

1

2

3

4

5

6

7

Abstract

8

9 In this second revision we are responding to the last issue you raised concerning our
10 incorrect statement that 50% of the expected correlations fall below the observed correlation.
11 We are extremely grateful you pointed out this incorrect statement, and have corrected it as
12 explained below. We have also made some minor changes after carefully re-reading the
13 manuscript that we detail at the end of this letter. We hope you think our submission is
14 ready to be accepted for publication in Meta-Psychology.

Reply to Reviews

Dear editor,

thank you for your comments on our revised manuscript. You rightly noted that our statement that “There is always a 50% probability of observing a correlation smaller or larger than the true correlation” was not correct. Our statement was true for the corresponding z-value when transforming the correlation while calculating confidence intervals, but not for correlations, which are not symmetrically distributed. We have deleted the following:

In other words, if the true effect size is the same as the equivalence bound, it is equally likely to find an effect more extreme than the equivalence bound, as it is to observe an effect that is less extreme than the equivalence bound.

We have made a number of small changes. First, we clearly explain that, although the z-distribution is symmetrical, the correlation is not. We note how the transformation has the goal to yield accurate error rates (which is important in the TOST procedure) on page 18, lines 250-255.

If the observed correlation falls exactly on the equivalence bound the p -value for the equivalence test is 0.5. In the equivalence test for correlations the p value is computed based on a z-transformation which better controls error rates (Goertzen & Cribbie, 2010). This transformation is computed as follows, where r is the observed correlation and ρ is the theoretical correlation under the null:

$$z = \frac{\frac{\log(\frac{1+r}{1-r})}{2} - \frac{\log(\frac{1+\rho}{1-\rho})}{2}}{\sqrt{\frac{1}{n-3}}}$$

Because the z -distribution is symmetric, the probability of observing the observed or more extreme z -score, assuming the equivalence bound is the true effect size, is 50%. However, because the r distribution is not symmetric, this does not mean that there is always a 50% probability of observing a correlation smaller or larger than the true correlation.

Second, we have added an explicit statement that in the extreme case, the probability of observing a correlation smaller than the equivalence bound is not 50%, but in our example, 36% on page 19, lines 271-275.

In the most extreme case (i.e., a sample size of 4, and equivalence bounds set to $r = -0.99$ and 0.99 , with a true correlation of 0.99) 97.60% of the confidence interval overlaps with the equivalence range, even though in the long run only 36% of the correlations observed in the future will fall in this range.

We again want to sincerely thank you for catching this error before our paper was published.

Additional Minor changes

We replaced some values in the document that were hard-coded with the R scripts that produced these values, making the document (as far as we can see) 100% reproducible. (note that this is not noticeable in the PDF)

We wrote we “replicated” Blume, but we now say we “build on” Blume - all our examples use a mean of 145, where upon rereading Blume et al we

noticed they use a mean of 146 - an inconsequential difference, but instead of changing all our examples to a mean of 146, we just changed the word “replicated” to “build on”.

We removed “if too few observations are collected” which was redundant with the “note that when sample sizes are small” earlier in the same sentence.

We added in Figure 6 that the standard deviation for the visualized example is 2.

We added some clarification when describing Figure 12, replacing “the two second generation p -values associated with the observed correlations at $r = -0.45$ and $r = 0.45$ are larger than 50%” with “the proportion of the confidence interval that overlaps with the equivalence range is larger than 50% when the observed correlations are $r = -.45$ and $r = .45$, meaning that the two second generation p -values associated with these correlations are larger than 50%.”

In the conclusion, we noted that p -values from TOST are more consistent, and easier to interpret. Upon re-reading our manuscript, we chose to delete the “and easier to interpret” part of that sentence, since this is an empirical claim, and we have no data to support it, and is not relevant for the points we make.

References

- 75
76 Goertzen, J. R., & Cribbie, R. A. (2010). Detecting a lack of association: An equivalence
77 testing approach. *British Journal of Mathematical and Statistical Psychology*, 63(3),
78 527–537. doi:10.1348/000711009X475853