Confidence Intervals

Jiří Neubauer

Department of Econometrics FVL UO Brno
office 69a, tel. 973 442029
email:Jiri.Neubauer@unob.cz

Confidence Intervals

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample and θ. If $T_L(x_1, x_2 \ldots, x_n)$ and $T_U(x_1, x_2 \ldots, x_n)$ are two functions of the sample such that

$$P(T_L < \theta < T_U) = 1 - \alpha,$$

we call the interval (T_L, T_U) a **confidence interval** for the parameter θ with confidence level α $(0 < \alpha < 1)$.

A confidence interval is thus a random interval that contains the parameter θ with the probability α.One we have numerical observations on our sample, we get numerical values of T_L and T_U and refer to this as a observed confidence interval.

Confidence Intervals

A confidence Interval is possible to define by an inequality $\theta > T_L$ or $\theta < T_U$. We call these interval one–sided confidence intervals.
A **both–sided confidence intervals** which fulfils

$$P(\theta \leq T_L) = P(\theta \geq T_U) = \frac{\alpha}{2}$$

we call **symmetric** confidence interval. We will use only symmetric C.I.

Confidence Intervals

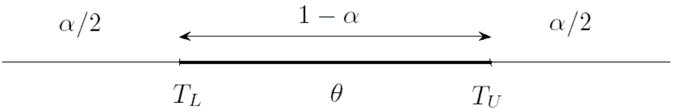
A Confidence Interval for a parameter θ

If

$$P(T_L < \theta < T_U) = 1 - \alpha,$$

$$P(\theta \leq T_L) = P(\theta \geq T_U) = \frac{\alpha}{2},$$

then the interval $T_L < \theta < T_U$ is called a *(both–sided) confidence interval for θ*.

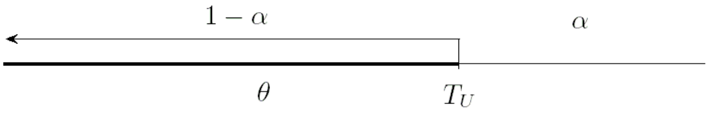


Confidence Intervals

A right–sided (upper) confidence interval for a parameter θ
If

$$P(\theta < T_U) = 1 - \alpha, P(\theta \geq T_U) = \alpha,$$

then the interval $\theta < T_U$ is called a *right–sided confidence interval for* θ.



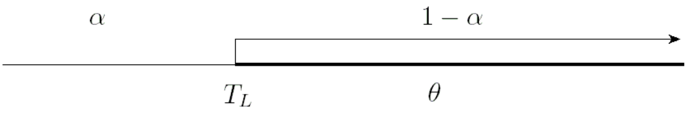
Confidence Intervals

A left–sided (lower) confidence interval for a parameter θ

If

$$P(\theta > T_L) = 1 - \alpha, P(\theta \leq T_L) = \alpha,$$

then the interval $\theta > T_L$ is called a *left–sided confidence interval for* θ.



C.I. for the Mean in the Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. A random variable

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1),$$

has a Student t-distribution with $\nu = n - 1$ degrees of freedom,

C.I. for the Mean in the Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. A random variable

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1),$$

has a Student t-distribution with $\nu = n - 1$ degrees of freedom,we can write

$$P(t_{\frac{\alpha}{2}}(\nu) < T < t_{1-\frac{\alpha}{2}}(\nu)) = 1 - \alpha,$$

C.I. for the Mean in the Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. A random variable

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1),$$

has a Student t-distribution with $\nu = n - 1$ degrees of freedom,we can write

$$P(t_{\frac{\alpha}{2}}(\nu) < T < t_{1-\frac{\alpha}{2}}(\nu)) = 1 - \alpha,$$

we get

$$P\left(t_{\frac{\alpha}{2}}(\nu) < \frac{\overline{X} - \mu}{S}\sqrt{n} < t_{1-\frac{\alpha}{2}}(\nu)\right) = 1 - \alpha.$$

C.I. for the Mean in the Normal Distribution

$$P\left(t_{\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}} < \overline{X} - \mu < t_{1-\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}}\right) = 1 - \alpha,$$

C.I. for the Mean in the Normal Distribution

$$P\left(t_{\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}} < \overline{X} - \mu < t_{1-\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}}\right) = 1 - \alpha,$$

$$P\left(-\overline{X} + t_{\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}} < -\mu < -\overline{X} + t_{1-\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}}\right) = 1 - \alpha,$$

C.I. for the Mean in the Normal Distribution

$$P\left(t_{\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}} < \overline{X} - \mu < t_{1-\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}}\right) = 1-\alpha,$$

$$P\left(-\overline{X} + t_{\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}} < -\mu < -\overline{X} + t_{1-\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}}\right) = 1-\alpha,$$

$$P\left(\overline{X} - t_{1-\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}} < \mu < \overline{X} - t_{\frac{\alpha}{2}}(\nu)\frac{S}{\sqrt{n}}\right) = 1-\alpha.$$

C.I. for the Mean in the Normal Distribution

Quantiles of a t-distribution fulfil $t_{\frac{\alpha}{2}} = -t_{1-\frac{\alpha}{2}}$, we get for measured values $\overline{x}$ and s of random variables $\overline{X}$ and S

$$P\left(\overline{x} - t_{1-\frac{\alpha}{2}}(\nu)\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{1-\frac{\alpha}{2}}(\nu)\frac{s}{\sqrt{n}}\right) = 1 - \alpha.$$

C.I. for the Mean in the Normal Distribution

Quantiles of a t-distribution fulfil $t_{\frac{\alpha}{2}} = -t_{1-\frac{\alpha}{2}}$, we get for measured values $\overline{x}$ and s of random variables $\overline{X}$ and S

$$P\left(\overline{x} - t_{1-\frac{\alpha}{2}}(\nu)\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{1-\frac{\alpha}{2}}(\nu)\frac{s}{\sqrt{n}}\right) = 1 - \alpha.$$

A confidence interval for the parameter μ is

$$\overline{x} - t_{1-\frac{\alpha}{2}}(\nu)\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{1-\frac{\alpha}{2}}(\nu)\frac{s}{\sqrt{n}}.$$

C.I. for the Mean in the Normal Distribution

A right–sided confidence interval for μ is

$$\mu < \overline{x} + t_{1-\alpha}(\nu)\frac{s}{\sqrt{n}}.$$

C.I. for the Mean in the Normal Distribution

A right–sided confidence interval for μ is

$$\mu < \overline{x} + t_{1-\alpha}(\nu)\frac{s}{\sqrt{n}}.$$

A left–sided confidence interval for μ is

$$\mu > \overline{x} - t_{1-\alpha}(\nu)\frac{s}{\sqrt{n}}.$$

C.I. for the Variance in the Normal Distribution

Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$. The random variable

$$\chi^2 = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

has a χ^2-distribution with $\nu = n - 1$ degrees of freedom.

C.I. for the Variance in the Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. The random variable

$$\chi^2 = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

has a χ^2-distribution with $\nu = n - 1$ degrees of freedom. We can derive

$$P\left(\chi^2_{\frac{\alpha}{2}}(\nu) < \frac{n-1}{\sigma^2} S^2 < \chi^2_{1-\frac{\alpha}{2}}(\nu)\right) = 1 - \alpha,$$

C.I. for the Variance in the Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. The random variable

$$\chi^2 = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

has a χ^2-distribution with $\nu = n-1$ degrees of freedom. We can derive

$$P\left(\chi^2_{\frac{\alpha}{2}}(\nu) < \frac{n-1}{\sigma^2} S^2 < \chi^2_{1-\frac{\alpha}{2}}(\nu)\right) = 1-\alpha,$$

$$P\left(\frac{1}{\chi^2_{\frac{\alpha}{2}}(\nu)} > \frac{\sigma^2}{(n-1)S^2} > \frac{1}{\chi^2_{1-\frac{\alpha}{2}}(\nu)}\right) = 1-\alpha,$$

C.I. for the Variance in the Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. The random variable

$$\chi^2 = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

has a χ^2-distribution with $\nu = n-1$ degrees of freedom. We can derive

$$P\left(\chi^2_{\frac{\alpha}{2}}(\nu) < \frac{n-1}{\sigma^2} S^2 < \chi^2_{1-\frac{\alpha}{2}}(\nu)\right) = 1-\alpha,$$

$$P\left(\frac{1}{\chi^2_{\frac{\alpha}{2}}(\nu)} > \frac{\sigma^2}{(n-1)S^2} > \frac{1}{\chi^2_{1-\frac{\alpha}{2}}(\nu)}\right) = 1-\alpha,$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(\nu)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(\nu)}\right) = 1-\alpha.$$

C.I. for the Variance in the Normal Distribution

For a observed value s^2 of a random variable S^2 we obtain a confidence interval for σ^2

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(\nu)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(\nu)}.$$

C.I. for the Variance in the Normal Distribution

For a observed value s^2 of a random variable S^2 we obtain a confidence interval for σ^2

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(\nu)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(\nu)}.$$

A right–sided confidence interval for σ^2 is

$$\sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha}(\nu)}.$$

C.I. for the Variance in the Normal Distribution

For a observed value s^2 of a random variable S^2 we obtain a confidence interval for σ^2

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(\nu)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(\nu)}.$$

A right–sided confidence interval for σ^2 is

$$\sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha}(\nu)}.$$

A left–sided confidence interval for σ^2 is

$$\sigma^2 > \frac{(n-1)s^2}{\chi^2_{1-\alpha}(\nu)}.$$

C.I. for the Mean – Large Samples

Let $X_1, X_2, \dots, X_n$ be a random sample from any distribution with a mean μ and a finite variance σ^2. One of the result of the central limit theorem is that a random variable

$$U = \frac{\overline{X} - \mu}{S}\sqrt{n}$$

has for $n \to \infty$ ($n > 30$) approximately normal distribution $N(0, 1)$.

C.I. for the Mean – Large Samples

We can write

$$P\left(u_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{S}\sqrt{n} < u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

C.I. for the Mean – Large Samples

We can write

$$P\left(u_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{S}\sqrt{n} < u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

A confidence interval for a parameter μ is

$$\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \mu < \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}.$$

C.I. for the Mean – Large Samples

We can write

$$P\left(u_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{S}\sqrt{n} < u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

A confidence interval for a parameter μ is

$$\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \mu < \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}.$$

A right–sided confidence interval for μ is

$$\mu < \overline{x} + u_{1-\alpha}\frac{s}{\sqrt{n}}.$$

C.I. for the Mean – Large Samples

We can write

$$P\left(u_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{S}\sqrt{n} < u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

A confidence interval for a parameter μ is

$$\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \mu < \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}.$$

A right–sided confidence interval for μ is

$$\mu < \overline{x} + u_{1-\alpha}\frac{s}{\sqrt{n}}.$$

A left–sided confidence interval for μ is

$$\mu > \overline{x} - u_{1-\alpha}\frac{s}{\sqrt{n}}.$$

Determining the Sample Size

We are going to show the connection between the sample size and the accuracy of estimation. We focus on the confidence interval for μ.

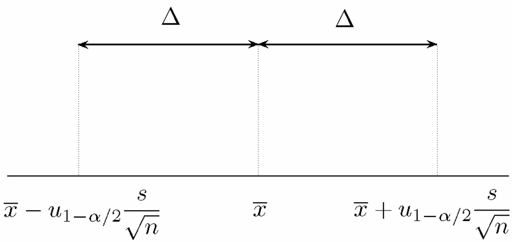
Definition

A **desired error margin** of estimation of μ is

$$\Delta = u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = u_{1-\frac{\alpha}{2}} \widehat{SE}.$$

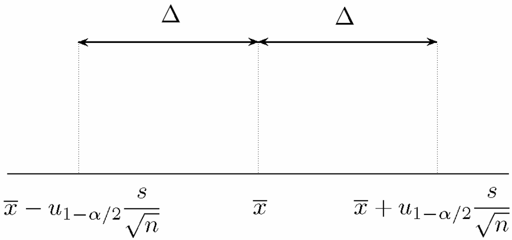
The desired error margin equals to the half of C.I.

Determining the Sample Size



How large must be the sample size n not to exceed Δ?

Determining the Sample Size



How large must be the sample size n not to exceed Δ?

$$P\left(|\overline{x} - \mu| < \Delta\right) = 1 - \alpha \Rightarrow u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \Delta \Rightarrow n > \left(u_{1-\frac{\alpha}{2}}\frac{s}{\Delta}\right)^2$$

C.I. for the Population Proportion – Large Samples

We would like to estimate the population proportion $\pi, 0 < \pi < 1$. Let us assume we have a random sample $X_1, \dots, X_n$, where $X_i, i = 1, \dots, n$ are Bernoulli random variables. An unbiased estimator of the population proportion is a sample proportion $\hat{\pi} = P$. A random variable

$$U = \frac{P - \pi}{\sqrt{\pi(1-\pi)/n}}$$

has for $n \to \infty$ approximately normal distribution $N(0, 1)$ (see the central limit theorem). For the observed value $\hat{\pi}$ is

$$P\left(u_{\frac{\alpha}{2}} < \frac{\hat{\pi} - \pi}{\sqrt{\hat{\pi}(1-\hat{\pi})/n}} < u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

C.I. for the Population Proportion – Large Samples

$$P\left(u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < \hat{\pi} - \pi < u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1 - \alpha,$$

C.I. for the Population Proportion – Large Samples

$$P\left(u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < \hat{\pi}-\pi < u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1-\alpha,$$

$$P\left(-\hat{\pi}+u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < -\pi < -\hat{\pi}+u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1-\alpha,$$

C.I. for the Population Proportion – Large Samples

$$P\left(u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < \hat{\pi}-\pi < u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1-\alpha,$$

$$P\left(-\hat{\pi}+u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < -\pi < -\hat{\pi}+u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1-\alpha,$$

$$P\left(\hat{\pi}-u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < \pi < \hat{\pi}-u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1-\alpha.$$

C.I. for the Population Proportion – Large Samples

$$P\left(u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < \hat{\pi}-\pi < u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1-\alpha,$$

$$P\left(-\hat{\pi}+u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < -\pi < -\hat{\pi}+u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1-\alpha,$$

$$P\left(\hat{\pi}-u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < \pi < \hat{\pi}-u_{\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}\right) = 1-\alpha.$$

Quantiles of standardized normal distribution fulfil $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$, the confidence interval for π is

$$\hat{\pi}-u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n} < \pi < \hat{\pi}+u_{1-\frac{\alpha}{2}}\sqrt{\hat{\pi}(1-\hat{\pi})/n}.$$

C.I. for the Population Proportion – Large Samples

By analogy we can derive
A right–sided confidence interval for π

$$\pi < \hat{\pi} + u_{1-\alpha}\sqrt{\hat{\pi}(1-\hat{\pi})/n}.$$

C.I. for the Population Proportion – Large Samples

By analogy we can derive
A right–sided confidence interval for π

$$\pi < \hat{\pi} + u_{1-\alpha}\sqrt{\hat{\pi}(1-\hat{\pi})/n}.$$

A left–sided confidence interval for π

$$\pi > \hat{\pi} - u_{1-\alpha}\sqrt{\hat{\pi}(1-\hat{\pi})/n}.$$