Verifying Confidence Intervals In The Wage-And-Hour Context

By **Brian Kriegler**

Brian Kriegler

In recent years, there have been numerous noteworthy opinions in which the courts discuss the use of sampling and statistics. Such rulings include Tyson Foodsv. Bouaphakeo before the U.S. Supreme Court and Duran v. U.S. Bank before the California Supreme Court. Following these decisions, there appears to be greater focus in the legal community on what conclusions can and cannot be drawn from representative evidence, e.g., a random sample of class members' testimony.

In this article, we explore one type of statistical output that frequently is presented to the trier of fact: confidence intervals for population (classwide) data characteristics. The examples presented below are in the context of wage and hour.[1]

Overview of Random Sampling and Confidence Intervals

Random samples have numerous advantageous features. Such datasets can be used to reach conclusions about a defined population. The statistical literature has shown random sampling to be a neutral and unbiased process.[2] There are also formal and scientific methods for deriving confidence intervals, or CIs, for measurements such as the population mean.[3]

By way of example, suppose a random sample of employees are asked how long they are at the workplace before their shift begins. Assume that the 95 % CI is 20 +/- seven minutes. The interpretation is that (1) the unbiased estimate for average is 20 minutes; and (2) we are 95% confident that the population average is between 13 and 27 minutes.

Figure 1 below shows a visual representation of a CI. Each horizontal line shows a CI based on a new random sample. Blue lines straddle the population mean and red lines miss the mean. Over a large number (e.g., thousands) of random samples of a specified size, we would expect 95% of the horizontal lines to cross the vertical (black) bar signifying the population mean. Equivalently, we can also expect that 5% of the horizontal lines will not cross the vertical bar. In this article we will refer to the percentage of intervals that contain the population mean as the "capture rate."

FIGURE 1
Graphical Illustration of 100 Distinct 95 Percent Confidence Intervals

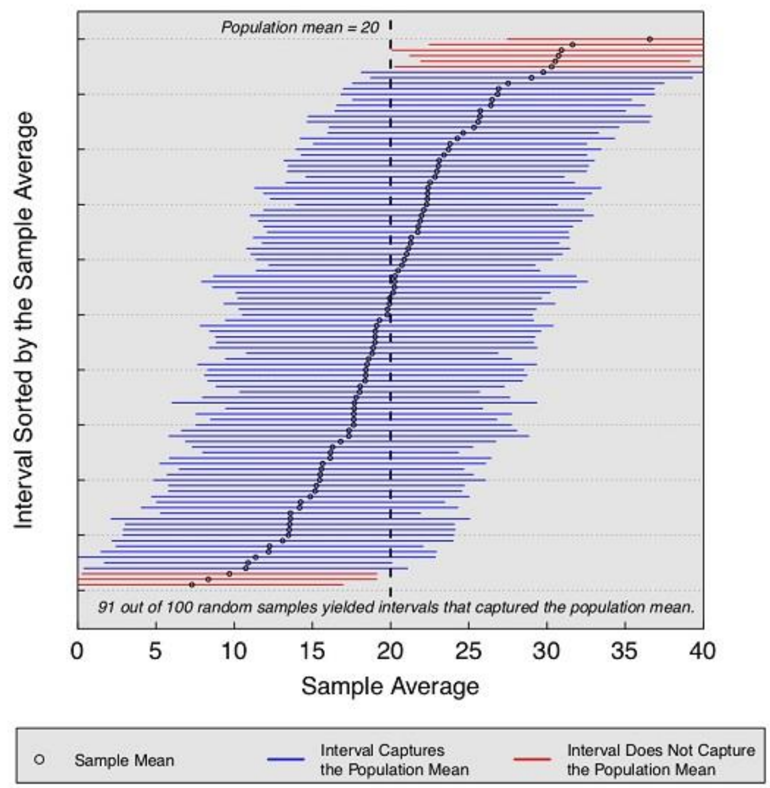


Figure 1 illustrates three key features of CIs and the capture rate. First, each random sample produces a different sample average and interval. Second, some intervals will not contain the population mean. Third, the capture rate does not necessarily equal the stated confidence level. In this instance the capture rate is 91% while the stated confidence level was 95%, i.e., a difference of 4%.

We will explore the implications of the distance between the capture rate and stated confidence level in more detail below. The capture rate provides valuable information about a confidence interval. It can be used as a basis for comparison to the desired level of confidence. It also signifies an estimate for the actual level of certainty.

Here are four questions pertaining to the use of CIs and the capture rate:

1. How do we design an experiment so that the capture rate is compared to the stated confidence level, e.g., 90 or 95%?

2. To what extent is the CI reliable if the difference between the capture rate and stated confidence level is "relatively large"?

3. How does the capture rate compare to the stated confidence level if the underlying data are not symmetric and/or the sample size is small?

4. How do we estimate the capture rate based on a random sample of data?

These questions will be addressed using computational simulations. Results will be shown across combinations of (1) four distinct populations; (2) four sample sizes; (3) two methods for calculating CIs; and (4) two stated confidence levels.[4]

The Experiment

Table 1 below lists nine steps needed to (1) compute the capture rate; and (2) gauge the distance between the capture rate and the stated confidence level. The steps shaded in grey apply to both the Central Limit Theorem-based and bootstrapped CIs.

Table 1: Steps Needed to Compare the Capture Rate to the Stated Confidence Level

Step	If Using the Central Limit Theorem	If Using Bootstrapping
1	*Define* a population and calculate its mean	
2	*State* the desired confidence level(s)	
3	*Select* a random sample of a specified size from the population defined in Step 1	
4	*Compute* the sample mean and sample standard deviation from Step 3	*Generate* thousands of samples with replacement ("resamples") from the random selection in Step 3
5	*Compute* the confidence interval(s) of interest using (i) established confidence interval formulas, and (ii) the sample mean and standard deviation from Step 4*	*Identify* the percentiles of resample means that align with the desired confidence level**
6	*Record* whether the confidence interval from Step 5 has captured the population mean computed in Step 1	
7	*Repeat* Steps 3 through 6 thousands of times	
8	*Calculate* the "Capture Rate," *i.e.*, percentage of CIs that include the population mean based on Step 7	
9	*Compare* this percentage to the target confidence level stated in Step 2	

* - The CLT-based formulas for two-sided 90 and 95 percent CIs are as follows:

- 90% - Sample Mean +/- 1.7 x (Sample Standard Deviation/√Sample Size)
- 95% - Sample Mean +/- 2.0 x (Sample Standard Deviation/√Sample Size)

** - For instance, if 1,000 resamples were taken, a two-sided 90 percent confidence interval would range from the 50th lowest resample mean to the 50th highest resample mean.

We will apply the steps listed in Table 1 using a wide range of statistical circumstances. Specifically, we will:

- Assume four hypothetical and distinct populations
- Select random samples of 25, 50, 100 and 300 observations

- Apply stated confidence levels of 90 and 95%
- Compute CIs using both the Central Limit Theorem, or CLT, and bootstrapping

For the time being we will assume that all values in each population are observable and known. The "Practical Considerations" section below includes a discussion about how the practitioner can apply the aforementioned nine steps using a random sample of data, i.e., when the whole population is not available.

Figure 2 below shows graphs of each of the four hypothetical and distinct populations. These populations are defined as follows:

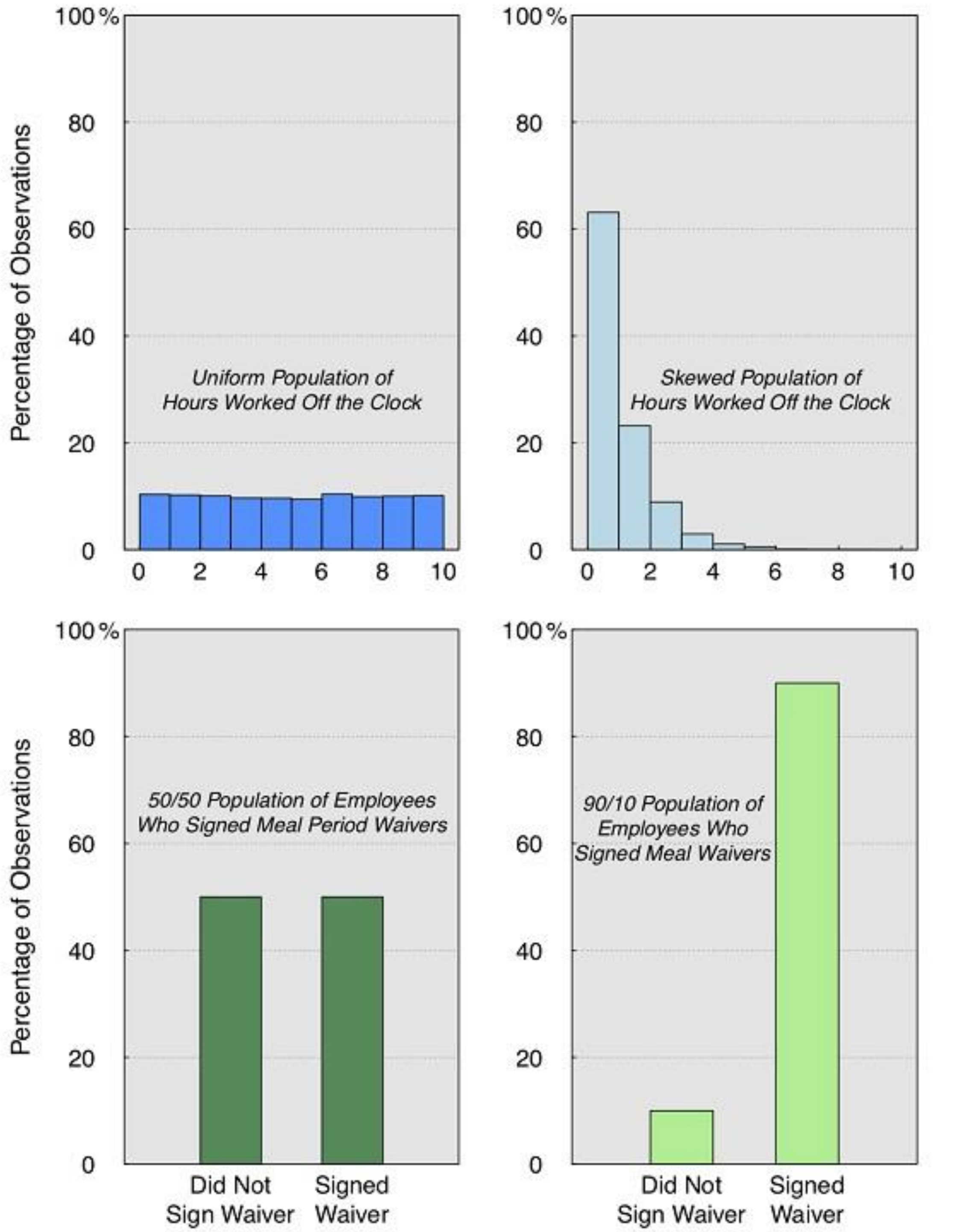
- ***Uniform Data:*** The first population consists of hours worked off the clock by hourly employees. These values range from 0 to 10 with an average of five hours. Each data value has roughly the same probability of occurrence.

- ***Skewed Data:*** The second population consists of hours worked off the clock under a different set of circumstances. These values range from 0 hours to 10 with an average of just one hour. Approximately 65% of values are less than an hour. The remaining 35% of values range from one to 10 hours.

- ***50/50 Data:*** The third population consists of nonexempt employees. Half of the employees signed allegedly invalid meal break waivers and the remaining employees did not sign meal break waivers.

- ***90/10 Data:*** The fourth population also consists of nonexempt employees under a different set of circumstances. 90% of the employees signed allegedly invalid meal break waivers and the remaining 10% of employees did not sign meal break waivers.

FIGURE 2
Four Distinct Populations:
Uniform, Skewed, 50/50, and 90/10 Data



Summary of Results and Graphical Representations

This section includes results pertaining to the distance between the capture rate and stated confidence level for each population and sample size. Such comparisons are performed using CLT-based and bootstrapped confidence intervals. In summary:

Results from CLT experiments:

- Capture rates for 90% CIs range from 87 to 92%
- Capture rates for 95% CIs range from 88 to 96%

Results from bootstrapping experiments:

- Capture rates for 90% CIs range from 86 to 92%
- Capture rates for 95% CIs range from 91 to 96%

Results across all experiments:

- Differences between the capture rate and stated confidence level tend to decrease as the sample size grows.[5]

- The skewed and 90/10 populations yield relatively larger differences between the capture rate and the stated confidence level.[6]

- The uniform distribution yields the smallest differences between the capture rate and the stated confidence level.[7]

- The largest difference between the capture rate is 6.7% and the stated confidence level.

Figure 3 and Figure 4 below show graphical representations of these results. The red dotted line denotes the stated confidence level, i.e., the "target." Figure 3 includes output using CLT-based formulas for CIs, and Figure 4 is based on bootstrapped CIs. These graphs use the same color scheme as in Figure 2 to denote each distinct population.[8] For a given population, sample size and CI method, the capture rates tend to straddle and/or be close to the stated level of confidence.

FIGURE 3
Capture Rates When Applying the CLT

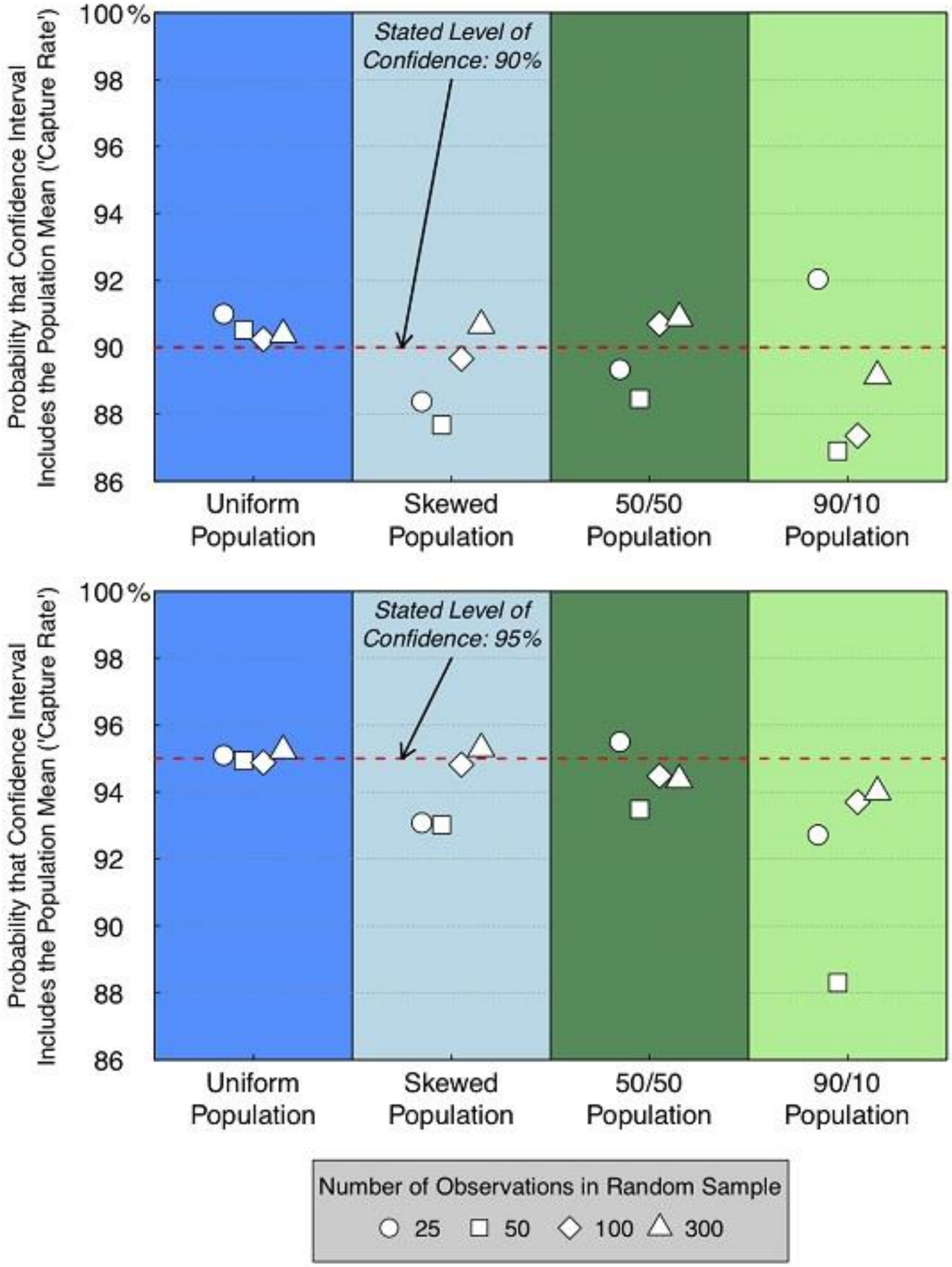
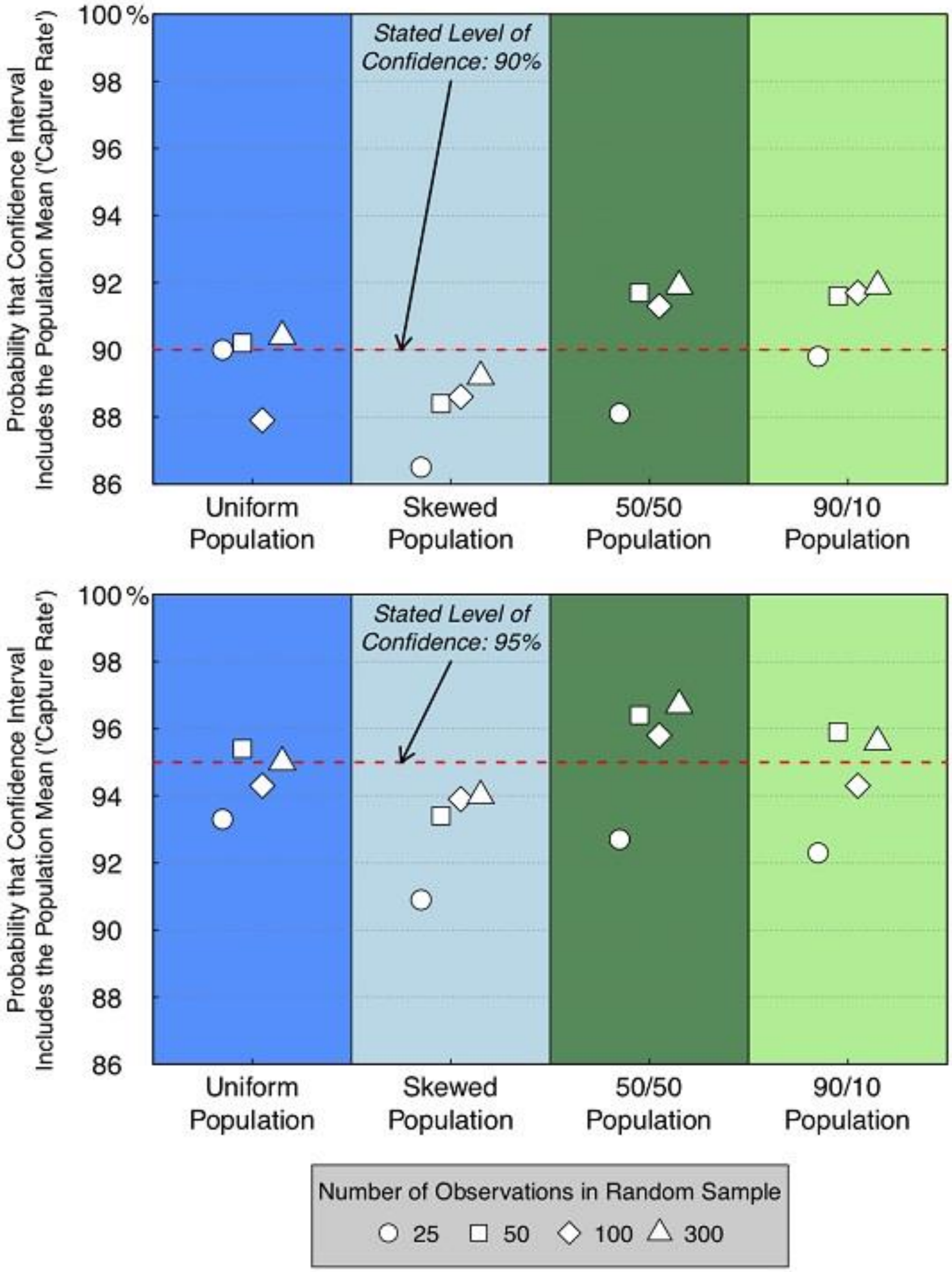


FIGURE 4
Capture Rates When Applying Bootstrapping



Practical Considerations When Deriving the Capture Rate and Interpreting Confidence Intervals

Thus far we have assumed that we know the values for every observation in the population. However, these circumstances are solely for illustrative purposes. In practice, it can be challenging to derive the capture rate because the practitioner only gets to collect/analyze one random sample, not hundreds or thousands. How might the experimental process and results described above be used given that only limited data are available? Below are four recommendations that can aid in the sampling process and/or CI interpretation.

Estimate the Capture Rate by Creating a Proxy for the Population and Proceeding With the Nine Steps Listed Above

A random sample can generate a "stand-in" for the population. Suppose the defined population includes 10,000 observations and a random sample of 100 observations has been analyzed. That sample can be used to simulate a dataset of 10,000 analyzed observations. Each observation can be selected multiple times, once or not at all.[9] Subsequently, the practitioner can calculate the capture rate and compare it to the stated confidence level.

Assess Whether the Likely Value of the Capture Rate Is "High Enough"

By way of example, consider a 95% CI based on a random sample of 25 observations. The results herein suggest that the actual level of confidence may be closer to 90 percent.[10] Depending on the circumstances and/or industry standards, 90 percent may be considered sufficiently high.[11] If so, then this random sample of 25 observations can be used to make inferences about the population mean and its range of likely values.

Recognize the Shape of Data Distribution in the Sample

If the random sample produces a nearly symmetric distribution of data, this supports the notion that the capture rate and the stated confidence level are relatively close to one another. Conversely, an asymmetric data distribution may imply that the actual level of confidence is lower than its target.

Increase the Sample Size so That the Capture Rate Is More Likely to Approach the Stated Confidence Level

The results herein suggest that random samples of several hundred observations likely yield a capture rate close to the stated confidence level. If time and cost constraints allow for such a sample size, this will enable the capture rate to approach the stated confidence level.

Concluding Remarks

This article includes a method that enables statistical practitioners to: (1) account for the fact that each random sample yields a unique estimate and margin of error; (2) estimate the rate at which the CI will capture the population mean; and (3) compare the capture rate to the stated confidence level. Numerous experiments, which run the gambit in terms of population shapes and sample sizes, have been performed. The computational process and results provide powerful information about CIs and their interpretation.

For some practitioners, it may be tempting to dismiss a random sample (e.g., of former/current employees) and corresponding CIs if the data are asymmetric and/or the sample size is small. The results presented above provide compelling evidence that neither of these issues impede the calculation of reliable confidence intervals. Ultimately, it may be informative to derive and report the capture rate in order to gauge the actual level of confidence. Doing so places the practitioner on solid ground when making statements about the defined population.

Brian Kriegler, Ph.D., is a managing director at Econ One Research Inc.

[1] This is the seventh installment in a series of articles on statistical sampling. Previous articles cover the following topics: (i) Validating random sampling as a neutral and unbiased process; (ii) Deriving confidence intervals when working with a small dataset; (iii) Working with data that includes missing and/or incomplete records; (iv) Common misperceptions about random sampling requirements; (v) When to consider using stratified sampling; and (vi) Common misperceptions about stratified sampling.

[2] See Practitioner's Guide to Statistical Sampling: Part 1

[3] See, Practitioner's Guide to Statistical Sampling: Part 2; See also, Moore, David S., McCabe, George P., and Craig, Bruce A. (2009). Introduction to the Practice of Statistics. 6th Ed. New York: W.H. Freeman. pp. 356-369, 420-421.

[4] In total there are 64 experiments, i.e., 2 methods x 2 CIs x 4 populations x 4 sample sizes.

[5] For the four sample sizes, the median differences between the Capture Rate and stated confidence level are as follows: 25 observations - 1.8 percent, 50 observations - 1.6 percent, 100 observations - 0.8 percent, and 300 observations - 0.7 percent.

[6] When the population is skewed, the median difference between the Capture Rate and stated confidence level is 1.5 percent. When the population is unbalanced, this difference is 1.65 percent.

[7] When the population is uniformly distributed, the median difference between the Capture Rate and stated confidence level is 0.3 percent.

[8] The uniform population is in blue, the skewed population is in light blue, the 50/50 population is in medium green, and the 90/10 population is in light green.

[9] Statisticians frequently refer to this type of selection and "sampling with replacement."

[10] Bootstrapping is likely a more appropriate CI method given a sample size of 25 observations. Applying a stated confidence level of 95 percent, the bootstrapped Capture Rates are as follows: Uniform Data - 93.3%, Skewed Data - 90.9%, 50/50 Data - 92.7%, and 90/10 Data - 92.3%.

[11] See, e.g., the CMS Medicare Program Integrity Manual ("[i]n most situations the lower limit of a one-sided 90 percent confidence interval shall be used as the amount of overpayment to be demanded for recovery from the provider or supplier."