Confidence Intervals

In order to construct a **confidence set**, which often takes the form of a **confidence interval**, or **CI**, we must (perhaps implicitly) invert a test.

We need a suitable **family of tests** for a set of point null hypotheses about a parameter, say θ.

- If we use a t test for the hypothesis that $\theta = \theta_0$ for any specified θ_0, then we have a family of t tests indexed by θ_0.
- A confidence set contains all values θ_0 for which the hypothesis that $\theta = \theta_0$ is not rejected by the appropriate test in the family.
- When the tests are at level α, we obtain an interval at **confidence level** $1 - \alpha$. Popular levels are .90, .95, and .99.

Let $\tau(\boldsymbol{y}, \theta_0)$ denote a test statistic for the hypothesis that $\theta = \theta_0$.

Assume for now that $\tau(\boldsymbol{y}, \theta_0) > 0$, like F and χ^2 statistics.

For each θ_0, the test compares $\tau(\boldsymbol{y}, \theta_0)$ with c_α, the level α critical value, which is the $1-\alpha$ **quantile** of some distribution.

This quantile is 3.8415 when $\tau \sim \chi^2(1)$ and $\alpha = 0.05$.

By the definition of c_α,

$$\Pr\big(\tau(\boldsymbol{y}, \theta_0) \leq c_\alpha\big) = 1 - \alpha. \tag{1}$$

For θ_0 to belong to the CI, it is necessary and sufficient that

$$\tau(\boldsymbol{y}, \theta_0) \leq c_\alpha. \tag{2}$$

The limits of the CI can be found by solving for θ the equation

$$\tau(\boldsymbol{y}, \theta) = c_\alpha. \tag{3}$$

This equation normally has two solutions. One solution is the upper limit, θ_u, and the other is the lower limit, θ_l.

We can write the interval as $[\theta_l, \;\; \theta_u]$.

Confidence intervals are random. The probability that the interval covers the true parameter value is called the **coverage probability**, or just the **coverage**, of the interval.

Confidence intervals may be **exact** (if based on an exact test) or **approximate** (if based on a test that is not exact).

- For an exact interval, the coverage is precisely the level of the interval, that is, $1 - \alpha$.
- For an approximate interval, the coverage may be larger or smaller than $1 - \alpha$.
- An approximate interval may over-cover, or under-cover, in both tails. Or it may over-cover in one tail and under-cover in the other.
- If a test over-rejects, then intervals based on it will under-cover.
- If a test under-rejects, then intervals based on it will over-cover.
- The most popular approximate intervals are usually based on either asymptotic theory or the bootstrap.

Exact Confidence Intervals

When every test in the family of tests is exact at level α, the coverage of the confidence interval (or confidence set) is exactly $1 - \alpha$.

Since θ_0 belongs to the confidence interval if and only if (2) holds, the confidence interval contains the true parameter value θ_0 with probability exactly equal to $1 - \alpha$.

If $F(x)$ is the CDF of X, and if $f(x) \equiv F'(x)$ exists and is strictly positive everywhere, then q_α, the **α quantile** of F, satisfies $F(q_\alpha) = \alpha$ for $0 \leq \alpha \leq 1$.

If the PDF is strictly positive, F is strictly increasing. Therefore, the inverse function $F^{-1}(\alpha)$ exists, and $q_\alpha = F^{-1}(\alpha)$. The function $F^{-1}(\alpha)$ is called the **quantile function**.

If F is not strictly increasing, or f does not exist (e.g. discrete distribution), then the α quantile may not exist, or it may not be uniquely defined, for every value of α.

Various quantiles of interest:

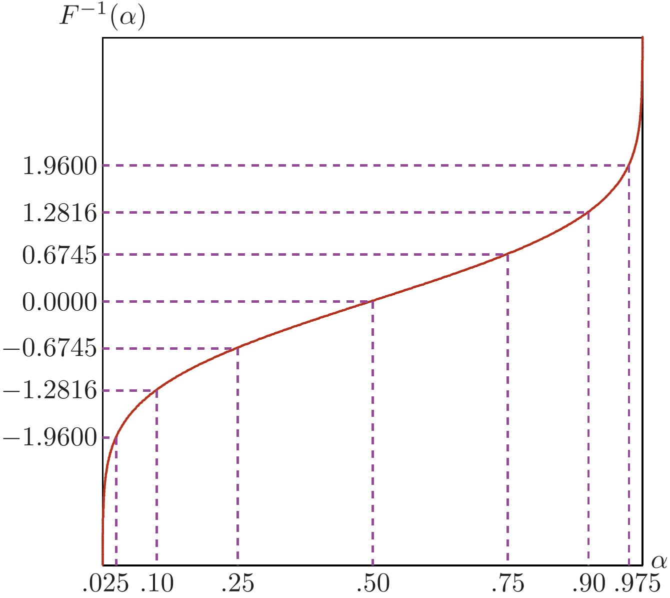
- The .5 quantile is called the **median**.
- For $\alpha = .25, .5$, and $.75$, the corresponding quantiles are called **quartiles**. For example, the .75 quantile is the **third quartile**.
- For $\alpha = .2, .4, .6$, and $.8$, they are called **quintiles**.
- For $\alpha = i/10$ with $1 \leq i \leq 9$, they are called **deciles**.
- For $\alpha = i/100$ with $1 \leq i \leq 99$, they are called **centiles**.

The quantile function of the standard normal distribution is shown in the following figure.

It is just the standard normal CDF rotated by 90 degrees, so that α is on the horizontal axis and the quantiles are on the vertical axis.

Finding the critical values for an interval at level $1 - \alpha$ based on the normal distribution requires us to find $\Phi^{-1}(1 - \alpha/2)$ and $\Phi^{-1}(\alpha/2) = -\Phi^{-1}(1 - \alpha/2)$.

For $\alpha = .05$, $\Phi^{-1}(\alpha/2) = -1.96$ and $\Phi^{-1}(1 - \alpha/2) = 1.96$.


$F^{-1}(\alpha)$
1.9600
1.2816
0.6745
0.0000
−0.6745
−1.2816
−1.9600
α
.025 .10 .25 .50 .75 .90 .975

Asymptotic Confidence Intervals

In principle, $\tau(\boldsymbol{y}, \theta_0)$ can be any sort of test statistic.

Ideally, we know its finite-sample distribution. If not, then we at least know its asymptotic distribution.

Suppose that

$$\tau(\boldsymbol{y}, \theta_0) \equiv \left(\frac{\hat{\theta} - \theta_0}{s_\theta}\right)^2, \tag{4}$$

where $\hat{\theta}$ is an estimate of θ, and s_θ is the corresponding standard error, that is, an estimate of the standard deviation of $\hat{\theta}$.

$\tau(\boldsymbol{y}, \theta_0)$ is the square of the t statistic for the hypothesis that $\theta = \theta_0$.

Under standard conditions, $\tau(\boldsymbol{y}, \theta_0) \overset{a}{\sim} \chi^2(1)$.

The asymptotic critical value c_α is the $1 - \alpha$ quantile of the $\chi^2(1)$ distribution.

To find the asymptotic interval for θ based on (4), we need to solve

$$\left(\frac{\hat{\theta}-\theta}{s_\theta}\right)^2 = c_\alpha. \tag{5}$$

Taking the square root of both sides and multiplying by s_θ then gives

$$|\hat{\theta}-\theta| = s_\theta c_\alpha^{1/2}. \tag{6}$$

There are two solutions to this equation. These are

$$\theta_l = \hat{\theta} - s_\theta c_\alpha^{1/2} \quad \text{and} \quad \theta_u = \hat{\theta} + s_\theta c_\alpha^{1/2}. \tag{7}$$

Thus the asymptotic $1-\alpha$ confidence interval for θ is

$$\left[\hat{\theta} - s_\theta c_\alpha^{1/2},\ \ \hat{\theta} + s_\theta c_\alpha^{1/2}\right]. \tag{8}$$

The lower limit θ_l is $\hat{\theta} - s_\theta c_\alpha^{1/2}$, and the upper limit θ_u is $\hat{\theta} + s_\theta c_\alpha^{1/2}$.

For $\alpha = .05$, the $1 - \alpha$ quantile of the $\chi^2(1)$ distribution is 3.8415. Since $\sqrt{3.8415} = 1.9600$, the confidence interval given by (8) becomes

$$\big[\hat{\theta} - 1.96 s_\theta,\ \ \hat{\theta} + 1.96 s_\theta\big]. \tag{9}$$

The construction of this interval is illustrated in the next figure.

The same interval arises if we invert the asymptotic t statistic $(\hat{\theta} - \theta_0)/s_\theta$ and use the $\mathrm{N}(0,1)$ distribution for a two-tailed test.

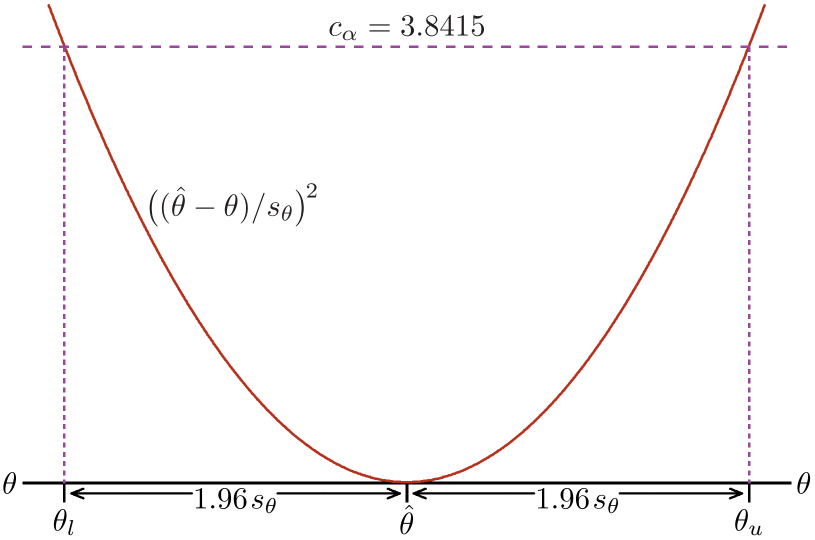
The critical values for such a test are the $\alpha/2$ and $1 - \alpha/2$ quantiles of the distribution. For the $\mathrm{N}(0,1)$ distribution, these are $z_{\alpha/2}$ and $z_{1-\alpha/2}$.

$z_{\alpha/2}$ is negative, since $\alpha/2 < 1/2$, and the median of the $\mathrm{N}(0,1)$ distribution is 0. By symmetry, it is easy to see that $z_{\alpha/2} = -z_{1-\alpha/2}$.

Thus the CI can can be written in two different ways:

$$\big[\hat{\theta} + s_\theta z_{\alpha/2},\ \ \hat{\theta} - s_\theta z_{\alpha/2}\big] \quad \text{and} \quad \big[\hat{\theta} - s_\theta z_{1-\alpha/2},\ \ \hat{\theta} + s_\theta z_{1-\alpha/2}\big]. \tag{10}$$

When $\alpha = .05$, we obtain the interval (9), since $z_{.025} = -1.96$ and $z_{.975} = 1.96$.


$c_\alpha = 3.8415$
$\left((\hat{\theta} - \theta)/s_\theta\right)^2$
θ
θ
$1.96\, s_\theta$
$1.96\, s_\theta$
θ_l
$\hat{\theta}$
θ_u

Asymmetric Confidence Intervals

The confidence interval (10) is **symmetric**, because θ_l is as far below $\hat{\theta}$ as θ_u is above it.

This is the case because the $\mathrm{N}(0,1)$ distribution is symmetric.

We can obtain an **asymmetric confidence interval** if we construct a confidence interval based on a two-tailed test when the distribution of the test statistic is not symmetric.

If we define the rejection region so that there is a probability mass of $\alpha/2$ in each tail, we obtain an **equal-tailed confidence interval**.

Two critical values are needed: The lower one, c_α^-, is the $\alpha/2$ quantile of the distribution, and the upper one, c_α^+, is the $1-\alpha/2$ quantile.

- Any value θ lies within the interval whenever $\tau(\boldsymbol{y},\theta) \geq c_\alpha^-$ and $\tau(\boldsymbol{y},\theta) \leq c_\alpha^+$.
- Values of θ for which $\tau(\boldsymbol{y},\theta)$ is below c_α^- or above c_α^+ lie outside the interval.

Exact Confidence Intervals for Regression Coefficients

Consider the classical normal linear model, where $\boldsymbol{\beta}$ has been partitioned as $[\boldsymbol{\beta}_1 \,\vdots\, \beta_2]$, with $\boldsymbol{\beta}_1$ a $(k-1)$-vector and β_2 a scalar.

If s_2 is the OLS standard error for $\hat{\beta}_2$, the t statistic for $\beta_2 = \beta_2^0$ is

$$\frac{\hat{\beta}_2 - \beta_2^0}{s_2} \sim t(N-k). \tag{11}$$

Since any DGP in the model satisfies $\beta_2 = \beta_2^0$ for some β_2^0,

$$\Pr\left(t_{\alpha/2} \le \frac{\hat{\beta}_2 - \beta_2^0}{s_2} \le t_{1-\alpha/2}\right) = 1 - \alpha, \tag{12}$$

where $t_{\alpha/2}$ and $t_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the $t(N-k)$ distribution.

We can use (12) to find a $1-\alpha$ confidence interval for β_2. The l.h.s. is

$$\begin{aligned}
&\Pr\big(s_2 t_{\alpha/2} \le \hat{\beta}_2 - \beta_2^0 \le s_2 t_{1-\alpha/2}\big) && (13)\\
&\quad = \Pr\big(-s_2 t_{\alpha/2} \ge \beta_2^0 - \hat{\beta}_2 \ge -s_2 t_{1-\alpha/2}\big) && (14)\\
&\quad = \Pr\big(\hat{\beta}_2 - s_2 t_{\alpha/2} \ge \beta_2^0 \ge \hat{\beta}_2 - s_2 t_{1-\alpha/2}\big). && (15)
\end{aligned}$$

Therefore, the confidence interval we are seeking is

$$\big[\hat{\beta}_2 - s_2 t_{1-\alpha/2},\ \ \hat{\beta}_2 - s_2 t_{\alpha/2}\big]. \tag{16}$$

This looks odd, because the upper limit is obtained by subtracting something from $\hat{\beta}_2$.

But what is subtracted is negative, because $t_{\alpha/2} < 0$. We could instead have added $s_2 t_{1-\alpha/2}$ and obtained the same interval.

The lower and upper limits of (16) depend, respectively, on the upper-tail and lower-tail quantiles of the $t(N-k)$ distribution.

This fact matters for asymmetric confidence intervals.

Because the t distribution is symmetric, the interval (16) is the same as

$$\left[\hat{\beta}_2 - s_2 t_{1-\alpha/2},\ \ \hat{\beta}_2 + s_2 t_{1-\alpha/2}\right]. \tag{17}$$

For concreteness, suppose that $\alpha = .05$ and $N - k = 32$. In this special case, $t_{1-\alpha/2} = t_{.975} = 2.037$.

The .95 confidence interval based on (17) extends from 2.037 standard errors below $\hat{\beta}_2$ to 2.037 standard errors above it.

We obtained the interval (16) by starting from the t statistic (11) and using the Student's $t(N-k)$distribution. Such an interval gets longer, for the same standard error, as $N - k$ decreases.

We would have obtained precisely the same interval if we had started from the square of (11) and used the $F(1, N-k)$ distribution.

- If we invert a test with the correct size, we obtain a CI with the correct coverage.
- The more powerful is the test, the shorter is the interval.

Confidence Regions

The **confidence region** for a set of k parameters, say the components of a k-vector $\boldsymbol{\theta}$, is often the k-dimensional analog of an ellipse constructed by inverting a joint test based on the F or χ^2 distributions.

For every point $\boldsymbol{\theta}_0$ in the confidence region, the joint hypothesis that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is not rejected at level α.

A confidence region covers the true values of the parameter vector $\boldsymbol{\theta}$ $100(1-\alpha)\%$ of the time, either exactly or approximately.

Suppose we are interested in $\boldsymbol{\theta}_2$, a subvector of $\boldsymbol{\theta}$. The k_2-vector of parameter estimates $\hat{\boldsymbol{\theta}}_2$ has covariance matrix estimated by $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_2)$.

The Wald statistic

$$(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^0)^\top \big(\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_2)\big)^{-1} (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^0) \tag{18}$$

can be used to test the joint null hypothesis that $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$.

Let c_α denote the $1-\alpha$ quantile of the $\chi^2(k_2)$ distribution. Then an approximate $1-\alpha$ confidence region is the set of all $\boldsymbol{\theta}_0$ such that

$$(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^0)^\top \big(\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_2)\big)^{-1} (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^0) \leq c_\alpha. \tag{19}$$

This **asymptotic confidence region** is elliptical or ellipsoidal.

Consider the classical normal linear model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{20}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are a k_1-vector and a k_2-vector with $k = k_1 + k_2$.

This can be rewritten as

$$\boldsymbol{y} - \boldsymbol{X}_2\boldsymbol{\beta}_2^0 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\gamma} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{21}$$

where $\boldsymbol{\gamma} = \boldsymbol{\beta}_2 - \boldsymbol{\beta}_2^0$. Testing $\boldsymbol{\gamma} = \mathbf{0}$ is equivalent to testing $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0$.

The F statistic is

$$\frac{(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)^\top \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)/k_2}{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}/(N-k)}. \tag{22}$$

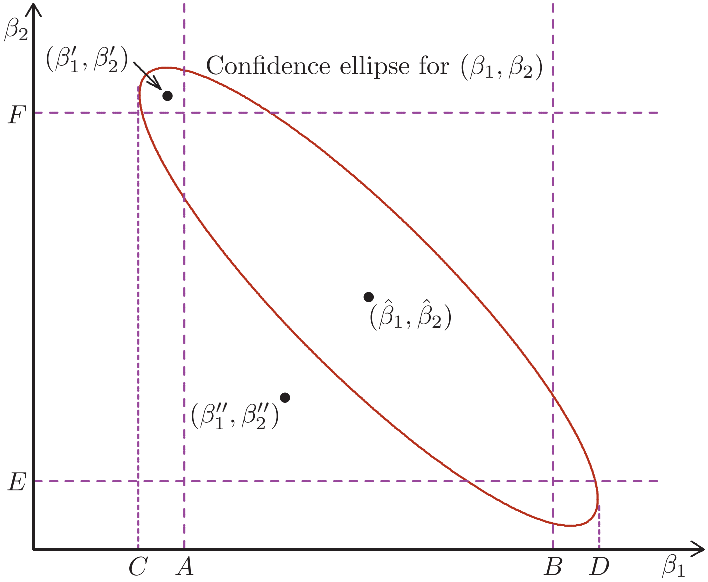
This is just $1/k_2$ times the Wald statistic (18), where $\boldsymbol{\beta}$ replaces $\boldsymbol{\theta}$ and

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_2) = s^2 \big(\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2\big)^{-1}. \tag{23}$$

If c_α denotes the $1-\alpha$ quantile of the $F(k_2, N-k)$ distribution, the $1-\alpha$ confidence region is the set of all $\boldsymbol{\beta}_2^0$ for which

$$(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)^\top \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0) \le c_\alpha k_2 s^2. \tag{24}$$

Because the left-hand side of (24) is quadratic in $\boldsymbol{\beta}_2^0$, the confidence region is the interior of an ellipse for $k_2 = 2$ and the interior of a k_2-dimensional ellipsoid for $k_2 > 2$.


β_2
(β_1', β_2')
Confidence ellipse for (β_1, β_2)
F
$(\hat{\beta}_1, \hat{\beta}_2)$
(β_1'', β_2'')
E
C
A
B
D
β_1

- Points like (β_1'', β_2'') lie outside the confidence ellipse.
- Points like (β_1', β_2') are contained in the ellipse but lie outside one or both of the confidence intervals.
- Negative slope arises when regressors are positively correlated. Think of a model for daily electricity demand in the summer, using temperature and humidity as regressors.
- When correlation between parameter estimates is zero, axes of the ellipse are parallel to the coordinate axes.
- When estimates are positively (negatively) correlated, ellipse is oriented from lower left to upper right (upper left to lower right).
- Variances of parameter estimates determine height and width.
- If the variances are equal and the correlation is zero, then the ellipse is a circle.
- The infinitely high rectangle bounded by the vertical lines through A and B must have probability mass $1 - \alpha$.