

TEACHING CONFIDENCE INTERVALS

JOHN M. HOLTE
NCS-MAA SPRING MEETING

ABSTRACT. An easy way to teach the notion of a “confidence interval” is presented. It involves a class in the simulation of confidence intervals, it provides striking visual imagery, and it lends itself to discussion of related issues. The basic idea is this: Simulate samples of size 3 from the population of digits with the uniform distribution, and form confidence intervals for the median from the minimum and maximum of each sample. Other confidence intervals for the median and other quantiles are also considered.

1. INTRODUCTION

What is meant by a “95% confidence interval”? Many students in introductory statistics classes cannot answer this question correctly. This is not surprising: Since its introduction by Jerzy Neyman in 1937, the notion of a confidence interval has been an elusive concept. Recently, inspired by an example in the original Minitab Student Handbook (Ryan, Joiner, and Ryan 1976), I found a way to present the concept that actively engages a class, provides a strong visual impression, illustrates several noteworthy features of confidence intervals, and offers an opportunity for classroom discussion related to sampling and simulation. Also, it is very easy to carry out.

2. BASIC LESSON

In order for students to appreciate this presentation, it is necessary that they understand at least the rudiments of probability theory. Of course, they must also have an idea of what probability means, or they will be unable to grasp the distinction between probability and confidence. If, in addition, the students understand the basics of the binomial probability model, they will be able to pursue many interesting extensions of the basic lesson, as we shall see.

Date: 26 April 2003.

Key words and phrases. Teaching statistics; interval estimation; median; quantiles; order statistics.

2.1. Probabilistic Preliminaries. We seek a confidence interval for the median η of a population. Assume, for convenience, that the probability is zero that a random value X sampled from the population is equal to the population median: $P(X = \eta) = 0$. Of course, this is automatically true when X is a continuous random variable, but our soon-to-be-given example will be discrete.

Now consider a random sample of size 3: X_1, X_2 , and X_3 . Following Ryan, Joiner, and Ryan (1976, pp. 87–89), our preliminary goal is to show that

$$P(\text{Min} \leq \eta \leq \text{Max}) = \frac{3}{4}$$

where Min is the smallest sample value and Max is the largest sample value.

Consider the complementary event. The sample minimum is greater than η if and only if all three sample values are greater than η , and the probability of this is $P(X_1 > \eta)P(X_2 > \eta)P(X_3 > \eta) = (\frac{1}{2})^3$. Similarly, the probability that the sample maximum is less than η is also $\frac{1}{8}$. Thus,

$$P(\text{Min} \leq \eta \leq \text{Max}) = 1 - \left(\frac{1}{8} + \frac{1}{8}\right) = \frac{3}{4}.$$

Accordingly, after one has taken a sample of size 3 and determined actual values for the sample minimum, min , and the sample maximum, max , the interval $[\text{min}, \text{max}]$ will be a 75% confidence interval for the population median.

A convenient population consists of the digits 0, 1, 2, \dots , 9 with the uniform distribution. What is its median? Students should readily see that if $\eta = 4.5$, then $P(X < \eta) = \frac{1}{2}$, $P(X > \eta) = \frac{1}{2}$, and $P(X = \eta) = 0$.

2.2. Simulated Confidence Intervals. Now the idea is this: Simulate samples of size 3 from the population of digits, and form confidence intervals for the median from the minimum and maximum of each sample.

The detailed procedure might be as follows. The instructor writes the digits 0, 1, 2, \dots , 9 horizontally across the top of the board and draws a line starting between 4 and 5 vertically down the board.

Using a table of random digits, the students, in turn, pick three random digits and report them. On the board the instructor writes each sample, possibly ordered, and draws a horizontal line representing the confidence interval. It starts at a point matching the minimum digit and extends to a position matching the maximum digit. (The probability of getting a degenerate interval is only 0.001.) If this line intersects the vertical line, the confidence interval succeeds in including

the median. Successive lines should be drawn close together so that many lines may be fitted into one panel of the board. A key feature of this simulation is that the samples may be processed rapidly.

After some predetermined number N of trials, say $N =$ the number of students in the class, the board will display N horizontal lines representing the simulated confidence intervals, and those that include the population median value of 4.5 will conspicuously cross the vertical line that marks η . The percentage of confidence intervals that include the true median is then calculated and compared to the theoretical 75% value.

2.3. Discussion. The following points regarding confidence intervals may be adduced in class discussion:

- Each confidence interval either succeeds or fails in capturing the true value of the population parameter. For any particular interval, the probability that it actually includes the true value is either 0 or 1; for example, $P(\eta \in [6, 9]) = 0$, while $P(\eta \in [4, 7]) = 1$. In the simulation, because we know the true value from the start, we can tell which intervals work and which do not, but in real applications we would not know for sure.
- The locations of the intervals vary, and, in particular, their centers vary, depending on the sample.
- The widths (lengths) of the intervals vary, depending on the sample. This is true of many confidence intervals that arise in applications, such as those based on t , χ^2 , and F distributions. (It is not true, though, for the confidence interval based on a sample of given size n for the mean μ of a normal population having a known standard deviation σ ; its width is $2z^*\sigma/\sqrt{n}$ for a value of z^* that depends only on the confidence level.)
- The observed proportion of successful confidence intervals should be approximately equal to the confidence level.

One may also discuss features of the simulation:

- The sequence of outcomes of the simulated confidence intervals—in each case, the success or failure in capturing the population parameter—may be viewed as a Bernoulli trials process.
- As noted above, the observed proportion of successes is an approximation to the true probability of success, which is the theoretical confidence level C . The expected value of the proportion \hat{p} of successes is equal to the true probability $p = C$ of success— $E(\hat{p}) = C$ —so \hat{p} is an unbiased estimator of the confidence level.

- The standard deviation of \hat{p} is inversely proportional to the number N of simulations:

$$\sigma_{\hat{p}} = \sqrt{\frac{C(1-C)}{N}}.$$

For example, if one simulates twenty 75% confidence intervals, $\sigma_{\hat{p}}$ is slightly less than 0.10, so one has a pretty good chance of getting \hat{p} within 0.10 of 0.75; in fact, $P(.65 \leq \hat{p} \leq .85) = .69$. (See equation (2) in Section 4.) For $N = 30$ this probability is .80.

3. VARIATIONS, EXTENSIONS, AND EMBELLISHMENTS

3.1. Other Confidence Levels. A confidence level of 75% is atypically low. A variation on the above procedure yields a confidence level of 93.75%—close to the traditional 95%. The minimum and maximum of a sample of 5 random digits are 93.75% confidence limits for the median. This is an exercise in Ryan, Joiner, and Ryan (1976, p. 90).

One might also like to have a confidence interval where the endpoints are not simply the minimum and maximum. An example is the 87.5% confidence interval for the median based on a sample of 7 random digits whose endpoints are the second smallest and second largest sample digits. Let us consider how such confidence levels are calculated.

3.2. Order-statistic-based Intervals for Quantiles. It is no more difficult to derive confidence intervals for general quantiles than it is for the median, or fiftieth percentile. For $0 < p < 1$, let ξ_p denote the p^{th} quantile, and assume that $P(X = \xi_p) = 0$. (Actually, the following assumes only that $P(X \leq \xi_p) = p$.) For $k = 1, 2, \dots, n$, let $X_{k:n}$ denote the k^{th} order statistic of the sample X_1, X_2, \dots, X_n , so that $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$. Let S_n denote the number of indices j for which $X_j \leq \xi_p$. Then, for $1 \leq r < s \leq n$,

$$\begin{aligned} P(X_{r:n} \leq \xi_p < X_{s:n}) &= P(X_{r:n} \leq \xi_p) - P(X_{s:n} \leq \xi_p) = P(S_n \geq r) - P(S_n \geq s) \\ &= \sum_{j=r}^n \binom{n}{j} p^j (1-p)^{n-j} - \sum_{j=s}^n \binom{n}{j} p^j (1-p)^{n-j}, \end{aligned}$$

so

$$(1) \quad P(X_{r:n} \leq \xi_p < X_{s:n}) = \sum_{j=r}^{s-1} \binom{n}{j} p^j (1-p)^{n-j},$$

a well-known result in nonparametric statistics (Walsh, 1962, p. 206).

The median is $\eta = \xi_{1/2}$. By (1) and the symmetry relation $\binom{n}{n-j} = \binom{n}{j}$,

$$P(X_{k:n} \leq \eta < X_{n-k+1:n}) = \sum_{j=k}^{n-k} \binom{n}{j} \left(\frac{1}{2}\right)^n = 1 - 2^{1-n} \sum_{j=0}^{k-1} \binom{n}{j}.$$

This formula yields the confidence level for $[x_{k:n}, x_{n-k+1:n})$ as a confidence interval for the population median; values for small n and k are given in Table 1.

n :	3	4	5	6	7	8	9	10
$k = 1$.7500	.8750	.9375	.9688	.9844	.9922	.9961	.9980
$k = 2$.0000	.3750	.6250	.7812	.8750	.9297	.9609	.9785
$k = 3$.0000	.0000	.0000	.3125	.5469	.7109	.8203	.8906

TABLE 1. Confidence levels for $[x_{k:n}, x_{n-k+1:n})$

4. HOW WELL WILL IT WORK?

One may predict in advance—probabilistically—how well the demonstration of the chosen confidence interval will work in the the classroom. As in Section 2.3, one simply models the outcomes (success or failure) of the simulated confidence intervals as a Bernoulli trials process. Let C be the chosen confidence level, and let N be the number of simulations to be done. Let T_N denote the number of confidence intervals that succeed in capturing the population parameter (η or ξ_p), and let $\hat{p} = T_N/N$. Then the expected number of successful confidence intervals is $E(T_N) = NC$, and the variance is $\text{Var}(T_N) = NC(1 - C)$. Also, $E(\hat{p}) = C$ and $\text{Var}(\hat{p}) = C(1 - C)/N$.

For example, if $N = 30$ and $C = .75$, the number of successful confidence intervals has expected value 22.5 with a standard deviation of 2.37. If $N = 30$ still but $C = .95$, then the number of successful confidence intervals has expected value 28.5 with a standard deviation of 1.19. Here one has to worry about too much success, i.e., every confidence interval covering the true parameter, because one wants the students to see that the confidence interval is not infallible. In this case the probability that all succeed is $C^N = .95^{30} = .21$ —a moderate level of risk. Compare this with $.75^{30} = .0002$. This is one reason for demonstrating lower confidence levels in class.

For a given value of N one may calculate in advance the probability that the observed proportion \hat{p} of successes will be within a given ϵ of

the actual probability, C :

$$(2) \quad P(|\hat{p} - C| \leq \epsilon) = \sum_{N(C-\epsilon) \leq j \leq N(C+\epsilon)} \binom{N}{j} C^j (1-C)^{n-j}.$$

For moderate and large values of N one may use the normal approximation to the binomial:

$$P\left(\frac{|\hat{p} - C|}{\sigma_{\hat{p}}} \leq \frac{\epsilon}{\sigma_{\hat{p}}}\right) \approx P\left(|Z| \leq \frac{\epsilon\sqrt{N}}{\sqrt{C(1-C)}}\right),$$

where Z is a standard normal random variable.

One may use this approximation to estimate the number of simulations one should do. Let z_α denote the value for which $P(Z \geq z_\alpha) = \alpha$. In order to achieve a probability p for the observed proportion of successful confidence intervals being within ϵ of the true value, C , one should simulate

$$N \approx C(1-C) \left(\frac{z_{(1-p)/2}}{\epsilon}\right)^2$$

confidence intervals. For example, if one wants to have a probability of at least .80 that the percentage of simulated 87.5% confidence intervals that capture the median is between 77.5% and 97.5%, one should simulate about $(.875)(.125)(1.282/.10)^2 = 18$ intervals—easily done. As a check, the exact formula (2) yields a probability of .84 in this case: There is a very good chance of achieving one's desire. But if one wants the observed percentage of successes to be within 5% of the same confidence level, one would have to do four times as many simulations—a tedious task.

These examples suggest what may realistically be accomplished in a classroom demonstration. A moderate number of simulations can be done quickly, and because the construction is so easy to understand and display, this scheme provides a clear and effective introduction to the concept of a confidence interval.

REFERENCES

- [1] Ryan Jr., T. A., Joiner, B. L., and Ryan, B. F. (1976), MINITAB Student Handbook, North Scituate, Mass.: Duxbury Press.
- [2] Walsh, J. E. (1962), Handbook of Nonparametric Statistics, Princeton: Van Nostrand.

GUSTAVUS ADOLPHUS COLLEGE, 800 W. COLLEGE AVE., ST. PETER, MN 56082, FAX 507-933-7041, HOLTE@GAC.EDU