# Using group projects to assess the learning of sampling distributions

Robert O. Neidigh
Shippensburg University

Jake Dunkelberger
Shippensburg University

**ABSTRACT**

In an introductory business statistics course, student groups used sample data to compare a set of sample means to the theoretical sampling distribution. Each group was given a production measurement with a population mean and standard deviation. The groups were also provided an excel spreadsheet with 40 sample measurements per week for 52 weeks for their production measurement. The sample data was generated using the Random Number Generation Tool in the Data Analysis Add-In in Microsoft Excel. Each group is asked to calculate sample means for each week's sample data and then describe the set of sample means using statistical measurements and/or graphical methods. The groups are then asked to compare the set of sample means to the theoretical sampling distribution for their production measurement.

These projects were assigned after sampling distributions were covered in class. Assessment results for these group projects over two years indicate there is considerable confusion about sampling distributions among the students. The assessment results and analysis of the submitted group projects are used to provide suggested changes to the timing and structure of the projects to enhance student learning.

Keywords: business statistics education, sampling distributions, student assessment, group projects, empirical research

## INTRODUCTION

A sampling distribution is the set of sample means selected from a population for a given sample size. When the sampling distribution is known to be normal the distribution of sample means follows the normal distribution. The teaching of normal sampling distributions is essential to any business statistics course because they provide the foundation to confidence interval estimation and hypothesis testing.

Typical questions in business statistics courses related to sampling distributions ask students to determine the probability of a particular sample mean being within a given range or the value such that a certain percentage of sample means are above or below that point. Most students learn to calculate the z-values needed to determine probabilities and/or look up a given probability to determine a z-value to determine a value. The students are able to answer required questions by determining these probabilities and finding the z-values. The ability to answer these types of questions gives the appearance that students comprehend sampling distributions.

Students commonly are more concerned about learning what is needed to score well on exams rather than fully comprehending the material. This is especially so for courses which are not well liked such as business statistics. Unfortunately, many students learn the process to answering the typical business statistics questions and can score well on exams without ever truly understanding what a sampling distribution is. These students have learned to use the equations and z-tables to successfully determine the correct probabilities and values, but do not necessarily comprehend the properties of sampling distributions. Even more disturbing, some students do not even know what a sampling distribution is. Pollatsek, Lima, and Well [7] state, “Learning a computational formula is a poor substitute for gaining an understanding of the basic underlying concept.”

A thorough understanding of sampling distributions is essential for a student to grasp the theory behind confidence interval estimation and hypothesis testing. Those students who have not grasped the fundamentals of sampling distributions will most likely struggle with confidence interval estimation and hypothesis testing.

## LITERATURE SEARCH

Most of the past research on teaching sampling distributions provides methods for the teaching or illustration of the sampling distribution concepts. Zerbolio [8] uses imaginary marbles, chips, and bags to create distributions of sample means. Dyck and Gee [4] use M&M candy to illustrate the sampling distribution and found that students who participated in the M&M demonstration scored better on an exam than students who only had textbook learning. Aberson, Berger, Healy, Kyle, and Romero [1] use a Web-based, interactive tutorial to present the sampling distribution and use pretests and posttests to show that the tutorial can be comparable in effectiveness to lectures. Aguinis and Branstetter [2] use an approach that stimulates both visual and auditory learning and engages students in the process of learning through problem solving and demonstrate that the approach enhanced learning by approximately 60 percent.

Much of the research in teaching the sampling distribution uses computer simulation. Kett [6] uses the computer program Autograph to illustrate the sampling distribution. Gourgey [5] teaches sampling distributions using a collaborative learning simulation based on political polling. Computer simulations alone do not guarantee conceptual change as demonstrated by delMas, Garfield, and Chance [3] using the *Sampling Distributions* computer simulation.

## ASSESSMENT DESIGN

A group project was developed to evaluate the groups' conceptual knowledge of sampling distributions. The project was not assigned until after sampling distributions were taught in lectures and assessed in an in-class exam. Each group was provided a production measurement with a population mean and standard deviation for the production process. For example, one production measurement provided is the volume of a soda bottle with a population mean of 20.05 ounces and a population standard deviation of 0.02 ounce.

Each group was also given an excel file with sample measurements from the production process. Each production measurement has 52 weeks of data with forty sample measurements for each week. The sample data was generated using the Random Number Generation Tool from the Data Analysis Toolpack in Microsoft Excel.

There are two deliverables for the project with Deliverable 1 worth 2 points and Deliverable 2 worth three points. In Deliverable 1, each group was asked to calculate the sample mean for each week's data and then describe the distribution of the 52 sample means. The description of the 52 sample means should include descriptive and graphical methods. In Deliverable 2, each group was asked to compare the empirical data set of 52 sample means to the theoretical sampling distribution of sample means. Examples should have been provided to highlight and support the groups' conclusions. Appendix I is the project description for the Spring 2012 semester.

## MINIMUM GROUP EXPECTATIONS

The groups know that the project assesses sampling distributions so their knowledge of the properties of sampling distributions should shape their answers. For Deliverable 1, the groups are expected to calculate the mean and standard deviation of the 52 sample means. Since the sample size is at least thirty, the students should recognize that the distribution of sample means will be normally distributed so the students should graph the fifty-two sample means with a histogram. The histogram will provide a quick visual check of normality.

For Deliverable 2, the groups are expected to determine the theoretical mean and standard deviation of the sampling distribution for the provided production measurement with a sample size of forty. The groups should then to compare the empirical mean of the fifty-two sample means to the sampling distribution mean. Typically, the empirical mean is close in value to the sampling distribution mean. The groups also should compare the empirical standard deviation of the fifty-two sample means to the sampling distribution standard deviation. Also, the empirical standard deviation of the sample means is usually close in value to the sampling distribution standard deviation. The groups should note how close the empirical values are to the theoretical values.

If the group graphed a histogram in Deliverable 1, the group should discuss how similar or dissimilar the shape of the histogram is to the normal shape. The groups should recognize that the sampling distribution is normally distributed so the histogram is expected to be somewhat normally shaped. With a small sample size of 52 sample means, the histogram will not necessarily be normally shaped.

In theory, half of the fifty-two sample means should be greater than the population mean and half should be less than the population mean. The groups should note how many of the fifty-two sample means are greater than the population mean and how many are less than the population mean.

**PROJECT RESULTS**

The results of seventy-three projects for three different semesters are summarized in Appendix II. The results vary somewhat among the three semesters given information provided by the instructor and feedback in response to questions, but in aggregate the results provide a sound basis for generalized conclusions.

In Deliverable 1, the groups should calculate the sample mean and sample standard deviation of the 52 sample means and then graph the sample means using a histogram. Only 39 groups correctly calculated the sample mean and sample standard deviation. Some of the remaining groups did not calculate the sample standard deviation but most incorrectly calculated the sample standard deviation. The most common mistake the groups made was calculating the 52 sample standard deviations of the forty measurements in each week and then averaging these 52 sample standard deviations. Since only 53% of the groups correctly calculated both the sample mean and sample standard deviation of the 52 sample means, there was obvious confusion among the groups about how the mean and standard deviation of a sampling distribution describes the distribution of sample means.

Seventy-three percent graphed the 52 sample means with a histogram, but this number was inflated in the final two semesters (31 of 39 and 17 of 20) due to “coaching” provided by the instructor. In the first semester, only 5 of 14 groups provided a histogram. The groups typically don’t recognize that the 52 sample means are a sample of a normal sampling distribution and therefore the shape of the sample means should be “somewhat” normally distributed.

For Deliverable 2, the results were much poorer than for Deliverable 1. A very small percentage, 18%, of the groups compared the empirical sample mean and standard deviation to the sampling distribution mean and standard deviation. Even most of the groups that correctly calculated the sample mean and standard deviation did not make the comparison. Less than half, 20 of 53, of the groups that provided a histogram discussed the normality of the histogram. Only 28 groups provided examples to show how the distribution of the sample means compared to the theoretical distribution.

The groups just did not recognize that the 52 sample means are a sample of a sampling distribution. This is extremely disappointing given that the project description plainly states, "Compare and contrast the empirical data set of sample means to the theoretical sampling distribution of sample means for the given population mean and standard deviation." Also, given that the instructor "coached" the groups into providing certain requirements the results are even more disappointing. The results clearly demonstrate that students generally do not comprehend the basics of sampling distributions even after covering the material in class!

## CONCLUSIONS AND FUTURE DIRECTIONS

The project results for three semesters clearly show that too many student groups do not understand the properties of a sampling distribution. More disturbing, the poor results seem to indicate that some student groups do not even understand what a sampling distribution is! Far too many groups had a difficult time simply describing the sample set of 52 sample means let alone comparing the distribution of sample means to the sampling distribution. These confusions and difficulties remain after much lecture time and an in-class exam on sampling distributions.

The project was originally designed and so far implemented only to assess student comprehension, not to enhance student learning. Some student learning probably occurred during completion of the project, but no assessment of this learning has been undertaken. The project was assigned a grade only so that the groups would complete the project. Upon completion of the project, the instructor spent little time discussing the project expectations with the class.

In the future, the project will be used during the learning process as a tool to enhance student comprehension. The project deliverables and expectations will remain the same, however the post-assessment analysis will change. After project completion and assessment, the instructor will explain to the class the project expectations; then select one of the production measurements and show the expected project deliverables for that production measurement. This post-assessment analysis will illustrate to the students how the concepts and properties of sampling distributions can be used to describe the expected distribution of a group of sample means. Hopefully, when the project is implemented and analyzed in this manner, student learning will be enhanced!

**REFERENCES**

[1] Aberson, Christopher L., Berger, Dale E., Healy, Michael R., Kyle, Diana J., and Romero, Victoria L. (2000), "Evaluation of an Interactive Tutorial for Teaching the Central Limit Theorem", *Teaching of Psychology* 27, 289-291.

[2] Aguinis, Herman and Branstetter, Steven A. (2007), "Teaching the Concept of the Sampling Distribution of the Mean", *Journal of Management Education* 31, 467-483.

[3] 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).

[4] Dyck, Jennifer L. and Gee, Nancy R. (1998), "A Sweet Way to Teach Students about the Sampling Distribution of the Mean", *Teaching of Psychology* 25, 192-195.

[5] Gourgey, Annette F. (2000), "A Classroom Simulation Based on Political Polling to Help Students Understand Sampling Distributions", *Journal of Statistics Education* [Online] 8(3).

[6] Kett, James R. (2011), "Teaching Sampling Distributions", *Mathematics Teacher* 105, 226-229.

[7] Lima, S., Pollatsek, A., and Well, A.D. (1981), "Concept or computation: Students' understanding of the mean", *Educational Studies in Mathematics* 12, 191-204.

[8] Zerbolio, Dominic J. (1989), "A "Bag of Tricks" for Teaching about Sampling Distributions", *Teaching of Psychology* 16, 207-209.