Confidence intervals

Rona Axelrod and Danny Kaplan

StatPREP Instructor Guide

Activities

- What is a confidence interval?
- Sampling bias and confidence intervals
- Comparing two groups
- Comparing two confidence intervals
- Two-sample t test

Learning objectives connected to confidence intervals

1. Master the relevant vocabulary ... population, sample, sample statistic, point estimate, margin of error, confidence interval, confidence level.
2. Recognize the estimation process that leads to a confidence interval:

   1. a sample is selected randomly from a population
   2. a statistic is calculated from that sample
   3. the margin of error and, correspondingly, the confidence interval are calculated from that sample.

3. Distinguish between the "summary interval" (which is a summary of the raw values of the variable) and the "confidence interval" (which describes uncertainty in a sample statistic or model parameter).
4. Understand that each time that estimation process is carried out, a new confidence interval will be created.
5. Recognize and appropriately make use of the confidence interval as a presentation of uncertainty due to sampling.
6. Appropriately compare confidence intervals not by whether their endpoints are approximately the same but by whether or not they overlap.

   i. Interpret a lack of overlap as an indication that the populations underlying the two samples are different.

7. Be able to translate the width of a confidence interval from a preliminary study of size n into a reasonable guess for the margin of error for a study of size N.
8. Recognize common sources of sampling bias – e.g. self-selection, pre-screening, survival – and that the possible existence of sampling bias is not incorporated into the confidence interval.

Additional resources

-

Orientation

A fundamental aspect of statistical reasoning is recognizing that estimates computed from a sample are, to some extent, random. That is, another sample will likely produce a different estimate.

Remarkably, we can estimate the amount of uncertainty due to the randomness of sampling from *just a single sample*. We quantify this uncertainty with a *margin of error*. As a rule, the margin of error is inversely proportional to the square root of the sample size: $\sqrt{n}$. That proportionality with n plays out against a base value that depends on the particular statistic being calculated, e.g. the mean, the median, the standard deviation, the 75th percentile, and so on.

There are two conventional notations for the margin of error:

- **In text** the margin of error is presented after the point estimate with a ± serving as punctuation. Example: the mean height of a group of 100 adults would be written as 55.6 ± 0.3 inches, where 55.6 inches is the point estimate and 0.3 inches is the margin of error. The whole assembly – that is, 55.6 ± 0.3 inches – is called a *confidence interval*.
- **In graphics**, the margin of error is the half-length of the I-shaped *error bar* drawn centered on the point estimate. The ends of the error-bar glyph demark the ends of the *confidence interval*.

The idea of a margin of error is closely tied to the use of the normal probability distribution to model the sampling variability of the sample statistic. But many important sample statistics do not have a sampling distribution that is meaningfully approximated as normal. An example is the risk ratio: the ratio of two proportions that is often used in medical or epidemiological communication but which is much more widely applicable. With such statistics, the confidence interval is not centered on the point estimate.

Role in statistical practice

Much of the use of margins of error and confidence intervals in statistical communication is based on a misconception. Consider this definition of confidence interval from the widely used *Elementary Statistics* text by Triola:

> A confidence interval is a range of values used to estimate the true value of a population parameter. – p. 300, 13th edition

This quote correctly captures how confidence intervals are generally used: as a prediction interval for the "true value of a population parameter." That is, many people interpret a confidence interval as marking the probability distribution of the "true value" given the sample. This is wrong, simply because the "true value" isn't random. But the interpretation isn't horrible and is arguably much better than ignoring the uncertainty in the sample statistic.

Perhaps it would be better simply to regard confidence intervals as a marker of an abstract quantity, *precision*. Then we can focus on how that abstract quantity can be used in practice, for instance,

1. Determine when another estimate of the same statistical quantity (using a different sample) is *inconsistent* with the estimate we have made with our sample. Such inconsistency is indicated when the confidence intervals of the two estimates do not overlap. Or, if we are looking for consistency with a point estimate, the inconsistency is marked when our confidence interval doesn't include the point estimate.
2. Anticipate how the precision would change if more data were collected or if an estimate were made using a new sample entirely.

Conceptual pitfalls

Many students (and professionals) will acquire a completely incorrect interpretation of confidence intervals and treat them as if they were summary intervals. For instance, consider an estimate of mean commute time and its confidence interval. Many people will mistakenly think that "most" commute times fall within the confidence interval. In reality the fraction that do can be vanishingly small for large n.

We suspect that one reason for this misconception is that conventional intro stats book do not have a name for the interval that contains "most" of the raw values. This is why we introduce the *summary interval* in these StatPREP 101 lessons.

Similarly, we need to deal with the usual failure of introductory statistics to include reference to Bayesian ideas. Usually this is justified by pointing out the dominance of frequentist inference ideas in published research. That may be so, but there is a strong connection between informal Bayesian ideas and the ways people think about knowledge and uncertainty.

Keep in mind that the way most people (mis-)interpret statistical inferential objects such as confidence intervals and p-values is fundamentally Bayesian. That is, they seek to know the probability of an event, given the data. That event might be "the 'true value' falls within the interval," or "the null hypothesis is false."

It can be tempting to try to correct the common misconceptions about confidence intervals and p-values. For instance, some textbooks include a process view of confidence intervals by imagining a large collection of samples, each of which produces its own confidence interval. Then, when the "true value" of the population parameter is revealed to us, we will find that 95% of those intervals include the "true value." Unfortunately, the "correct" interpretation of frequentist objects of inference does not inform practice any differently than the "incorrect" Bayesian interpretation. Perhaps correctives to such misconceptions should be saved for courses about the development of statistical methods rather than their application.

We often forget that confidence intervals are an expression of *uncertainty* and seek to make them precise beyond any reasonable use. This is perhaps why many educators favor 1.96 over 2 or spend so much time on special techniques for small data (such as t-intervals). The search for extra precision is also what prevents the introduction of simple techniques for comparing estimates – do the confidence intervals overlap? – in favor of more complex techniques such as the t-test.

The confidence interval can be calculated from data and is a reliable measure of the amount of uncertainty created by a random sample of size n. There are other sources of uncertainty – biases – which can be important but which are not captured by a confidence interval. In particular, the canonical example used for teaching confidence intervals, survey or poll results, are in today's world so heavily influenced by self-selection bias to the extent that the confidence interval itself does not provide a valid indication of the reliability of results.

Student prerequisites

1. Familiarity with how the process of random sampling introduces uncertainty: the sample statistic will likely be different from one sampling trial to another.
2. Acquaintance with the idea of an interval representing the *vast majority* (taken usually as 95%) of the possibilities.

Creating an active classroom

Some specific discussion topics/themes for confidence intervals:

- Where is the mean located relative to the lower and upper bounds of the confidence interval?

- How do the following factors affect the distance between the lower and upper bounds of the confidence interval: sample size, spread of the data, and confidence level?
- Do you notice any overlap of the confidence intervals? Generate new samples and notice whether the same behaviors occur. What does the overlap or lack of overlap tell us?

Assessment items

- Given a population parameter and a single confidence interval for a sample of size n . . .

  1. Draw 20 plausible confidence intervals for new samples of size n
  2. Draw 20 plausible confidence intervals for new samples of size 4n

- Given a population parameter and a confidence interval on the mean for size n, estimate and draw a 95% summary interval.
- Given a small set of data, construct a bootstrap trial

Pushing the envelope/advancing the field

Author info