

REVIEWER A

Paper reviewed: On using a pilot sample variance for sample size determination in the detection of difference between two means: Power consideration.

The paper previously cited is an interesting research work that explains the effect of the use of sample variance in sample size determination, and the adequacy of adjusting the sample variance in order to determine the sample size in the detection of difference between two means. This work has a good derivation of the statistical reasoning and with some data to support that statistical derivation. In consequence I think that this work could be published in *Psicológica*. However, in order to be published in this journal, I consider that some aspects should be revised.

One of the first things to say is that in the keywords it isn't the term sample size and I think that is an important one for the title of the paper and the development of the work presented.

The reference: Kraemer, Mintz, Noda, Tinklenberg and Yesavage (2006) appears the first time in the paper cited as Kraemer et al. (2006) (p. 2), but following APA guidelines, this first time should appear with the surnames of all the authors because have less than 6 authors.

Another aspect to consider is related with the order of the paper, maybe it could be interesting to differentiate a method section, and in this section do more explanation about how the data analyzed in tables 1 to 3 were generated, because the first time that a reference appears to those tables not all the statistical expressions were explained (for example equation 11). Moreover, in this point I consider important to point out the necessity of a further explanation about how the data analyzed in those tables were generated, basically the sample variance ($\hat{\sigma}_p^2$), it is derived from expression 6? Or is fixed to 100? as it seems in appendix A and B (SIGSQP=100).

I think that the example of Aronson et al. (1999) is irrelevant in this paper and maybe the example explained in the pp. 16-17 could be more illustrative before the explanation of tables 1 and 3.

Finally, maybe it's important to cite more recent works related with this one, as for example:

Guo, J.H., Chen, H.J., & Luh, W.M. (2011). Sample size planning with the cost constraint for testing superiority and equivalence of two independent groups. *British Journal of Mathematical and Statistical Psychology*, 64 (3), 439-461.

Jan, S-L., & Shieh, G. (2011). Optimal sample sizes for Welch's test under various allocation and cost considerations. *Behavior Research Methods*, 43(4), 1014-1022.

Lawson, C.A., & Fisher, A.V. (2011). It's in the sample: The effects of sample size and sample diversity on the breadth of inductive generalization. *Journal of Experimental Child Psychology*, 110(4), 499-519.

Maribel Peró
University of Barcelona (Spain)

REVIEWER B

This paper is a valuable contribution to the literature concerning determination of sample size needed for suitable power in two-sample t-tests of differences in means. It combines two existing approaches to the problem into a unified development that seems to have promise. The review of previous literature is good, and the explanations of previous findings by other authors are easy to follow. The mathematical development and symbolism is nicely presented. Most of the writing is very good. In my opinion, the problem of sample size determination is one in which new research is welcome. I have a few concerns about the format and some of the writing.

The material is presented in a series of very long paragraphs that are somewhat difficult to read. Would the text perhaps be more readable if it were presented in shorter paragraphs that cover smaller “chunks” of material?

There seems to be unnecessary repetition of material at some places throughout the paper. For example, in the bottom paragraph on page 1, the name and date Browne (1995) occurs three times in a row (6th, 4th, and 2nd lines from the bottom of the page). Perhaps this could be modified so that (1995) occurs just once and then “he showed ..” or something similar is said in the following sentences. The same problem happens in the second paragraph on page 2.

In the title of the paper, I would suggest saying either “detection of a difference between two means ...” or “detection of differences between two means”

The “keywords” on the title page mentions distributions but does not give much information about the content of the paper. How about: sample size determination, t-test, pilot sample, etc.?

There are some errors in grammar at various places, and unfortunately I have not kept a list. For example, at the end of the second paragraph on page 10, the sentence contains “... both Browne and Kieser and Wassmer did not conduct ...” It is better to say “... neither Browne nor Kieser and Wassmer conducted”

At the bottom of page 11: Perhaps replace “... an alternative criterion for sample size determination is to ensure ...” with “... an alternative criterion for sample size determination is agreement of expected actual power with planned power.”

Should the title of the last section be “Concluding Remarks” instead of “Concluding Remark”?

Should the equations at the bottom of page 8 and at the top of page 12 be numbered? If not, should they be included inline in the text along with the equations that are not numbered?

I do not know if it is necessary to include computer programs in SAS/IML code as Appendix A and Appendix B, in light of the fact that people doing similar research may be using different platforms. But that may be a matter of the policy of the journal.