Confidence Intervals

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
University of Minnesota

October 17, 2020

Table of Contents

Table of Contents

Definition of a Confidence Interval

Given a confidence level $\alpha \in (0, 1)$, the probabilistic statement

$$P\left(a(\hat{\theta}) < \theta < b(\hat{\theta})\right) = 1 - \alpha$$

defines a $100(1-\alpha)\%$ confidence interval for the unknown parameter θ.

The confidence interval endpoints $a(\cdot)$ and $b(\cdot)$ are functions of the estimate $\hat{\theta}$, which is a random variable.

The CI provides a range of values depending on $\hat{\theta}$ such that the probability of θ being within the interval is $1 - \alpha$.

Example 1: Confidence Interval for Mean

Suppose that $x_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ for $i = 1, \ldots, n$ and we want to form a confidence interval for μ. We will use $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ to estimate μ.

- $\bar{x} \sim N(\mu, \sigma^2/n) \quad \longrightarrow \quad \sqrt{n}(\bar{x} - \mu)/\sigma \sim N(0, 1)$
- $P\left(z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1 - \alpha$

Rearranging the terms inside the probability statement gives:

$$
\begin{aligned}
1 - \alpha &= P\left(z_{\alpha/2}\sigma/\sqrt{n} < \bar{x} - \mu < z_{1-\alpha/2}\sigma/\sqrt{n}\right) \\
&= P\left(z_{\alpha/2}\sigma/\sqrt{n} - \bar{x} < -\mu < z_{1-\alpha/2}\sigma/\sqrt{n} - \bar{x}\right) \\
&= P\left(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} > \mu > \bar{x} - z_{1-\alpha/2}\sigma/\sqrt{n}\right)
\end{aligned}
$$

which implies that $a(\bar{x}) = \bar{x} - z_{1-\alpha/2}\sigma/\sqrt{n}$ and $b(\bar{x}) = \bar{x} - z_{\alpha/2}\sigma/\sqrt{n}$.

Example 1: Confidence Interval for Mean (continued)

Given that $-z_{\alpha/2} = z_{1-\alpha/2}$ we can write the two endpoints of the confidence interval as

$$\bar{x} \pm z_{1-\alpha/2} \mathrm{SE}(\bar{x})$$

where $\mathrm{SE}(\bar{x}) = \sigma/\sqrt{n}$ is the standard error of the sample mean.

In practice, it is typical to form a...

- 90% confidence interval (i.e., $\alpha = 0.1$), where $z_{0.95} \approx 1.65$
- 95% confidence interval (i.e., $\alpha = 0.05$), where $z_{0.975} \approx 1.96$
- 99% confidence interval (i.e., $\alpha = 0.01$), where $z_{0.995} = 2.58$

Example 2: Confidence Interval for Variance

Suppose that $x_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ for $i = 1, \ldots, n$ and we want to form a confidence interval for σ^2. Use $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$ as an estimate.

- $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$ Let $q_{n-1;\alpha}$ denote α quantile of χ^2_{n-1}
- $P\left(q_{n-1;\alpha/2} < (n-1)\frac{s^2}{\sigma^2} < q_{n-1;1-\alpha/2}\right) = 1 - \alpha$

Rearranging the terms inside the above probability statement gives

$$\begin{aligned} 1 - \alpha &= P\left(\frac{q_{n-1;\alpha/2}}{n-1} < \frac{s^2}{\sigma^2} < \frac{q_{n-1;1-\alpha/2}}{n-1}\right) \\ &= P\left(\frac{q_{n-1;\alpha/2}}{s^2(n-1)} < \frac{1}{\sigma^2} < \frac{q_{n-1;1-\alpha/2}}{s^2(n-1)}\right) \\ &= P\left(\frac{s^2(n-1)}{q_{n-1;\alpha/2}} > \sigma^2 > \frac{s^2(n-1)}{q_{n-1;1-\alpha/2}}\right) \end{aligned}$$

which implies that $a(s^2) = \frac{s^2(n-1)}{q_{n-1;1-\alpha/2}}$ and $b(s^2) = \frac{s^2(n-1)}{q_{n-1;\alpha/2}}$

Table of Contents

Misinterpretation of Confidence Intervals

Confidence intervals are often misinterpreted in scientific literature.

The most common misinterpretation is that there is a $100(1-\alpha)\%$ chance that $a(\hat{\theta}) < \theta < b(\hat{\theta})$ for a given estimate $\hat{\theta}$.

This interpretation is incorrect because for any given estimate $\hat{\theta}$ and corresponding confidence interval $[a(\hat{\theta}), b(\hat{\theta})]$, the inequality statement $a(\hat{\theta}) < \theta < b(\hat{\theta})$ is either true or false.

Correct Interpretation of Confidence Intervals

Suppose that we repeat our experiment a large number of independent times, i.e., we collect R independent samples of data each of size n.

- Let $\hat{\theta}_r$ denote the estimate of θ for the r-th sample of data
- Let $[a(\hat{\theta}_r), b(\hat{\theta}_r)]$ denote confidence interval formed from $\hat{\theta}_r$

As the number of replications $R \to \infty$, we have that

$$\frac{1}{R}\sum_{r=1}^{R} I\left(a(\hat{\theta}_r) < \theta < b(\hat{\theta}_r)\right) = 1 - \alpha$$

where $I(\cdot)$ is an indicator function, i.e., $I(\cdot) = 1$ if the inequality statement is true, and $I(\cdot) = 0$ otherwise.

For the r-th replication, the true parameter θ is either in the interval or not, i.e., $I(\cdot)$ is either equal to 1 or 0.

Example 3: Interpreting a Confidence Interval

Assume $x_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ and we want to form a 95% CI for μ (as in Example 1). The below R code uses $R = 10000$ replications with $n = 25$ observations. Note that $\mu = 0$ and $\sigma^2 = 1$ in the below example.

```
> R <- 10000
> n <- 25
> set.seed(1)
> xbar <- replicate(R, mean(rnorm(n)))
> ci.lo <- xbar - qnorm(.975) / sqrt(n)     # 95% CI lower bound
> ci.up <- xbar - qnorm(.025) / sqrt(n)     # 95% CI upper bound
> ci.in <- (ci.lo <= 0) & (0 <= ci.up)
> summary(ci.in)
   Mode   FALSE    TRUE
logical     499    9501
> mean(ci.in)
[1] 0.9501
```

Table of Contents

Confidence Intervals or Bounds?

A confidence interval involves finding both the lower bound $a(\hat{\theta})$ and the upper bound $b(\hat{\theta})$ that provides a range (or interval) of values such that the probability of θ being within the interval is $1 - \alpha$.

However, in some cases, it may be preferable to define a lower or upper bound, i.e., just one of the endpoints, instead of an interval.

Note that lower and upper confidence bounds are appropriate if there is a specific direction of interest that is important for making a decision.

Definition of a Confidence Bound

Given a confidence level $\alpha \in (0, 1)$, the probabilistic statement

$$P\left(a(\hat{\theta}) < \theta\right) = 1 - \alpha$$

defines a $100(1-\alpha)\%$ lower confidence bound for θ.

Given a confidence level $\alpha \in (0, 1)$, the probabilistic statement

$$P\left(\theta < b(\hat{\theta})\right) = 1 - \alpha$$

defines a $100(1-\alpha)\%$ upper confidence bound for θ.

When to Use Confidence Bounds

Confidence bounds are frequently used in studies that are attempting to establish the effectiveness of a treatment, e.g., clinical trials.

- Establishing that a treatment is effective (in a hypothesized direction of interest) only requires using $a(\hat{\theta})$ or $b(\hat{\theta})$ — not both.

Example: Suppose that a drug company has created a new medication that is designed to treat depression.

- X is pre-treatment depression severity and Y is post-treatment
- $Z = Y - X$ denotes the difference and assume $Z \sim N(\theta, \sigma^2)$
- Observing $b(\hat{\theta}) < 0$ would suggest an effective treatment

Table of Contents

Coverage Rates

Given a procedure for forming confidence interval, the coverage rate refers to the proportion of times that the parameter θ is included within the interval using some specified number of replications $R \gg 1$.

- Coverage rate in the example was 0.9501 using $R = 10000$.

The confidence interval procedure is said to be…

- Conservative if coverage rate is greater than $1 - \alpha$
- Liberal is the coverage rate is less than $1 - \alpha$
- Accurate if the coverage rate is approximately equal to $1 - \alpha$

Width and Shape

The width of the confidence interval refers to the distance between the upper and lower bounds, i.e., width $= b(\hat{\theta}) - a(\hat{\theta})$.

- Prefer the narrowest CI with an accurate coverage rate

The shape of the confidence interval refers to the ratio of the distance between the bounds and the estimate, i.e.,
shape $= [b(\hat{\theta}) - \hat{\theta}]/[\hat{\theta} - a(\hat{\theta})]$.

- shape > 1 indicates that the CI is wider on the right side
- shape < 1 indicates that the CI is wider on the left side
- shape $= 1$ indicates that the CI is symmetric around $\hat{\theta}$

Width and shape are only important if the coverage rate is accurate.

Table of Contents

Frameworks for Forming Confidence Intervals

There are three general frameworks that can be used to form confidence intervals:

- *Parametric*: If the distribution of the estimate $\hat{\theta}$ can be exactly derived, then an exact confidence interval can be formed.
- *Asymptotic*: If the distribution of the estimate $\hat{\theta}$ can be asymptotically derived, then an (asymptotically) approximate confidence interval can be formed.
- *Nonparametric*: If the distribution of the estimate $\hat{\theta}$ is unknown, then resampling methods can be used to estimate $F_{\hat{\theta}}$, which can be used to form approximate confidence intervals.

Parametric Confidence Intervals

The first procedure (i.e., parametric) was used in the previous examples. We knew the exact probability distributions of $\bar{x}$ and s^2 under the specified data generation assumptions.

Using these known distributions, we were able to construct and manipulate probability statements that enabled us to derive the lower and upper bounds of the confidence interval.

This procedure always depends on assumptions about the parametric nature of the data—and sometimes it isn't possible to derive the distribution of $\hat{\theta}$ even if the probabilistic nature of X is known.

Asymptotic Confidence Intervals

The second procedure (i.e., asymptotic) can be used to construct confidence intervals for large samples of data when we know the asymptotic (i.e., limiting) distribution of $\hat{\theta}$.

If we are interested in forming a confidence interval for the population mean $\mu = E(X)$, then the central limit theorem (CLT) can be used to form an asymptotic confidence interval.

- $\bar{x} \dot\sim N(\mu, \frac{\sigma^2}{n})$ for large n

Of course, this procedure depends on having a large n, but it is not typically clear how large n needs to be. And for some statistics the limiting distribution of $\hat{\theta}$ will not be known.

Nonparametric Confidence Intervals

The third procedure (i.e., nonparametric) is more computationally intensive than the other two procedures, but is much more general.

The first two procedures (i.e., parametric and asymptotic) can only be used in the small number of circumstances where we know the exact or asymptotic distribution of $\hat{\theta}$, e.g., for the sample mean.

To form confidence intervals for generic parameters, the nonparametric bootstrap can be used (Efron, 1979; Efron and Tibshirani, 1993).

Table of Contents

Basic Idea

Suppose that $x_i \stackrel{\text{iid}}{\sim} F$, and assume that the distribution F depends on some parameter $\theta = t(F)$.

Given $\mathbf{x} = (x_1, \ldots, x_n)^\top$, suppose that we can compute an estimate $\hat{\theta} = s(\mathbf{x})$, and assume that the distribution of $\hat{\theta}$ is unknown.

To approximate the distribution of $\hat{\theta}$, the nonparametric bootstrap uses the empirical cumulative distribution function $\hat{F}_n$ as a surrogate for F.

- As $n \to \infty$, we have that $\hat{F}_n \to F$ and $\hat{F}_{\hat{\theta}} \to F_{\hat{\theta}}$
- For small n, the nonparametric bootstrap may not work well

Bootstrap Distribution

The bootstrap distribution consists of $R \gg 1$ replications of the estimate (or statistic) $\hat{\theta}$. For $r = 1, \ldots R$, the bootstrap:

1. Defines $\mathbf{x}_r = (x_{1r}, \ldots, x_{nr})^\top$ where x_{ir} is sampled *with replacement* from the original sample $\{x_1, \ldots, x_n\}$
2. Calculates $\hat{\theta}_r = s(\mathbf{x}_r)$

Note that $\hat{\theta}_r$ is referred to as the r-th replicate of the statistic, and the collection $\{\hat{\theta}_r\}_{r=1}^R$ is referred to as the bootstrap distribution.

Typically $R \geq 10000$, but it may be necessary to set R even larger when F is relatively skewed or leptokurtic (Hesterberg, 2015).

Bootstrap Standard Error and Bias

The bootstrap distribution $\{\hat{\theta}_r\}_{r=1}^R$ can be used to estimate properties of the estimate $\hat{\theta}$ (Helwig, 2017a).

The bootstrap uses $\widehat{\mathrm{Var}}(\hat{\theta}) = \frac{1}{R-1}\sum_{r=1}^R (\hat{\theta}_r - \bar{\theta})^2$ to estimate $\mathrm{Var}(\hat{\theta})$.

- $\bar{\theta} = \frac{1}{R}\sum_{r=1}^R \hat{\theta}_r$ is the mean of the bootstrap distribution
- $\widehat{\mathrm{SE}}(\hat{\theta}) = \left(\widehat{\mathrm{Var}}(\hat{\theta})\right)^{1/2}$ is the corresponding standard error estimate

The bootstrap uses $\widehat{\mathrm{Bias}}(\hat{\theta}) = \bar{\theta} - \hat{\theta}$ to estimate $\mathrm{Bias}(\hat{\theta})$.

- Reminder: the true bias of $\hat{\theta}$ is defined as $\mathrm{Bias} = E(\theta) - \theta$
- Nonparametric bootstrap is using $\hat{F}_n$ to approximate unknown F

Bootstrap Confidence Intervals

The nonparametric bootstrap can be used to form a variety of different types of confidence intervals (Helwig, 2017b).

Some types are simple to compute but less accurate:

- *Normal approximation*: $\hat{\theta} \pm z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{\theta})$
- *Percentile method*: $[\hat{\theta}_{(R\alpha/2)}, \hat{\theta}_{(R(1-\alpha/2))}]$
- *Basic (reverse percentile)*: $[2\hat{\theta} - \hat{\theta}_{(R\alpha/2)}, 2\hat{\theta} - \hat{\theta}_{(R(1-\alpha/2))}]$

Other types are harder to compute but more accurate:

- *Studentized (t table)*: $[T_{(R\alpha/2)}, T_{(R(1-\alpha/2))}]$ with $T_r = \frac{\hat{\theta}_r - \hat{\theta}}{\widehat{\mathrm{SE}}(\hat{\theta}_r)}$
- *Bias-corrected and accelerated (BCa)*: $[\hat{\theta}_{(R\alpha_1/2)}, \hat{\theta}_{(R(1-\alpha_2/2))}]$

Simulation Study Design

Generate $n \in \{10, 20, 50, 100, 200\}$ observations from a χ^2_1 distribution, and form a confidence interval for the mean μ.

The nonparametric bootstrap was implemented using the `np.boot` function in the **nptest** R package (Helwig, 2020)

Repeat data generating process 10,000 times, and define the coverage rate as the proportion of the times where the given confidence interval method contained the true parameter $\mu = 1$.

Simulation Study Results

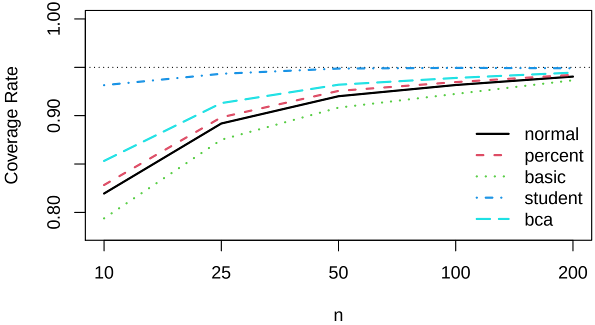

Figure 1: Coverage rates for different bootstrap confidence interval methods.

References

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statistics 7*(1), 1–26.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. Boca Raton: Chapman & Hall.

Helwig, N. E. (2017a). Bootstrap Resampling. `http://users.stat.umn.edu/~helwig/notes/boot-Notes.pdf`.

Helwig, N. E. (2017b). Bootstrap Confidence Intervals. `http://users.stat.umn.edu/~helwig/notes/bootci-Notes.pdf`.

Helwig, N. E. (2020). *nptest: Nonparametric Tests*. R package version 1.0-2.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician 69*(4), 371–386.