Week 02: New statistics (effect sizes, CIs, simulation)

Foundations & reproducibility (Part 2)

Bartosz Maćkiewicz

Confidence intervals are a useful way to convey the meaning of an experimental result that goes beyond the simple hypothesis test.

The sample mean ($\bar{X}$), as you already know, is an unbiased estimate of μ. When we have one specific estimate of a parameter, we call this a **point estimate**.

There are also **interval estimates**, which are attempts to set limits that have a high probability of encompassing the true (population) value of the mean [the mean (μ) of a whole population of observations]. What we want here are confidence limits on μ. These limits enclose what is called a confidence interval.

Confidence intervals around a sample statistic can be understood a alternative (or a complementary) practice to the hypothesis testing.

Handbooks and courses on statistics - and as a result also researchers in their practice - focus on the hypothesis testing. However, the confidence intervals gives us a much richer information. Null hypothesis significance testing allows us to answer the question about some particular hypothetical value of the population mean. Confidence interval on the other hand allows us to determine a whole range of values $(\mu L, \mu U)$, which with some probability includes the population mean.

For example, when we tested the null hypothesis about the moon illusion that $\mu = 1.00$ we rejected that hypothesis.

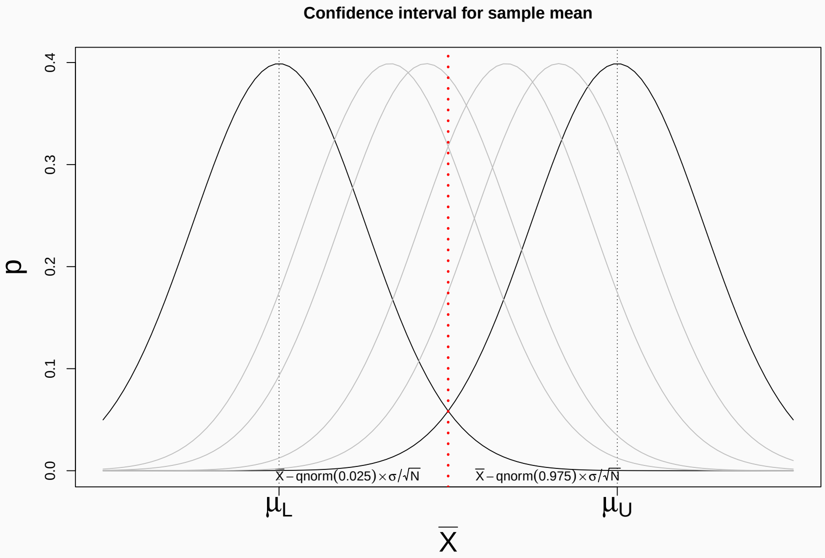
What if we tested the null hypothesis that $\mu = 0.93$? We would again reject that null.

We can keep decreasing the value of m to the point where we just barely do not reject H_0, and that is the smallest value of μ for which we would be likely to obtain our data at $p > .025$.

Then we could start with larger values of m (e.g., 2.2) and keep increasing μ until we again just barely fail to reject H_0. That is the largest value of μ for which we would expect to obtain the data at $p > .025$.

Now any estimate of μ that fell between those lower and upper limits would lead us to retain the null hypothesis.

Confidence Intervals: a graphical representation



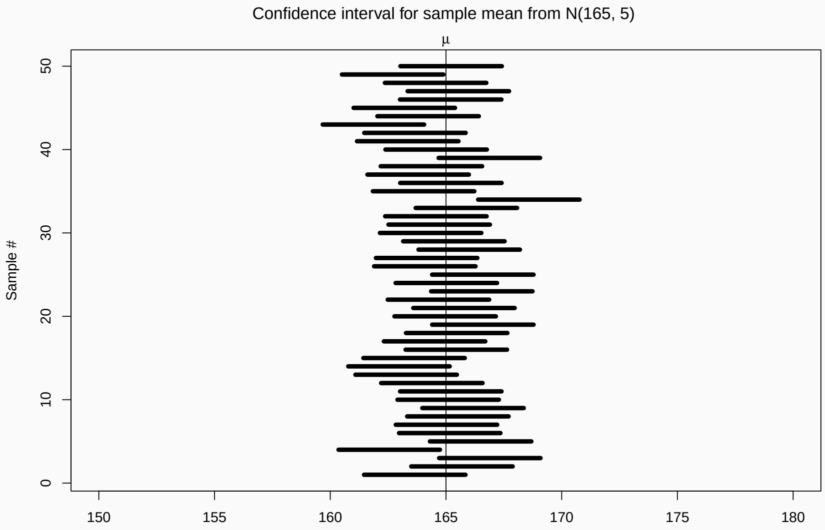
But What Is a Confidence Interval?

In a certain sense the confidence intervals are “reverse” of the null hypothesis significance testing: we are determining the interval of such values of the population mean for which we would not reject the null.

95% confidence interval is an interval which in 95% of cases **will** include the real population mean.

This code (next slide) draws k samples from a given population and for each of them draws a segment representing 95% confidence interval for the mean.

Confidence Intervals: a simulation study



Confidence Intervals: a simulation study

This is a simplified version of the simulation: instead of the plot it returns only one number.

Each iteration of the expression enclosed in the `replicate()` function returns a boolean value - TRUE of FALSE - depending on whether a confidence interval includes the population mean or not.

As a result we obtain a vector with 100 000 booleans. `sum` function converts it to zeroes and ones and returns a number of “hits”. Thanks to that we come to know how many times the confidence interval encompassed the population mean. By dividing this number by 100 000 we obtain a proportion of “hits”.

```
i <- 100000
sum(replicate(i, {
  sample <- rnorm(N, mu, sigma)
  mu >= mean(sample) - qnorm(.975) * sigma/sqrt(N) &
  mu <= mean(sample) + qnorm(.975) * sigma/sqrt(N)
})) / i
```

Confidence Intervals: a simulation study

An interactive version of this simulation is available on the course page:

- Week 02 → **Interactive apps (Quarto + Shiny): confidence intervals & power**

How does manipulation of N and σ influences the confidence intervals?

- The larger N, the narrower the confidence interval.
- The larger σ, the wider the confidence interval.
- The larger N, the more often the confidence interval includes population mean.
- The larger σ, the more often the confidence interval includes population mean.

Both the population standard deviation and the sample size influence standard error and as a consequence the precision of the estimation. This determines a width of the confidence interval. Those two factors however do not change confidence level (that's the point) because we fix it somewhat arbitrary (e.g. to 95)

Confidence Intervals: a simulation study with sample SD

Constructing confidence intervals we have assumed that we know the population standard deviation.

Such situations in fact sometimes occur, but more often than not we do not know this parameter. If we do not know the population mean, why would we know the population standard deviation?

Confidence Intervals: a simulation study with sample SD

What can we do? What would happen if we change the unknown population standard deviation to a standard deviation that we estimate from a sample?

Modify the simulation code and check what does happen. You can also explore this variant in the interactive app.

- The width of the confidence interval changes from sample to sample.
- The width of the confidence interval stays the same from sample to sample.
- Given its nominal 95% confidence level, the confidence interval includes the population mean slightly more often that it should.
- The confidence interval includes the population mean slightly more often than 50%

Confidence Intervals: a simulation study with sample SD

It appears that the sample standard deviation is a quite good estimator of the population standard deviation: confidence intervals work.

However, because that for each replication we calculate the standard deviation and it changes from the sample, the width of the confidence interval also changes: its “arm” instead of `qnorm(.975) * sigma/sqrt(N)` is `qnorm(.975) * sd(sample)/sqrt(N)`.

It might seem that the length of the “arm” sometimes would be overestimated and sometimes would be underestimated.

Meanwhile it turns out that by using the sample standard deviation we are systematically narrowing down the confidence interval whose real confidence level becomes lower than its nominal 95%. Do this simulation for $N = 10$ and $N = 100$ and check yourself.

Why is this the case?

Confidence Intervals: a simulation study with sample SD and t

If we do not know σ and rely on s, the t distribution is a better approximation than the normal distribution. You can change the code to use `qt` instead of `qnorm` or explore it in the interactive app.