OPEN ACCESS JOURNALS

Indiana Journal of Arts & Literature
Abbriviate Tittle- Ind J Arts Lite
ISSN *(Online)*- 2582-869X
Journal Homepage Link- https://indianapublications.com/journal/IJAL

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-03|Issue-03|2022

Programming-Based Activity Illustrating the Essence of Normality of Data Assumptions in Teaching Z- Confidence Intervals for Population Mean

Gilbert Alerta*

Philippine Science High School-Caraga Region Campus, Philippine

Article History
Received: 20.02.2022
Accepted: 03.03.2022
Published: 05.03.2022

Citation
Alerta, G. (2022). Programming-Based Activity Illustrating the Essence of Normality of Data Assumptions in Teaching Z-Confidence Intervals for Population Mean. *Indiana Journal of Arts & Literature, 3*(3), 1-4.

Abstract: The (1- α) Confidence Interval (CI) estimation is one of the important topics in basic statistics under the K-12 curriculum yet available instructional materials used in many schools fail to illustrate the essence of the normality of data on the performance of T and Z confidence interval formulas. This study integrates computer and statistical skills of the learners through a programming-based classroom activity.
In the activity, students are tasked to develop a computer program that generates all possible samples (without replacement) of sizes n = 5, 10, and 15 from normal and non-normal population of size N= 20 with known μ . The program also computes all the Z- confidence interval estimates and checked if (1- α) of these intervals contain the known μ or not, considering the level of significance (α).
Programming results exposed students to empirical evidences that at least (1- α) confidence intervals contained the known μ only when the data are from normal population. Further, the learners are also able to visualize that when data are from non-normal population, generated confidence interval estimates become lesser than (1- α). Through classroom experiences, students enjoyed the learning process and begin to raise questions related to test on normality of data.

Keywords: Programming-Based Activity, Normality of Data, Z- Confidence Interval.

INTRODUCTION

Confidence interval (CI) estimation is considered as one of the essential concepts in inferential statistics particularly in the estimation theory. The duality principle between CI estimation and hypothesis testing provides a school of thought that substantiates the implied importance of CI in statistical decision making. Wang *et al.* (2017) in fact argued that confidence intervals are one of the most commonly used statistical methods to summarize uncertainty in parameter estimates from data analyses. Hence, efforts in teaching confidence interval estimation must be given emphasis to build students' strong understanding of the concept and to foster adequate foundation in dealing with uncertainties.

Several studies have already proposed methods and approaches in teaching confidence interval estimation for easy understanding. Fidler & Cumming (2005) presented interactive figures and simulations that lead to insightful interpretations of results with lesser misconceptions. Confidence intervals are difficult to teach especially in high school (De Veaux *et al.*, 2011; Diez *et al.*, 2012; & Kalinowski, 2010; and Kaplan *et al.*, 2010) and one way to address this difficulty is by using simulation (Hagtvedt *et al.*, 2008). Wang *et al.* (2017) used a ten-minute trivia game-based activity that addresses misconceptions of the topic by exposing students to confidence intervals from a personal perspective. Inspite of the known methods and materials in teaching confidence intervals in the literature, there is no approach that gives emphasis on the essence of the normality of data when estimating population mean using the formulas $\overline{X} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$ and $\overline{X} \pm z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)$, where $\overline{X}$, σ , s , n and $Z_{\frac{\alpha}{2}}$ denote for the sample mean, population standard deviation, sample standard deviation, and the Z value at a given α , respectively.

Theoretically, when data sample are non-normal, constructing confidence interval estimates for the population mean becomes less reliable. We say that confidence interval estimation is less reliable when the number of interval estimates that contain the true population mean is lesser than (1- α)%. In this study, the approach in teaching confidence interval includes cases of sampling from normal and non-normal population. The goal is to inculcate knowledge, as early as high school, the necessity of the normality assumption before using confidence interval formulas that are mostly written in textbooks and other learning materials.

In the Philippine Science High School System Curriculum, introductory course to statistics is taught in the ninth grade along with basics on computer programming. This study utilizes the integration of the basic programming skills of the students in

***Corresponding Author:** Gilbert Alerta

understanding the confidence interval estimation considering the assumptions of normal distribution. The suggested pedagogical technique will allow students visualize the property of confidence interval estimates for the population mean when data are non-normal. The information that will be gained by the students through this approach will certainly prepare them in the future discussions on parametric test procedures that strictly demand for normality assumption.

Objectives of the Study

This study generally aims to introduce a programming-based activity or approach in understanding the essence of normality of data assumption when performing Z- confidence interval estimation for a population mean. It specifically exposes students to simulation exercise to visualize uncertainties in estimation procedures when necessary assumption is violated.

MATERIALS AND METHODS

In this study, the approach used in teaching confidence interval estimation utilizes the skills of the ninth grade students on the basic computer programming using either C or C++ languages. In specific, the integration of programming to easily understand the uncertainties in statistical decision making is embedded in the steps below.

Step 1: Normal and Non-normal data of size N=20 are given as two independent populations and both have known mean $\mu = 10$ and $\sigma = 5$. Figures 1 and 2 display the histograms of the normal and non-normal populations for this exercise.

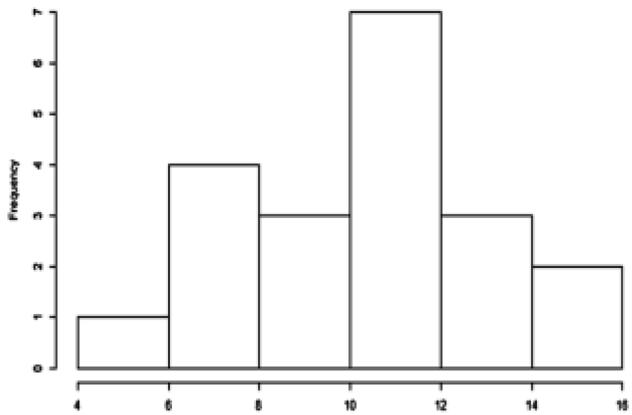

Figure 1. Histogram for Normal Population

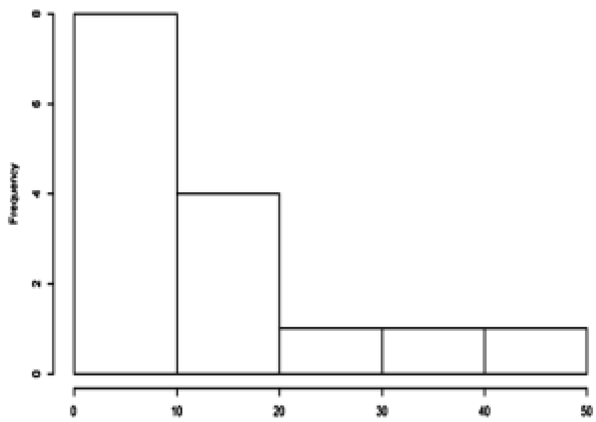

Figure 2. Histogram for Non-normal Population

Step 2: The students will be asked to create a computer program that will generate all confidence interval estimates of μ with varying sample sizes (5, 10, and 15) at each level of significance $\alpha = 0.10, 0.05$, and 0.01. In particular, the task is to make a program that will prompt the user to input the sample size and α, then returns back the percentage of confidence intervals that contain the true population mean which is in this case 10. Algorithm shown below serves as guide in the development of the desired program.

1st **Step:** Declare the array containing all the elements of the particular population with $\mu = 10$ and $\sigma = 5$. Further, declare variables count and percentage.

2nd **Step:** Input α, the corresponding $Z_{\frac{\alpha}{2}}$ value, and sample size n.

3rd **Step:** Generate all samples of size n from the declared population via Simple Random Sampling without Replacement. The total number of samples at a specific size n can be computed by ${}_N C_n$, the combinatorial formula in choosing n from N. Take note that in this exercise N=20.

4th step: For all the samples generated in the 3rd step, compute and display all the confidence intervals using the formula;

$$\overline{X} \pm z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}} \right) * \sqrt{\frac{N-n}{N}}$$

5th **Step:** If $\mu = 10$ is contained in the $\overline{X} \pm z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}} \right) * \sqrt{\frac{N-n}{N}}$, then count=1, otherwise count=0. The $\sqrt{\frac{N-n}{N}}$ is defined as the correction factor for finite population.

6th **step:** Print $\frac{\sum count}{{}_N C_n}$. This presents the percentage of confidence intervals that contain the true population mean $\mu = 10$.

Step 3: Complete the matrix below by filling out each cell with the output of the program. Further, discuss the table.

***Corresponding Author:** Gilbert Alerta

Table 1. Matrix for the percentage of confidence intervals that contain the true population mean with known

α	Normal Population			Non-normal Population		
	${}_{20}C_5$	${}_{20}C_{10}$	${}_{20}C_{15}$	${}_{20}C_5$	${}_{20}C_{10}$	${}_{20}C_{15}$
0.10						
0.05						
0.01						

RESULTS AND DISCUSSION

This section presents both the numerical results from the computer program and the actual observations when implementing the approach in the classroom setting.

Table 2. Completed matrix for the percentage of confidence intervals that contain the true population mean with known

α	Normal Population			Non-normal Population		
	${}_{20}C_5$	${}_{20}C_{10}$	${}_{20}C_{15}$	${}_{20}C_5$	${}_{20}C_{10}$	${}_{20}C_{15}$
0.10	90.33%	90.47%	90.50%	77.29%	84.57%	89.08%
0.05	95.37%	95.35%	95.68%	84.12%	87.26%	90.84%
0.01	99.57%	99.17%	99.25%	94.93%	91.95%	95.54%

Table 2 shows confirmatory evidences on the theory regarding the confidence interval estimates when samples are taken from a normal population. On the contrary, the same confidence interval formula becomes less reliable when normality assumption is not met. At $\alpha = 0.10$, it is expected that at least 90% of the CIs should contain the population mean $\mu = 10$ if data are normally distributed. Indeed, the given table unfolds at least 90% of the CIs in different sample sizes 5, 10, and 15. On the other note, percentages of CIs containing $\mu = 10$ are lesser than 90% when data are sampled from the non-normal population. For example, when n=5 at $\alpha = 0.10$, only 77.29% (which is supposedly at least 90%) of the CIs contained the true $\mu = 10$ when data are obtained from the non-normal population. Similar trends of findings are exposed at $\alpha = 0.05$ and $\alpha = 0.10$. It is further noted that increasing sample size while sampling from the non-normal population increases the reliability of CIs.

Majority of the students were able to execute the program particularly when applying nested loops in generating all possible samples of size n from the population with size N. This is the most crucial part of the algorithm which students have demonstrated high programming skills. Although it was just by mere observations, students were able to recognize the essential role of the normality of data in ensuring the accuracy and precision of the confidence interval estimate. The learning outcome implies that students are indeed already prepared to encounter parametric statistical procedures as they already have experienced the important role of the normality assumption. Another remarkable experience includes the moment when many students begin to raise questions regarding tests for the normality of data.

CONCLUSION AND RECOMMENDATION

This study concludes the usefulness of integrating computer programming exercises in visualizing statistical uncertainties considering the information regarding the normality of data. Particularly in visualizing the reliability of confidence interval estimation for normal and non-normal population, programming integration provides students better appreciation and understanding on the concept.

The approach that is used in this study is hereby recommended for classroom use. In teaching confidence interval estimation, it is very important that normality of data shall be given emphasis along with the uncertainties in statistical inference. Future researchers may also look into the possibility of the inclusion of confidence interval estimation for non-normal data in the high school curriculum.

REFERENCES

1. De Veaux, R. D., Velleman, P., & Bock, D. E. (2011). *Stats: Data and Models* (3rd Ed.). Pearson.
2. Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2012). *OpenIntro statistics* (pp. 174-175). Boston, MA, USA: Open Intro.
3. Fidler, F., & Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings of the 55th international statistics institute session*, *50*.
4. Hagtvedt, R., Jones, G. T., & Jones, K. (2008). Teaching confidence intervals using simulation. *Teaching Statistics*, *30*(2), 53-56.
5. Kalinowski, P. (2010, July). Identifying misconceptions about confidence intervals.

***Corresponding Author:** Gilbert Alerta

In *Proceedings of the eighth international conference on teaching statistics* (Vol. 50).

6. Kaplan, J., Fisher, D. G., & Rogness, N. T. (2010). Lexical ambiguity in statistics: how students use and define the words: association, average, confidence, random and spread. *Journal of Statistics Education*, *18*(2).
7. Wang, X., Reich, N. G., & Horton, N. J. (2019). Enriching students' conceptual understanding of confidence intervals: An interactive trivia-based classroom activity. *The American Statistician*, *73*(1), 50-55.

***Corresponding Author:** Gilbert Alerta