger the sample size *n*, the closer $s^2_{n,N}$ tends to be to σ^2/n. Further, for any two sample sizes, n_1 and n_2, the ratio of the variances of $S^2_{n_1,N}$ and $S^2_{n_2,N}$ does not depend on *N*,

$$\frac{Var\left(S^2_{n_1,N}\right)}{Var\left(S^2_{n_2,N}\right)} = \frac{\dfrac{2\sigma^4}{n_1^2(N-1)}}{\dfrac{2\sigma^4}{n_2^2(N-1)}} = \frac{n_2^2}{n_1^2}$$

Thus, as with the mean, increasing the number of samples, *N*, does not change the pattern students see.

8. Discussion

When using the properties of the sampling distribution of the mean, students must understand that, for all sample sizes, the mean of the SDM is (exactly) equal to the mean of the population and the standard deviation is (exactly) equal to $\sigma/\sqrt{n}$. However, we have shown that from observing the patterns in a typical series of simulated SDMs constructed using increasing sample sizes, students are led to conclude that the mean tends to get closer to the population mean as the sample size increases and the standard deviation tends to get closer to $\sigma/\sqrt{n}$ as the sample size increases. This creates a mismatch between the theory we want to teach and what students observe from their simulations.

There are at least two reasons, other than simulation, why students may develop this misunderstanding. One is that students have a tendency to believe that everything gets better with a larger sample size, which generally is a useful belief to hold. A second reason is the ambiguous summary of the properties of the SDM commonly found in textbooks:

> *When n is sufficiently large, the sampling distribution of the mean is approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$.*

This statement inadvertently reinforces what students are likely to observe in their simulation, that *n* must be sufficiently large for *each* property to hold.

What can be done by an instructor who wishes to use simulation to illustrate the properties of the SDM? We offer three choices, none of them ideal. First, an instructor who believes that honesty is the best policy could warn students that the pattern they see is an artifact of using simulation to construct an approximate SDM. The simple and intuitive argument that the mean of a simulated SDM, constructed from *N* random samples each of size *n*, can be found either by averaging the means of the *N* samples or by averaging the *nN* individual values makes it clear why the variance of a simulated SDM depends on both the sample size, *n,* and the number of samples, *N*. However, even this small amount of theory may be problematic for an introductory statistics class as the discussion would be time consuming and largely irrelevant to the goals of the course.

Second, an instructor could push the bounds of the teaching axiom "tell the truth, but not the whole truth" and use a very large number of samples, *N*, to construct each simulated sampling distribution. As we have seen, this does not change the pattern that, when *N* is fixed, the larger the sample size *n*, the more precise the estimate of $\mu_{\bar{X}_n} = \mu$ tends to be and the closer the standard deviation of the simulated sampling distribution tends to be to $\sigma/\sqrt{n}$. However with large enough *N*, the program could be set to display a small number of decimal places in summary statistics so that round-off error will obscure the pattern.

Third, an instructor could use simulation only to introduce the Central Limit Theorem, justifying the first two properties by example or mathematical methods. For example, students could construct an exact sampling distribution from a small population, verifying that $\mu_{\bar{X}_n} = \mu$ and $\sigma_{\bar{X}_n} = \sigma/\sqrt{n}$. A caveat about this approach is that, when students list all possible samples from a finite population to justify that $\sigma_{\bar{X}_n} = \sigma/\sqrt{n}$, their instructor would be forced to be explicit that the sampling must be done with replacement, which students find unrealistic and hence unconvincing. The finite population issue can be dodged by using a probability distribution such as the mean of the rolls of two dice, but introductory students see such a distribution as different from a sampling distribution. While mathematical proofs are beyond the introductory course, some students are convinced by the argument that it is obviously true that $\mu_{\bar{X}_n} = \mu$ and $\sigma_{\bar{X}_n} = \sigma/\sqrt{n}$ for the smallest possible sample size, $n = 1$.

Whatever strategy the instructor chooses, it is especially important that he or she summarize the properties of the SDM to emphasize the role of sample size:

> *No matter what the sample size, n, the sampling distribution of the mean has mean μ and standard deviation $\sigma/\sqrt{n}$. The sampling distribution will be normal in shape if the population is normal; for other populations, the shape becomes more normal as n increases.*

In this paper, we have presented the mathematical reason why students who observe simulated sampling distributions of the mean may develop or reinforce the misunderstanding that the formulas for its mean and standard deviation, $\mu_{\bar{X}_n} = \mu$ and $\sigma_{\bar{X}_n} = \sigma/\sqrt{n}$, are exactly true only in the limit as n becomes large. While we know that this misunderstanding has occurred with some of our own students, we do not know the extent to which it occurs in general or whether it can develop solely as a result of simulation. Such a misunderstanding is unlikely to affect student performance in introductory classes because the sample size in textbook problems involving skewed distributions must be large enough for the Central Limit Theorem to ensure approximate normality of the SDM. However, such a misunderstanding may contribute to unwarranted distrust of statistical inference, especially when using small samples. Thus, we look forward to future research on students' thinking about the SDM and about how instructors can use simulation to teach the SDM without fostering such misunderstandings.

We are sorry to present another instance where eternal vigilance is the price of teaching statistics. However, we hope we have convinced instructors, especially those who use simulation to illustrate the properties of the sampling distribution of the mean, to be alert to the fact that students may apply the heuristic that otherwise serves them well, "bigger samples are better," to decide when they can use the formulas for the mean and standard deviation of a sampling distribution.

Acknowledgments

The work of the authors was partially supported by the National Science Foundation, Grant No. 1119016. They would like to thank Mark Schilling and William Watkins for their helpful advice about how best to present this topic.

References

Chance, B., delMas, R., and Garfield, J. (2004), "Reasoning about Sampling Distributions," in *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, eds. D. Ben-Zvi and J. Garfield, Dordrecht: Kluwer, pp. 295–323.

Committee on the Undergraduate Program in Mathematics (CUPM) (1972), "Introductory Statistics without Calculus," in *Compendium of CUPM Recommendations* (Vol. 2), Washington, DC: Mathematical Association of America, pp. 472-525. Available online at www.maa.org/programs/faculty-and-departments/curriculum-department-guidelines-recommendations/cupm/cupm-deep-history.

delMas, R. C., Garfield, J., and Chance, B. L. (1999), "A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning," *Journal of Statistics Education* [online], 7 (3). Available at www.amstat.org/publications/jse/jse_archive.htm#1999.

De Veaux, R. D., Velleman, P. F., and Bock, D. E. (2012), *Stats: Data and Models* (4th ed.), Boston: Pearson Education.

Doerr, H. M. and Jacob, B. (2011),"Investigating Secondary Teachers' Statistical Understandings," *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education,* eds. M. Pytlak, T. Rowland, and E. Swoboda, pp. 776-786. Available online at www.cerme7.univ.rzeszow.pl/WG/5/CERME_Doerr-Jacob.pdf.

ESPN.com (2013) [online]. Available at *espn.go.com/nba/salaries*.

Garfield, J. and Ben-Zvi, D. (2007), "How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics," *International Statistical Review*, 75, 372-396.

Gentleman, J. F. (1977), "It's All a Plot (Using Interactive Computer Graphics in Teaching Statistics)," *The American Statistician*, 31, 166-175.

Hesterberg, T. C. (1998), "Simulation and Bootstrapping for Teaching Statistics," *American Statistical Association Proceedings of the Section on Statistical Education,* pp. 44–52. Available online at www.timhesterberg.net/articles.

Hodgson, T. and Burke, M. (2000), "On Simulation and the Teaching of Statistics," *Teaching Statistics,* 22 (3), 91-96.

Koehler, M. H. (2006), "Using Graphing Calculator Simulations in Teaching Statistics," in *Thinking and Reasoning with Data and Chance, Sixty-eighth Yearbook,* eds. G. F. Burrill and P. C. Elliott, Reston VA: National Council of Teachers of Mathematics, pp. 257-272.

Jowett, G. H. and Davies, H. M. (1960), "Practical Experimentation as a Teaching Method in Statistics," *Journal of the Royal Statistical Society, Series A (General),* 123, 10-36.

Lane, D. M. (project leader, Rice University) (2014), "Online Statistics Education: An Interactive Multimedia Course of Study," [online]. Available at onlinestatbook.com/2/sampling_distributions/samp_dist_mean.html.

Lunsford, M. L., Rowell, G. H., and Goodson-Espy, T. (2006), "Classroom Research: Assessment of Student Understanding of Sampling Distributions of Means and the Central Limit Theorem in Post-Calculus Probability and Statistics Classes, *Journal of Statistics Education* [online], 14 (1). Available at www.amstat.org/publications/jse/jse_archive.htm.

Meletiou-Mavrotheris, M. (2003), "Technological Tools in the Introductory Statistics Classroom: Effects on Student Understanding of Inferential Statistics," *International Journal of Computers for Mathematical Learning,* 8, 265-297.

Mills, J. D. (2002), "Using Computer Simulation Methods to Teach Statistics: A Review of the Literature," *Journal of Statistics Education* [online], 10 (1). Available at www.amstat.org/publications/jse/jse_archive.htm#2002.

Mulekar, M. S. and Siegel, M. H. (2009), "How Sample Size Affects a Sampling Distribution," *The Mathematics Teacher*, 103, 35-42.

Noll, J. and Sharma, S. (2014), "Qualitative Meta-Analysis on the Hospital Task: Implications for Research, *Journal of Statistics Education* [online], 22 (2). Available at http://www.amstat.org/publications/jse/v22n2/noll.pdf.

Pfaff, T. and Weinberg, A. (2009), "Do Hands-On Activities Increase Student Understanding? A Case Study," *Journal of Statistics Education* [online], 17 (3). Available at www.amstat.org/publications/jse/jse_archive.htm#2009.

Renolls, K. and Massay, M. (1991), "Discovering the Laws of Sampling," *Teaching Statistics*, 13, 71-73.

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W, and Onghena, P. (2007), "Students' Misconceptions of Statistical Inference: A Review of the Empirical Evidence from Research on Statistics Education," *Educational Research Review*, 2, 98–113.

Ann E. Watkins
Department of Mathematics
California State University, Northridge
Northridge, CA 91330-8313
ann.watkins@csun.edu

Anna Bargagliotti
Department of Mathematics
Loyola Marymount University
Los Angeles, CA 90045
abargagl@lmu.edu

Christine Franklin
Department of Statistics

University of Georgia
Athens, GA 30602
codycory@uga.edu

Volume 22 (2014) | Archive | Index | Data Archive | Resources | Editorial Board | Guidelines for Authors | Guidelines for Data Contributors | Guidelines for Readers/Data Users | Home Page | Contact JSE | ASA Publications