

On the Misuse of Slovin's Formula

Jeffrey J. Tejada and Joyce Raymond B. Punzalan
University of the Philippines Diliman

In a number of research studies involving surveys, the so-called Slovin's formula is used to determine the sample size. Unfortunately, many of these studies use the formula inappropriately, giving the wrong impression that it can be used in just about any sampling problem. This paper provides a careful examination of the formula, showing that it is applicable only when estimating a population proportion and when the confidence coefficient is 95%. Moreover, it is optimal only when the unknown population proportion is believed to be close to 0.5.

Keywords and phrases: Slovin's formula, sample size, margin of error

1. Introduction

A number of research studies use the so-called Slovin's (or sometimes Sloven's) formula for obtaining the sample size. Denoting by n the sample size, Slovin's formula is given by

$$n = \frac{N}{1 + Ne^2}$$

where N is the population size and e is the margin of error (see Almeda, Capistrano, Sarte, 2010).

Researchers are inclined to use this formula because of its simplicity. However, a careful examination of the formula reveals a lack of basis for its usage in some of the literature. For instance, some papers do not state the degree of confidence $1-\alpha$ nor take into account the population variance. It would be unthinkable to take the same n from two populations of the same N but with differing variability.

David and Maligalig in their article “Are We Teaching Statistics Correctly to our Youth?” published in *The Philippine Statistician* in 2006 presented the results of a review by professional statisticians of locally-authored elementary statistics textbooks used in the tertiary level. There were eight major findings, but the most glaring of them is: Some statistical concepts were not presented correctly. Of the nine books reviewed, the one written by Pagoso and Montaña in 1985 dedicated several pages to Slovin’s formula.

To help set the record straight, this paper traces the derivation of Slovin’s formula in order to identify the limitations of its use. This is done in Section 2. In Section 3, we present some literature that utilizes the formula inappropriately, based on the findings discussed in Section 2. In Section 4, we give some conclusions as well as recommendations as to when the formula may be used. Also included as an appendix to this paper is a list of some books, articles, and websites that either use Slovin’s formula incorrectly or present the formula without mentioning the instances when it is applicable.

2. Derivation of Slovin’s Formula

To make inferences on the population proportion P under simple random sampling without replacement (SRSWOR), Cochran (1977) presents the following formula for sample size when working within a finite population:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

where

$$n_0 = \frac{z^2 p(1-p)}{e^2},$$

N is the population size, z is the standard normal variate based on the confidence coefficient, p is the estimate for P , and e is a specified margin of error.

To arrive at Slovin’s formula, we first assume a 95% degree of confidence, so that z is approximately equal to 2. Also, in the absence of any prior knowledge about P , the conservative approach is to maximize

$$p(1-p) = \frac{1}{4} - \left(\frac{1}{2} - p\right)^2,$$

Notice that this quantity is maximized when the subtrahend is 0, that is, when $p = 0.5$. Plugging in $p = 0.5$ and $z = 2$ in the equation for n_0 yields

$$n_0 = \frac{2^2 (0.5)(1-0.5)}{e^2} = \frac{1}{e^2}$$

so that the formula for n becomes

$$n = \frac{1/e^2}{1 + \frac{1/e^2}{N}} = \frac{N}{1 + Ne^2}$$

which is Slovin's formula. Hence, Cochran's and Slovin's formulas coincide when estimating P using a 95% confidence coefficient and $p = 0.5$.

Some textbooks also derive Slovin's formula, though perhaps in a slightly different manner. Of the older ones is Yamane (1967) while of the newer ones we find Almeda, Capistrano, Sarte (2010).

The derivations above show that Slovin's formula is applicable only when estimating a population proportion using a confidence coefficient of 95%. Furthermore, because of the derivation assumption that $p = 0.5$, we remark that using Slovin's formula even under the correct inferential problem could yield an unnecessarily high sample size. Indeed, if we had some belief that P is close to 0 or 1, then Cochran's formula would give the optimal sample size, one that is smaller than what Slovin's formula would yield.

3. Misuse of Slovin's Formula

If the inferential assumptions stated in the previous section hold, then Slovin's formula conveniently gives the correct minimum sample size for estimating a population proportion. However, if any of the assumptions are not satisfied, then the use of the formula would be inappropriate.

For instance, if the estimand is not a population proportion, then using Slovin's formula assumes an unreasonably small population variance, one that would have come from supposedly binary data used in estimating a proportion. Also, it should be clear from Section 2 above that using Slovin's formula restricts the confidence coefficient to 95%. Furthermore, when estimating a proportion that is suspected to be far from 0.5, using Slovin's formula yields a sample size that is unnecessarily large.

It is unfortunate that many studies that involve computing sample sizes using Slovin's formula do not check whether the assumptions are valid. Worse, such studies use the formula even when they are not estimating a population proportion.

In Pagoso and Montaña (1985), the following example is given on page 23: "A researcher would want make a socioeconomic survey of a school with a population of 5,000 students. If he allows a 5% margin of error, how many students must he take into his sample?" (Slovin's) formula was used to get $n = 370$. (The authors did not attribute the formula to Slovin.) On pages 52-53,

(Slovin's) formula is again used for a population with $N = 6924$ to "have a 95% confidence coefficient OR a 5% margin of error" so that $n = 378$. (Notice the incorrect interpretation of margin of error.)

However, Pagoso and Montaña do not specify the situations when Slovin's formula is applicable. They do not specify any proportion as the estimand and use 95% confidence level as interchangeable with 5% margin of error. Unfortunately, their book is available at the Filipiniana Section of UP Diliman's Main Library and at the library of UP College of Human Kinetics.

In the study of Sangcap (2010), which was presented in an international conference, the following research questions were addressed: (1) to determine Filipino college students' positive and negative beliefs about mathematics and mathematical problem solving by administering the 36-item (six scales) self-report questionnaire through stratified random sampling and (2) to analyze possible significant differences in mathematics-related beliefs related to gender, year level, and field of specialization. She uses correlations, t-tests and one-way ANOVA for her inferences. The sample size was determined using the Slovin's formula with two percent (2%) margin of error. The study of Suderio (2010) sought to find the level of mathematics achievement and its correlation, if any, with overall performance of PMA cadets in the academy. Slovin's formula was used with a margin of error of 0.027.

In the above two studies, nowhere in the objectives do we find an estimand that is a population proportion and so they are assuming standard deviations that are unreasonably small. Also, we argue that the 2% and 2.7% margins of error are too small given that they are estimating means, among others.

In the study of Widiyanti and Handajani (2010), the main purpose was to determine the mean amount of water consumption per day. Based on $N = 2,329,928$ and $e = 0.06$, n was determined at 280. Since they were estimating a population mean instead of a proportion, they should have specified not only the margin of error but also the population standard deviation (or an estimate of it), which shall serve as inputs to sample size formulas found in many sampling textbooks from well-established authors, say, Cochran (1977) and Kish (1965).

Estimating a mean. This above case corresponds to an inferential problem where the estimand is a population mean and not a proportion. Using Slovin's formula in this inappropriate manner assumes that the population standard deviation is 0.5. To see this, take the case of estimating the population mean under

SRSWOR: Cochran's formula is $n = \frac{n_0}{1 + \frac{n_0}{N}}$, where $n_0 = \frac{z^2 s^2}{e^2}$ and s is the

estimate for the population standard deviation S . Assuming again a 95% confidence

coefficient, we get $n_0 = \frac{4s^2}{e^2}$ so that

$$n = \frac{4s^2 / e^2}{1 + \frac{4s^2 / e^2}{N}} = \frac{4s^2 N}{4s^2 + Ne^2}.$$

This sample size formula derived for estimating a mean will be equal to Slovin's formula if $4s^2 = 1$, that is, when $s^2 = 0.25$. This variance is certainly almost impossible to obtain in practice. Hence, it is difficult to justify the use of Slovin's formula if the estimand is a population mean.

Furthermore, recall that in estimating \bar{Y} using the sample mean \bar{y} the margin of error is $e = |\bar{y} - \bar{Y}|$. Thus a margin of error similar to the ones mentioned in the above studies is too small in estimating the mean. They could have used a more appropriate (higher) margin of error but we feel that they were forced (by using Slovin's formula) to specify a margin of error that is less than 1 for if not, the resulting sample size is less than 1. Again this goes back to the problem of using Slovin's formula even when the estimand is not a proportion.

In these examples and in most articles in the web especially Google Scholar (please see the Appendix), it is not clear what the estimands are. Most users think that knowing and specifying e are enough to justify the use of Slovin's formula. This was also mentioned in David and Maligalig (2006) where a review of locally-authored Elementary Statistics textbooks in the tertiary level [Pagoso and Montaña (1985) included] was done. The following entry is found in the said review:

“The book introduced the “Slovin's formula” as a method of sample size determination. In here, the authors never bothered to explain, much less simplify, what the margin of error means, which is a vital component of the formula. I still cannot see the theoretical merits of such formula.”

In addition to the literature mentioned in this section, we list in the Appendix some books, articles, websites, and online fora that unfortunately used (or advised to use) Slovin's formula even when it is not appropriate.

4. Conclusions and Recommendations

Slovin's formula is applicable only when estimating a population proportion and when the confidence coefficient is 95%. Additionally, it is optimal only when the population proportion is suspected to be close to 0.5. Hence, it is not advisable to use Slovin's formula if any of the abovementioned assumptions do not hold.

It is then recommended that the formula be used only when the assumptions are met. For other inferential problems such as mean estimation, one can always refer to the more credible textbooks available.

If it is possible, we also recommend the review of books, articles, and other materials that discuss Slovin's formula and make the necessary corrections of those that present or use the formula inappropriately.

As a last note, from our literature review, there does not seem to be a person named Slovin who put forward the formula. It seems that Yamane (1967) is the oldest reference in which the formula can be found.

References

- ALMEDA, J., T. CAPISTRANO, G. SARTE, 2010, *Elementary Statistics*, Quezon City: UP Press.
- COCHRAN, W., 1977, *Sampling Techniques* 3rd Ed., New York: John Wiley and Sons, Inc.
- DAVID, I. and D. MALIGALIG, 2006, Are We Teaching Statistics Correctly to our Youth?, *The Philippine Statistician*, 55(3 and 4):1-28.
- KISH, L., 1965, *Survey Sampling*, New York: John Wiley and Sons, Inc.
- YAMANE, T., 1967, *Elementary Sampling Theory*, New Jersey: Prentice-Hall, Inc.

Appendix

Here is a list of some books, articles, and websites that either used Slovin's formula inappropriately or presented the formula without mentioning its assumptions.

- CHEN TSE-PIN, The Impact of Human Capital, Research and Development, Technology Management and Knowledge Management on the Innovation Performance of Employees of Top IC Design Companies in Taiwan <http://140.130.135.23:8080/dspace/bitstream/TTCIR/358/1/JTTC1709.pdf>
- DALUMAY, F., 2007, Science Attitude and Involvement Of Stakeholders In The Diocese Of La Union: An Input To A Science Learning Enhancement Structural Model. http://www.eisrjc.com/journals/journal_1/coe-rd-jan-jun-07.pdf#page=150
- GRAGASIN, A., 2007, Library Marketing in Selected Private Academic Libraries in Manila: Its Impact on the Academic Community, *Journal of Philippine Librarianship* 27 (1&2): 203-204.
- PAGOSO, C. and R. MONTAÑA, 1985, *Introductory Statistics*, Manila: Rex Book Store.
- RAMOS, E., An Evaluation of the University of Rizal System Morong's (URSM) University Library through User's Assessment, *Journal of Philippine Librarianship* 27 (1&2): 231-232.
- SANGCAP, P., 2010, Mathematics-related Beliefs of Filipino College Students: Factors Affecting Mathematics and Problem Solving Performance, *Procedia Social and Behavioral Sciences* 8, 465-475
- SUDERIO, E., 2010, Mathematics Achievements: Their Relationships with the Academic Performance of the Philippine Military Academy Cadets, *Integritas: Research Journal of the Philippine Military Academy* 1 (1), 35-43
- WIDIANTI, D. and M. HANDAJANI, 2010, Greywater Characterization to Know the Potential Utilization of Greywater Reuse in Bandung City, http://www.ftsl.itb.ac.id/kk/rekayasa_air_dan_limbah_cair/wp-content/uploads/2010/11/pe-dini-widianti-15305024-ww1.pdf

Websites:

- http://www.ftsl.itb.ac.id/kk/air_waste/wp-content/uploads/2010/11/PE-SW2-Yeni-Rahmawati-15305074.pdf
- <http://www.waset.ac.nz/journals/ijhss/v1/v1-4-33.pdf>
- http://www.ftsl.itb.ac.id/kk/rekayasa_air_dan_limbah_cair/wp-content/uploads/2010/11/pe-dini-widianti-15305024-ww1.pdf
- <http://repository.ipb.ac.id/handle/123456789/40618>
- <http://repository.petra.ac.id/2642/>
- http://www.eisrjc.com/journals/journal_1/bsu-2008-19.pdf
- <http://www.journal.uui.ac.id/index.php/JSB/article/view/2014>
- <http://www.springerlink.com/content/y63302465j673053/>
- [http://www.knu.edu.tw/lecture/%E6%AD%B7%E5%B9%B4%E6%95%99%E5%AD%B8%E8%B3%87%E6%96%99/2008%20APIEMS\(BALI\)/PAPER/87-130.pdf](http://www.knu.edu.tw/lecture/%E6%AD%B7%E5%B9%B4%E6%95%99%E5%AD%B8%E8%B3%87%E6%96%99/2008%20APIEMS(BALI)/PAPER/87-130.pdf)
- <http://www.afbe.biz/main/wp-content/uploads/AFBEConfPapers2010.pdf#page=298>

<http://www.ejournals.ph/index.php?journal=MSEUFRS&page=article&op=viewFile&path%5B%5D=1872&path%5B%5D=1974>
http://research.smciligan.edu.ph/downloads/Priv_School_Standards.pdf
<http://journals.smu.edu.ph/index.php/partuat/article/view/97>
<http://eprints.umm.ac.id/592/>
<http://ejournals.ph/index.php?journal=LJHER&page=article&op=viewArticle&path%5B%5D=596>
<http://eprints.undip.ac.id/14974/>
<http://answers.yahoo.com/question/index?qid=20060623001827AAaAYk>
<http://ellinahs.blogspot.com/2006/01/surveys-slovins-formula.html>
http://wiki.answers.com/Q/What_is_slovin's_formula
<http://www.scribd.com/doc/7119782/Sampling>
http://ivythesis.typepad.com/term_paper_topics/2009/05/research-and-analysis-report-on-the-identification-of-key-factors-and-indicators-in-the-motivation-o.html
http://www.ehow.com/way_5475547_slovins-formula-sampling-techniques.html
<http://www.istorya.net/forums/campus-talk/206347-slovins-formula-need-help.html>
<http://team.zobel.dlsu.edu.ph/sites/students/H1/Lists/Announcements/DispForm.aspx?ID=116>