

Confidence intervals

1 General idea

The $p\%$ confidence interval $[a, b]$ for the parameter, \mathcal{Y} , is the interval that contains the true parameter $p\%$ of the time. In math speak,

$$a \leq \mathcal{Y} \leq b \quad (1)$$

For better or worse, you will find that the $p = 95\%$ confidence interval is used the majority of the time.

With real data, if you have enough of it, you can calculate the confidence interval by calculating the 2.5%-tile and 97.5%-tile, and then you have your confidence interval.

2 Relationship with α and the t-score and z-score

Recall from our discussion of cumulative probability density function F , that

$$Pr(a \leq x \leq b) = \int_a^b f(x) dx \quad (2)$$

$$Pr(a \leq x \leq b) = F(b) - F(a). \quad (3)$$

For a standard normal random variable z , we determined that

$$Pr(-1 \leq z \leq 1) = 68.27\% \quad (4)$$

$$Pr(-2 \leq z \leq 2) = 95.45\% \quad (5)$$

These are confidence intervals for z . The first is the 68.27% confidence interval, and the second is the 95.45% interval.

To find the 95% confidence interval for z , or $\alpha = 0.05$, we need to use our $F(z)$ table, and find the upper-bound b where

$$Pr(z) \geq b = 1 - \frac{\alpha}{2} = 0.975. \quad (6)$$

Recall, we look for 0.975 because we want the total area to add to 5% ($\alpha = 0.05$), and we want the interval to be symmetric - so, 2.5% ($\alpha/2$) on the upper and 2.5% ($\alpha/2$) on the lower.

If you look at your table, you will find that $z_{\alpha/2} = b = 1.96$. Thus, the 95% confidence bounds on z are

$$-z_{\alpha/2} \leq z \leq z_{\alpha/2} \quad (7)$$

$$-z_{0.025} \leq z \leq z_{0.025} \quad (8)$$

$$-1.96 \leq z \leq 1.96 \quad (9)$$

A useful way to interpret this confidence interval is that 95% of the time, the true z falls within -1.96 and 1.96, so, the probability of measuring a z outside of this range is 5% by random chance. This is often considered *small odds*. Thus, one often interprets this as interesting, and potentially implying that the measured z has something interesting going on, different from what might be expected from random chance.

For the standard t-statistic t , the confidence intervals are different than for z when the number in the sample $N \leq 30$. Also, recall that the t-statistic requires the degrees of freedom ν as well. However, these slight differences withstanding, calculating the confidence interval is exactly the same as for the z-score.

For example, say we have a random variable t with $N = 12$ (degrees of freedom $\nu = 11$). Using our table, we find that the 95% confidence interval ($\alpha = 0.05$) for t is

$$-t_{\alpha/2} \leq t \leq t_{\alpha/2} \quad (10)$$

$$-t_{0.025} \leq t \leq t_{0.025} \quad (11)$$

$$-2.20 \leq t \leq 2.20 \quad (12)$$

Likewise, for the 90% confidence interval with $\nu = 11$

$$-t_{\alpha/2} \leq t \leq t_{\alpha/2} \quad (13)$$

$$-t_{0.05} \leq t \leq t_{0.05} \quad (14)$$

$$-1.80 \leq t \leq 1.80 \quad (15)$$

3 Applying confidence intervals to non-standardized variables

The above examples, are a relatively straight-forward application of the t-score and z-score because we are dealing with standardized data. However, what if your data is not standardized? You have two options, (1) you can standardize your data and then do all of your analysis using standard normal variables (as above), or, (2) you can use a modified equation for the confidence interval that takes into consideration the real data's mean μ and standard deviation σ .

Recall that for the sample mean, we know that the standardized z-score is

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}} \quad (16)$$

So, if you want to calculate the 95% confidence interval using the z-statistic for the *true* distribution mean μ given your measured sample mean \bar{x} (N samples), you would write the following

$$\bar{x} - z_{0.025} \cdot \frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{x} + z_{0.025} \cdot \frac{\sigma}{\sqrt{N}} \quad (17)$$

Note that while this equation looks messy, it is really just the z-distribution shifted over by \bar{x} and scaled by σ/\sqrt{N} . As before $z_{\alpha/2} = 1.96$.

For the t-statistic, the true distribution mean has the following 95% confidence interval

$$\bar{x} - t_{0.025} \cdot \frac{\sigma}{\sqrt{N-1}} \leq \mu \leq \bar{x} + t_{0.025} \cdot \frac{\sigma}{\sqrt{N-1}} \quad (18)$$

Finally, since the normal distribution is symmetric, the above confidence intervals can be re-written more simply as

$$\mu = \bar{x} \pm z_{0.025} \cdot \frac{\sigma}{\sqrt{N}} \quad (19)$$

$$\mu = \bar{x} \pm t_{0.025} \cdot \frac{\sigma}{\sqrt{N-1}} \quad (20)$$