Psychonomic Bulletin & Review - Rejection letter

25-Sep-2018

Dear Dr. Pastore:

After reading your Manuscript PBR-Otr-18-017 entitled "Measuring similarities among samples: a distribution-free overlapping index" that you submitted to Psychonomic Bulletin & Review, I have decided not to send out your manuscript for review.

Although I find the topic of your paper interesting and in principle PBR would welcome contributions like yours on distribution measures of effect, I believe your paper

1. does not seem to address how to test hypotheses using your index (but your abstract introduces your paper in the context of hypothesis testing), particularly directional hypotheses; for instance, this is well possible with Cliff's delta, another distribution-free measure
2. calculates the index using kernel estimation, but I believe an approach solely based on data is preferable
3. is a bit too statistical in nature, with still a significant gap from theory to application by potential readers
4. does not yet provide convincing useful practical applications in my opinion (see my first point); hypothesis testing (directional or not), standard error and confidence intervals of the measure, reasonable effect size values.

Sorry to bring the bad news, particularly because I value non-parametric or distribution-free statistics.

ACTION: I am rejecting the manuscript. I think the concerns are sufficient to preclude publication in this journal. I hope the reviewer comments will be helpful in your future work.

Thank you for considering Psychonomic Bulletin & Review for the publication of your research.

Sincerely, . . .

Action Editor, Psychonomic Bulletin & Review

Our response:

Dear ...,

Thank you for directly rejecting our paper avoiding us to wait for revision; We have just some comments as a response of yours:

1. In the abstract we wrote: "... we illustrate the use of a distribution-free overlapping measure as an alternative way to quantify sample differences ..." referring to an effect size approach and not to a hypothesis testing context. **This index is an effect size and not a statistic for testing**; from our conclusions: *This index can be considered as an alternative measure of classical effect size indices, such as for example, Cohen's d, Cohen's U or McGraw and Wong's CL.*
2. What does it means: "I believe an approach solely based on data is preferable"?? Kernel estimation **is based on data**; its formula is:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \tag{0.1}$$

   where $x_i (i = 1, \ldots, n)$ are observed data!
3. The idea of overlapped area is too statistical and too difficult to understand. Maybe You think that it is more easy understanding the idea of p-value and statistical testing? See for example: Aschwanden (2015, 2016); Baker (2016); Bakker and Wicherts (2011); Berger and Sellke (1987); Chavalarias, Wallach, Li, and Ioannidis (2016); Cohen (1994); Gelman (2013, 2016); Gelman and Carlin (2017); Gelman and Stern (2006); Gigerenzer, Krauss, and Vitouch (2004); Goodman (2008); Greenland et al. (2016); Hubbard and Lindsay (2008); Krantz (1999); Nuzzo (2014); Resnick (2019); Rothman (2014); Wagenmakers (2007); Wasserstein, Lazar, et al. (2016); Verdam, Oort, and Sprangers (2014); Ziliak and McCloskey (2008); Zhu (2012).
4. We are sorry to only present four different examples in the paper, and cite five papers using the index only (Lionetti, Mastrotheodoros, & Palladino, 2018; Marci et al., 2018; Altoè, D'Amore, & Scalfari, 2018; Pluess et al., 2018; Lionetti, Aron, et al., 2018). More recently, the index was used (and cited) in other papers: Bean, Huffaker, and Migliaccio (2018); Brauns, Brabender, Gehre, Rinke, and Weitere (2019); D'Amario et al. (2019); Drury et al. (2019); Ford, Waldner, Sanchez, and Bharadwaj (2019); Giofrè, Pastore, Cornoldi, and Toffalini (2019); Giuntoli et al. (2019); McMunn (2018); Schank et al. (2017). It seems that there are no useful practical applications.

NOTE: Since You decided to not send manuscript for review, how can we consider "reviewer's comments", as You write in ACTION?

Sorry for delay in our response.

Sincerely
MP & AC

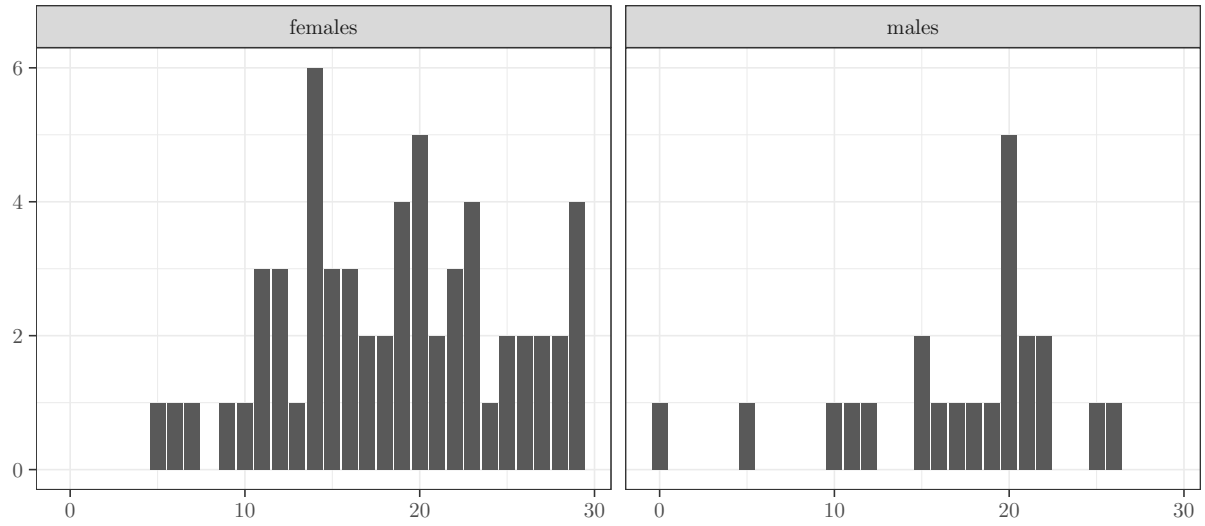

Fig. 0.1. Exemplary case study: Frequency distributions of exam scores for males and females.

P.S. just for fun

If you want to test something, consider this example:

Let's suppose to have a sample of 81 students'scores on a given exam (22 males and 59 females). Figure 0.1 shows the frequency distributions of the scores for females (left panel) and males (right panel).

Your idea is that the difference in the scores distributions between males and females is relevant only if the overlapped density area is lower than .9[1]. Formally: $H_0 : \eta > .9$ (i.e., there is no relevant difference) vs $H_1 : \eta < .9$[2].

You can proceed in such way:

```
library( overlapping )
dataList <-
    list( males = c( 12, 20, 5, 20, 25, 15, 18, 26, 20, 20, 0, 15, 22, 21, 11, 10, 17,
                 22, 16, 21, 19, 20 ),
          females = c( 6, 11, 12, 28, 14, 19, 25, 21, 25, 18, 17, 21, 16, 19, 23, 29,
                14, 23, 14, 29, 5, 11, 13, 15, 27, 28, 23, 14, 10, 24, 15, 11, 20, 18,
                14, 29, 22, 22, 12, 20, 19, 20, 17, 15, 12, 26, 22, 27, 14, 29, 20, 23,
                 19, 16, 16, 26, 7, 9, 20 ))
```

First, bootstrap the data and compute the index repeatedly:

```
> out <- boot.overlap( dataList )
```

Then, evaluate the estimated 95% confidence interval:

```
> quantile( out$OVboot_dist, probs = c(.025,.975) )

      2.5%     97.5%
0.3663276 0.8009690
```

Given that the target value .9 is outside the interval, you can reject your null hypothesis.

[1] Obviously, you can change this value depending on your specific purposes.
[2] Should be this a directional hypothesis?

References

Altoè, G., D'Amore, G., & Scalfari, F. (2018). Skulls and transvariation. In M. D. Bacco & F. Scalfari (Eds.), *Biostat at 25 invited essays in theoretical, biomedical and social statistics.* Edizioni ETS, PISA.

Aschwanden, C. (2015, Nov.). *Not even scientists can easily explain p-values.* https://fivethirtyeight.com/. Retrieved from `https://fivethirtyeight.com/features/not-even-scientists-can-easily-explainp-values/`

Aschwanden, C. (2016, Mar.). *Statisticians found one thing they can agree on: Its' time to stop misusing p-values.* https://fivethirtyeight.com/. Retrieved from `https://fivethirtyeight.com/features/statisticians-found-one-thing-they-can-agree-on-its-time-to-stop-misusing-p-values/`

Baker, M. (2016). Statisticians issue warning over misuse of p values. *Nature News*, *531*(7593), 151.

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*(3), 666–678.

Bean, E. Z., Huffaker, R. G., & Migliaccio, K. W. (2018). Estimating field capacity from volumetric soil water content time series using automated processing algorithms. *Vadose Zone Journal*, *17*(1).

Berger, J. O., & Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association*, *82*(397), pp. 112–122.

Brauns, M., Brabender, M., Gehre, M., Rinke, K., & Weitere, M. (2019). Organic matter resources fuelling food webs in a human-modified lowland river: importance of habitat and season. *Hydrobiologia*, 1–11.

Chavalarias, D., Wallach, J. D., Li, A. H. T., & Ioannidis, J. P. (2016). Evolution of reporting P values in the biomedical literature, 1990-2015. *Jama*, *315*(11), 1141–1148.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist*, *49*, 997–1003.

D'Amario, S. C., Rearick, D. C., Fasching, C., Kembel, S. W., Porter-Goff, E., Spooner, D. E., ... Xenopoulos, M. A. (2019). The prevalence of nonlinearity and detection of ecological breakpoints across a land use gradient in streams. *Scientific reports*, *9*(1), 3878.

Drury, J. P., Anderson, C. N., Cabezas Castillo, M. B., Fisher, J., McEachin, S., & Grether, G. F. (2019). A general explanation for the persistence of reproductive interference. *The American Naturalist*, *194*(2), 000–000.

Ford, L., Waldner, C., Sanchez, J., & Bharadwaj, L. (2019). Risk perception and human health risk in rural communities consuming unregulated well water in saskatchewan, canada. *Risk Analysis*.

Gelman, A. (2013). Commentary: P values and statistical practice. *Epidemiology*, *24*(1), 69–72.

Gelman, A. (2016). The problems with p-values are not just with p-values. *The American Statistician*, *70*, 10.

Gelman, A., & Carlin, J. (2017). Some natural solutions to the p-value communication problem—and why they won't work. *Journal of the American Statistical Association*, *112*(519), 899–901.

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *American Statistician*, *60*, 328–331. doi: 10.1198/000313006X152649

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual. In D. Kaplan (Ed.), *The sage handbook of quantitative methodology for the social sciences* (pp. 391–408). Sage Thousand Oaks, CA.

Giofrè, D., Pastore, M., Cornoldi, C., & Toffalini, E. (2019). Lumpers vs. splitters: Intelligence in children with specific learning disorders. *Intelligence*, *76*, 101380.

Giuntoli, L., Marchetti, I., Panzeri, A., Spoto, A., Vidotto, G., & Caudek, C. (2019). Measuring cognitive vulnerability to depression: Further evidence on the factorial and predictive validity of negative cognitive style. *Journal of behavior therapy and experimental psychiatry*, *65*, 101479.

Goodman, S. (2008). A dirty dozen: twelve p-value misconceptions. In *Seminars in hematology* (Vol. 45, pp. 135–140).

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*, *31*(4), 337–350.

Hubbard, R., & Lindsay, R. M. (2008). Why P values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology*, *18*(1), 69–88.

Krantz, D. H. (1999). The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association*, *94*(448), 1372–1381.

Lionetti, F., Aron, A., Aron, E. N., Burns, G. L., Jagiellowicz, J., & Pluess, M. (2018). Dandelions, tulips and orchids: evidence for the existence of low-sensitive, medium-sensitive and high-sensitive individuals. *Translational psychiatry*, *8*(1), 24. doi: 10.1038/s41398-017-0090-6

Lionetti, F., Mastrotheodoros, S., & Palladino, B. E. (2018). Experiences in close relationships revised child version (ecr-rc): Psychometric evidence in support of a security factor. *European Journal of Developmental Psychology*, *15*(4), 452–463. doi: 10.1080/17405629.2017.1297228

Marci, T., Lionetti, F., Moscardino, U., Pastore, M., Calvo, V., & Altoé, G. (2018). Measuring attachment security via the Security Scale: Latent structure, invariance across mothers and fathers and convergent validity. *European Journal of Developmental Psychology*, *15*(4), 481–492. doi: 10.1080/17405629.2017.1317632

McMunn, M. S. (2018). Predicted asymmetrical effects of warming on nocturnal and diurnal ectotherms. *bioRxiv*, 441501.

Nuzzo, R. (2014). Scientific method: statistical errors. *Nature News*, *506*(7487), 150.

Pluess, M., Assary, E., Lionetti, F., Lester, K. J., Krapohl, E., Aron, E. N., & Aron, A. (2018). Environmental sensitivity in children: Development of the highly sensitive child scale and identification of sensitivity groups. *Developmental psychology*, *54*(1), 51. doi: 10.1037/dev0000406

Resnick, B. (2019). *800 scientists say it's time to abandon "statistical significance".* https://www.vox.com/. Retrieved from `https://www.vox.com/latest-news/2019/3/22/18275913/statistical-significance-p-values-explained`

Rothman, K. J. (2014). Six persistent research misconceptions. *Journal of general internal medicine*, *29*(7), 1060–1064.

Schank, C. J., Cove, M. V., Kelly, M. J., Mendoza, E., O'Farrill, G., Reyna-Hurtado, R., ... others (2017). Using a novel model approach to assess the distribution and conservation status of the endangered baird's tapir. *Diversity and Distributions*, *23*(12), 1459–1471.

Verdam, M. G., Oort, F. J., & Sprangers, M. A. (2014). Significance, truth and proof of p values: reminders about common misconceptions regarding null hypothesis significance testing. *Quality of Life Research*, *23*(1), 5–7.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, *14*(5), 779–804.

Wasserstein, R. L., Lazar, N. A., et al. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, *70*(2), 129–133.

Zhu, W. (2012). Sadly, the earth is still round ($p < 0.05$). *Journal of Sport and Health Science*, *1*(1), 9–11.

Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance.* Ann Arbor, MI: University of Michigan Press.