# EFFECTS OF TEACHING THE SAMPLING DISTRIBUTION OF THE MEANS USING SIMULATION WITH AND WITHOUT STATING THE CENTRAL LIMIT THEOREM

Rini Oktavia, Nurmaulidar, Intan Syahrini, & Hafnani
Department of Mathematics, Syiah Kuala University, Indonesia
rini_oktavia@unsyiah.ac.id

*Teaching sampling distribution of the means (SDM) using simulation has the potential to mislead students who might falsely believe that the mean of SDM will more closely approximate the population mean (μ) as the sample size (n) increases. A teaching experiment was conducted involving two Introductory Statistics classes. Both classes were taught the concept of SDM using simulation but the central limit theorem (CLT) was only stated in one class. A questionnaire on assessing students' thinking and possible misunderstanding about CLT and sampling distribution was administered before and after the experiment to both groups. A statistical analysis comparing both groups were conducted. It was found that the instructions affect students' understanding of SDM, however, there is no significant difference in students' understanding of sampling distribution for both classes after the instructions.*

INTRODUCTION

The revised Guidelines for Assessment and Instruction in Statistics Education (GAISE) recommendations includes 1) Teach statistical thinking; 2) Focus on conceptual understanding; 3) Integrate real data with a context and a purpose; 4) Foster active learning; 5) Use technology to explore concepts and analyze data; and 6) Use assessments to improve and evaluate student learning (GAISE College Report ASA Revision Committee, 2016). In order to apply the revised guidelines, teaching important statistical concepts using technological tools such as computers, graphing calculators, software, and the internet has been an increasingly used approach to help students to gain conceptual understanding on several important statistical concepts. Sampling Distribution is one of these important concepts that relate the distribution of a random sample with the distribution of the population from which the sample is drawn to make inferences about the population. Several studies on the use of simulation to develop students' reasoning about sampling distribution have been conducted (e. g. Lane & Tang, 2000; Chance, delMas, & Garfield, 2004; Lunsford, Rowell, & Goodson-Espy, 2006; Lane & Peres, 2006).

However, the use of simulation on teaching sampling distribution of the means (SDM) has the potential to mislead the students who might falsely believe that the mean of SDM will more closely approximate the population mean (μ) as sample size (n) increases (Watkins, Bargagliotti, & Franklin, 2014). Some students fail to understand that regardless of the sample size, the mean of SDM is always equal to the mean of the population. Watkins, Bargagliotti, and Franklin (2014) reported that, even though they asked students to increase the number of repetitions of the simulated sampling distribution of the mean, the students still saw the pattern that the means of the simulated sampling distributions will be closer to the mean of the population as the sample size increase. David M. Lane (2015) suggested that despite the fact that the pattern is not a function of the number of repetitions in the simulated sampling distributions, using a very large number of samples will make the difference between the simulated sampling distributions of various sample sizes and the mean of the population will be invisible. Using a sampling distribution simulation written in JavaScript (available at http://onlinestatbook.com/stat_sim/sampling_dist/index.html), Lane (2015) showed that the pattern that the mean of simulated sampling distributions approaching the mean of the population did not appear if the number of repetitions is very large. This approach, however, could not be done if there is no internet connection that allows us to access the applet.

Problems of having no internet connection are still relevant for many developing countries, including Indonesia. Therefore, it is reasonable if some instructors in developing countries chose to use a statistical package that can be used offline. One such package is StatKey, a statistical package accompanied the textbook Statistics: Unlocking the Power of Data by Lock, Lock, Lock, Lock, and Lock (2017) that can be used freely without an internet connection as a google chrome app. In this study, we would like to see whether students' misunderstanding found by Watkins, Bargagliotti, and Franklin (2014) can be eliminated if the instructor purposely states the CLT when teaching the concept of SDM using simulation with StatKey.

METHODOLOGY

Two groups of first-year university students who were taking a non-calculus based Introductory Statistics course participated in this study. One group of 20 students was randomly assigned as the control group, meanwhile, another group of 23 students was assigned as the experimental group. Based on students' responses on a preliminary assessment, the instructor found that most students could be categorized in level A or B of statistical development level suggested by GAISE Pre-K-12 report (see Oktavia & Sorto, 2014).

One instructor taught both groups about the sampling distribution of the mean (SDM) using simulation. The Google Chrome app StatKey was used to simulate the sampling distributions of the mean for various sample size. The data set of 2015 US Baseball player salary consisting of 868 cases were used as the population data. The students worked in pairs and the instructor started the lesson by asking each pair of students to take a simple random sample with sample size $n = 4$. Then, they were asked to plot the mean and standard deviation of their samples on the board. This activity was conducted to help the students understand about the simulation that they will do in StatKey. A discussion on the concept of sampling distribution followed the activity. The instructor emphasized that now, they will only focus on one type of sampling distribution, which is called the sampling distribution of the means (SDM).

The instructor asked students to do some simulations for various sample sizes ($n$) using multiples of 1000 repetitions ($N$) and take notes on the patterns and information that they observed. A hand out consisting a list of information that they need to focus on was given to the students. After the simulations, a discussion was initiated by the instructor. The students were asked the patterns that they saw regarding the means and standard errors of simulated sampling distributions and their connection with the mean ($\mu$) and the standard deviation of the population ($\sigma$). In the experimental group class, the instructor summarized the results of the discussion and stated that what the students found was actually described the CLT that the shapes of simulated sampling distributions became more normal as the sample size increased, the mean of the SDM is equal to the mean of the population ($\mu$), and the standard error of the SDM is equal to $\sigma/\sqrt{n}$. The statement of the CLT was not discussed in the control group. Even though the conclusions of both groups were directed to be the same, the missing of the CLT statement was predicted to make the students remember the conclusions as some disconnected facts.

---

Which of the following statements are NOT true? Select all that apply.

A. The larger number of the sampling repetition the closer the mean of the SDM to the mean of the population

B. An increase in sample size from *n* = 16 to *n* = 25 will produce a sampling distribution with a smaller standard deviation.

C. The mean of a sampling distribution of sample means is equal to the population mean divided by the square root of the sample size.

D. The larger the sample size, the more the sampling distribution of sample means resembles the shape of the population.

E. The mean of the sampling distribution of sample means for samples of size *n* = 15 will be the same as the mean of the sampling distribution for samples of size *n* = 100.

F. The larger the sample size, the more the sampling distribution of sample means will resemble a normal distribution.

---

Figure 1. Example of Pre-test and Post-test Question

A pre-test and a post-test were administered before and after the lessons. The pre-test consists of 11 questions that were given again in the post-test, which was administered 2 weeks after the lessons. Five additional questions were given in the post-test. The test questions were adapted from Assessment Resource Tools for Improving Statistical Thinking (ARTIST, 2006) assessment builder (https://apps3.cehd.umn.edu/artist/user/login.asp) by translating the questions in Bahasa Indonesia and

adding the aspects of misconception found by Watkins, Bargagliotti, and Franklin (2014) into the items. The English version of the questions is shown in Figure 1.

Student responses on the pre- and post-tests were coded and scored and a descriptive statistical analysis was conducted. Each response was scored on a scale from 0 to 4, making the maximum total score possible 44. Some simple statistical inference tests were also conducted to compare students' performances in pre- and post-tests, and to investigate whether there are significant differences in students performances between control and experimental groups. After post-test, a review on sampling distribution lesson was given to both groups to make sure that all students were exposed to the statement of the CLT.

RESULTS AND DISCUSSIONS

Comparing all students' test scores on the pre-test and post-test, it was found that students achieved better scores on the post-test. The descriptive statistics of all students' scores on both tests are given in Table 1 below. There are three students who did not take the post-test so that the analysis was conducted only for 40 cases. The paired t-test conducted for total scores on pre-test and post-test for all students (Control group and Experimental group combined) showed that the difference of means on both test scores is significant ($n = 40$, p-value < 0.000). When we analyzed students' test scores on the pre-test and post-test for each group separately using the paired t-test, the result showed that the difference of means on the pre-test and post-test scores of the Control group is also significant with a p-value of 0.005 ($n = 19$). Another paired t-test conducted in the Experimental group also showed a significant difference between total scores in pre-test and post-test ($n = 21$, p-value = 0.007). These results infer that in general, students of both classes gained a better understanding of SDM after the lessons.

Table 1. Descriptive Statistics of Total Scores of Pre- and Post-test

|                      | Total Scores of Pre-test | Total Scores of Post-test |
|----------------------|:------------------------:|:-------------------------:|
| Number of cases      | 40                       | 40                        |
| Mean                 | 17.43                    | 21.85                     |
| Minimum              | 5                        | 13                        |
| Median               | 17.50                    | 21.50                     |
| Maximum              | 33                       | 32                        |
| Standard Deviation   | 6.046                    | 5.395                     |

The Pearson correlation between both total scores is 0.385 with a p-value of 0.014, suggested that both variables are moderately correlated. This result infers that students' previous understanding linearly related to their abilities to gain more understanding from the instruction, however, the relation is not strong.

Table 2. Descriptive Statistics of Total Scores for Experimental and Control Groups

|                      | Total Scores of Pre-test | | Total Scores of Post-test | |
|----------------------|:-------:|:----------:|:-------:|:----------:|
|                      | Control | Experiment | Control | Experiment |
| Number of cases      | 19      | 21         | 19      | 21         |
| Mean                 | 18.68   | 16.29      | 23.63   | 20.24      |
| Minimum              | 5       | 5          | 13      | 13         |
| Median               | 20      | 17         | 24      | 21         |
| Maximum              | 26      | 33         | 32      | 30         |
| Standard Deviation   | 5.229   | 6.619      | 4.844   | 5.467      |

In comparing the statistics for the total scores on both pre-test and post-test by factoring into experimental and control groups, it was found that the differences are quite clear. The summary statistics of both groups are presented in Table 2. We can see in Table 2 that in general, the control group has better scores than the experimental group for both pre-test and post-test.
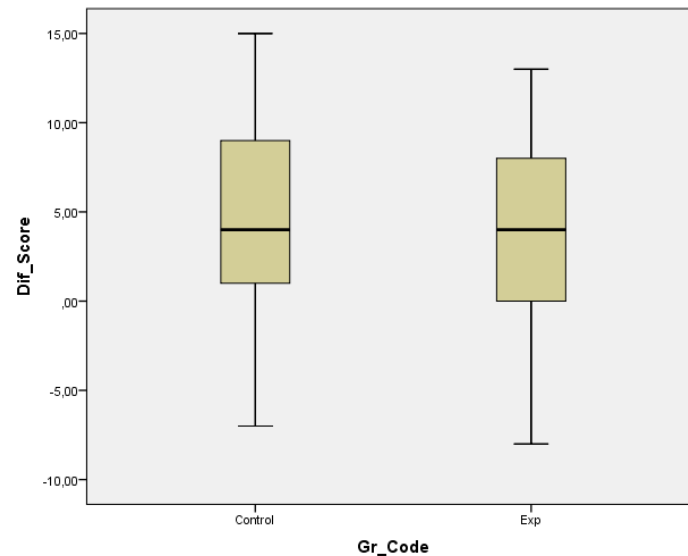
Figure 2. Box Plots of the Difference Between Total Scores on Pre-test and Post-test Factored by Group

By subtracting the total score on the pre-test from the total score on the post-test for each participant, we got that the means of the differences are 4.95 for the Control group and 3.75 for the Experimental group. The box plots of the difference between total scores on the pre-test and post-test factored by group are shown in Figure 2. The result of the one-way ANOVA test to compare the difference of total scores in pre-test and post-test between both groups showed that there is no significant difference among means of the difference between the total scores in the pre-test and the total scores in the post-test (p-value = 0.568). It is more likely that the increasing total scores from the pre-test to the post-test of both groups showed in the paired t-tests were affected by the instruction on sampling distribution only. We cannot conclude that stating the CLT affects students' understanding of the SDM.

From the one-way ANOVA test's result, we realized that the total scores cannot explain further regarding the effects of stating the CLT in the instruction on students' understanding of sampling distribution. We tried to look at the students' responses for particular questions to see whether we can analyze students' understanding for some important concepts related to sampling distribution by examining their responses on particular questions as well as by analyzing particular individual students' responses. The discussion of the analysis will be described below.

If we looked at the maximum total scores at the Pre-test, we found that the student who got a total score of 33 in the pre-test, got 28 on the post-test. This student was from the experimental group. The student answered 9 questions consistently at the pre- and post-tests but changed the responses on two items, one of them is the question showed in Figure 1. In this particular item, the student answered C on the pre-test but changed it into E on the post-test. The correct answer for this item is A, C, and D. Since the student got one of the three answers correctly in the pre-test, the student got a partial credit of 1 for this item. However, since the answer on the post-test was wrong, the score for this item was 0 on the post-test. This student did not gain a better understanding of the sampling distribution concept since this item is crucial in measuring students' conceptual understanding regarding sampling distribution and the CLT. By choosing that "The mean of the sampling distribution of sample means for samples of size n = 15 will be the same as the mean of the sampling distribution for samples of size n = 100" as a wrong statement, this student still has a misconception regarding the value of the mean of SDM. Another item that this particular student answer incorrectly is also an item measuring students' basic conceptual understanding regarding the shape of the sampling distribution. This student failed to see that if the sample size is one, then the shape of the simulated sampling distribution will be similar to the shape of the distribution of the population. Interestingly, the student answered this item correctly on the pre-test. This result shows that some students still did not get the conceptual understanding of sampling distribution after the instruction. Further revision of the instructional approach needs to be done in the future.

When we analyzed the distribution of students' responses to the item shown in Figure 1, we found that the responses are quite disturbing. More students answered that the statement "The mean of the sampling distribution of sample means for samples of size n = 15 will be the same as the mean of the sampling distribution for samples of size n = 100" was wrong in the post-test than in the pre-test. Further investigation is needed to be conducted. Interviews with the students should be initiated to investigate whether the students choose this option due to their misunderstanding of the values of the mean of SDM or they misunderstood the question. Our students are used to having multiple choice questions with a single correct answer.

Analyzing students' responses on an item to measure whether students have the misconception that sampling distributions created by larger repetitions of samples have mean values that are closer to the mean of the population, we found that students observed that the number of repetitions affected the mean of the SDM. This misconception could be avoided if we used an applet that allows students to create a simulated sampling distribution with a large number of repetitions in one step as done by Lane ( 2015). In StatKey, students created a simulated sampling distribution gradually by generating at most 1000 samples at a time. Instructors need to address this misconception in future lessons.

The negative nature of the question can be misleading too. Many students are used to choosing a single correct statement instead of choosing more than one incorrect statements. Based on the analysis of students' responses on items that have more than one correct responses such as items showed in Figure 1, we suspected that those items have some psychometric problems that should be addressed. Further instrument validation process must be conducted that could lead to the revision of the instrument for the pre-test and post-test.

One exciting result that we can take from analyzing students' responses to particular items is their response to the item shown in Figure 3. In this item, 18 of 21 students in the Experimental group responded to this item correctly, meanwhile, 17 of 19 students in the Control group answered this item correctly. The same excitement was experienced when we analyze the students' responses to the item showed in Figure 4. Fifteen (15) of 18 students in the Control group responded to this item correctly on the post-test, meanwhile, only 7 of 19 students in the group answered this item correctly in the pre-test. However, the big difference in the number of correct responses between the pre-test and post-test did not exist in the Experimental group. There were 14 of 21 students answered the item correctly in pre-test and 17 of 21 students answered correctly in the post-test. This result showed that stating the CLT did not affect students' understanding of the shape of the sampling distribution. Students in the Control group increased their level of understanding about the shape of SDM without being exposed to the statement of the CLT.

| If you take all samples of a particular size from a particular population, find the mean of each sample, and then plot the means, what have you created? |
| --- |
| A. Sample distribution        B. Population distribution     C. Sampling distribution |

Figure 3. Pre-test and Post-test Item that is Correctly Answered by Most Students 1

| If sampling distributions of sample means are examined for samples of size 1, 5, 10, 16 and 50 you will notice that as *n* increases in size, the shape of the sampling distribution appears more like that of the: |
| --- |
| A. normal distribution            C. uniform distribution <br> B. population distribution       D. even distribution |

Figure 4. Pre-test and Post-test Item that is Correctly Answered by Most Students 2

Other items that most students responded to correctly on the post-test compared to their responses in the pre-test are items regarding the properties of the sample following the law of large numbers. Most students correctly responded that the statistics of a sample with a very small sample size are most likely to differ than the population parameters. More than 80% of the students responded

correctly for items measuring this concept. There is no significant difference between the Experimental and Control group responses.

CONCLUSION

The instructional approach applied in this study has been proved to be effective in increasing students' general understanding of sampling distributions. This proof also infers that teaching sampling distribution using simulation has a positive impact on students' learning. However, based on the quantitative analysis, there is not enough evidence to conclude that stating the CLT in the instruction influence students' understanding of sampling distribution of the means. Misconceptions found in previous studies still existed after the instructions.

The instructional approach applied in this study needs to be revised, especially in addressing the values of the mean and standard error of SDM. The instrument for the pre-test and the post-test need to be validated and revised further. The importance of having a sample that is big enough in size has been understood well by the students. Based on the analysis conducted on students' responses on individual items in the pre-test and post-test, students' conceptual understanding of SDM was not satisfactory, but their basic understanding of the definition of a sampling distribution and its shape was quite promising.

The use of simulation helped students understand about definition and shape of the sampling distribution of the means. However, to help students understand other concepts related to the sampling distribution of the means and the CLT, more carefully designed instructions are needed. The fact that most students have not been in the statistical development level C, meaning that most students have not yet developed sound statistical thinking, should be considered in designing future instructions. Simplifying the instructions and slowing the pace of learning sessions could be effective. A wide opportunity is open for statistics educators in Indonesia to improve their instructions following the GAISE College Report's suggestion.

REFERENCES

ARTIST Project. (2006). *ARTIST Assessment Builder*. Retrieved from ARTIST: https://apps3.cehd. umn.edu/artist/user/login.asp

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi, & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy* (pp. 295–323). Dordrecht, The Netherlands: Kluwer Academic Publisher.

GAISE College Report ASA Revision Committee. (2016, July). *Guidelines for Assessment and Instruction in Statistics Education College Report 2016*. Retrieved from http://www.amstat.org/ education/gaise

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*(2), http://www.amstat.org/ publications/jse/v23n2/lane.pdf.

Lane, D. M., & Peres, S. C. (2006). Interactive simulations in the teaching of statistics. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute.

Lane, D. M., & Tang, Z. (2000). Effectiveness of simulation training on transfer of statistical concepts. *Journal of Educational Computing Research, 22*(4), 383–396.

Lock, R. H., Lock, P. F., Morgan, K. L., Lock, E. F., & Lock., D. F. (2017). *Statistics: Unlocking the Power of Data 2nd Ed.* John Wiley & Sons.

Lunsford, M., Rowell, G., & Goodson-Espy, T. (2006). Classroom research: Assessment of student understanding of sampling distributions of means and the CLT in post-calculus probability and statistics classes. *Journal of Statistics Education, 14*(3).

Oktavia, R., & Sorto, M. A. (2014). Confirmatory and Exploratory Factor Analyses of Students' Developmental Levels in Learning Statistic. In K. Makar, B. de Sousa, & R. Gould (Eds.)*, Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July 2014), Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute.

Watkins, A. E., Bargagliotti, A., & Franklin, C. (2014). Simulation of the Sampling Distribution of the Mean Can Mislead. *Journal of Statistics Education, 22*(3), www.amstat.org/publications/jse/ v22n3/watkins.pdf.