

Encyclopedia of Research Design

Confidence Intervals

Contributors: Qiaoyan Hu & Shi Zhao & Jie Yang

Editors: Neil J. Salkind

Book Title: Encyclopedia of Research Design

Chapter Title: "Confidence Intervals"

Pub. Date: 2010

Access Date: October 14, 2013

Publishing Company: SAGE Publications, Inc.

City: Thousand Oaks

Print ISBN: 9781412961271

Online ISBN: 9781412961288

DOI: <http://dx.doi.org/10.4135/9781412961288.n68>

Print pages: 212-216

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781412961288.n68>

A confidence interval is an interval estimate of an unknown population parameter. It is constructed according to a random sample from the population and is always associated with a certain confidence level that is a probability, usually presented as a percentage. Commonly used confidence levels include 90%, 95%, and 99%. For instance, a confidence level of 95% indicates that 95% of the time the confidence intervals will contain the population parameter. A higher confidence level usually forces a confidence interval to be wider.

Confidence intervals have a long history. Using confidence intervals in statistical inference can be tracked back to the 1930s, and they are being used increasingly in research, especially in recent medical research articles. Researchers and research organizations such as the American Psychological Association suggest that confidence intervals should always be reported because confidence intervals provide information on both significance of test and variability of estimation.

Interpretation

A confidence interval is a range in which an unknown population parameter is likely to be included. After independent samples are randomly selected from the same population, one confidence interval is constructed based on one sample with a certain confidence level. Together, all the confidence intervals should include the population parameter with the confidence level.

Suppose one is interested in estimating the proportion of bass among all types of fish in a lake. A 95% confidence interval for this proportion, [25%, 36%], is constructed on the basis of a random sample of fish in the lake. After more independent random samples of fish are selected from the lake, through the same procedure more confidence intervals are constructed. Together, all these confidence intervals will contain the true proportion of bass in the lake approximately 95% of the time.

The lower and upper boundaries of a confidence interval are called *lower confidence limit* and *upper confidence limit*, respectively. In the earlier example, 25% is the lower 95% confidence limit, and 36% is the upper 95% confidence limit.

Confidence Interval versus Significance Test

A significance test can be achieved by constructing confidence intervals. One can conclude whether a test is significant based on the confidence intervals. Suppose the null hypothesis is that the population mean μ equals 0 and the predetermined significance level is α . Let I be the constructed $100(1-\alpha)\%$ confidence interval for μ . If 0 is included in the interval I , then the null hypothesis is accepted; otherwise, it is rejected. For example, a researcher wants to test whether the result of an experiment has a mean 0 at 5% significance level. After the 95% confidence interval $[-0.31, 0.21]$ is obtained, it can be concluded that the null hypothesis is accepted, because 0 is included in the confidence interval. If the 95% confidence interval is $[0.11, 0.61]$, it indicates that the mean is significantly different from 0.

For a predetermined significance level or confidence level, the ways of constructing confidence intervals are usually not unique. Shorter confidence intervals are usually better because they indicate greater power in the sense of significance test.

One-sided significance tests can be achieved by constructing one-sided confidence intervals. Suppose a researcher is interested in an alternative hypothesis that the population mean is larger than 0 at the significance level .025. The researcher will construct a one-sided confidence interval taking the form of $(-\infty, b]$ with some constant b . Note that the width of a one-sided confidence interval is infinity. Following the above example, the null hypothesis would be that the mean is less than or equal to 0 at the .025 level. If the 97.5% one-sided confidence interval is $(-\infty, 3.8]$, then the null hypothesis is accepted because 0 is included in the interval. If the 97.5% confidence interval is $(-\infty, -2.1]$ instead, then the null hypothesis is rejected because there is no overlap between $(-\infty, -2.1]$ and $[0, \infty)$.

Examples

Confidence Intervals for a Population Mean with Confidence Level $100(1-\alpha)\%$

Confidence intervals for a population mean are constructed on the basis of the sample mean distribution.

The Population Follows a Normal Distribution with a Known Variance σ^2

After a random sample of size N is selected from the population, one is able to calculate the sample mean

\bar{x}

, which is the average of all the observations. Thus the confidence interval is

$$[\bar{x} - z_{\alpha/2} \times \sigma / \sqrt{N}, \bar{x} + z_{\alpha/2} \times \sigma / \sqrt{N}]$$

that is centered at

\bar{x}

with half of the length

$$z_{\alpha/2} \times \sigma / \sqrt{N}. z_{\alpha/2}$$

is the upper $\alpha/2$ quantile, meaning $P(Z > z_{\alpha/2}) = \alpha/2$

$P(Z \leq z) = \alpha/2$. Here Z is a standard normal random variable. To find z

, one can either refer to the standard normal distribution table or use statistical software. Nowadays most of the statistical software, such as Excel, R, SAS, SPSS (an IBM company, formerly called PASW[®] Statistics), and Splus, have a simple command that will do the job. For commonly used confidence intervals, 90% confidence level corresponds to z

$z = 1.645$, 95% corresponds to z

$z = 1.96$, and 99% corresponds to z

$z = 2.56$.

The above confidence interval will shrink to a point if N goes to infinity. So the interval estimate turns into a point estimate. It can be interpreted as if the whole population is taken as the sample; the sample mean is actually the population mean.

The Population Follows a Normal Distribution with Unknown Variance.

Suppose a sample of size N is randomly selected from the population with observations

$x_1,$

,

x_2, \dots, x_N . Let $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$

. This is the sample mean. The sample variance is defined as

$$\sum_{i=1}^N (x_i - \bar{x})^2 / (N - 1)$$

, denoted by s^2 . Therefore the confidence interval is

$$[\bar{x} - t_{N-1, \alpha/2} \times s / \sqrt{N} \quad \bar{x} + t_{N-1, \alpha/2} \times s / \sqrt{N}]$$

. Here

$$t_{N-1, \alpha/2}$$

is the upper #/2 quantile, meaning

$$P(t_{N-1, \alpha/2}$$

and T
 $N-1$

follows the t distribution with degree of freedom $N-1$. Refer to the t -distribution table or use software to get t
 $N-1, \alpha/2$.

If N is greater than 30, one can use z
 $\alpha/2$

instead of t
 $N-1, \alpha/2$

in the confidence interval formula because there is little difference between them for large enough N .

As an example, the students' test scores in a class follow a normal distribution. One wants to construct a 95% confidence interval for the class average score based on an available random sample of size $N = 10$. The 10 scores are 69, 71, 77, 79, 82, 84,

80, 94, 78, and 67. The sample mean and the sample variance are 78.1 and 62.77, respectively. According to the t -distribution table, $t_{9,0.25}$

= 2.26. The 95% confidence interval for the class average score is [72.13, 84.07].

The Population Follows a General Distribution Other than Normal Distribution

This is a common situation one might see in practice. After obtaining a random sample of size N from the population, where it is required that $N \geq 30$, the sample mean and the sample variance can be computed as in the previous subsections, denoted by \bar{x}

and s^2 , respectively. According to the central limit theorem, an approximate confidence interval can be expressed as

$$[\bar{x} - z_{\alpha/2} \times s / \sqrt{N}, \bar{x} + z_{\alpha/2} \times s / \sqrt{N}]$$

Confidence Interval for Difference of Two Population Means with Confidence Level $100(1-\alpha)\%$

One will select a random sample from each population. Suppose the sample sizes are N_1

and N_2

2

. Denote the sample means by

\bar{x}

1

and

\bar{x}

2

[p. 213 ↓] and the sample variances by s

1

s and s

2

s , respectively, for the two samples. The confidence interval is

$$[\bar{x}_1 - \bar{x}_2 - z_{\alpha/2} \times \sqrt{s_1^2/N_1 + s_2^2/N_2}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} \times \sqrt{s_1^2/N_1 + s_2^2/N_2}]$$

.

If one believes that the two population variances are about the same, the confidence interval will be

$$[\bar{x}_1 - \bar{x}_2 - t_{N_1+N_2-2, \alpha/2} \times s_p, \bar{x}_1 - \bar{x}_2 + t_{N_1+N_2-2, \alpha/2} \times s_p]$$

, where

$$s_p = \sqrt{(N_1 + N_2)[(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2] / [(N_1 + N_2 - 2)N_1N_2]}$$

Continuing with the above example about the students' scores, call that class Class A: If one is interested in comparing the average scores between Class A and another class, Class B, the confidence interval for the difference between the average class scores will be constructed. First, randomly select a group of students in Class B. Suppose the group size is 8. These eight students' scores are 68, 79, 59, 76, 80, 89, 67, and 74. The sample mean is 74, and the sample variance is 86.56 for Class B. If Class A and Class B are believed to have different population variances, then the 95% confidence interval for the difference of the average scores is $[-4.03, 12.17]$ by the first formula provided in this subsection. If one believes these two classes have about the same population variance, then the 95% confidence interval will be changed to $[-4.47, 12.57]$ by the second formula.

Confidence Intervals for a Single Proportion and Difference of Two Proportions

Sometimes one may need to construct confidence intervals for a single unknown population proportion. Denote

\hat{p}

the sample proportion, which can be obtained from a random sample from the population. The estimated standard error for the proportion is

$$se(\hat{p}) = \sqrt{\hat{p}(1 - \hat{p})/N}$$

. Thus the confidence interval for the unknown proportion is

$$[\hat{p} - z_{\alpha/2} \times se(\hat{p}), \hat{p} + z_{\alpha/2} \times se(\hat{p})]$$

This confidence interval is constructed on the basis of the normal approximation (refer to the central limit theorem for the normal approximation). The normal approximation is not appropriate when the proportion is very close to 0 or 1. A rule of thumb is that when

$$N\hat{p} > 5$$

and

$$N(1 - \hat{p}) > 5$$

, usually the normal approximation works well.

For example, a doctor wants to construct a 99% confidence interval for the chance of having a certain disease by studying patients' x-ray slides. $N = 30$ x-ray slides are randomly selected, and the number of positive slides follows a distribution known as the binomial distribution. Suppose 12 of them are positive for the disease. Hence

$$\hat{p} = 12/30$$

, which is the sample proportion. Since

$$N\hat{p} = 12$$

and

$$N(1 - \hat{p}) = 18$$

are both larger than 5, the confidence interval for the unknown proportion can be constructed using the normal approximation. The estimated standard error for

$$\hat{p}$$

is 0.09. Thus the lower 99% confidence limit is 0.17 and the upper 99% confidence limit is 0.63. So the 99% confidence interval is [0.17, 0.63].

The range of a proportion is between 0 and 1. But sometimes the constructed confidence interval for the proportion may exceed it. When this happens, one should

truncate the confidence interval to make the lower confidence limit 0 or the upper confidence limit 1.

Since the binomial distribution is discrete, a correction for continuity of $0.5/N$ may be used to improve the performance of confidence intervals. The corrected upper limit is added by $0.5/N$, and the corrected lower limit is subtracted by $0.5/N$.

One can also construct confidence intervals for the proportion difference between two populations based on the normal approximation. Suppose two random samples are independently selected from the two populations, with sample sizes N_1

and N_2

and sample proportions

\hat{p}_1

and \hat{p}_2

respectively. The estimated population proportion difference is the sample proportion difference,

$\hat{p}_1 - \hat{p}_2$

and

$\hat{p}_1 - \hat{p}_2$

respectively.

The estimated population proportion difference is the sample proportion difference,

$\hat{p}_1 - \hat{p}_2$

and

$\hat{p}_1 - \hat{p}_2$

respectively.

2

, and the estimated standard error for the proportion difference is

$$se(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}_1(1 - \hat{p}_1) / N_1 + \hat{p}_2(1 - \hat{p}_2) / N_2}$$

. The confidence interval for two-sample proportion difference is

$$[(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \times se(\hat{p}_1 - \hat{p}_2), (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \times se(\hat{p}_1 - \hat{p}_2)]$$

.

Similar to the normal approximation for a single proportion, the approximation for the proportion difference depends on sample sizes and sample proportions. The rule of thumb is that

$$N\hat{p}_1$$

, $N(1 -$

$$\hat{p}_1$$

1

),

$$N\hat{p}_2$$

, and $N(1 -$

$$\hat{p}_2$$

2

) should each be larger than 10.

Confidence Intervals for Odds Ratio

An *odds ratio* (OR) is a commonly used effect size for categorical outcomes, especially in health [p. 214 ↓] science, and is the ratio of the odds in Category 1 to the odds in Category 2. For example, one wants to find out the relationship between smoking and lung cancer. Two groups of subjects, smokers and nonsmokers, are recruited. After a few years' follow-ups, N

11

subjects among the smokers are diagnosed with lung cancer and N

21

subjects among the nonsmokers. There are N

12

and N

22

subjects who do not have lung cancer among the smokers and the nonsmokers, respectively. The *odds* of having lung cancer among the smokers and the non-smokers are estimated as N

11

/ N

12

and N

21

/ N

22

, respectively. The OR of having lung cancer among the smokers compared with the nonsmokers is the ratio of the above two odds, which is

$$(N_{11}/N_{12})/(N_{21}/N_{22}) = (N_{11}N_{22})/(N_{21}N_{12}).$$

For a relatively large total sample size, $\ln(OR)$ is approximated to a normal distribution, so the construction for the confidence interval for $\ln(OR)$ is similar to that for the normal distribution. The standard error for $\ln(OR)$ is defined as $se(\ln(OR))$

$$= \sqrt{1/N_{11} + 1/N_{12} + 1/N_{21} + 1/N_{22}}$$

. A 95% confidence interval for $\ln(OR)$ is $[\ln(OR) - 1.96 \times se(\ln(OR)), \ln(OR) + 1.96 \times se(\ln(OR))]$. As the exponential function is mono-tonic, there is one-to-one mapping between the OR and $\ln(OR)$. Thus a 95% confidence interval for the OR is $[\exp(\ln(OR) - 1.96 \times se(\ln(OR))), \exp(\ln(OR) + 1.96 \times se(\ln(OR)))]$.

The confidence intervals for the OR are not symmetric about the estimated OR. But one can still tell the significance of the test on the basis of the corresponding confidence interval for the OR. For the above example, the null hypothesis is that there is no difference between smokers and non-smokers in the development of lung cancer; that is, $OR = 1$. If 1 is included in the confidence interval, one should accept the null hypothesis; otherwise, one should reject it.

Confidence Intervals for Relative Risk

Another widely used concept in health care is *relative risk* (RR), which is the risk difference between two groups. *Risk* is defined as the chance of having a specific outcome among subjects in that group. Taking the above example, the risk of having lung cancer among smokers is estimated as N

N_{11}

$/(N_{11} + N_{12})$

N_{21}

$+ N_{22}$

N_{12}

), and the risk among nonsmokers is estimated as N_{21}

$$1/(N_{21})$$

$$+ N_{22}$$

). The RR is the ratio of the above two risks, which is $[N_{11}/N_{21}] / [N_{12}/N_{22}]$.

$$1/(N_{11})$$

$$+ N_{12}$$

$$)]/[N_{21}]$$

$$1/(N_{21})$$

$$+ N_{22}$$

)].

Like the OR, the sampling distribution of $\ln(RR)$ is approximated to a normal distribution. The standard error for $\ln(RR)$ is $se(\ln(RR))$

$$se(\ln(RR)) = \sqrt{1/N_{11} - 1/N_{1.} + 1/N_{21} - 1/N_{2.}}$$

. A 95% confidence interval for $\ln(RR)$ is

$$[\ln(RR) - 1.96 \times se(\ln(RR)), \ln(RR) + 1.96 \times se(\ln(RR))].$$

Table 1 Lung Cancer Among Smokers and Nonsmokers

<i>Smoking Status</i>	<i>Lung Cancer</i>	<i>No Lung Cancer</i>	<i>Total</i>
Smokers	N_{11}	N_{12}	N_1
Nonsmokers	N_{21}	N_{22}	N_2
Total	N_1	N_2	N

Thus the 95% confidence interval for the RR is $[\exp(\ln(RR) - 1.96 \times se(\ln(RR))), \exp(\ln(RR) + 1.96 \times se(\ln(RR)))]$.

The confidence intervals for the RRs are not symmetric about the estimated RR either. One can tell the significance of a test from the corresponding confidence interval for the RR . Usually the null hypothesis is that $RR = 1$, which means that the two groups have the same risk. For the above example, the null hypothesis would be that the risks of developing lung cancer among smokers and non-smokers are equal. If 1 is included in the confidence interval, one may accept the null hypothesis. If not, the null hypothesis should be rejected.

Confidence Intervals for Variance

A confidence interval for unknown population variance can be constructed with the use of a central chi-square distribution. For a random sample with size N and sample variance s^2 , an approximate two-sided $100(1-\alpha)\%$ confidence interval for population variance σ^2 is

$$[(N - 1)s^2 / \chi_{N-1}^2(\alpha/2), (N - 1)s^2 / \chi_{N-1}^2(1 - \alpha/2)]$$

. Here

$$\chi_{N-1}^2(\alpha/2)$$

is at the upper $\alpha/2$ quantile satisfying the requirement that the probability that a central chi-square random variable with degree of freedom $N-1$ is greater than X
 $N-1$

$^2 (\alpha/2)$ is $\alpha/2$. Note that this confidence interval may not work well if the sample size is small or the distribution is far from normal.

[p. 215 ↓]

Bootstrap Confidence Intervals

The bootstrap method provides an alternative way for constructing an interval to measure the accuracy of an estimate. It is especially useful when the usual confidence interval is hard or impossible to calculate. Suppose $s(x)$ is used to estimate an unknown population parameter θ based on a sample x of size N . A bootstrap confidence interval for the estimate can be constructed as follows. One randomly draws another sample x^* of the same size N with replacement from the original sample x . The estimate $s(x^*)$ based on x^* is called a *bootstrap replication* of the original estimate. Repeat the procedure for a large number of times, say 1,000 times. Then the α th quantile and the $(1-\alpha)$ th quantile of the 1,000 bootstrap replications serve as the lower and upper limits of a $100(1-2\alpha)\%$ bootstrap confidence interval. Note that the interval obtained in this way may vary a little from time to time due to the randomness of bootstrap replications. Unlike the usual confidence interval, the bootstrap confidence interval does not require assumptions on the population distribution. Instead, it highly depends on the data x itself.

Simultaneous Confidence Intervals

Simultaneous confidence intervals are intervals for estimating two or more parameters at a time. For example, suppose μ
1

and μ

2

are the means of two different populations. One wants to find confidence intervals I_1

and I_2

simultaneously such that

$$P(\mu_1 \in I_1 \text{ and } \mu_2 \in I_2) = 1 - \alpha.$$

If the sample x_1

used to estimate μ_1

is independent of the sample x_2

for μ_2

, then I_1

and I_2

can be simply calculated as

$$100\sqrt{1 - \alpha}\%$$

confidence intervals for μ_1

and μ_2

, respectively.

The simultaneous confidence intervals I_1

and I_2

can be used to test whether μ_1

and μ_2

are equal. If I_1

and I_2

are nonoverlapped, then μ_1

and μ_2

are significantly different from each other at a level less than α .

Simultaneous confidence intervals can be generalized into a *confidence region* in the multidimensional parameter space, especially when the estimates for parameters are not independent. A $100(1-\alpha)\%$ confidence region D for the parameter vector θ satisfies $P(\theta \in D) = 1-\alpha$, where D does not need to be a product of simple intervals.

Links with Bayesian Intervals

A Bayesian interval, or *credible interval*, is derived from the posterior distribution of a population parameter. From a Bayesian point of view, the parameter θ can be regarded as a random quantity, which follows a distribution $P(\theta)$, known as the *prior distribution*. For each fixed θ , the data x are assumed to follow the conditional distribution $P(x|\theta)$, known as the *model*. Then Bayes's theorem can be applied on x to get the adjusted distribution $P(\theta|x)$ of θ , known as the *posterior distribution*, which is proportional to $P(x|\theta) \cdot P(\theta)$. A $1-\alpha$ Bayesian interval I is an interval that satisfies $P(\theta \in I|x) = 1-\alpha$ according to the posterior distribution. In order to guarantee the optimality or uniqueness of I , one may require that the values of the posterior density function inside I always be greater than any one outside I . In those cases, the Bayesian interval I is called the *highest posterior density region*. Unlike the usual confidence intervals, the level $1-\alpha$ of a Bayesian interval I indicates the probability that the random θ falls into I .

Qiaoyan Hu, Shi Zhao and Jie Yang

<http://dx.doi.org/10.4135/9781412961288.n68>

See also

Further Readings

Blyth, C. R., & Still, H. A. Binomial confidence intervals. *Journal of the American Statistical Association*, (1983). vol. 78, pp. 108–116.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.

Fleiss, J. R., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions*. Hoboken, NJ: Wiley.

Newcombe, R. G. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, (1998). vol. 17, pp. 857–872.

Pagano, M., & Gauvreau, K. (2000). Principles of biostatistics. Pacific Grove, CA: Duxbury.

Smithson, M. (2003). Confidence intervals. Thousand Oaks, CA: Sage.

Stuart, A., Ord, K., & Arnold, S. (1998). Kendall's advanced theory of statistics: Vol. 2A. Classical inference and the linear model (6th ed.). London: Arnold.