Simulation of the Sampling Distribution of Mean—Controversy and Misconceptions in Teaching Statistics

Sadia Sarkar and Khairul Islam

Department of Mathematics and Statistics
Eastern Michigan University, Ypsilanti, MI 48197

Abstract
While Watkins et al. (2014) concluded that the "Simulation of the Sampling Distribution of the Mean Can Mislead," Lane (2015) concluded that the "Simulations of the Sampling Distribution of the Mean Do Not Necessarily Mislead and Can Facilitate Learning." Given such controversy in conclusions, instructors have to rethink about specific learning objectives and examples for better outcomes in the simulation of the sampling distribution of means while teaching. This paper proposes some specific objectives and examples which is expected to reduce controversy and misconceptions in teaching statistics.

Key Words: Teaching Statistics, Sampling distribution of means, Law of large number, Central limit theorem, Misconceptions

1. Introduction

The simulation of the sampling distribution of the mean (SSDM) can mislead students (Watkins, Bargagliotti, and Franklin, 2014), and can facilitate learning (Lane, 2015). Due to these contradictory conclusions, instructors interested in implementing SSDM in pedagogy require specific learning objectives and examples that could reduce controversy and misconceptions. It appears that most of the reported misconceptions could be explained via the implication of the law of large numbers. Therefore, it is good idea to introduce the law of large number with examples and demonstration before approaching the SSDM. It is also important to address the distinction of exact sampling distribution of mean (SDM) from the SSDM. In addition, expectation should be made clear via supervised and guided engagement in the SSDM.

In this paper, we only aim at some specific learning objectives and examples for an SSDM in pedagogy.

2. An Overview of an SDM

Let us introduce the concept of sampling distribution of mean for a finite population, along with its properties

2.1 Useful notation

Let us start with some notations:
P=a population,
μ_x =Mean of P,
σ_x=Standard deviation of P and
n(P)=Number of all possible samples of size n from the population P, with replacement.
$\bar{x}$ = Mean of a sample of size n from the population P

For example, let P be a finite population P= $\{1, 2, 3, 4\}$ with $\mu_x = 2.5$, σ_x=1.118, approximately.
The sampling distribution of mean (SDM) is the probability distribution of sample means ($\bar{x}$) over all possible samples of a given size (n) from the population. It appears that with P= $\{1, 2, 3, 4\}$, the number of all possible samples of size n, with replacement, is $n(\text{P}) = 4^n$. The SDM in this case is the probability distribution of 4^n sample means $\bar{x}$ from P. Clearly, if $n = 2$ or $n = 3$, we will 4^2 =16 or 4^3 =64 possible samples. Therefore, the SDM is the probability distribution 16 or 64 sample means $\bar{x}$ of size 2 or 3, respectively.
Let $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ denote the mean and standard deviation of the SDM over all possible samples of size n. Since $\sigma_{\bar{x}}$ describes the accuracy of $\bar{x}$ in estimating μ_x, the term standard error (SE) is preferable to standard deviation $\sigma_{\bar{x}}$ for an SDM. Then, the SDM has the following three important properties:

- *(i)* $\mu_{\bar{x}} = \mu_x$, *which is invariant with respect to the sample size n and the distribution of P.*
- *(ii)* $\sigma_{\bar{x}} = \sigma_x/\sqrt{n}$, *which means that the standard error of $\bar{x}$ due to the estimation of μ_x by $\bar{x}$ can be decreased by increasing the sample size n.*
- *(iii) The SDM is either (a) normal if P is normal, irrespective of the sample size n, or (b) approximately normal if the sample size n large,* i.e., $n \geq 30$, *given P is not normal. This property is called the Central Limit Theorem (CLT).*

2.2. Properties of SDM for P

Given P=$\{1, 2, 3, 4\}$ with $\mu_x = 2.5$ and σ_x=1.118, approximately, the summary of **Table 1** follows immediately by properties (i-iii) for n =1, 2, 3, 4.

Table 1. Summary of SDMs for varying n for the population P

Sample size n	Number of samples in SDM, n(P)	Mean of SDM, $\mu_{\bar{x}}$	SE of SDM $\sigma_{\bar{x}} = \sigma_x/\sqrt{n}$
1	$1(P) = 4^1 = 4$	2.5	$1.118/\sqrt{1} = 1.118$
2	$2(P) = 4^2 = 16$	2.5	$1.118/\sqrt{2} = 0.791$
3	$3(P) = 4^3 = 64$	2.5	$1.118/\sqrt{3} = 0.6455$
4	$4(P) = 4^4 = 256$	2.5	$1.118/\sqrt{4} = 0.559$

Let $f_{\bar{x}}$ be the frequency distribution or $p(\bar{x})$ be the probability distribution the SDM. Then, one can verify properties (i)-(ii) using the following two formulas:

$$\mu_{\bar{x}} = \frac{\sum f_{\bar{x}} * \bar{x}}{\sum f_{\bar{x}}} \text{ or } \mu_{\bar{x}} = \sum \bar{x} * p(\bar{x})$$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum f_{\bar{x}} * (\bar{x} - \mu_{\bar{x}})^2}{\sum f_{\bar{x}}}} \text{ or } \sigma_{\bar{x}} = \sqrt{\sum (\bar{x} - \mu_{\bar{x}})^2 * p(\bar{x})}$$

For example, when $n = 3$, an SDM with $3(\text{P}) = 4^3 = 64$ possible samples will have the frequency ($f_{\bar{x}}$) and probability distribution, $p(\bar{x})$, as in **Table 2**.

Table 2. Frequency ($f_{\bar{x}}$) and probability distribution, $p(\bar{x})$, of $\bar{x}$ for 3(P)

$\bar{x}$	$f_{\bar{x}}$	$p(\bar{x})$
1.000	1	0.01563
1.333	3	0.04688
1.667	6	0.09375

2.000	10	0.15625
2.333	12	0.18750
2.667	12	0.18750
3.000	10	0.15625
3.333	6	0.09375
3.667	3	0.04688
4.000	1	0.01563
Total	64	1.00000

Therefore,

(i) $\mu_{\bar{x}} = \frac{\sum f_{\bar{x}} * \bar{x}}{\sum f_{\bar{x}}} = \frac{160}{64} = 2.5.$

(ii) $\sigma_{\bar{x}} = \sqrt{\frac{\sum f_{\bar{x}} * (\bar{x} - \mu_{\bar{x}})^2}{\sum f_{\bar{x}}}} = \sqrt{\frac{1*(1-2.5)^2 + 3*(1.333-2.5)^2 + \cdots + 1*(4-2.5)^2}{64}} = 0.6455.$

(iii) The shape of the SDM for 3(P) is symmetric (see Figure 1(c)).

Similarly, one can verify that when $n = 4$, an SDM with $4(\mathrm{P}) = 4^4 = 256$ possible samples will have the frequency ($f_{\bar{x}}$) and probability distribution, $p(\bar{x})$, as in **Table 3**.

Table 3. Frequency ($f_{\bar{x}}$) and probability distribution, $p(\bar{x})$, of $\bar{x}$ for 4(P)

$\bar{x}$	$f_{\bar{x}}$	$p(\bar{x})$
1	1	0.004
1.25	4	0.016
1.5	10	0.039
1.75	20	0.078
2	31	0.121
2.25	40	0.156
2.5	44	0.172
2.75	40	0.156
3	31	0.121
3.25	20	0.078
3.5	10	0.039
3.75	4	0.016
4	1	0.004
Total	256	1

Therefore,

(i) $\mu_{\bar{x}} = \frac{\sum f_{\bar{x}} * \bar{x}}{\sum f_{\bar{x}}} = \frac{(1*1) + (4*1.25) + \cdots + (4*3.75) + (1*4)}{256} = \frac{640}{256} = 2.5.$

(ii) $\sigma_{\bar{x}} = \sqrt{\frac{\sum f_{\bar{x}} * (\bar{x} - \mu_{\bar{x}})^2}{\sum f_{\bar{x}}}} = \sqrt{\frac{1*(1-2.5)^2 + 4*(1.25-2.5)^2 + \cdots + 1*(4-2.5)^2}{256}} = 0.559.$

(iii) The shape of the SDM for 4(P) is symmetric (see Figure 1(d)).

Figures 1(a)-1(d), represents the shape of four SDMs for n=1, 2, 3 and 4 for the population P. Note that different sample sizes lead to a different SDMs with identical means (=2.5) but different SEs, which gets smaller as the sample size gets larger. Since four SDMs for n=1, 2, 3, 4 have only a finite number of elements, namely, 4, 16, 64 and 256, respectively, they cannot be normal.

Figure 1(a). Probability Histogram of the SDM for 1(P)

Figure 1(b). Probability Histogram of the SDM for 2(P)

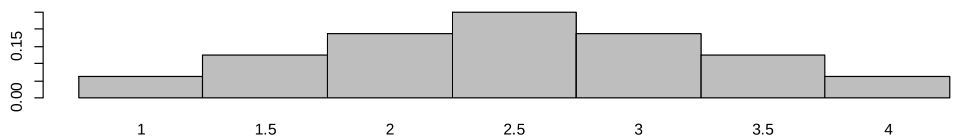


Figure 1(c). Probability Histogram of the SDM for 3(P)

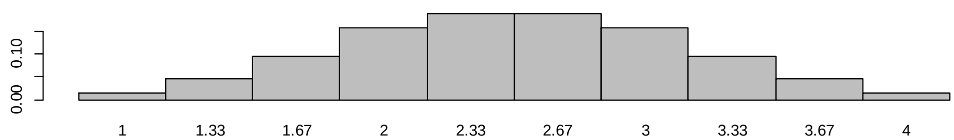
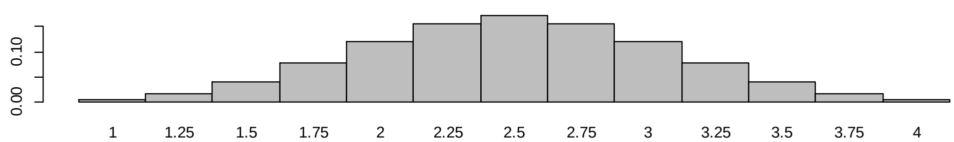


Figure 1(d). Probability Histogram of the SDM for 4(P)



3. Recommendations, Warnings and Misconceptions in Using an SSDM

The “Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report” recommend “*greater use of computer based simulations and the use of resampling methods to teach concepts of inference*” (GAISE College Report ASA Revision Committee 2016, p. 5). Many previous studies evaluated and recommended the use of SSDM while teaching (e.g., Jowett and Davies 1960; Gentleman 1977; Weir, McManus, and Kiely 1991; Schwarz and Sutherland 1997; delMas, Garfield, and Chance 1999; Lane and Tang 2000; Ziemer and Lane 2000; Mills 2004; Chance et al. 2007; Garfield and Ben-Zvi 2007; Lane 2015; Hancock and Rummerfeld 2020). Schwarz and Sutherland (1997) note the ease with which students can compare computer simulated sampling distributions from different populations using summary statistics and data visualization. Garfield and Ben-Zvi (2007) concluded that simulations can play a significant role in enhancing students’ ability to study random processes and statistical concepts. Chance, delMas, and Garfeld (2004) concluded that mere exposure to sampling distribution simulations is unlikely to significantly change students’ deep understanding. It is suggested that computer simulation methods can be made more effective if the instructor asks students to predict the shape, center, and spread of a particular sampling distribution prior to performing the simulation, and then asks students to reflect on what they observed afterward.

No difference or a modest difference in students' understanding of sampling distribution following instruction with or without simulation is also available in literature (Mills 2002; Meletiou-Mavrotheris 2003; Chance et al. 2004; Pfaff and Weinberg 2009). Various forms of misconceptions and challenges in using an SSDM are evident in a number of studies (e.g., Pfaff and Weinberg 2009; Hesterberg 1998; Hodgson and Burke 2000; Lunsford, Rowell and Goodson-Espy 2006; Watkins, Bargagliotti, and Franklin, 2014). Pfaff and Weinberg (2009) with a card-based simulation did not find any evidence that the simulation was effective. 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." Hodgson and Burke (2000, p. 94) observed that the 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." Some researchers argue that the computer simulation methods help students develop a more intuitive understanding of sampling distributions (e.g., Mills 2002; Wood 2005; Pfaff and Weinberg 2009; Beckman, DelMas, and Garfeld 2017).

Watkins, Bargagliotti, and Franklin (2014) discovered that simulations of the sampling distribution of the mean can mislead and the misleading pattern is persuasive to students plotting their estimated means against the sample sizes. Misunderstandings reported in Watkins et al. are similar, in full or part, to those in Renolls and Massay (1991, p. 72), Mulekar and Siegel (2009, p. 37 and 40) and Lunsford, Rowell, and Goodson-Espy (2006). While Watkins et al. (2014) concluded that misconceptions due to simulation cannot be fixed by increasing the number of samples, Lane (2015) argued that the problem can be fixed by increasing the number of samples.

4. An SSDM from a Finite Population

In this section, we consider an SSDM from a finite population with specific learning objectives. We consider P={167, 150, 125, 120, 150, 150, 40, 136, 120, 150} consisting of heart rates, measured by beats per minute, of ten asthmatic patients in a state of respiratory arrest (Pagano and Gauvreau, 2000) where $\mu_x = 130.8$ and $\sigma_x = 33.65$. In order to assess the shape of the population P, we provide a boxplot (Figure 2(a)) and histogram (Figure 2(b)).

Figure 2(a). Boxplot of P

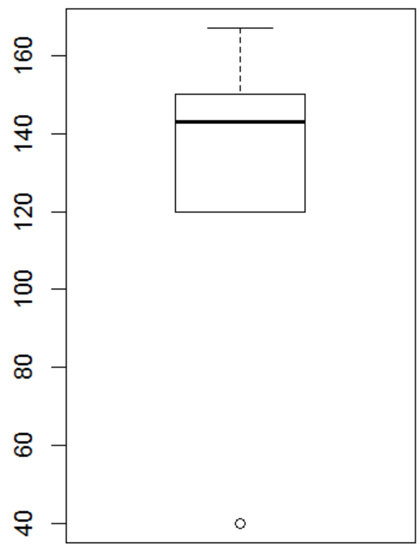
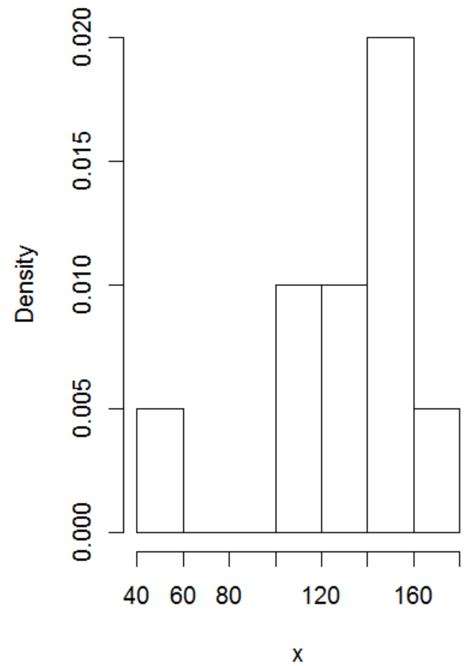


Figure 2(b). Probability histogram of P



Figures 2(a)-2(b) suggest that the population P is not normal. Also, according to the properties of SDM specified in Section 2, the SDM will be non-normal with mean $\mu_{\bar{x}} = \mu_x = 130.8$ and $\sigma_{\bar{x}} = \sigma_x/\sqrt{n} = 33.65/\sqrt{n}$ for any given sample size n.

4.1 Specific Learning Objectives for an SSDM from P

Imagine an SDM of size 10 from the population P, where we will have $10(P) = 10^{10}$ possible samples. How reasonable or feasible is it to investigate all possible means towards investigating the properties of the SDM? Of course, it is not reasonable to try to get an SDM of P of size $n = 10$. Here is where an SSDM comes handy—instead of trying to generate the exact SDM using 10^{10} sample means, we could just generate an SSDM using a suitable number of sample replications, say $M = 1000$, each of size 10 with replacement, via simulation, with some specific learning objectives. We wish to verify properties of the SDM via an SSDM with the aid of the test of hypothesis approach.

This approach leads to an inquiry-based learning, with the following specific learning objectives:

(a) Investigate the shape of the SSDM for the normality using boxplot, QQ-plot and histogram.
(b) Investigate skewness and kurtosis for the normality of the SSDM.
(c) Verify if mean of the SSDM conforms to the mean of the SDM by employing a test of the null hypothesis H_0: $\tilde{\mu}_{\bar{x}} = \mu_{\bar{x}}$, where $\tilde{\mu}_{\bar{x}}$ is the mean of the SSDM and $\mu_{\bar{x}} = 130.8$ is known for the exact SDM.
(d) Verify if SE of the SSDM conforms to the SE of the SDM by employing a test of the null hypothesis H_0: $\tilde{\sigma}_{\bar{x}} = \sigma_{\bar{x}}$ where $\tilde{\sigma}_{\bar{x}}$ is the SE of the SSDM and $\sigma_{\bar{x}} = \frac{33.65}{\sqrt{10}} = 10.64$ is the known SE for an exact SDM.

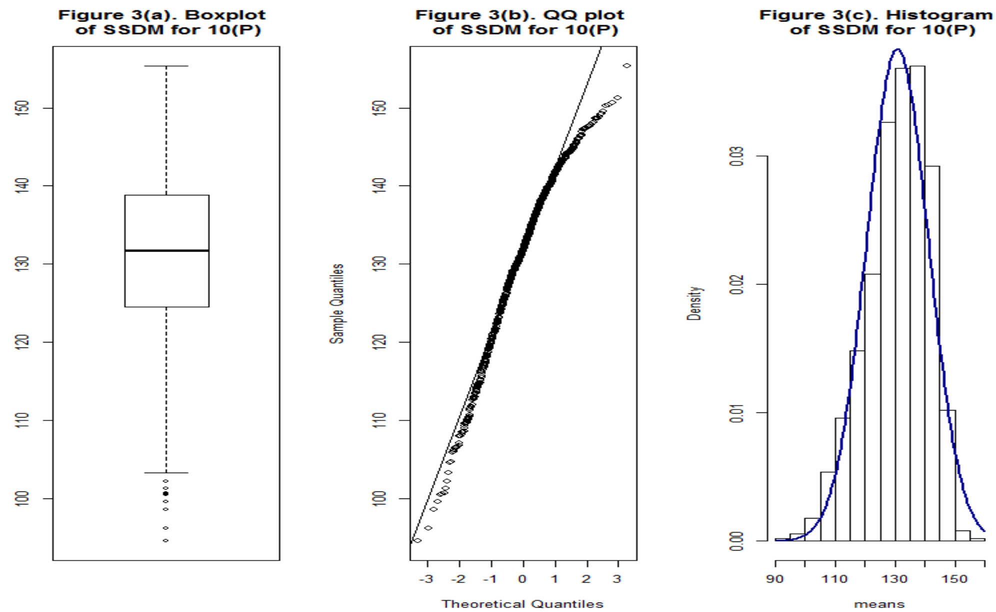
(e) Test the normality of the SSDM by any formal test such as Shapiro-Wilk test or Anderson-Darling test.

4.2 Evaluation of objectives

To evaluate objectives (a)-(e), we consider an SSDM of 1000 samples (M=1000) with $n = 10$. For objective (a), we produce specified graphs. For objective (b)-(e), consider five runs of "SSDM of 1000 samples" and report corresponding results. Let us perform an evaluation of all specified objectives in a chronological order.

Evaluation of objective (a)

We construct a boxplot, QQ plot and histogram with super imposed normal curve with the mean and SE of the SSDM and is presented in **Figures 3(a)-3(c)**.



Conclusions: (i) The boxplot has long tail on the left with evidence of outlying observations in the SSDM (ii) The Q-Q plot does not form a line. (iii) The histogram is not shape symmetric. (iv) The normal curve misses a lot of observations of the SSDM on the right, suggesting that the shape of the SSDM is left-skewed. These facts provide evidence against the normality of the SSDM, which indeed makes sense due to the fact that the actual population P is not normal.

Evaluation of objective (b)

A normal distribution has skewness of 0 and kurtosis of 3. If an SSDM is normal, it will have skewness close to 0 and kurtosis close to 3. Look at results from five runs of the "SSDM":

Runs	Skewness	Kurtosis
1	-0.560	2.949
2	-0.438	2.853
3	-0.577	3.261
4	-0.516	3.158
5	-0.610	3.391

Conclusions: (i) The skewness suggests that the shape of five SSDMs are left skewed. (ii) The values of kurtosis also provide evidence against normality of the five SSDMs. These results are expected since P itself is not normal.

Evaluation of objective (c)

To test H_0: $\tilde{\mu}_{\bar{x}} = \mu_{\bar{x}}$, we carry out t-test and z-test to five SSDMs. The t-test statistic given by

$$T = \frac{\left(\bar{\bar{x}}_{n,M} - \mu_{\bar{x}}\right)}{se\left(\bar{\bar{x}}_{n,M}\right)}$$

where $\bar{\bar{x}}_{n,M}$ is the mean of all sample means over M replications in the SSDM and $se(\bar{\bar{x}}_{n,M})$ is an estimate of $\sigma_{\bar{x}}$ given by

$$se\left(\bar{\bar{x}}_{n,M}\right) = \tilde{s}/\sqrt{M}$$

with $\tilde{s} = \sqrt{\frac{1}{M-1}\sum_{j=1}^{M}\left(\bar{x}_{n,j} - \bar{\bar{x}}_{n,M}\right)^2}$. The test statistic T is assumed to follow a ***t*** distribution with $(M-1)$ degrees of freedom.

The z-test statistic given by

$$Z = \frac{\bar{\bar{x}}_{n,M} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

which is assumed to follow a standard normal distribution. The results of tests from five runs of the SSDM have been reported in **Table 4**.

Table 4. p-values for z and t tests for objective (c) in five runs of SSDM for 10(P)

Runs	$\tilde{\mu}_{\bar{x}}$	$t.test$	$z.test$	Decisions
1	130.84	**0.916**	**0.918**	Accept H_0: $\tilde{\mu}_{\bar{x}} = \mu_{\bar{x}}$
2	131.13	**0.317**	**0.326**	Accept H_0: $\tilde{\mu}_{\bar{x}} = \mu_{\bar{x}}$
3	130.36	**0.182**	**0.189**	Accept H_0: $\tilde{\mu}_{\bar{x}} = \mu_{\bar{x}}$
4	131.02	**0.513**	**0.521**	Accept H_0: $\tilde{\mu}_{\bar{x}} = \mu_{\bar{x}}$
5	130.66	**0.683**	**0.673**	Accept H_0: $\tilde{\mu}_{\bar{x}} = \mu_{\bar{x}}$

Conclusions: On the basis of observed p-values of t.test and z.test, it may be concluded that the mean of the SSDM conforms to the mean of the SDM, 130.8 for all five runs at 5% level of significance.

Evaluation of objective (d)

To the test H_0: $\tilde{\sigma}_{\bar{x}} = \sigma_{\bar{x}}$, we conduct a Chi-square test for a specified variance given by

$$\chi^2 = \frac{(M-1)\tilde{s}_{\bar{x}}^2}{\sigma_{\bar{x}}^2}$$

where $\tilde{s}_{\bar{x}}^2 = \frac{\sum_{j=1}^{M}(\bar{x}_{n,j} - \bar{\bar{x}}_{n,M})^2}{M-1}$. The test statistic χ^2 is assumed to follow a chi-square distribution with $(M-1)$ degrees of freedom. We can implement this test in R using the **sigma.test()** available from the package TeachingDemos. The result of the test from five runs of the "SSDM" are given in **Table 5**.

Table 5. p-values for Chi-squared tests for objective (d) in five runs of SSDM for 10(P)

Runs	$\tilde{\sigma}_{\bar{x}}$	$chi2.test$	Decisions
1	10.41	**0.351**	Accept H_0: $\tilde{\sigma}_{\bar{x}} = \sigma_{\bar{x}}$
2	10.43	**0.402**	Accept H_0: $\tilde{\sigma}_{\bar{x}} = \sigma_{\bar{x}}$
3	10.48	**0.523**	Accept H_0: $\tilde{\sigma}_{\bar{x}} = \sigma_{\bar{x}}$
4	10.43	**0.385**	Accept H_0: $\tilde{\sigma}_{\bar{x}} = \sigma_{\bar{x}}$

5	10.99	**0.140**	Accept H_0: $\tilde{\sigma}_{\bar{x}} = \sigma_{\bar{x}}$

Conclusions: On the basis of the observed p-values, it may be concluded that the SE of SSDM conforms to that of the SDM in all five runs of the SSDM at 5% level of significance.

Evaluation of objective (e)

To test H_0: "SSDM is normal", we employ Shapiro-Wilk, and Anderson-Darling tests of normality to five runs of "SSDM". These two tests are implemented in R using shapiro.test() and ad.test(). The results of these tests have been reported in **Table 6.**

Table 6. p-values of tests of normality for objective (e) in five runs of SSDM for 10(P)

Runs	$Shapiro.test$	$ad.test$	Decisions
1	5.273264e-12	4.907891e-15	Reject H_0: "SSDM is normal
2	6.644692e-09	1.085329e-09	Reject H_0: "SSDM is normal
3	6.028231e-11	1.085623e-10	Reject H_0: "SSDM is normal
4	5.437490e-10	1.062992e-09	Reject H_0: "SSDM is normal
5	1.035714e-11	3.754409e-12	Reject H_0: "SSDM is normal

Conclusion: On the basis of the observed p-values, it may be concluded that the SSDMs are not normal at 5% level of significance since $p.value < 0.05$ for all five runs.

By all means and measures we observe via objectives (a)-(e) that an SSDM is not misleading and it facilitates learning, with only a few exceptions where students confused about an SSDM.

5. An SSDM from a Normal Population

Similar learning objective as in Section 4 can be targeted for an SSDM from a normal population with any given mean and standard deviation. In this section, we consider a N(μ=1.5, σ=2) distribution, where the parameters are chosen arbitrarily. We wish to generate "Different SSDMs with M=1000" with selected values of sample sizes n between 5 and 1000, and report mean and SE for the SDM and SSDM, along with the difference in the SEs (see **Table 7**).

Table 7. Estimated means and SEs for the SSDMs with varying n and M=1000 and 5000

	M=1000				M=5000			
n	$\tilde{\mu}_{\bar{x}}$	$\tilde{\sigma}_{\bar{x}}$	$\sigma_{\bar{x}}$	SE_{diff}	$\tilde{\mu}_{\bar{x}}$	$\tilde{\sigma}_{\bar{x}}$	$\sigma_{\bar{x}}$	SE_{diff}
5	1.518	0.827	0.894	0.068	1.504	0.844	0.894	0.051
10	1.509	0.614	0.632	0.019	1.499	0.616	0.632	0.017
50	1.495	0.280	0.283	0.003	1.507	0.282	0.283	0.001
100	1.504	0.199	0.200	0.001	1.505	0.200	0.200	0.000
150	1.504	0.163	0.163	0.000	1.503	0.163	0.163	0.000
200	1.501	0.142	0.141	0.000	1.497	0.141	0.141	0.000

250	1.503	0.126	0.126	0.000	1.498	0.126	0.126	0.000
300	1.500	0.115	0.115	0.000	1.502	0.115	0.115	0.000
400	1.498	0.100	0.100	0.000	1.500	0.100	0.100	0.000
500	1.503	0.089	0.089	0.000	1.500	0.089	0.089	0.000
600	1.504	0.082	0.082	0.000	1.501	0.082	0.082	0.000
700	1.498	0.076	0.076	0.000	1.500	0.076	0.076	0.000
900	1.503	0.067	0.067	0.000	1.498	0.067	0.067	0.000
1000	1.497	0.063	0.063	0.000	1.501	0.063	0.063	0.000

From the results of Table 7, it appears that the means are close to 1.5, irrespective of the sample size n. The SEs of SDM and SSDM are also close for any specific sample size; they decrease with increasing n. The results of the SDMs and SSDMs conform to the properties (i)-(ii) of sections 2.

6. Remarks in Reference to Previous Misunderstandings

As reported in Watkins et al. (2014, p. 11) "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 examples, cited as Renolls and Massay (1991, p. 72) and Mulekar and Siegel (2009, p. 37 and 40), showing a clear trend for the means of the simulated SDMs to get closer to the population mean as the sample size increases. Also, Watkins et al. (2014, p. 8) reported "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." These misunderstandings are interpretational deficiencies, which can be explained easily by the LLN which states that *as the size of the sample n increases, the sample mean $\bar{x}$ gets closer and closer to the population mean μ*. Due to this fact if $n_1 > n_2$, then samples of size n_1 are expected to provide an estimate closer to the mean μ *than the samples of size n_2. Naturally, then, the average of all (N) samples of size n_1 will give a close estimate of μ than the average of all (N) samples of size n_2. In other words, an SSDM of N samples of size n_1 will have mean close to μ than will have an SSDM of samples N of size n_2. Therefore, the misunderstandings of the types reported in Watkins et al. is just an interpretational deficiency.*

7. Discussions and Concluding Remarks

By learning objectives (a)-(e) in Section 4, we mean to signify that an SSDM should be aimed at a guided manner so as to minimize misconceptions. Some important aspects of an SSDM should be emphasized:

- The mean and SE of an SDM are parameters, while the mean and SE of an SSDM are estimates of parameters of an SDM and hence are random variables. As such, an SSDM should not be expected to be identically equal to an SDM. Indeed, it is impossible to ever generate an SSDM identical to an SDM because exactly the same set of samples in an SDM cannot be ensured in an SSDM.
- Expectations from an SSDM should be communicated to students via examples, activities or at least by simulated outputs with necessary explanations or interpretations before engaging students in an SSDM.

- Instructors should be aware against any misunderstanding and misconceptions while proceeding for an SSDM so as to avoid misunderstandings, for example, of the types reported in Watkins et al.
- The concept of the SDM can be better delivered by undertaking a very simple and finite population, where instructors can generate all possible samples easily by means of computer programming. It is a very good practice to let students investigate or justify the properties of an SDM given simple hands-on-activity, printouts or homework before engaging in an SSDM. As students feel comfortable with SDM, they could be exposed to interesting issues in relation to an SSDM that would make them appreciate simulation and thereby get them involved in computational approaches.

Overall, an SSDM can be made effective if it is administered in a guided manner by the instructor with specific learning objectives. This prepares students to critically evaluate their simulated results without any misconception.

References

Beckman, M. D., DelMas, R. C., and Garfeld, J. (2017), "Cognitive Transfer Outcomes for a Simulation-Based Introductory Statistics Curriculum," *Statistics Education Research Journal*, 16, 419–440.

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, 295–323.

Chance, B., Ben-Zvi, D., Garfeld, J., and Medina, E. (2007), "The Role of Technology in Improving Student Learning of Statistics," *Technology Innovations in Statistics Education*, 1. Available at https://escholarship.org/uc/item/8sd2t4rr.

delMas, R., Garfield, J., and Chance, B. (1999), "A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning," *Journal of Statistics Education*, 7, http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm.

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

GAISE College Report ASA Revision Committee (2016), "Guidelines for Assessment and Instruction in Statistics Education College Report 2016," Technical Report.

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

Hancock, S.A. and Rummerfeld, W. (2020), "Simulation Methods for Teaching Sampling Distributions: Should Hands-on Activities Precede the Computer?," *Journal of Statistics Education,* 28:1, 9-17. Available at https://www.tandfonline.com/doi/full/10.1080/10691898.2020.1720551

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.

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. (2015), "Simulations of the Sampling Distribution of the Mean Do Not

Necessarily Mislead and Can Facilitate Learning," *Journal of Statistics Education*, 23. Available at https://doi.org/10.1080/10691898.2015.11889738.

Lane, D. M. and Tang, Z. (2000), "Effectiveness of Simulation Training on Transfer of statistical concepts" *Journal of Educational Computing Research*, 22, 383-396.

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 (3). Available at ww.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.

Mills, J. D. (2004), "Learning abstract statistics concepts using simulation," *Educational Research Quarterly*, 28, 18-33.

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

Pagano, M. and Gauvreau, K. (2000), "Principles of Biostatistics", Second Edition, Duxbury Thomson Learning.

Pfaff, T. J. and Weinberg, A. (2009), "Do hands-on activities increase student understanding?: A case study," *Journal of Statistics Education* 17, http://www.amstat.org/publications/jse/v17n3/pfaff.html

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

Schwarz, C. J., and Sutherland, J. (1997), "An On-Line Workshop Using a Simple Capture Recapture Experiment to Illustrate the Concepts of a Sampling Distribution," *Journal of Statistics Education*, 5. Available at https://doi.org/10.1080/10691898.1997.11910523.

Watkins, A. E., Bargagliotti, A., and Franklin, C. (2014), "Simulation of the Sampling Distribution of the Mean Can Mislead," *Journal of Statistics Education*, 22. Available at https://doi.org/10.1080/10691898.2014.11889716

Weir, C. G., McManus, I. C., and Kiely, B. (1991), "Evaluation of the Teaching of Statistical Concepts by Interactive Experience With Monte Carlo Simulations," *British Journal of Educational Psychology*, 61, 240–247.

Wood, M. (2005), "The Role of Simulation Approaches in Statistics," *Journal of Statistics Education*, 13. Available at *https://doi.org/10.1080/10691898.2005.11910562*

Ziemer, H. and Lane, D. (2000), "Evaluating the Efficacy of the Rice University Virtual Statistics Lab," Poster presented at the 22nd Annual Meeting of the National Institute on the Teaching of Psychology, St. Petersburg Beach, FL, January.