A First Step to Understanding the Difficulty in Teaching Sampling Distributions, Sampling Error, and Statistical Inference

S. David Kriska[1], Mark C. Fulcomer[2], and Marcia M. Sass[3]
[1] , The Ohio State University, Fisher College of Business
[2] Richard Stockton College of New Jersey
[3] University of Medicine and Dentistry of New Jersey

Abstract
Even after having taken an introductory statistics class, many students continue to struggle with the concepts of sampling distributions, sampling error, and statistical inference. As a first step in identifying ways of more effectively approaching these topics, the poster will examine methods currently used in popular text books and propose an alternative based on using binomial and Poisson distributions. The examination of textbooks will identify variability or lack of variability in the approaches to teaching these topics with particular attention being paid to the use of statistical language. The poster will also illustrate how exercises useful in teaching the three topics of sampling distributions, sampling error, and statistical inference.

Key Words: Confidence intervals, sampling error, margin of error, sample variability, introductory statistics

1.Motivation for the Study

As long-time instructors of introductory statistics, we believe that teaching the concept of sampling distributions is most challenging. Delmas, Garfield, Ooms, and Chance (2007) support this view and note that even after taking a statistics class, students demonstrate difficulty with understanding sampling variability and sampling distributions and do not demonstrate an understanding of confidence intervals. (p. 49) The Delmas et al. conclusions may or may not be surprising, but they certainly are challenging.

What can be done to enable students to grasp important basic concepts? Certainly, statistics instructors strive to have students develop critical thinking skills. Using formulae to make calculations is not the challenge. Requiring students to demonstrate a through understanding of statistical inference, however, is challenging, and the challenge provides an excellent opportunity to teach quantitative reasoning skills. The logic of statistical inference, i.e., constructing confidence intervals and testing hypotheses, depends upon the concept of a sampling distribution. This paper proposes that discussions about confidence intervals are aided by using the terms margin of error and sampling error. As a first step in understanding the challenge of teaching the concepts underlying interval estimation, we reviewed the use of three terms: sampling distribution, sampling error, and margin of error in a variety of textbooks.

Experienced statisticians realize that a confidence interval for the population mean depends upon a sample mean, the standard deviation for a sampling distribution for the sample means, and the probability distribution or probability density of the sampling distribution itself. The sampling distribution and level of confidence determine values to be used in calculating the upper and lower limits of the confidence interval.

In the simplest case, the population variance is known, the sampling distribution of sample means is normal, and the confidence interval for μ is calculated as follows:

$$(1-\alpha)\%\ CI = \overline{X} \pm Z_{1-\alpha/2}\sigma_{\overline{X}} \quad . \qquad (1)$$

We define the difference between the population mean and unknown sample mean, $\left|\overline{X}-\mu\right|$, to be the sampling error (S.E.) and the term $Z_{1-\alpha/2}\sigma_{\overline{X}}$ to be the margin of error (M.E.). The probabilistic relationship between sampling error and margin of error can be easily be shown (e.g., Hayes, 1973, p.376) to be

$$P(\left|\overline{X}-\mu\right| \le Z_{1-\alpha/2}\sigma_{\overline{X}}) = P(S.E. \le M.E.) = 1-\alpha \qquad (2)$$

While possibly not obvious to the students, for a given sample, $1-\alpha$ is the probability the sampling error is less than or equal to the margin of error. Our belief is that if a student is to demonstrate an understanding of confidence intervals, the relationship between sampling error and margin of error along with the concept of sampling distribution are important

Because of the Delmas et al. (2007) result, the logic that seems so clear to experienced teachers may not be understood by students. As a first step in gaining an understanding of students' difficulty, we examined the several statistics text and their use of the terms sampling distribution, sampling error, and margin of error.

2. Method

First a Google search was conducted for the terms sampling distribution, sampling error, hypothesis testing, and margin of error. While hypothesis testing is not the focus of this paper, it provides a reference by which to judge the frequency of use of the three terms that are the focus. The Google search was undertaken to obtain an easy approximation of the relative importance of the three terms according to usage on the internet. Obviously there are better ways of examining the usage of the terms in various disciplines, but a Google search is quick and easy.

Second, a sample of 12 text books was examined. The text books were readily available and all but two were beyond the first edition indicating some success in class room use. The texts were targeted at business, psychology, and health science students. The textbook indexes were reviewed to see which of the three terms were listed. Then the minimal difference in page numbers between sampling error and margin of error was determined and then the page distance between sampling error and sampling distribution. Finally, comments in the text books about sampling error and margin of error were noted.

3. Results

The Google search was conducted by the first author on July 25, 2008 for the three targeted terms and the term hypothesis testing. Margin of error clearly resulted in the most hits and sampling distribution the fewest. Table 1 shows the results for all four terms. Quite possibly, the frequent use of the margin of error term is due to political, business, and government surveys being reported on the internet.

Table 1:

Google Search Results for the Targeted Terms

Margin of error	about 1.96 million hits
Hypothesis testing	about 1.24 million hits
Sampling error	about 1.19 million hits
Sampling distribution	about 212,000 hits

The first element of the text book review was to see whether the three terms used in teaching confidence intervals appeared in each text book. Table 2 shows that all but one of the textbooks contained the term sampling distribution in the index. On the other hand, only four of the textbooks used both sampling error and margin of error.

Table 2

Results of Index Search

Textbook	Targeted Term Contained in Index		
	Margin of error	Sampling error	Sampling Distribution
Anderson et al.	Yes	Yes	Yes
Berenson et al.	Yes	Yes	Yes
Black	No	Yes	Yes
Gerstman	Yes	No	Yes
Gravetter & Wallnau	Yes	Yes	Yes
Groebner et al.	Yes	Yes	Yes
Hamburg	No	Yes	Yes
Hays	No	No	Yes
Keller	No	Yes	Yes
Moore et al.	Yes	No	Yes
Witte & Witte	Yes	No	Yes
Zar	No	No	No

To see whether the terms sampling error and margin of error were used together, the indexes were searched to find the minimal distance in pages between these two terms. Table 3 presents these results for the four texts that included the two terms in their indexes. Table 4 presents similar results for the seven textbooks that included both sampling distribution and sampling error in their indexes. Note that often there were multiple references to the terms in the indexes, but the results show only the pages used to calculate the minimum distance.

Table 3

Clo*seness of Use: Sampling Error and Margin of Error*

Textbook	Sampling Error Page	Margin of Error Page	Minimum Distance
Groebner et al.	330	337	7
Berenson et al.	264	265	1
Anderson et al.	287	286	1
Gravetter & Wallnau	7	7	0

Table 4

Closeness of Use: Sampling Distribution and Sampling Error

Textbook	Sampling Distribution Page	Sampling Error Page	Minimum Distance
Keller	277	152	125
Groebner et al.	296	292	4
Berenson et al.	267	264	3
Anderson et al.	252	250	2
Hamburg	212	214	2
Black	232	231	1
Gravetter & Wallnau	197	198	1

In addition to examining the indexes, the text books were examined to obtain comments made about the terms sampling error and margin of error. Table 5 includes comments thought to be of interest. One noteworthy result is inconsistency in defining sampling error, some texts use the absolute value of the difference between the value of the statistic and parameter being estimated, other texts do not use the absolute value. Occasionally, the texts are vague and portray sampling error as the sample not representing the population, e.g., the Black textbook.

The margin of error concept is most interesting. Two of the textbooks, Berenson, et al. and Gravetter and Wallnau, equate the margin of error and sampling error. The Gertsman text portrays margin of error as “wiggle room” involved in making an estimate. Keller does not use the term margin of error, but does define W which equates to others’ use of the margin of error term. While we claim the importance of knowing the level of confidence is the probability that the sampling error is less than the margin of error, only the Anderson et al. text uses the terms sampling error and margin of error in consecutive sentences.

Table 5

Interesting Comments about Margin of Error and Sampling Error

Textbook	Comment
Anderson et al.	"The absolute value of the difference between an unbiased point estimate and the corresponding population parameter is called **sampling error.** P. 250
Anderson et al.	"There is a .95 probability that the sample mean will provide a sampling error of 3.92 or less. … The value 3.92 … is referred to as the *margin of error."* P. 288
Berenson, et al.	"… a statement regarding the margin of error such as 'the results of the poll are expected to be within ± 4 percentage points of the actual value.' This **margin of error** is the sampling error." P. 264-5
Black	"**Sampling error** occurs *when the sample is not representative of the population.*" P. 231
Gerstman	We may think of the margin of error as the "wiggle room" surrounding the point estimate. P. 197
Gravetter & Wallnau	"The 'margin of error' is the sampling error." P. 7
Gravetter & Wallnau	"The general concept of sampling error is that the sample typically will not provide a perfectly accurate representation of its population." P. 209
Groebner et al.	"Sampling error. The difference between a measure computed from a sample (a statistic) and the corresponding measure computed from the population (a parameter)." P. 286 and 1033
Groebner et al.	$Sampling\ \ error = \bar{x} - \mu$. P. 287
Hamburg	"The difference between the statistic $\bar{x}$ and the parameter μ, $\bar{x} - \mu$, is referred to as *sampling error.* P. 214
Keller	"…let W represent the sampling error we are willing to tolerate, which is the quantity following the ± sign." P. 321

4. Discussion

The intent of this study was to attempt to try to understand why students take statistics classes and still fail to understand the concepts of sampling variability, sampling distributions, and confidence intervals. There is no intent to disparage any of the textbooks that were reviewed nor is any one textbook being promoted. Textbooks have many features, cover many topics, and can be evaluated on many criteria. Clearly there are differences in how the topics reviewed here are presented to students, and having variation in textbooks allows for instructor choice when developing a class. On the other hand, we believe we have identified a general weakness in the textbooks that may in part explain the result that Delmas et al. (2007) lament.

A premise of this paper is that understanding the concepts of sampling error, sampling distributions, and margin of error need to be understood if a student is to demonstrate an understanding of confidence intervals and their relationship to sampling variability and sampling error. Textbooks are often the students' primary tool to understanding difficult concepts, and hence, the presentation and definition of terms are important. There are three specific points that deserve comment.

A first point to consider is the use and definition of margin of error. Even though margin of error is the term with the most hits in the Google search, not all of the textbooks included it in their indexes. The textbooks that were reviewed are used in disciplines outside of statistics and hence, the courses are often considered service courses. From a view of developing statistical literacy, students are better served when the term is included. The frequent use of margin of error in media and on the internet suggests failing to include the term does a disservice to students. We believe that equating margin of error to sampling error does another disservice because two distinct concepts are involved, i.e. setting the limits of the confidence interval and knowing that the value of the sample statistic does not equal the population parameter.

Second, the definitions of sampling error may be a point for discussion. Should the absolute value be involved in defining sampling error? Is merely saying that samples differ from populations because of sampling error sufficiently precise to enable students to understand confidence intervals? Sampling error was defined above using the absolute value, but there are reasons not to use it, e.g., the distribution of sampling errors would then have the same shape as the distribution of the sample statistic if absolute value were not used. On the other hand, using a vague definition of sampling error, does not set a firm basis for the critical understanding of confidence intervals that students are often expected to demonstrate.

Finally, the statement that the level of confidence is the probability that the sampling error is less than the margin of error is an important concept. The news media often reports survey results with a point estimate and a margin of error. Interestingly, the level of confidence is often not reported. The use of the margin of error in the media on the internet support the argument that understanding relationship between sampling error and margin of error is important from a statistical literacy perspective. Only the Anderson et al. text came close to making a clear statement using the two terms together.

Obviously the study has several limitations. The Google search was easy, but may not reflect the actual importance of the terms in understanding confidence intervals. For example, we certainly think that understanding all of inference depends on mastering sampling distributions, yet of the terms we searched, sampling distribution had the fewest hits. A second potential criticism concerns the textbooks. Several of the textbooks that were reviewed may have impressive track records, but we do not know their overall circulation or relative frequency of adoption. Perhaps there are many other textbooks that clearly integrate the concepts being studied and have great circulation. Another possible limitation is that the textbook may not be the primary source of information for the students. Instructors often have detailed notes that supplement the textbooks and these notes have not been considered. In spite of these limitations, instructors should review how they encourage students to learn the relationships among confidence intervals, sampling error, and margin of error.

Because the relationship among the concepts is challenging, teaching the relationships provides a good opportunity to develop critical thinking skills and "stress conceptual understanding rather than mere knowledge of procedures" as encouraged by the GAISE report (2005). Toward this end, we are developing a set of exercises (see Appendix) that focuses on using the definitions of margin of error and sampling error along with the relevant sampling distributions. Software (Fulcomer, Kriska, Sass, and Jauregui, 2008) is Excel based and calculates exact sampling distributions for sample means and proportions for several parent populations. The exercises are available from the first author (davidkriska@sbcglobal.net) and the software is available from the second author (mcfulcomer@aol.com).

We believe we have taken the first step in understanding the results that Delmas et al. have brought to our attention. As a next step, we would like the thoughts expressed in this paper to be discussed among textbook authors, statistics instructors, and statistics education researchers. Clearly there will be views that differ with ours. There are also opportunities for empirical research to see if the ideas that have been suggested may actually make a difference.

5, Reference List

References marked with an asterisk indicate textbooks included in the review.

*Anderson, D. R., Sweeney, D. J., & Williams/Thomas A. (2002). *Statistics for Business and Economics, 8th* ed. Cinncinnati, OH: South-Western, Thomson Learning.

*Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2009). *Basic Business Statistics: Concepts and Applications 11th ed*. Upper Saddle River, NJ: Pearson Prentice Hall.

*Black, K. (2004). Business *Statistics for Contemporary Decision Making, 4th* ed. Hoboken, NJ: Wiley.

Delmas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6(2),* 28-58.

Fulcomer, M.C., Kriska, S.D., Sass, M.M., & Jauregui, M. (2008). Classroom Illustrations of the Construction of Discrete Sampling Distributions. Poster presented at JSM, Denver, CO August 4, 2008.

GAISE College Report (2005). *http://www.amstat.org/education/gaise/GAISECollege.htm.*

**Gerstman, B. B. (2008). Basic* Statistics: Statistics for Public Health Practice. Boston: Jones and Bartlett Publishers.

*Gravetter, F. J., & Wallnau, L. B. (2007). *Statistics for the Behavioral Sciences, 7th ed*. Belmont, CA: Thomson Wadsworth.

*Groebner, D. K., Shannon, P. W., Fry, P. C., & Smith, K. D. (2008). *Business Statistics: A Decision - Making Approach, 7th ed*. Upper Saddle River, NJ: Pearson Prentice Hall.

*Hamburg, M. (1977). *Statistical Analysis for Decisioni Making, 2nd ed.* Chicago: Harcourt, Brace, Jovanovich, Inc.

*Hays, W. L. (1973). *Statistics for the Social Sciences, 2nd ed*. Chicago: Holy, Rinehart and Winston, Inc.

*Keller, G. (2005). *Statistics for Management and Economics, 7th ed*. Belmont, CA: Thomson Brooks/Cole.

*Moore, D. S. , McCabe, G. P., Duckworth, W. M., & Sclove, S. L. (2003). *The Practice of Business Statistics: Using Data for Decisions.* New York: W.H. Freeman and Company.

*Witte, R. S. , & Witte, J. S. (2007). *Statistics, 8th ed*. Hoboken, NJ: John Wiley & Sons, Inc.

*Zar, J. H. (1999). *Biostatistical Analysis*. Prentice Hall.

Appendix

Exercise Integrating Terms—simple sampling distribution and sampling error.

1. For the sake of this problem, presume that we know that 30% (*p)* of all US adult citizens approved of the president's current performance. A researcher randomly selects four US adult citizens and asks each person in the sample whether he or she approves of the president's performance.
   a. (Bloom's comprehension) In the first sample, exactly 50% of the sample (2 of 4, ($\hat{p} = .50$)) approved of the president's current performance. Identify the following terms as they apply to this problem:
      i. Population
      ii. Sample
      iii. Parameter
      iv. Statistic
      v. Sampling error
   b. (Bloom's Application/Analysis) Part a is the result of one only sample. Now, identify every possible value for the sample proportion ($\hat{p}$) for samples of size four. For each value of the sample statistic ($\hat{p}$), calculate the probability of the value and then the probabilities for the various values of sampling error.
      Hint: complete the following tables. To calculate $P(\hat{p})$, use the binomial formula.

   c. (Bloom's analysis) For random samples of size four, what is the probability that the sampling error for the sample proportion is greater than .25? What is the probability the absolute value of the sampling error is greater than .25?

   d. Note how the following computer output corresponds to your work above. The computer will make calculations easy as we change the sample size.

Number of Trials (the maximum value to be displayed; see #2 below) **N =** **4**

Probability of Success on a Typical Trial (*p*) **Model I** **0.30000000**

Number of Successes	$\hat{p}$ Proportion of Successes	Number of Combinations	Probability of a Typical Sequence		$P(\hat{p})$ Probability	
X = r ("count")	X = r ("mean")	$S_r \binom{n}{X}$ ("sequences")	T_r Model I		S_rT_r Model I	
-----	-----	-----	-----		-----	
0	0.00000000	1	0.24010000		0.24010000	
1	0.25000000	4	0.10290000		0.41160000	
2	0.50000000	6	0.04410000		0.26460000	
3	0.75000000	4	0.01890000		0.07560000	
4	1.00000000	1	0.00810000		0.00810000	