Developing a Hypothetical Learning Trajectory for the Sampling Distribution of the Sample Means

Syafriandi
Mathematics Department, Universitas Negeri Padang, Jl. Prof. Hamka, Padang, Sumatera Barat, Indonesia 25131
syafriandi_math@fmipa.unp.ac.id

Abstract. Special types of probability distribution are sampling distributions that are important in hypothesis testing. The concept of a sampling distribution may well be the key concept in understanding how inferential procedures work. In this paper, we will design a hypothetical learning trajectory (HLT) for the sampling distribution of the sample mean, and we will discuss how the sampling distribution is used in hypothesis testing.

1. Introduction
Sampling distribution is an important topic in studying inferential statistics, which is one of topics in the elementary statistics course. Sampling distributions are studied after students learn the foundational concepts of measure of central tendency, measure of variability, probability, and normal distribution. A sampling distribution of a statistic is the probability distribution of the statistic computed from all possible random samples of the same size from the same population [1]. Especially, a sampling distribution of sample means is a distribution using the means computed from all possible random samples of a specific size taken from a population [2]. The ultimate goal of studying sampling distributions is that students are able to develop reasonable ideas of the Central Limit Theorem (CLT).

In studying the sampling distribution, the students is not only able to use the sampling distribution in inferential statistics, but the students must be able to examine how the sampling distribution is obtained. The students are expected to be able to rediscover the concept of distribution sampling as the experts first found it. However, many studies suggest that students have difficulty in understanding the concept of sampling distribution in hypothesis testing. The research undertaken by [3, 4] reveals that students often see hypothesis testing only as a set of calculation procedures, without such procedures hypothesis testing does not have much meaning for them. Furthermore, the results of research [5,6,7] reveal that students have difficulty identifying null hypotheses and alternative hypotheses, and fail to understand the role of sampling distribution in hypothesis testing.

This also happened to the students of the third semester in Mathematics Education program of Universitas Negeri Padang. Based on the data about 50 percent of the students fail is lack of understanding of a sampling distribution of sample means. This happen because most of students tend to focus on procedures without strengthening their understanding on sampling distribution concept itself. There are some misconceptions found when students study the sampling distribution, that are: the sampling distribution should look like the population, sampling distribution for large samples have more variability, and a sampling distribution is not a distribution of sample statistics. Lecturers should not rush to explain the procedure without a good understanding of the concept. If the students to make conjectures that they test with data, than they become more engaged in reasoning about the sampling
distribution concepts than if the concepts were only presented to them. For that, the lecturers should use appropriate approaches to develop conceptual thinking before students are introduced to the formal ideas of sampling distribution.

Based on this situation, the researcher is interested to design a Hypothetical Learning Trajectory (HLT) which is series of integrated classroom activities in a design research method. The hypothetical learning trajectory is made up of three components: the learning goal that defines the direction, the learning activities, and the hypothetical learning process-a prediction of how the students’ thinking and understanding will evolve in the context of the learning activities [8]. In order to support students development in understanding the sampling distribution concepts researcher to apply Realistic Mathematics Education (RME). The choice of RME as an underlying theory for this study relates to its potential to address issues of instructional design [9]. Realistic Mathematics Education emphasizes that of mathematics as a human activity connected to reality. Mathematical learning is doing mathematics [10]. Furthermore, [11] reveals that there are five characteristics of RME, that are (1) phenomenological exploration: constructions stimulated by concreteness; (2) bridging by vertical instrument: developing mathematical tools to move from concreteness to abstraction; (3) student contribution: students own construction or production assumed will be meaningful for them; (4) interactivity: stimulating the social activity of learning by interaction; and (5) intertwining learning strands in order to get mathematical idea structured.

2. Method and Design

This research use the educational design research as a research approach. According to [12] educational design research is the systematic study of designing, developing, and evaluating educational interventions as solutions to such problems. Educational design research aims at advancing our knowledge about the characteristics of these interventions and the processes to design and develop them. The reasons to choose this methodology are: (1) the design research provide a productive perspective for theory development; (2) the design research has typical usefulness of its results, and (3) design research directly involved the researcher in the improvement in mathematics education.

Design research consist of three phases that are done cyclically [13], (1) preparing for the experiment: drafting a preliminary HLT; (2) a teaching experiment: trying, elaborating and improving the instructional design or local instructional theory and developing an understanding of how it works, and (3) retrospective analysis: studying the entire data set to contribute to the development of a local instructional theory and more encompassing theories.

3. Findings and Discussion

The first phase in design research is preparing for the experiment. In this phase, researchers designed preliminary HLT consisting of: learning goal, learning activities: planned instructional activities and the tools that will be used, and the hypothetical learning process or conjectured of learning processes. The general overview of the hypothetical learning trajectory of the sampling distribution of the sample means showed in Table 1.

| Table 1. General overview of HLT of the sampling distribution of the sample means. |
|-------------------------------|-----------------|-----------------|---------------------------------|-------------------------|
| Name of activity               | Learning goal                                           | Learning activities                                      | Hypothetical learning process                                      | Mathematical Idea                  |
| Activity 1: let's find the mean of the sampling distribution | The students can find that the mean of the sampling distribution of the sample mean is exactly equal to the mean of the | Doing to simulation by taking all possible random samples of the same size from the same population and | Able to distinguish the mean of the sampling distribution and the mean of the population. | $\mu_{\bar{X}} = \mu$ |
population from which the sample were drawn. | computed the mean of the sampling distribution and than compare with the mean of the population.
---|---
**Activity 2:** let's find the standard deviation of the sampling distribution | The students can find that the standard deviation of the sampling distribution of the sample mean is exactly equal to the standard deviation of the population divided by the square root of the sample size. | Doing to simulation by taking all possible random samples of the same size from the same population and computed the standard deviation of the sampling distribution and than compare with the standard deviation of the population. | Able to distinguish the standard deviation of the sampling distribution and the standard deviation of the population. \[ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \]
**Activity 3:** Constructing a sampling distribution of the sample mean | The students are able to constructing the sampling distribution of the sample mean using the central limit theorem | Construct the frequency distribution for the sample mean, and using the normal distribution as a model. | Utilizing the frequency histogram of the sample mean for the various sample sizes \[ z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \]

In the teaching experiment phase, the lecturer introduced the sampling distribution, and he said that sampling distribution is an important topic in studying inferential statistics. Furthermore, the lecturer invites the students to understand that the sampling distribution of a statistic is the probability distribution of the statistic computed from all possible random samples of the same size from the same population. The students worked in groups of four and there were five groups in elementary statistics class. Students worked in groups to do the simulation takes all possible random samples with the specific size of the same population. In pairs in one group, the first pair takes a sample of two and the second pair takes a sample of three with replacement from the same population. Suppose the population consists of 1, 2, 3, and 4, the sampling results as shown in Figure 1.

![Figure 1. Random sample of population](image-url)
The second step is to compute the statistic (the sample mean) for each random sample, the results as shown in Figure 2.

![Figure 2. The sample mean of random sample](image)

In calculating of the sample mean, the students pair having a sample size of four has a shortage of time, as there are 64 samples that are calculated of mean. For that, the lecturer gives a conjecture so that students use the software to calculate the sample mean.

The next step is to computed the mean and the standard deviation of the sampling distribution of the sample mean and then compare with the mean and the standard deviation of the population. All students are able to calculate the mean and standard deviation of the population, but the students found a problem in calculating the standard deviation of the sampling distribution of the sample mean. The Students are confused in using the standard deviation concept of the sampling distribution of the sample mean, as shown in Figure 3.

![Figure 3. One of the student's answers to calculate the standard deviation of the sampling distribution.](image)

Figure 3 shows that the student uses statistic in calculating the standard deviation of the sampling distribution of the sample mean, in which case the standard deviation of the sampling distribution of sample mean is a parameter. After the student group presents in front of the class, another group of students responds that the standard deviation of the sampling distribution is a parameter. The argument used is the sampling distribution is the probability distribution of the statistic computed from all possible random samples, so this group believes that the concept used is \( \sigma_{X} = \sqrt{\frac{\sum (X - \mu_{X})^2}{N_{X}}} \). Finally, the groups of four discussed the simulation results that are summarized in tabular form.
Table 2. Simulation results

| Sample size | Population Mean ($\mu_X$) | Compare | Standard deviation | Compare |
|-------------|---------------------------|---------|--------------------|---------|
| n = 2       | 2.5                       | $\mu_X = \mu_{\bar{X}}$ | $\sqrt{5}/4$ | $\sqrt{5}/8$ |
| n = 3       | 2.5                       | $\mu_X = \mu_{\bar{X}}$ | $\sqrt{5}/4$ | $\sqrt{5}/12$ |

Conclusion: $\mu_X = \mu_{\bar{X}}$ and $\sigma_{\bar{X}} = \sigma_x / \sqrt{n}$

In activity 3: the students constructing a sampling distribution of the sample mean using the frequency histogram of the sample mean for the various sample sizes (n = 2 and n = 3), the results are shown in Figure 4.

Figure 4. Histogram of population and sampling distribution of n = 2 and n = 3
The students in his group discussed the three histograms obtained. Some groups cannot find the concept differences contained in Figure (a) and (b). The students in his group discussed the three histograms obtained. Some groups cannot find the concept differences contained in Figure (b) and (c). They said that the two histograms is an approximate normal distribution, then the lecturer asks the students to observe the number of boxes of each histogram if the sample size increases. Another group finds that the larger the sample size the histogram gets closer, so the resulting frequency polygon is an approximate the normal curve. Finally, the class discussion concludes that although initially the population is not normally distributed, the sampling distribution of the sample mean approximates the normal distribution, so that the central limit theorem applies, that is \[ z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \]

4. Conclusion

The research found that the three activities designed in a Hypothetical Learning Trajectory (HLT) building students understanding on the concept of the sampling distribution of the sample mean. The students need a real condition (the simulation takes all possible random samples with the specific size of the same population) and they find themselves three amazing things, that is (1) \( \mu_x = \mu_\bar{x} \), (2) \( \sigma_\bar{x} = \sigma_x / \sqrt{n} \), and the central limit theorem: \[ z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \].

References

[1] Glenberg, Arthur M. & Andrzejewski, M. E. 2008. Learning from data: an introduction to statistical reasoning. New York: Taylor & Francis Group, LLC.
[2] Bluman, A. G. 2012. Elementary statistics: A step by step approach. 8th ed. New York: McGraw Hill.
[3] Mendez, H. 1991. Understanding the central limit theorem. (Doctoral dissertation, University of California, Santa Barbara, 1991). Dissertation Abstracts International, 53, 2722
[4] Batanero, C. 2000. Controversies around the Role of Statistical Tests in Experimental Research. Mathematical Thinking and Learning, 2 (1-2): 75–98.
[5] Castro Sotos, A. E., Vanhooft, S., Noortgate, W. V. den, & 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.
[6] Garfield, J., & Ahlgren, A. 1988. Difficulties in learning basic concepts in probability and statistics: implications for research. Journal for Research in Mathematics Education, 19, 44-63.
[7] Thompson, P., Liu, Y., & Saldanha, L. 2007. Intricacies of statistical inference and teachers’ understandings of them. In M. C. Lovett & P. Shah (Eds.), Thinking with Data (1st ed.;: 207 – 231). Mahwah, NJ: Psychology Press.
[8] Simon, M. A. 1995. Reconstructing mathematics pedagogy from a constructivist perspective. Journal for Research in Mathematics Education, 26, 114-145.
[9] Simpson, J. A., & Weiner, E. S. C. (Eds.). (1989). Average. In: The Oxford English dictionary
[10] Bakker, A. 2004. Design research in statistics education: On symbolizing and computer tools. Dissertation Utrecht University. Nederlands: Freudenthal Institute.
[11] Gravemeijer, K.P.E. 1994. Developing Realistic Mathematics Education. Utrecht: Freudenthal Institute.
[12] Plomp, Tjeerd. 2013. Educational Design Research: An Introduction”, in Tjeerd Plomp dan Nienke Nieven (Eds.), Educational Design Research. Netherlands: Institute for Curriculum Development (SLO).
[13] Gravemeijer, K., dan Paul Cobb. 2006. “Design Research from A Learning Design Perspective”, in Jan van den Akker, Keno Gravermeijer, Susan McKenney, dan Nienke Nieven (Eds.), Educational Design Research. New York: Routledge