A Systematic Review on Randomization and Permutation Tests in the Educational and Behavioral Sciences

Ming Huo
ming.huo@ped.kuleuven.be

Mieke Heyvaert
mieke.heyvaert@ped.kuleuven.be

Wim Van den Noortgate
wim.vandennoortgate@kuleuven-kortrijk.be

Patrick Onghena
patrick.onghena@ped.kuleuven.be

Katholieke Universiteit Leuven
Versaliusstraat 2, Leuven, Belgium

BACKGROUND

In many educational and behavioral studies, the assumptions of the classical parametric hypothesis tests (e.g., normality, homogeneity of variance, independence of errors) are often considered implausible [1, 2, 3, 4]. An alternative to the traditional statistical methods that does not rely on such strict assumptions is to use a randomization test (RT) or permutation test (PT). RTs and PTs constitute a set of distribution-free statistical tests that calculate the probability of getting a value as extreme or more extreme than an obtained value of a test statistic under a null hypothesis by recalculating the test statistic for all or many permutations of the data. They do not depend on a specific error distribution, and they use the original values of the data instead of the ranks. RTs and PTs were proposed in the early twentieth century, but were not widely used until much later. This is mostly because (a) they were too computationally intensive, (b) their applicability was limited to simple scenarios, (c) and they could be replaced by the available classical nonparametric tests based on ranks [5].

Over the past two decades, RTs and PTs have received much attention in the educational and behavioral sciences, with an accompanying multitude of applications. However, a general overview of the theoretical development and applications of RTs and PTs in the educational and behavioral sciences is still lacking, mainly because articles on RTs and PTs are spread out over the literature. Accordingly, a systematic review is called for.

Aim

There are three objectives of this paper: 1) to provide an overview of the theoretical development of RTs and PTs and summarize several key areas of theoretical research; 2) to summarize several active areas of educational and behavioral applications of RTs and PTs; 3) to identify the experimental designs in which RTs and PTs have been applied.

METHODS

In order to realize the above-mentioned three objectives, the databases ERIC, PsycINFO and Web of Science were searched for articles on RTs and PTs, published in the educational and behavioral journals between 1989 and 2008. Searches were performed by using the keywords [randomization tests] and [permutation tests]. Articles written in languages other than English were excluded. Abstracts were read to identify the relevant articles. This review was carried out in two tracks: a theoretical track and an application track. The first track intends to summarize the theoretical evolution of RTs and PTs. The second track focuses on applications of RTs and PTs and intended to summarize the experimental designs as well as areas in which these methods have been applied.

RESULTS

124 articles were identified, which included 87 theoretical articles and 37 application articles. In the theoretical articles, seven major topics were identified: a) introduction to and instruction of RTs and PTs (e.g., [3], [6]); b) algorithms, programs and software for RTs and PTs (e.g., [7], [8]); c) RTs and PTs for group designs (e.g., [9], [10]); d) RTs for single-case designs (e.g., [11], [12]); e) multivariate RTs and PTs (e.g., [13], [14]); f) performance of RTs and PTs (e.g., [1], [15]); and g) advanced topics (e.g., [16], [17]). In the application articles, RTs and PTs were applied in the following three active areas: a) models of vocational interest structure (e.g., [18], [19]); b) event-related potential (ERP) and electroencephalogram (EEG) (e.g., [20], [21]); and c) animal behaviors (e.g., [22], [23]). Meanwhile, RTs and PTs were found to apply for the following experimental designs: a) one-group design and association analysis (e.g., [24]); b) paired-group designs

(e.g., [25]); c) multivariate designs and multiple comparisons (e.g., [14]); d) distance matrices analysis (e.g., [22]); and e) single-case design (e.g., [26]).

DISCUSSION

From a theoretical perspective, methodological possibilities of RTs and PTs have been extended over the last 20 years. Among those articles, a majority of them focused on the implementation of RTs and PTs, RTs for single-case design and RTs and PTs for group designs, whereas multivariate RTs and PTs as well as advanced topics were less discussed.

From an application perspective, RTs and PTs have been employed in several active areas of behavioral research. However, compared with the range of topics discussed in the theoretical part, both the application areas and types of experimental design are limited. Some applications of RTs and PTs can be attributed to the theoretical development of RTs and PTs in the educational and behavioral sciences, such as the area of models of vocational interest structure. Among those application articles, RTs and PTs were used as alternatives to parametric hypothesis tests in order to avoid the stringent distributional assumptions.

CONCLUSIONS

The methodological possibilities of RTs and PTs have been extended substantially during the last years, RTs and PTs are not only applied in simple (e.g., two-group design) but also in complex contexts (e.g., multivariate designs). Moreover, recent developments of RTs and PTs have widened their application scope. RTs and PTs are employed in some more complex and exciting areas, such as ERP and EEG, since RTs and PTs are powerful tools to solve the multiple comparisons problem in these areas. While many theoretical articles were published in educational journals, more applications were published in behavioral journals.

REFERENCES

1. Adams, D. C., & Anthony, C. D. Using randomization techniques to analyse behavioural data. *Animal Behaviour*, *51* (1996), 733–738.
2. Edgington, E. S., & Onghena, P. *Randomization tests* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC (2007).
3. Hunter, M. A., & May, R. B. Statistical testing and null distributions: What to do when samples are not random. *Canadian Journal of Experimental Psychology, 57* (2003), 176–188.
4. Manly, B. F. J. *Randomization, bootstrap, and Monte Carlo methods in biology* (2nd ed.). New York: Chapman & Hall, 1997.
5. Welch, W. J. Construction of permutation tests. *Journal of the American Statistical Association*, ***85*** (1990), 693–698.
6. Bear, G. Computationally intensive methods warrant reconsideration of pedagogy in statistics. *Behavior Research Methods, Instruments & Computers*, ***27*** **(**1995), 144-147.
7. Cai, L. Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation. *Behavior Research Methods, **38*** (2006), 51–59.
8. Chen, R. S., & Dunlap, W. P. SAS procedures for approximate randomization tests. *Behavior Research Methods, Instruments & Computers, 25* (1993), 406–409.
9. Mielke, P. W., & Berry, K. J. Permutation tests for common locations among samples with unequal variances. *Journal of Educational and Behavioral Statistics*, *19* (1994), 217–236.
10. Mielke, P. W., & Berry, K. J. Data-dependent analyses in psychological research. *Psychological Reports*, *91* (2002), 1225–1234.
11. Onghena, P., & Edgington, E. S. Randomization tests for restricted alternating treatments designs. *Behaviour Research and Therapy*, *32* (1994), 783–786.
12. Onghena, P., & Edgington, E. S. Customization of pain treatments: Single-case design and analysis. *Clinical Journal of Pain*, *21* (2005), 56–68.
13. Blair, R. C., Higgins, J. J., Karniski, W., & Kromrey, J. D. A study of multivariate permutation tests which may replace Hotelling's T^2 test in prescribed circumstances. *Multivariate Behavioral Research*, *29* (1994), 141–163.
14. Blair, R. C., & Karniski, W. An alternative method for significance testing of wave-form difference potentials. *Psychophysiology*, *30* (1993), 518–524.
15. Peres-Neto, P. R., & Olden, J. D. Assessing the robustness of randomization tests: Examples from behavioural studies. *Animal Behaviour*, *61* (2001), 79–86.
16. Johnston, J. E., Berry, K. J., & Mielke, P. W. Permutation tests: Precision in estimating probability values. *Perceptual and Motor Skills, 105* (2007), 915–920.
17. Manly, B. F. J. Comments on a note on permutation tests of significance for multiple regression coefficients by long, et al. *Psychological Reports*, *101* (2007), 1041–1042.
18. Hedrih, V. Structure of vocational interests in Serbia: Evaluation of the spherical model. *Journal of Vocational Behavior*, *73* (2008), 13–23.
19. Yang, W. W., Stokes, G. S., & Hui, C. H. Cross-cultural validation of Holland's interest structure in Chinese population. *Journal of Vocational Behavior*, *67* (2005), 379–396.

20. Maris, E. Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, *41* (2004), 142–151.

21. Henderson, L. M., Yoder, P. J., Yale, M. E., & McDuffie, A. Getting the point: Electrophysiological correlates of protodeclarative pointing. International *Journal of Developmental Neuroscience*, *20* (2002), 449–458.

22. Mitani, J. C., & Brandt, K. L. Social-factors influence the acoustic variability in the long-distance calls of male chimpanzees. *Ethology*, *96* (1994), 233–252.

23. Sendova-Franks, A., & Franks, N. Spatial relationships within nests of the ant Leptothorax-unifasciatus (Latr) and their implications for the division-of-labor. *Animal Behaviour*, *50* (1995), 121–136.

24. Boyd, J. P., Fitzgerald, W. J., & Beck, R. J. Computing core/periphery structures and permutation tests for social relations data. *Social Networks*, *28* (2006)., 165–178.

25. Murrihy, R., & Byrne, M. K. Training models for psychiatry in primary care: A new frontier. *Australasian Psychiatry*, *13* (2005), 296–301.

26. Cicchetti, D., Rosenheck, R., Showalter, D., Charney, D., & Cramer, J. Interrater reliability levels of multiple clinical examiners in the evaluation of a Schizophrenic patient: Quality of life, level of functioning, and neuropsychological symptomatology. *Clinical Neuropsychologist*, *13* (1999), 157–170.