Improving the interpretation of confidence and credible intervals

Rink Hoekstra*, Richard Morey**, & Eric-Jan Wagenmakers*

*University of Groningen // **Cardiff University

Interpreting Confidence intervals

Confidence procedure (CP) and Confidence intervals (CIs)

An X% CP for a parameter θ is a procedure that has an X% probability, in repeated sampling, of generating an interval containing the true value of θ, for all possible values of θ (Neyman, 1937). **A confidence interval is a realization from such a procedure.**

What does this mean for the interpretation of a single CI?

A CI is randomly drawn from all possible CIs. The confidence coefficient (X%) implies nothing about how likely *this particular interval* is to include the parameter.

Three common fallacies (from Morey et al., 2015):

A 95% CI **cannot** in general be interpreted as…

- … having a 95% probability of including θ (*fundamental confidence fallacy*)
- … indicating the precision of the parameter estimate (*precision fallacy*)
- … indicating the range of plausible values for θ (*plausibility fallacy*)

Nevertheless, researchers often make such mistakes:

Hoekstra et. al (2014) found that **115 out of 118 researchers (>97%) endorsed fallacious statements about CIs.** Researchers were no better than first-year students

CIs often promoted as a sound alternative for the significance test. Nevertheless, **they are not as easy to interpret as popularly thought.**

Textbooks: Part of the problem

*"The 95 percent confidence interval spans a segment on the horizontal axis that **we are 95 percent certain** contains the population mean." (Frankel & Wallen, 2009; p. 221)*

*"Confidence interval: An interval, with limits at either end, having a **specified probability of including the parameter** being estimated." (Howell, 2010)*

*"The interval is made up of numbers that are **the most believable values for the unknown parameter**, based on the data observed" (Agresti & Franklin, 2013; p. 352)*

"Confidence interval – An interval estimate, based on the sample statistic, that includes the population mean a certain percentage of the time, were we to sample from the population repeatedly" (Nolan & Heinzen, 2010)

How often?

We analyzed **23 regularly used textbooks** with respect to their definitions of CIs. The frequency of fallacies was determined by how often the common fallacies were included.

Fallacy	Frequency* (%)
Fundamental confidence fallacy	16 (70%)
Plausibility fallacy	6 (26%)
Precision fallacy	2 (9%)
Correct definition	3 (13%)

** The percentages do not add to 100%, since some books included several definitions. Two books included only correct definitions*

It is not surprising that researchers misinterpret CIs: **students are taught the fallacies from the start of their education.**

Solution: Contrast frequentism and Bayes

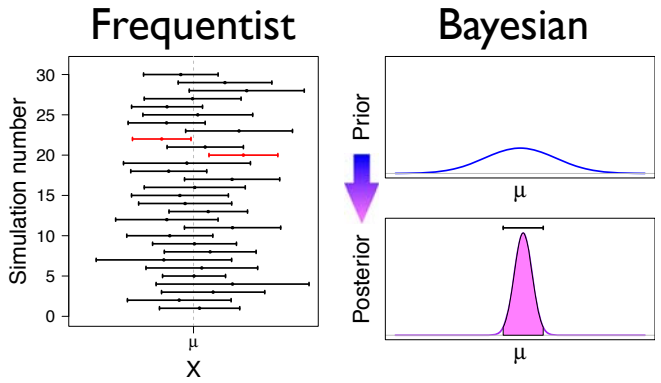
Without correct textbooks and proper training, CIs will remain frequently misunderstood. The most common fallacy (the fundamental confidence fallacy) is to interpret CIs as if they were credible intervals.

"If you can find even one [published analysis] where the confidence interval is not interpreted as a credible interval, then I will eat your hat." (Briggs, 2012)

Killing two birds with one stone

We should teach our students CIs *and* credible intervals. Arguably, contrasting the two would improve the understanding of both techniques (e.g., Haller & Kraus, 2002).

Concept	Logic	Teaching tool
Confidence interval	Frequentist, pre-data	Long-run simulations
Credible interval	Bayesian, post-data	Bayes' theorem



Teaching both CIs and credible intervals would **clarify confidence intervals**, help **prepare students** for future contact with Bayesian ideas, and generally **improve their statistical reasoning.**

More information on website

http://goo.gl/dJRGnY