A Note on Confidence Interval Estimation and Margin of Error

Dennis Gilliland
Vince Melfi
Michigan State University

Key Words: Confidence interval estimation; Margin of error; Interpretations; Misinterpretations.

Abstract

Confidence interval estimation is a fundamental technique in statistical inference. Margin of error is used to delimit the error in estimation. Dispelling misinterpretations that teachers and students give to these terms is important. In this note, we give examples of the confusion that can arise in regard to confidence interval estimation and margin of error.

1. Introduction

Confidence interval estimation is a widely used method of inference and *margin of error* is a commonly used term, and these occupy a large part of introductory courses and textbooks in Statistics. It is well-known that these concepts are often misused and misunderstood. Examples of incorrect interpretations from a variety of sources, some "authoritative," are given in Thornton and Thornton (2004). Recent doctoral theses of Liu (2005) and Noll (2007) address teacher and teaching assistant understanding of the concepts. Misunderstandings arise for a variety of reasons, some as simple as confusing population parameters and sample statistics. Some of the confusion in regard to the meaning and interpretation of these terms stems from the lack of appreciation of the difference between a random variable (a function) and its realization

(evaluation). This is illustrated in the following sections. For simplicity we discuss confidence intervals of the form $\hat{\theta} \pm me$ where $\hat{\theta}$ denotes a statistic and *me* denotes *margin of error*.

2. Confidence interval.[1]

Elementary methods in statistics include confidence interval estimation of a population mean μ and population proportion π. The interpretation of a confidence interval derives from the sampling process that generates the sample from which the confidence interval is calculated. With a probability distribution over possible samples, the (random) interval is constructed to have a specified probability of covering the "true" parameter value. For example, assuming that the observations are normally distributed with mean μ and standard deviation σ, $\Pr_{\mu}(\bar{y} \pm 2\sigma/\sqrt{n} \text{ captures } \mu) = 0.954$ for all μ.

There is a certain awkwardness in interpreting the interval once the data are available and the endpoints are calculated. Returning to the above example, suppose that $n = 25$, $\sigma = 10$ and the simple random sampling process results in a sample with mean $\bar{y} = 50$. The realized interval estimate is 50 ± 4. Many writers tell the readers that a statement like $\Pr(46 < \mu < 54) = 0.954$ is to be avoided since "μ is either in the interval (46, 54) or not." This reluctance is natural to the frequentist since the statement $\Pr(46 < \mu < 54) = 0.954$ might suggest that there is a probability distribution over the parameter values. Therefore, we refer to a *level of confidence* 95.4% for the interval $46 < \mu < 54$.

As expected, textbooks use reasonable language in describing confidence intervals. Consider De Veaux, et al (2009, p. 489). Here an interpretation based on repeated sampling is given for the ideal 95% confidence interval for a population proportion π: "Formally what we mean is that '95% of the samples of this size will produce confidence intervals that capture the true proportion.' This is correct, but a little long winded, so we sometimes say, 'we are 95% confident that the true proportion lies in our interval.'"

Moore, et al (2009, p. 367) state the following in regard to the realization 461 ± 9 for the random interval estimate $\bar{x} \pm 9$: "We cannot know whether our sample is one of the 95% for which the interval $\bar{x} \pm 9$ captures μ or one of the unlucky 5% that does not catch μ. The statement that we are 95% confident is shorthand for saying. 'We arrived at these numbers by a method that gives correct results 95% of the time.'"

As another example, consider discussion of the (random) interval estimate $\bar{y} \pm 1.96\sigma/\sqrt{n}$ for the mean μ of a normal distribution based on random sampling found in Devore (2008, Chapter 7). In Example 7.2, an interval (79.3, 80.7) is realized. The author concludes: "That is, we can be highly confident, at the 95% confidence level, that $79.3 < \mu < 80.7$." However, on pages 257-258 we find a reluctance to use "confidence" to describe the

---

[1] A general theory of confidence set estimation was developed by Neyman (1937). We will not get into Fisher's fiducial probability and Bayesian approaches.

realized interval: "the confidence level 95% is not so much a statement about any particular interval such as (79.3, 80.7). Instead it pertains to what would happen if a very large number of like intervals were to be constructed using the same CI formula".

The above quotes illustrate the commonly accepted use of *frequentist probability* when it comes to the process by which a confidence interval estimate is generated and the term *confidence level* when it comes to the realized interval estimate[2].

Although it is incorrect to make a probability statement such as $Pr(46 < \mu < 54) = 0.954$ about a realized confidence interval, there is an important way in which the interpretation of the realized interval is the same as the interpretation of the random interval. First consider a simpler example: Steve Nash is one of the NBA's best free throw shooters, making 90% of his free throws. With the sample space {Make, Miss} we can write $Pr\{Make\} = 0.9$. Consider a bettor Mr. X who will pay c dollars if Steve Nash misses his next free throw and collect d dollars if Steve Nash makes the free throw. Then Mr. X has an expected profit of $0.9d - 0.1c$. So before Steve Nash shoots his next free throw Mr. X, with mathematical expectation governing his behavior (and ignoring questions of utility and special circumstances), would be indifferent between betting nothing and making the above bet with $c = 9$ and $d = 1$. If Steve Nash already has shot his free throw but Mr. X does not know the result, Mr. X would still be indifferent between betting nothing and betting \$9 versus \$1 that the free throw was made, even though it may not now make sense to write $Pr\{Make\} = 0.9$.

Next consider a 90% confidence interval for a mean in the context of a normal population with known standard deviation. Mr. X would be indifferent between betting nothing and betting \$9 versus \$1 that the random interval will capture μ, since $Pr_\mu(\bar{y} \pm 1.65\sigma/\sqrt{n} \text{ captures } \mu) = 0.9$ for all μ. Now suppose that the data are gathered and the interval is calculated to be (10, 17). Mr. X would still be indifferent between betting nothing and betting \$9 versus \$1 that the value of μ is between 10 and 17 (assuming that in the future the true value of μ could be revealed), even though the probability statement $Pr(10 < \mu < 17) = 0.9$ may not make sense.

Neyman's (1937) construction of confidence intervals via the inversion of a family of hypothesis tests provides an alternative interpretation that has found favor in some legal proceedings. In oral testimony in a legal proceeding, Rothman (2007) presented a 95% confidence interval for a parameter as the set of values of the parameter not rejected by the data at level 0.05. In another legal proceeding, Katz (1975) presented the confidence set as the set of hypotheses (values of π) that are not rejected by the data at some pre-assigned level and used 0.05. These experts felt that, even for an audience with minimal statistical training, it is meaningful to interpret a confidence interval estimate as the set of values of the parameter not rejected by the data. We agree that this can be an illuminating way to view a confidence interval even for statistically naïve users, and hope that its effectiveness will be further studied.

---

[2] Noll (2007, p. 218) seems to take a contrary view in regard to the term *confidence* and the realized interval estimate.

In intermediate and advanced courses, the following examples may serve to remind students that there are situations where the confidence level derived from the marginal probability distribution induced by the sampling is problematic.

Example 1. (D. Basu). In this example, the sample mean $\bar{Y}$ does not contain all the useful information concerning the unknown parameter. Consider the family of discrete probability distributions P_θ where $P_\theta(Y = \theta - 1) = P_\theta(Y = \theta) = P_\theta(Y = \theta + 1) = 1/3$, where θ is an integer. Then for random sampling with $n = 2$, $P_\theta(\bar{Y} \pm 3/4 \text{ captures } \theta) = 7/9$ for all θ; yet conditional on any realized sample with range $R = 2$, the probability is 1. For example with this model and the realized sample $\{17, 19\}$, the parameter θ must be 18, and the imputation of confidence level 7/9 to the interval estimate $18 \pm \frac{3}{4}$ would be foolish. This example is unusual in that the parameter value is determined with certainty for certain easily identified sample outcomes.

Example 2. Medicare Overcharges. This is an example where the parameter of interest is known to satisfy certain conditions. Consider an audit of $N = 5{,}000$ charges to Medicare totaling \$10,000,000. Let y_i denote the <u>overcharge</u> to Medicare in billing i, $i = 1, 2, \ldots, N$ and let μ_y denote their average. Thus, μ_y is known to be no greater than \$2,000. Suppose that simple random sampling is used with $n = 25$ and that the sample statistics are $\bar{y} = \$2{,}150$ and $s = \$400$. The nominal 90% t-interval estimate of μ_y is $\$2{,}150 \pm 1.711\sqrt{(5{,}000 - 25)/25(5{,}000 - 1)}\$400 = (\$2013, \$2287)$ yet $\mu_y \leq \$2{,}000$. <u>Edwards et al (2003)</u> show that such examples do arise in practice. This prompted their development of the minimum sum method. In this example, for some realizations of the sample, the known constraints on the parameter show that the interval estimate does not cover the parameter.

3. Margin of error

<u>Liu (2005)</u> and <u>Noll (2007)</u> suggest that it is important to have an alternative way to look at confidence intervals (which is formally equivalent to the usual approach). The alternative approach begins, for example, with an elaboration of the statement $\Pr_\theta(\hat{\theta} \pm me \text{ captures } \theta) \geq 0.95$. The equivalence of the events $\hat{\theta} - me \leq \theta \leq \hat{\theta} + me$, $\theta - me \leq \hat{\theta} \leq \theta + me$ and $|\hat{\theta} - \theta| \leq me$ allows for such elaborations. For example, consider $\bar{y} \pm 1.96\sigma/\sqrt{n}$, a case where me is not random. Of course, $\bar{y} - 0.7 < \mu < \bar{y} + 0.7$ is equivalent to $\mu - 0.7 < \bar{y} < \mu + 0.7$, and <u>Liu (2005)</u> suggests that students may be better served by thinking in terms of the latter, namely, that there is a 95% probability that the random $\bar{y}$ falls within 0.7 of the population mean. <u>Liu (2005)</u> limits this approach to cases where me is not sample-based. In <u>Noll (2007</u>) the same idea is used in contexts where me itself changes from sample to sample, and calls the approach an "alternative construal."

There seems to be some confusion created in attempts to use either approach discussed in the preceding paragraph to elaborate on the meaning of a confidence interval estimate and margin of error. This arises when the margin of error is sample-based, the common case in t-intervals for the mean and in intervals for proportions. Consider the example of <u>De Veaux et al (2009, page 490)</u>:

FOR EXAMPLE. Polls and margin of error.
On January 30-31, 2007, Fox News/Opinion Dynamics polled 900 registered voters nationwide. When asked, "Do you believe global warning exists?" 82% said "Yes". Fox reported their margin of error to be $\pm$ 3%.

Question: It is standard among pollsters to use a 95% confidence interval unless otherwise stated. Given that, what does Fox News mean by claiming a margin of error of $\pm$ 3% in this context?

If this polling were done repeatedly, 95% *of all random samples would yield estimates that come within* $\pm$ 3% *of the true proportion of all registered voters who believe that global warning exists.*

The interpretation (italicized wording in the FOR EXAMPLE) is wrong if the margin of error is a function of the sample of 900 (for example, $\sqrt{\hat{\pi}(1-\hat{\pi})/900}$). (It is also wrong if it is calculated from $\sqrt{0.5(1-0.5)/900}$.)

Example 3. For estimation of π for a large population, suppose that simple random sampling was used with $n = 900$ with result $\hat{\pi} = 0.82$. The nominal 95% large sample confidence interval estimate of π is 0.82 ± 0.025. Assume for the purposes of illustration that the sample count $900\hat{\pi}$ is Binomial with $n = 900$, $\pi = 0.79$. If this polling were done repeatedly, about 93.5% (<u>not</u> 95%) of all random samples of size $n = 900$ would yield estimates $100\hat{\pi}\%$ that come within $\pm$ 2.5% of 79%. The reason the percentage is not 95% is that the margin of error of the interval is computed based on the sample value $\hat{\pi} = 0.82$ which is not equal to the population value of $\pi = 0.79$.

Example 4. A study was conducted of *usable space y* on utility poles in Michigan to which cable television lines were attached. The population consisted of over one million poles. Suppose that a sampling frame existed that allowed selection of a simple random sample of $n = 50$ poles and that sample statistics were $\bar{y} = 15.3\,ft$ and $s = 4.7\,ft$. The nominal 95% *t*-interval estimate of μ is $15.3\,ft \pm 1.34\,ft$.   Assume for purposes of illustration that y is distributed normally in the population with mean $16.0\,ft$ and standard deviation $6.0\,ft$ so that the distribution of $\bar{y}$ is normal with mean $16.0\,ft$ and standard deviation $0.85\,ft$. If this sampling were done repeatedly, about 88.5% (<u>not</u> 95%) of all random samples of size $n = 50$ would yield estimates $\bar{y}$ that come within $\pm$ 1.34 ft of the population mean usable space 16.0 ft. The percentage is 88.5% rather than 95% because the margin of error is computed based on the sample value $s = 4.7$ which is not equal to the population value of $\sigma = 6.0$.

4. Conclusion

Educators should be aware of the technical points raised in the preceding sections. But it is also important that they convey to students the fact that these technical points assume the correctness of a model, and that understanding the possible weaknesses of the model is also of great value. In

particular, educators should promote a healthy level of skepticism in reading statistical reports. The consumer of statistical information and the practitioner must be concerned with nonsampling errors and biases. It is prudent not to accept a realized confidence interval at face value until there is an audit of the processes that both generated the sample and the evaluations of *y* that shows conformance to accepted practice.

To make the point, consider the Gallup Poll Task used by Noll (2007, Figure 35, p. 215).

> **Gallup Poll Task**
> Your statistics class was discussing a Gallup poll of 500 Oregon voters' opinions regarding the creation of a state sales tax. The poll stated, "… the survey showed that 36% of Oregon voters think a sales tax is necessary to overcome budget problems". The poll had a margin of error of ± 4%. Discuss the meaning of margin of error in this context.

Here are issues that we would like to see students raise, in addition to technical issues: *Notice that the process by which the sample of 500 was generated is not given. There is no indication as to what sampling frame was used, how the voters were contacted, what defines a voter, what instrument was read or sent, what frame issues exist, and what protocols were used in regard to follow-up after initial nonresponse. There is no mention of nonresponse rate. There is no mention of documentation that allows for an audit of the sampling process, there is no indication as to how the sample was selected. (An educated guess is that it was not by simple random sampling, although standard errors based on simple random sampling commonly agree with values reported for polls in the media.) There is no mention of the processes used to control nonsampling errors and biases. All of these issues are important for a meaningful and defensible interpretation of the stated margin of error.*

However, assuming that Gallup followed accepted practice, used a probability sample, and a nominal 95% level confidence interval, then there is a rational basis for acting according to approximate 95% confidence in the statement that 36% ± 4% captured $100\pi\%$. Thus, we fall back on the usual interpretation, e.g. from De Veaux et al (2009, p. 489) that "we are 95% confident that the true proportion lies in our interval," with "95% confident" being appropriately interpreted.

We conclude with an analogy that seems to convey the essence of the probability features and inherited confidence level in confidence interval estimation.

Example 5. Pitching Horseshoes[3]**.** Consider a person so skilled that he/she can pitch a horseshoe blind-folded and with ear plugs and make ringers 95% of the time. After pitching the shoe one time, the person does not know whether a ringer was made or not. Yet, he/she is so confident that he/she is willing to bet at about 19:1 that a ringer was made.

Rarely in the practice of statistics is it possible to determine with certainty whether an interval estimate does or does not contain the parameter value, i.e., whether a "ringer was or was not

---

[3] See Wonnacott and Wonnacott (1990, pages 256-257).

made." It is the process that generated the interval estimate and the documentation of that process that allow the user and decision-maker to have confidence in the interval, i.e., that a "ringer was made."

Acknowledgment

References

De Veaux, R. D., Velleman, P. F. and Bock, D. E. (2009). Intro Stats, Instructor's Edition, Third Edition, Pearson.

Devore, Jay L. (2008). Probability and Statistics for Engineering and the Sciences, Seventh Edition, Thompson.

Edwards, D., Ward-Besser, G., Lasecki, J., Parker, B., Wieduwilt, K., Wu, F. and Moorehead, P. (2003). The minimum sum method: a distribution-free sampling procedure for Medicare fraud investigations. *Health Services and Outcomes Research Methodolgy*. **4** 241-263.

Katz, Leo (1975). Presentation of a confidence interval estimate as evidence in a legal proceeding. *The Am. Statist.* **29** 138-142.

Liu, Yan (2005). *Teachers' understandings of probability and statistical inference and their implications for professional development*. PhD Dissertation, Vanderbilt University. www.stat.auckland.ac.nz/~iase/publications/dissertations/dissertations.php

Moore, David S., McCabe, George P. and Craig, Bruce A. (2009). Introduction to the Practice of Statistics, Sixth Edition, Freeman.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London, Series A. Mathematical and Physical Sciences*. **236** 333-380.

Noll, Jennifer Ann (2007). *Graduate teaching assistants' statistical knowledge for teaching*. PhD Dissertation, Portland State University. www.stat.auckland.ac.nz/~iase/publications/dissertations/dissertations.php

Rothman, Edward (2007), Oral testimony in US Federal Court, Western District of Michigan, Case No. 1: 03-CR-00094 – GJQ.

Thornton, Robert J. and Thornton, Jennifer A. (2004) Erring on the Margin of Error. *Southern Economic Journal*. **71** 130-135.

Wonnacott, Thomas H. and Wonnacott, Ronald J. (1990). Introductory Statistics, Fifth Edition, John Wiley and Sons.

Dennis Gilliland
Department of Statistics and Probability
Michigan State University
mgilliland@stt.msu.edu

Vince Melfi
Department of Statistics and Probability
Division of Science and Mathematics Education
Michigan State University
melfi@stt.msu.edu