I HIGHLIGHTED SOME KEY POINTS

Confidence interval

From Wikipedia, the free encyclopedia

In statistics, a **confidence interval** (**CI**) is a type of interval estimate (of a population parameter) that is computed from the observed data. The **confidence level** is the frequency (i.e., the proportion) of possible confidence intervals that contain the true value of their corresponding parameter. In other words, if confidence intervals are constructed using a given confidence level in an infinite number of independent experiments, the proportion of those intervals that contain the true value of the parameter will match the confidence level.[1][2][3]

Confidence intervals consist of a range of values (interval) that act as good estimates of the unknown population parameter. However, the interval computed from a particular sample does not necessarily include the true value of the parameter. Since the observed data are random samples from the true population, the confidence interval obtained from the data is also random. If a corresponding hypothesis test is performed, the confidence level is the complement of the level of significance, i.e. a 95% confidence interval reflects a significance level of 0.05.[4] If it is hypothesized that a true parameter value is 0 but the 95% confidence interval does not contain 0, then the estimate is significantly different from zero at the 5% significance level.

The desired level of confidence is set by the researcher (not determined by data). Most commonly, the 95% confidence level is used.[5] However, other confidence levels can be used, for example, 90% and 99%.

Factors affecting the width of the confidence interval include the size of the sample, the confidence level, and the variability in the sample. A larger sample size normally will lead to a better estimate of the population parameter.

Confidence intervals were introduced to statistics by Jerzy Neyman in a paper published in 1937.[3]

## Contents

Conceptual basis

Introduction

Interval estimates can be contrasted with point estimates. A point estimate is a single value given as the estimate of a population parameter that is of interest, for example, the mean of some quantity. An interval estimate specifies instead a range within which the parameter is estimated to lie. Confidence intervals are commonly reported in tables or graphs along with point estimates of the same parameters, to show the reliability of the estimates.

For example, a confidence interval can be used to describe how reliable survey results are. In a poll of election–voting intentions, the result might be that 40% of respondents intend to vote for a certain party. A 99% confidence interval for the proportion in the whole population having the same intention on the survey might be 30% to 50%. From the same data one may calculate a 90% confidence interval, which in this case might be 37% to 43%. A major factor determining the length of a confidence interval is the size of the sample used in the estimation procedure, for example, the number of people taking part in a survey.
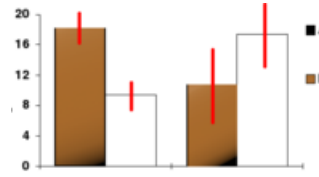
In this bar chart, the top ends of the brown bars indicate observed means and the red line segments ("error bars") represent the confidence intervals around them. Although the error bars are shown as symmetric around the means, that is not always the case. It is also important that in most graphs, the error bars do not represent confidence intervals (e.g., they often represent standard errors or standard deviations)

Meaning and interpretation

Various interpretations of a confidence interval can be given (taking the 90% confidence interval as an example in the following).

- The confidence interval can be expressed in terms of samples (or repeated samples): "*Were this procedure to be repeated on numerous samples, the fraction of calculated confidence intervals (which would differ for each sample) that encompass the true population parameter would tend toward 90%.*"[1]
- The confidence interval can be expressed in terms of a single sample: "*There is a 90% probability that the calculated confidence interval from some future experiment encompasses the true value of the population parameter.*" Note this is a probability statement about the confidence interval, not the population parameter. This considers the probability associated with a confidence interval from a pre-experiment point of view, in the same context in which arguments for the random allocation of treatments to study items are made. Here the experimenter sets out the way in which they intend to calculate a confidence interval and to know, before they do the actual experiment, that the interval they will end up calculating has a particular chance of covering the true but unknown value.[3] This is very similar to the "repeated sample" interpretation above, except that it avoids relying on considering hypothetical repeats of a sampling procedure that may not be repeatable in any meaningful sense. See Neyman construction.
- The explanation of a confidence interval can amount to something like: "*The confidence interval represents values for the population parameter for which the difference between the parameter and the observed estimate is not statistically significant at the 10% level*".[6] In fact, this relates to one particular way in which a confidence interval may be constructed.

In each of the above, the following applies: If the true value of the parameter lies outside the 90% confidence interval, then a sampling event has occurred (namely, obtaining a point estimate of the parameter at least this far from the true parameter value) which had a probability of 10% (or less) of happening by chance.

Misunderstandings

Confidence intervals are frequently misunderstood, and published studies have shown that even professional scientists often misinterpret them.[7][8][9][10]

- A 95% confidence interval does not mean that for a given realized interval there is a 95% probability that the population parameter lies within the interval (i.e., a 95% probability that the interval covers the population parameter).[11] Once an experiment is done and an interval calculated, this interval either covers the parameter value or it does not; it is no longer a matter of probability. The 95% probability

relates to the reliability of the estimation procedure, not to a specific calculated interval.[12] Neyman himself (the original proponent of confidence intervals) made this point in his original paper:[3]

> "It will be noticed that in the above description, the probability statements refer to the problems of estimation with which the statistician will be concerned in the future. In fact, I have repeatedly stated that the frequency of correct results will tend to α. Consider now the case when a sample is already drawn, and the calculations have given [particular limits]. Can we say that in this particular case the probability of the true value [falling between these limits] is equal to α? The answer is obviously in the negative. The parameter is an unknown constant, and no probability statement concerning its value may be made..."

Deborah Mayo expands on this further as follows:[13]

> "It must be stressed, however, that having seen the value [of the data], Neyman-Pearson theory never permits one to conclude that the specific confidence interval formed covers the true value of 0 with either $(1 - \alpha)100\%$ probability or $(1 - \alpha)100\%$ degree of confidence. Seidenfeld's remark seems rooted in a (not uncommon) desire for Neyman-Pearson confidence intervals to provide something which they cannot legitimately provide; namely, a measure of the degree of probability, belief, or support that an unknown parameter value lies in a specific interval. Following Savage (1962), the probability that a parameter lies in a specific interval may be referred to as a measure of final precision. While a measure of final precision may seem desirable, and while confidence levels are often (wrongly) interpreted as providing such a measure, no such interpretation is warranted. Admittedly, such a misinterpretation is encouraged by the word 'confidence'."

- A 95% confidence interval does not mean that 95% of the sample data lie within the interval.
- A confidence interval is not a definitive range of plausible values for the sample parameter, though it may be understood as an estimate of plausible values for the population parameter.
- A particular confidence interval of 95% calculated from an experiment does not mean that there is a 95% probability of a sample parameter from a repeat of the experiment falling within this interval.

Philosophical issues

The principle behind confidence intervals was formulated to provide an answer to the question raised in statistical inference of how to deal with the uncertainty inherent in results derived from data that are themselves only a randomly selected subset of a population. There are other answers, notably that provided by Bayesian inference in the form of credible intervals. Confidence intervals correspond to a chosen rule for determining the confidence bounds, where this rule is essentially determined before any data are obtained, or before an experiment is done. The rule is defined such that over all possible datasets that might be obtained, there is a high probability ("high" is specifically quantified) that the interval determined by the rule will include the true value of the quantity under consideration. The Bayesian approach appears to offer intervals that can, subject to acceptance of an interpretation of "probability" as Bayesian probability, be interpreted as meaning that the specific interval calculated from a given dataset has a particular probability of including the true value, conditional on the data and other information available. The confidence interval approach does not allow this since in this formulation and at this same stage, both the bounds of the interval and the true values are fixed values, and there is no randomness involved. On the other hand, the Bayesian approach is only as valid as the prior probability used in the computation, whereas the confidence interval does not depend on assumptions about the prior probability.

The questions concerning how an interval expressing uncertainty in an estimate might be formulated, and of how such intervals might be interpreted, are not strictly mathematical problems and are philosophically problematic.[14] Mathematics can take over once the basic principles of an approach to 'inference' have been established, but it has only a limited role in saying why one approach should be preferred to another: For example, a confidence level of 95% is often used in the biological sciences, but this is a matter of convention or arbitration. In the physical sciences, a much higher level may be used.[15]

Relationship with other statistical topics

Statistical hypothesis testing

Confidence intervals are closely related to statistical significance testing. For example, if for some estimated parameter θ one wants to test the null hypothesis that $\theta = 0$ against the alternative that $\theta \neq 0$, then this test can be performed by determining whether the confidence interval for θ contains 0.

More generally, given the availability of a hypothesis testing procedure that can test the null hypothesis $\theta = \theta_0$ against the alternative that $\theta \neq \theta_0$ for any value of θ_0, then a confidence interval with confidence level $\gamma = 1 - \alpha$ can be defined as containing any number θ_0 for which the corresponding null hypothesis is not rejected at significance level α.[16]

If the estimates of two parameters (for example, the mean values of a variable in two independent groups) have confidence intervals that do not overlap, then the difference between the two values is more significant than indicated by the individual values of α.[17] So, this "test" is too conservative and can lead to a result that is more significant than the individual values of α would indicate. If two confidence intervals overlap, the two means still may be significantly different.[18][19][20] Accordingly, and consistent with the Mantel-Haenszel Chi-squared test, is a proposed fix whereby one reduces the error bounds for the two means by multiplying them by the square root of ½ (0.707107) before making the comparison.[21]

While the formulations of the notions of confidence intervals and of statistical hypothesis testing are distinct, they are in some senses related and to some extent complementary. While not all confidence intervals are constructed in this way, one general purpose approach to constructing confidence intervals is to define a $100(1 - \alpha)\%$ confidence interval to consist of all those values θ_0 for which a test of the hypothesis $\theta = \theta_0$ is not rejected at a significance level of $100\alpha\%$. Such an approach may not always be available since it presupposes the practical availability of an appropriate significance test. Naturally, any assumptions required for the significance test would carry over to the confidence intervals.

It may be convenient to make the general correspondence that parameter values within a confidence interval are equivalent to those values that would not be rejected by a hypothesis test, but this would be dangerous. In many instances the confidence intervals that are quoted are only approximately valid, perhaps derived from "plus or minus twice the standard error," and the implications of this for the supposedly corresponding hypothesis tests are usually unknown.

It is worth noting that the confidence interval for a parameter is not the same as the acceptance region of a test for this parameter, as is sometimes thought. The confidence interval is part of the parameter space, whereas the acceptance region is part of the sample space. For the same reason, the confidence level is not the same as the complementary probability of the level of significance.

Confidence region

Confidence regions generalize the confidence interval concept to deal with multiple quantities. Such regions can indicate not only the extent of likely sampling errors but can also reveal whether (for example) it is the case that if the estimate for one quantity is unreliable, then the other is also likely to be unreliable.

Confidence band

A confidence band is used in statistical analysis to represent the uncertainty in an estimate of a curve or function based on limited or noisy data. Similarly, a **prediction band** is used to represent the uncertainty about the value of a new data point on the curve, but subject to noise. Confidence and prediction bands are often used as part of the graphical presentation of results of a regression analysis.

Confidence bands are closely related to confidence intervals, which represent the uncertainty in an estimate of a single numerical value. "As confidence intervals, by construction, only refer to a single point, they are narrower (at this point) than a confidence band which is supposed to hold simultaneously at many points."[22]

Basic steps

The basic breakdown of how to calculate a confidence interval for a population mean is as follows:

1. Identify the sample mean, $\bar{x}$.

2. Identify whether the standard deviation is known, σ, or unknown, s.

- If standard deviation is known then z* is used as the critical value. This value is only dependent on the confidence level for the test. Typical two sided confidence levels are:[23]

C	z*
99%	2.576
98%	2.326
95%	1.96
90%	1.645

- If the standard deviation is unknown then Student's t distribution is used as the critical value. This value is dependent on the confidence level (C) for the test and degrees of freedom. The degrees of freedom is found by subtracting one from the number of observations, $n - 1$. The critical value is found from the t-distribution table. In this table the critical value is written as $t_\alpha(r)$, where r is the degrees of freedom and $\alpha = \dfrac{1 - C}{2}$.

3. Plug the found values into the appropriate equations:

- For a known standard deviation: $\left(\bar{x} - z^* \dfrac{\sigma}{\sqrt{n}}, \bar{x} + z^* \dfrac{\sigma}{\sqrt{n}} \right)$

- For an unknown standard deviation: $\left(\bar{x} - t^* \dfrac{s}{\sqrt{n}}, \bar{x} + t^* \dfrac{s}{\sqrt{n}} \right)$

4. The final step is to interpret the answer. Since the found answer is an interval with an upper and lower bound it is appropriate to state that based on the given data we are __ % (dependent on the confidence level) confident that the true mean of the population is between __ (lower bound) and __ (upper bound).[24]

Statistical theory

Definition

Let X be a random sample from a probability distribution with statistical parameters θ, which is a quantity to be estimated, and ϕ, representing quantities that are not of immediate interest. A *confidence interval* for the parameter θ, with confidence level or confidence coefficient γ, is an interval with random endpoints $(u(X), v(X))$, determined by the pair of random variables $u(X)$ and $v(X)$, with the property:

$$\Pr_{\theta,\varphi}(u(X) < \theta < v(X)) = \gamma \text{ for all } (\theta, \varphi).$$

The quantities ϕ in which there is no immediate interest are called nuisance parameters, as statistical theory still needs to find some way to deal with them. The number γ, with typical values close to but not greater than 1, is sometimes given in the form $1 - \alpha$ (or as a percentage $100\% \cdot (1 - \alpha)$), where α is a small non-negative number, close to 0.

Here $\Pr_{\theta,\phi}$ indicates the probability distribution of X characterised by (θ, ϕ). An important part of this specification is that the random interval $(u(X), v(X))$ covers the unknown value θ with a high probability no matter what the true value of θ actually is.

Note that here $\Pr_{\theta,\phi}$ need not refer to an explicitly given parameterized family of distributions, although it often does. Just as the random variable X notionally corresponds to other possible realizations of x from the same population or from the same version of reality, the parameters (θ, ϕ) indicate that we need to consider other versions of reality in which the distribution of X might have different characteristics.

In a specific situation, when x is the outcome of the sample X, the interval $(u(x), v(x))$ is also referred to as a confidence interval for θ. Note that it is no longer possible to say that the (observed) interval $(u(x), v(x))$ has probability γ to contain the parameter θ. This observed interval is just one realization of all possible intervals

for which the probability statement holds.

Approximate confidence intervals

In many applications, confidence intervals that have exactly the required confidence level are hard to construct. But practically useful intervals can still be found: the rule for constructing the interval may be accepted as providing a confidence interval at level γ if

$$\Pr_{\theta,\varphi}(u(X) < \theta < v(X)) \approx \gamma \text{ for all } (\theta, \varphi)$$

to an acceptable level of approximation. Alternatively, some authors[25] simply require that

$$\Pr_{\theta,\varphi}(u(X) < \theta < v(X)) \geq \gamma \text{ for all } (\theta, \varphi)$$

which is useful if the probabilities are only partially identified, or imprecise.

Desirable properties

When applying standard statistical procedures, there will often be standard ways of constructing confidence intervals. These will have been devised so as to meet certain desirable properties, which will hold given that the assumptions on which the procedure rely are true. These desirable properties may be described as: validity, optimality, and invariance. Of these "validity" is most important, followed closely by "optimality". "Invariance" may be considered as a property of the method of derivation of a confidence interval rather than of the rule for constructing the interval. In non-standard applications, the same desirable properties would be sought.

- *Validity*. This means that the nominal coverage probability (confidence level) of the confidence interval should hold, either exactly or to a good approximation.
- *Optimality*. This means that the rule for constructing the confidence interval should make as much use of the information in the data-set as possible. Recall that one could throw away half of a dataset and still be able to derive a valid confidence interval. One way of assessing optimality is by the length of the interval so that a rule for constructing a confidence interval is judged better than another if it leads to intervals whose lengths are typically shorter.
- *Invariance*. In many applications, the quantity being estimated might not be tightly defined as such. For example, a survey might result in an estimate of the median income in a population, but it might equally be considered as providing an estimate of the logarithm of the median income, given that this is a common scale for presenting graphical results. It would be desirable that the method used for constructing a confidence interval for the median income would give equivalent results when applied to constructing a confidence interval for the logarithm of the median income: specifically the values at the ends of the latter interval would be the logarithms of the values at the ends of former interval.

Methods of derivation

For non-standard applications, there are several routes that might be taken to derive a rule for the construction of confidence intervals. Established rules for standard procedures might be justified or explained via several of these routes. Typically a rule for constructing confidence intervals is closely tied to a particular way of finding a point estimate of the quantity being considered.

Descriptive statistics
This is closely related to the method of moments for estimation. A simple example arises where the quantity to be estimated is the mean, in which case a natural estimate is the sample mean. The usual arguments indicate that the sample variance can be used to estimate the variance of the sample mean. A naive confidence interval for the true mean can be constructed centered on the sample mean with a width which is a multiple of the square root of the sample variance.

Likelihood theory
Where estimates are constructed using the maximum likelihood principle, the theory for this provides two ways of constructing confidence intervals or confidence regions for the estimates. One way is by using Wilks's theorem to find all the possible values of θ that fulfill the following restriction:[26]

$$\ln(L(\theta)) \geq \ln(L(\hat{\theta})) - \frac{1}{2}\chi^2_{1,1-\alpha}$$

Estimating equations
The estimation approach here can be considered as both a generalization of the method of moments and a

generalization of the maximum likelihood approach. There are corresponding generalizations of the results of maximum likelihood theory that allow confidence intervals to be constructed based on estimates derived from estimating equations.

Via significance testing

If significance tests are available for general values of a parameter, then confidence intervals/regions can be constructed by including in the $100p\%$ confidence region all those points for which the significance test of the null hypothesis that the true value is the given value is not rejected at a significance level of $(1 - p)$.[16]

Bootstrapping

In situations where the distributional assumptions for that above methods are uncertain or violated, resampling methods allow construction of confidence intervals or prediction intervals. The observed data distribution and the internal correlations are used as the surrogate for the correlations in the wider population.

Examples

Practical example

A machine fills cups with a liquid, and is supposed to be adjusted so that the content of the cups is 250 g of liquid. As the machine cannot fill every cup with exactly 250.0 g, the content added to individual cups shows some variation, and is considered a random variable X. This variation is assumed to be normally distributed around the desired average of 250 g, with a standard deviation, σ, of 2.5 g. To determine if the machine is adequately calibrated, a sample of $n = 25$ cups of liquid is chosen at random and the cups are weighed. The resulting measured masses of liquid are $X_1, ..., X_{25}$, a random sample from X.
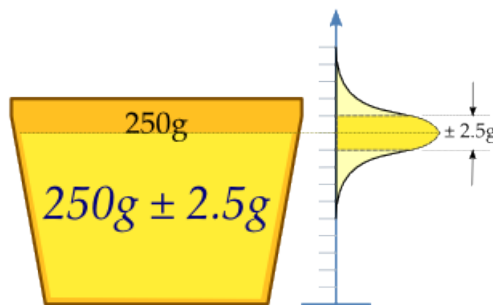
To get an impression of the expectation μ, it is sufficient to give an estimate. The appropriate estimator is the sample mean:

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

The sample shows actual weights $x_1, ..., x_{25}$, with mean:

$$\bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i = 250.2 \text{ grams.}$$

If we take another sample of 25 cups, we could easily expect to find mean values like 250.4 or 251.1 grams. A sample mean value of 280 grams however would be extremely rare if the mean content of the cups is in fact close to 250 grams. There is a whole interval around the observed value 250.2 grams of the sample mean within which, if the whole population mean actually takes a value in this range, the observed data would not be considered particularly unusual. Such an interval is called a confidence interval for the parameter μ. How do we calculate such an interval? The endpoints of the interval have to be calculated from the sample, so they are statistics, functions of the sample $X_1, ..., X_{25}$ and hence random variables themselves.

In our case we may determine the endpoints by considering that the sample mean $\bar{X}$ from a normally distributed sample is also normally distributed, with the same expectation μ, but with a standard error of:

$$\frac{\sigma}{\sqrt{n}} = \frac{2.5 \text{g}}{\sqrt{25}} = 0.5 \text{ grams}$$

By standardizing, we get a random variable:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{0.5}$$

dependent on the parameter μ to be estimated, but with a standard normal distribution independent of the parameter μ. Hence it is possible to find numbers $-z$ and z, independent of μ, between which Z lies with probability $1 - \alpha$, a measure of how confident we want to be.

We take $1 - \alpha = 0.95$, for example. So we have:

$$P(-z \leq Z \leq z) = 1 - \alpha = 0.95.$$

The number z follows from the cumulative distribution function, in this case the cumulative normal distribution function:

$$\Phi(z) = P(Z \leq z) = 1 - \frac{\alpha}{2} = 0.975,$$

$$z = \Phi^{-1}(\Phi(z)) = \Phi^{-1}(0.975) = 1.96,$$

and we get:

$$0.95 = 1 - \alpha = P(-z \leq Z \leq z) = P\left(-1.96 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 1.96\right)$$

$$= P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right).$$

In other words, the lower endpoint of the 95% confidence interval is:

$$\text{Lower endpoint} = \bar{X} - 1.96\frac{\sigma}{\sqrt{n}},$$

and the upper endpoint of the 95% confidence interval is:

$$\text{Upper endpoint} = \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}.$$

With the values in this example, the confidence interval is:

$$0.95 = \Pr(\bar{X} - 1.96 \times 0.5 \leq \mu \leq \bar{X} + 1.96 \times 0.5)$$

$$= \Pr(\bar{X} - 0.98 \leq \mu \leq \bar{X} + 0.98).$$

As the standard deviation of the population σ is known in this case, the distribution of the sample mean $\bar{X}$ is a normal distribution with μ the only unknown parameter. In the theoretical example below, the parameter σ is also unknown, which calls for using the Student's t-distribution.

Interpretation

This might be interpreted as: with probability 0.95 we will find a confidence interval in which the value of parameter μ will be between the stochastic endpoints

$$\bar{X} - 0.98$$

and

$$\bar{X} + 0.98.$$

This does not mean there is 0.95 probability that the value of parameter μ is in the interval obtained by using the currently computed value of the sample mean,

$$(\bar{x} - 0.98, \ \bar{x} + 0.98).$$

Instead, every time the measurements are repeated, there will be another value for the mean $\bar{X}$ of the sample. In 95% of the cases μ will be between the endpoints calculated from this mean, but in 5% of the cases it will not

be. The actual confidence interval is calculated by entering the measured masses in the formula. Our 0.95 confidence interval becomes:

$$(\bar{x} - 0.98; \bar{x} + 0.98) = (250.2 - 0.98; 250.2 + 0.98) = (249.22; 251.18).$$

In other words, the 95% confidence interval is between the lower endpoint 249.22 g and the upper endpoint 251.18 g.
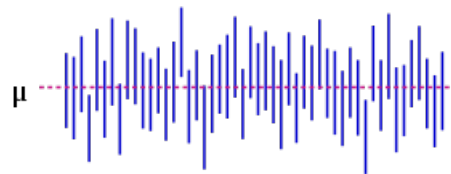
As the desired value 250 of μ is within the resulted confidence interval, there is no reason to believe the machine is wrongly calibrated.

The calculated interval has fixed endpoints, where μ might be in between (or not). Thus this event has probability either 0 or 1. One **cannot** say: "with probability $(1 - \alpha)$ the parameter μ lies in the confidence interval." One only knows that by repetition in $100(1 - \alpha)$ % of the cases, μ will be in the calculated interval. In

$100\alpha\%$ of the cases however it does not. And unfortunately one does not know in which of the cases this happens. That is (instead of using the term "probability") why one can say: "with **confidence level** $100(1 - \alpha)$ %, μ lies in the confidence interval."

The maximum error is calculated to be 0.98 since it is the difference between the value that we are confident of with upper or lower endpoint.

The figure on the right shows 50 realizations of a confidence interval for a given population mean μ. If we randomly choose one realization, the probability is 95% we end up having chosen an interval that contains the

The blue vertical line segments represent 50 realizations of a confidence interval for the population mean μ, represented as a red horizontal dashed line; note that some confidence intervals do not contain the population mean, as expected.

parameter; however, we may be unlucky and have picked the wrong one. We will never know; we are stuck with our interval.

Theoretical example

Suppose $\{X_1, ..., X_n\}$ is an independent sample from a normally distributed population with unknown (parameters) mean μ and variance σ^2. Let

$$\bar{X} = (X_1 + \cdots + X_n)/n,$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Where $\bar{X}$ is the sample mean, and S^2 is the sample variance. Then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has a Student's t-distribution with $n - 1$ degrees of freedom.[27] Note that the distribution of T does not depend on the values of the unobservable parameters μ and σ^2; i.e., it is a pivotal quantity. Suppose we wanted to calculate a 95% confidence interval for μ. Then, denoting c as the 97.5th percentile of this distribution,

$$\Pr(-c \le T \le c) = 0.95$$

("97.5th" and "0.95" are correct in the preceding expressions. There is a 2.5% chance that T will be less than $-c$ and a 2.5% chance that it will be larger than $+c$. Thus, the probability that T will be between $-c$ and $+c$ is 95%.)

Consequently,

$$\Pr\left(\bar{X} - \frac{cS}{\sqrt{n}} \le \mu \le \bar{X} + \frac{cS}{\sqrt{n}}\right) = 0.95$$

and we have a theoretical (stochastic) 95% confidence interval for μ.

After observing the sample we find values $\bar{x}$ for $\bar{X}$ and s for S, from which we compute the confidence interval

$$\left[\bar{x} - \frac{cs}{\sqrt{n}}, \bar{x} + \frac{cs}{\sqrt{n}}\right],$$

an interval with fixed numbers as endpoints, of which we can no longer say there is a certain probability it contains the parameter μ; either μ is in this interval or isn't.

Alternatives and critiques

Confidence intervals are one method of interval estimation, and the most widely used in frequentist statistics. An analogous concept in Bayesian statistics is credible intervals, while an alternative frequentist method is that of prediction intervals which, rather than estimating *parameters,* estimate the outcome of *future* samples. For other approaches to expressing uncertainty using intervals, see interval estimation.

Comparison to prediction intervals

A prediction interval for a random variable is defined similarly to a confidence interval for a statistical parameter. Consider an additional random variable Y which may or may not be statistically dependent on the random sample X. Then $(u(X), v(X))$ provides a prediction interval for the as-yet-to-be observed value y of Y if

$$\Pr_{\theta,\varphi}(u(X) < Y < v(X)) = \gamma \text{ for all } (\theta, \varphi).$$

Here $\Pr_{\theta,\phi}$ indicates the joint probability distribution of the random variables (X, Y), where this distribution depends on the statistical parameters (θ, ϕ).

Comparison to tolerance intervals

Comparison to Bayesian interval estimates

A Bayesian interval estimate is called a credible interval. Using much of the same notation as above, the definition of a credible interval for the unknown true value of θ is, for a given γ,[28]

$$\Pr(u(x) < \Theta < v(x) \mid X = x) = \gamma.$$

Here Θ is used to emphasize that the unknown value of θ is being treated as a random variable. The definitions of the two types of intervals may be compared as follows.

- The definition of a confidence interval involves probabilities calculated from the distribution of X for a given (θ, ϕ) (or conditional on these values) and the condition needs to hold for all values of (θ, ϕ).
- The definition of a credible interval involves probabilities calculated from the distribution of Θ conditional on the observed values of $X = x$ and marginalised (or averaged) over the values of Φ, where this last quantity is the random variable corresponding to the uncertainty about the nuisance parameters in ϕ.

Note that the treatment of the nuisance parameters above is often omitted from discussions comparing confidence and credible intervals but it is markedly different between the two cases.

In some simple standard cases, the intervals produced as confidence and credible intervals from the same data set can be identical. They are very different if informative prior information is included in the Bayesian analysis, and may be very different for some parts of the space of possible data even if the Bayesian prior is relatively uninformative.

There is disagreement about which of these methods produces the most useful results: the mathematics of the computations are rarely in question–confidence intervals being based on sampling distributions, credible

Confidence intervals for proportions and related quantities

An approximate confidence interval for a population mean can be constructed for random variables that are not normally distributed in the population, relying on the central limit theorem, if the sample sizes and counts are big enough. The formulae are identical to the case above (where the sample mean is actually normally distributed about the population mean). The approximation will be quite good with only a few dozen observations in the sample if the probability distribution of the random variable is not too different from the normal distribution (e.g. its cumulative distribution function does not have any discontinuities and its skewness is moderate).

One type of sample mean is the mean of an indicator variable, which takes on the value 1 for true and the value 0 for false. The mean of such a variable is equal to the proportion that has the variable equal to one (both in the population and in any sample). This is a useful property of indicator variables, especially for hypothesis testing. To apply the central limit theorem, one must use a large enough sample. A rough rule of thumb is that one should see at least 5 cases in which the indicator is 1 and at least 5 in which it is 0. Confidence intervals constructed using the above formulae may include negative numbers or numbers greater than 1, but proportions obviously cannot be negative or exceed 1. Additionally, sample proportions can only take on a finite number of values, so the central limit theorem and the normal distribution are not the best tools for building a confidence interval. See "Binomial proportion confidence interval" for better methods which are specific to this case.

Counter-examples

Since confidence interval theory was proposed, a number of counter-examples to the theory been developed to show how the interpretation of confidence intervals can be problematic, at least if one interprets them naïvely.

Confidence procedure for uniform location

Welch [29] presented example which clearly shows the difference between the theory of confidence intervals and other theories of interval estimation (including Fisher's fiducial intervals and objective Bayesian intervals). Robinson [30] called this example "[p]ossibly the best known counterexample for Neyman's version of confidence interval theory." To Welch, it showed the superiority of confidence interval theory; to critics of the theory, it shows a deficiency. Here we present a simplified version.

Suppose that X_1, X_2 are independent observations from a Uniform$(\theta - 1/2, \theta + 1/2)$ distribution. Then the optimal 50% confidence procedure[31] is

$$\bar{X} \pm \begin{cases} \dfrac{|X_1 - X_2|}{2} & \text{if } |X_1 - X_2| < 1/2 \\[2ex] \dfrac{1 - |X_1 - X_2|}{2} & \text{if } |X_1 - X_2| \geq 1/2. \end{cases}$$

A fiducial or objective Bayesian argument can be used to derive the interval estimate

$$\bar{X} \pm \frac{1 - |X_1 - X_2|}{4},$$

which is also a 50% confidence procedure. Welch showed that the first confidence procedure dominates the second, according to desiderata from confidence interval theory; for every $\theta_1 \neq \theta$, the probability that the first procedure contains θ_1 is *less than or equal to* the probability that the second procedure contains θ_1. The average width of the intervals from the first procedure is less than that of the second. Hence, the first procedure is preferred under classical confidence interval theory.

However, when $|X_1 - X_2| \geq 1/2$, intervals from the first procedure are *guaranteed* to contain the true value θ: Therefore, the nominal 50% confidence coefficient is unrelated to the uncertainty we should have that a specific interval contains the true value. The second procedure does not have this property.

Moreover, when the first procedure generates a very short interval, this indicates that X_1, X_2 are very close

together and hence only offer the information in a single data point. Yet the first interval will exclude almost all reasonable values of the parameter due to its short width. The second procedure does not have this property.

The two counter-intuitive properties of the first procedure — 100% coverage when X_1, X_2 are far apart and almost 0% coverage when X_1, X_2 are close together — balance out to yield 50% coverage on average. However, despite the first procedure being optimal, its intervals offer neither an assessment of the precision of the estimate nor an assessment of the uncertainty one should have that the interval contains the true value.

This counter-example is used to argue against naïve interpretations of confidence intervals. If a confidence procedure is asserted to have properties beyond that of the nominal coverage (such as relation to precision, or a relationship with Bayesian inference), those properties must be proved; they do not follow from the fact that a procedure is a confidence procedure.

Confidence procedure for ω^2

Steiger[32] suggested a number of confidence procedures for common effect size measures in ANOVA. Morey et al.[11] point out that several of these confidence procedures, including the one for ω^2, have the property that as the F statistic is becomes increasingly small — indicating misfit with all possible values of ω^2 — the confidence interval shrinks and can even contain only the single value $\omega^2=0$; that is, the CI is infinitesimally narrow (this occurs when $p \geq 1 - \alpha/2$ for a $100(1 - \alpha)\%$ CI).

This behavior is consistent with the relationship between the confidence procedure and significance testing: as F becomes so small that the group means are much closer together than we would expect by chance, a significance test might indicate rejection for most or all values of ω^2. Hence the interval will be very narrow or even empty (or, by a convention suggested by Steiger, containing only 0). However, this does *not* indicate that the estimate of ω^2 is very precise. In a sense, it indicates the opposite: that the trustworthiness of the results themselves may be in doubt. This is contrary to the common interpretation of confidence intervals that they reveal the precision of the estimate.

See also

- Cumulative distribution function-based nonparametric confidence interval
- CLs upper limits (particle physics)
- Confidence distribution
- Credence (statistics)
- Error bar
- Estimation statistics
- p-value
- Robust confidence intervals
- Confidence region

Confidence interval for specific distributions

- Confidence interval for binomial distribution
- Confidence interval for exponent of the power law distribution
- Confidence interval for mean of the exponential distribution
- Confidence interval for mean of the Poisson distribution
- Confidence intervals for mean and variance of the normal distribution

References

1. Cox D.R., Hinkley D.V. (1974) *Theoretical Statistics*, Chapman & Hall, p49, p209
2. Kendall, M.G. and Stuart, D.G. (1973) *The Advanced Theory of Statistics*. Vol 2: Inference and Relationship, Griffin, London. Section 20.4
3. Neyman, J. (1937). "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability" (http://www.jstor.org/stable/91337). *Philosophical Transactions of the Royal Society A*. **236**: 333–380. doi:10.1098/rsta.1937.0005 (https://doi.org/10.1098%2Frsta.1937.0005).
4. Field, Andy (2013). *Discovering statistics using SPSS*. London: SAGE.
5. Zar, J.H. (1984) *Biostatistical Analysis*. Prentice-Hall International, New Jersey, pp 43–45.
6. Cox D.R., Hinkley D.V. (1974) Theoretical Statistics, Chapman & Hall, p214, 225, 233

6. Cox D.R., Hinkley D.V. (1974) Theoretical Statistics, Chapman & Hall, p214, 225, 233

7. [1] (http://iase-web.org/documents/papers/icots8/ICOTS8_C104_KALINOWSKI.pdf)

8. [2] (http://irt.com.ne.kr/data/researchers%20misunderstand%20ci%20and%20error%20bars.pdf)

9. Hoekstra, R., R. D. Morey, J. N. Rouder, and E-J. Wagenmakers, 2014. Robust misinterpretation of confidence intervals. Psychonomic Bulletin Review, in press. [3] (http://www.ejwagenmakers.com/inpress/HoekstraEtAlPBR.pdf)

10. Scientists' grasp of confidence intervals doesn't inspire confidence (https://www.sciencenews.org/blog/context/scientists%E2%80%99-grasp-confidence-intervals-doesn%E2%80%99t-inspire-confidence), Science News, July 3, 2014

11. Morey, R. D.; Hoekstra, R.; Rouder, J. N.; Lee, M. D.; Wagenmakers, E.-J. (2016). "The Fallacy of Placing Confidence in Confidence Intervals" (http://doi.org/10.3758/s13423-015-0947-8). *Psychonomic Bulletin & Review*. **23** (1): 103–123. doi:10.3758/s13423-015-0947-8 (https://doi.org/10.3758%2Fs13423-015-0947-8).

12. "1.3.5.2. Confidence Limits for the Mean" (http://www.itl.nist.gov/div898/handbook/eda/section3/eda352.htm). *nist.gov*.

13. Mayo, D. G. (1981) "In defence of the Neyman-Pearson theory of confidence intervals" (http://www.phil.vt.edu/dmayo/PhilStatistics/In%20Defense%20of%20the%20Neyman-Pearson%20Theory%20of%20Confidence%20Intervals.pdf), *Philosophy of Science*, 48 (2), 269–280. JSTOR 187185 (https://www.jstor.org/stable/187185)

14. T. Seidenfeld, *Philosophical Problems of Statistical Inference: Learning from R.A. Fisher*, Springer-Verlag, 1979

15. "Statistical significance defined using the five sigma standard" (http://www.graphpad.com/www/data-analysis-resource-center/blog/statistical-significance-defined-using-the-five-sigma-standard/).

16. Cox D.R., Hinkley D.V. (1974) Theoretical Statistics, Chapman & Hall, Section 7.2(iii)

17. Pav Kalinowski, "Understanding Confidence Intervals (CIs) and Effect Size Estimation (http://www.psychologicalscience.org/index.php/publications/observer/2010/april-10/understanding-confidence-intervals-cis-and-effect-size-estimation.html)", *Observer* Vol.23, No.4 April 2010.

18. Andrea Knezevic, "Overlapping Confidence Intervals and Statistical Significance (https://www.cscu.cornell.edu/news/statnews/stnews73.pdf)", *StatNews # 73:* Cornell Statistical Consulting Unit, October 2008.

19. Goldstein, H.; Healey, M.J.R. (1995). "The graphical presentation of a collection of means". *Journal of the Royal Statistical Society*. **158**: 175–77. JSTOR view/2983411 (https://www.jstor.org/stable/view/2983411). doi:10.2307/2983411 (https://doi.org/10.2307%2F2983411).

20. Wolfe R, Hanley J (Jan 2002). "If we're so different, why do we keep overlapping? When 1 plus 1 doesn't make 2" (http://www.cmaj.ca/cgi/pmidlookup?view=long&pmid=11800251). *CMAJ*. **166** (1): 65–6. PMC 99228 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC99228). PMID 11800251 (https://www.ncbi.nlm.nih.gov/pubmed/11800251).

21. Daniel Smith, "Overlapping confidence intervals are not a statistical test (http://www.amstat.org/chapters/sacramento/Smithpresentation.pdf)", California Dept of Health Services, 26th Annual Institute on Research and Statistics, Sacramento, CA, March, 2005.

22. p.65 in W. Härdle, M. Müller, S. Sperlich, A. Werwatz (2004), Nonparametric and Semiparametric Models, Springer, ISBN 3-540-20722-8

23. "Checking Out Statistical Confidence Interval Critical Values – For Dummies" (http://www.dummies.com/how-to/content/checking-out-statistical-confidence-interval-criti.html). *www.dummies.com*. Retrieved 2016-02-11.

24. "Confidence Intervals" (http://www.stat.yale.edu/Courses/1997-98/101/confint.htm). *www.stat.yale.edu*. Retrieved 2016-02-11.

25. George G. Roussas (1997) A Course in Mathematical Statistics, 2nd Edition, Academic Press, p397

26. Abramovich, Felix, and Ya'acov Ritov. Statistical Theory: A Concise Introduction. CRC Press, 2013. Pages 121–122

27. Rees. D.G. (2001) *Essential Statistics*, 4th Edition, Chapman and Hall/CRC. ISBN 1-58488-007-4 (Section 9.5)

28. Bernardo JE, Smith, Adrian (2000). *Bayesian theory*. New York: Wiley. p. 259. ISBN 0-471-49464-X.

29. Welch, B. L. (1939). "On Confidence Limits and Sufficiency, with Particular Reference to Parameters of Location". *The Annals of Mathematical Statistics*. Institute of Mathematical Statistics. **10** (1): 58–69. JSTOR 2235987 (https://www.jstor.org/stable/2235987). doi:10.1214/aoms/1177732246 (https://doi.org/10.1214%2Faoms%2F1177732246).

30. Robinson, G. K. (1975). "Some Counterexamples to the Theory of Confidence Intervals". *Biometrika*. Oxford University Press. **62** (1): 155–161. JSTOR 2334498 (https://www.jstor.org/stable/2334498).

31. Pratt, J. W. (1961). "Book Review: Testing Statistical Hypotheses. by E. L. Lehmann". *Journal of the American Statistical Association*. Taylor & Francis, Ltd. **56** (293): 163–167. JSTOR 2282344 (https://www.jstor.org/stable/2282344).

32. Steiger, J. H. (2004). "Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis" (http://psycnet.apa.org/journals/met/9/2/164/). *Psychological Methods*. American Psychological Association. **9** (2): 164–182. doi:10.1037/1082-989x.9.2.164 (https://doi.org/10.1037%2F1082-989x.9.2.164).

Bibliography

- Fisher, R.A. (1956) *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh. (See p. 32.)
- Freund, J.E. (1962) *Mathematical Statistics* Prentice Hall, Englewood Cliffs, NJ. (See pp. 227–228.)
- Hacking, I. (1965) *Logic of Statistical Inference*. Cambridge University Press, Cambridge. ISBN 0-521-05165-7
- Keeping, E.S. (1962) *Introduction to Statistical Inference*. D. Van Nostrand, Princeton, NJ.
- Kiefer, J. (1977). "Conditional Confidence Statements and Confidence Estimators (with discussion)". *Journal of the American Statistical Association*. **72** (360a): 789–827. JSTOR 2286460 (https://www.jstor.org/stable/2286460). doi:10.1080/01621459.1977.10479956 (https://doi.org/10.1080%2F01621459.1977.10479956).
- Mayo, D. G. (1981) "In defence of the Neyman-Pearson theory of confidence intervals" (http://www.phil.vt.edu/dmayo/PhilStatistics/In%20Defense%20of%20the%20Neyman-Pearson%20Theory%20of%20Confidence%20Intervals.pdf), *Philosophy of Science*, 48 (2), 269–280. JSTOR 187185 (https://www.jstor.org/stable/187185)
- Neyman, J. (1937) "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability" (http://www.jstor.org/stable/91337) *Philosophical Transactions of the Royal Society of London A*, **236**, 333–380. (Seminal work.)
- Robinson, G.K. (1975). "Some Counterexamples to the Theory of Confidence Intervals". *Biometrika*. **62**: 155–161. JSTOR 2334498 (https://www.jstor.org/stable/2334498). doi:10.1093/biomet/62.1.155 (https://doi.org/10.1093%2Fbiomet%2F62.1.155).
- Savage, L. J. (1962), *The Foundations of Statistical Inference*. Methuen, London.
- Smithson, M. (2003) *Confidence intervals*. Quantitative Applications in the Social Sciences Series, No. 140. Belmont, CA: SAGE Publications. ISBN 978-0-7619-2499-9.

- Mehta, S. (2014) *Statistics Topics* ISBN 978-1-4992-7353-3
- Hazewinkel, Michiel, ed. (2001) [1994], "Confidence estimation" (https://www.encyclopediaofmath.org/index.php?title=p/c024620), *Encyclopedia of Mathematics*, Springer Science+Business Media B.V. / Kluwer Academic Publishers, ISBN 978-1-55608-010-4

External links

- The Exploratory Software for Confidence Intervals tutorial programs that run under Excel (http://www.latrobe.edu.au/psy/esci/)
- Confidence interval calculators for R-Squares (http://www.danielsoper.com/statcalc/calc28.aspx), Regression Coefficients (http://www.danielsoper.com/statcalc/calc26.aspx), and Regression Intercepts (http://www.danielsoper.com/statcalc/calc27.aspx)
- Weisstein, Eric W. "Confidence Interval" (http://mathworld.wolfram.com/ConfidenceInterval.html). *MathWorld*.

- CAUSEweb.org (http://www.causeweb.org/) Many resources for teaching statistics including Confidence Intervals.
- An interactive introduction to Confidence Intervals (http://www.usablestats.com/tutorials/CI)
- *Confidence Intervals: Confidence Level, Sample Size, and Margin of Error (http://demonstrations.wolfram.com/ConfidenceIntervalsConfidenceLevelSampleSizeAndMarginOfError/)* by Eric Schulz, the Wolfram Demonstrations Project.
- *Confidence Intervals in Public Health (http://health.utah.gov/opha/IBIShelp/ConfInts.pdf)*. Straightforward description with examples and what to do about small sample sizes or rates near 0.

Online calculators

- GraphPad QuickCalcs (http://www.graphpad.com/quickcalcs/index.cfm)
- TAMU's Confidence Interval Calculators (http://www.stat.tamu.edu/~jhardin/applets/index.html)
- MBAStats confidence interval and hypothesis test calculators (http://www.mbastats.net/)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Confidence_interval&oldid=795014764"