Testing for single-case designs by combined permutation tests with multivariate ordinal data

Rosa Arboretti Giancristofaro * Riccardo Ceccato † Luigi Salmaso ‡

Abstract

Single-case designs are experimental designs widely used in behavioral sciences, which involve only a single entity under observation.
Traditionally these designs are analyzed through visual analysis techniques, but several different approaches have emerged.
Among these approaches, permutation tests, also called randomization tests, deserve particular attention, because of the desirable property of being distribution free.
In addition, they can be used to analyze different type of data and both univariate and multivariate problems.
For this reason, we propose an extension of permutation tests to the analysis of a particular type of single-case designs, the completely randomized single-case designs, when the outcome is multivariate and ordinal.
A simulation study is hence performed to evaluate the power of this testing procedure and a real application is also provided.

Key Words: Non-parametric tests, Permutation tests, Single-case designs, Multivariate tests, NPC tests

1. Introduction

Single-case designs (Onghena and Edgington, 2005; Heyvaert and Onghena, 2014) are experimental designs in which a single entity is observed repeatedly during a certain period of time under different levels of a treatment variable, in order to assess the effect of this variable on the specific unit of interest.
These designs have been widely used in behavioral sciences, such as psychology and education (Heyvaert and Onghena, 2014; Smith, 2012; Plavnick and Ferreri, 2013), where personalized solutions are needed or generalization is not of primary importance.

As in multi-case experimental designs, randomization is fundamental to achieve internal validity and draw reliable inferences from the data (Kratochwill and Levin, 2010).
However, differently from multi-case experimental designs, only a single unit is observed and so randomization is not achieved by randomly assigning individuals to treatment groups, but by randomly assigning measurement instances.
Different randomization schemes have been introduced in literature (Onghena, 2005) allowing to differentiate the two main types of single-case designs: alternation designs and phase designs. In alternation designs the treatment alternation is random, while in phase designs it is the moment of intervention to be random.

To analyze these experimental designs, researchers traditionally rely on visual analysis techniques (Kratochwill, Hitchcock, et al., 2013; Onghena and Edgington, 2005), which, however, present several drawbacks. The main criticisms regard their reliability, due to the low agreement among and within users, and the higher risk of type I error (Park, Marascuilo, and Gaylord-Ross, 1990; Onghena and Edgington, 2005; Heyvaert and Onghena,

*Civil, Environmental and Architectural Engineering, University of Padova, Padua, Italy
†Department of Management and Engineering, University of Padova, Vicenza, Italy
‡Department of Management and Engineering, University of Padova, Vicenza, Italy

2014).
Several statistical procedures have been proposed to overcome these limitations (Onghena and Edgington, 2005; Park, Marascuilo, and Gaylord-Ross, 1990).
In addition to traditional t-test and F-test, non-parametric solutions have been presented. In particular, permutation (randomization) tests are gaining increasing attention, being distribution-free and easy to adapt to different types of single-case designs (Onghena and Edgington, 2005).
In addition, they can be used to analyze both univariate and multivariate problems (Solmi and Onghena, 2014), through the NonParametric Combination (NPC) procedure (Pesarin and Salmaso, 2010).

In this paper, we propose an adaptation of permutation tests within the NPC framework for the analysis of completely randomized single-case designs (CRD) with a multivariate ordinal outcome. Finally, a simulation study is conducted and a real application is shown.

2. Methodology

To analyze completely randomized single-case designs (CRD) with a multivariate ordinal outcome, permutation tests offer a suitable solution to overcome difficulties arising with traditional methods, such as visual analysis.
Let us consider two treatment levels, A and B, and the following system of hypotheses to test:

$$\begin{cases} H_0 : F_A(\mathbf{x}) = F_B(\mathbf{x}) \\ H_1 : F_A(\mathbf{x}) \neq F_B(\mathbf{x}) \end{cases}$$

where $\{\mathbf{x}_{Ai}\}_1^n$ and $\{\mathbf{x}_{Bi}\}_1^m$ are two independent samples drawn from F_A and F_B respectively.
Our goal is to test whether the multivariate ordinal dependent variable $\mathbf{X}$ is equally distributed among the two treatment levels.

In CRD only the number of measurement instances for each level of the treatment variable is relevant to determine the treatment sequence. There are not constraints over the alternation of the levels of the dependent variable.
In other words, n measurement instances for each level (A and B) of a dependent variable X result in $\binom{2n}{n}$ possible assignments.
This random assignment of the measurement instances enables to control over eventual confounding factors (Onghena and Edgington, 2005) and allows to use permutation tests (Sierra, Solanas, and Vicenç, 2005; HaardÖrfer and Gagné, 2010).
It is worth also noting that, under the null hypothesis, permutation tests are conditioned with respect to a set of sufficient statistics and require the assumption of exchangeability of data with respect to groups (Pesarin and Salmaso, 2010). These two characteristics indeed make them independent of the underlying likelihood model related to the unknown population distribution P (Pesarin and Salmaso, 2010) and the exchangeability condition is sufficient for being exact when using a randomization distribution to obtain valid statistical significance (Sierra, Solanas, and Vicenç, 2005; Good, 2001).

If these conditions are met, permutation tests are not constrained by the unknown population distribution P, thus facilitating their application also to different types of data, such us continuous, ordinal or mixed data, through the choice of an appropriate test statistic.
If X is an ordinal variable, a suitable test statistic for the comparison between two independent groups is the Anderson-Darling test statistic (Pesarin and Salmaso, 2010):

$$T_{AD} = \sum\nolimits_i [\hat{F}_2^*(X_i) - \hat{F}_1^*(X_i)]^2 \,/ \{\bar{F}(X_i)[1 - \bar{F}(X_i)]\}$$

Table 1: Representation of a two-sample multivariate permutation

$X_1(1)$	$\cdots$	$X_1(n_1)$	$X_1(1+n_1)$	$\cdots$	$X_1(n)$		T_1^o
$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\longrightarrow$	$\vdots$
$X_V(1)$	$\cdots$	$X_V(n_1)$	$X_V(1+n_1)$	$\cdots$	$X_V(n)$		T_V^o

$X_1(u_1^*)$	$\cdots$	$X_1(u_{n_1}^*)$	$X_1(u_{1+n_1}^*)$	$\cdots$	$X_1(u_n^*)$		T_1^*
$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\longrightarrow$	$\vdots$
$X_V(u_1^*)$	$\cdots$	$X_V(u_{n_1}^*)$	$X_V(u_{1+n_1}^*)$	$\cdots$	$X_V(u_n^*)$		T_V^*

where

$$\hat{F}_j^*(t) = \sum\nolimits_i \mathbb{I}(X_{ji}^* \leq t)/n_j,$$

$t \in \mathcal{R}^1$, $j = 1, 2$ and $\bar{F}(t) = \sum_{ji} \mathbb{I}(X_{ji} \leq t)/n$, $t \in \mathcal{R}^1$.
Other test statistics proposed in literature are Cramér-Von Mise's $T_{CM}^* = \sum_i (\hat{F}_2^*(X_i) - \hat{F}_1^*(X_i))$ and Kolmogorov-Smirnov's $T_{Ks}^* = max_i(\hat{F}_2^*(X_i) - \hat{F}_1^*(X_i))$.

With multivariate ordinal data, the NonParametric Combination (NPC) procedure represents certainly a suitable solution.
With a V-variate outcome $\mathbf{X}$ and 2 treatment levels, the NPC procedure decomposes the global hypotheses into V sub-hypotheses, as follows:

$$\begin{cases} H_0 : \bigcap_{h=1}^V H_{h0} = \bigcap_{h=1}^V [F_h(x) = F_{0h}(x)] \\ H_1 : \bigcup_{h=1}^V H_{h1} = \bigcup_{h=1}^V [F_h(x) \neq F_{0h}(x)] \end{cases}$$

where each sub-hypothesis can be tested separately and combined with the others later to solve the global problem.
The procedure can be essentially divided in two phases: at first each sub-hypothesis is tested, then combined tests are carried out.

The NPC first phase algorithm for estimating the V-variate distribution of $\mathbf{T}$ includes the following steps:

1. Calculate the vector of the observed values for tests $\mathbf{T} : \mathbf{T}^o = \mathbf{T}(\mathbf{X})$.

2. Consider a random permutation $\mathbf{X}^* \in \mathcal{X}_{/\mathbf{X}}$ of $\mathbf{X}$ and the values of vector statistics $\mathbf{T}^* = \mathbf{T}(\mathbf{X}^*)$, where $\mathbf{X}^*$ is obtained by first considering a random permutation $(u_1^*, \ldots, u_n^*)$ of $(1, \ldots, n)$ and then assigning the related individual data vectors to the proper group: $\mathbf{X}^* = \{\mathbf{X}(u_i^*) = [X_1(u_i^*), \ldots, X_V(u_i^*)], i = 1, \ldots, n\}$.

3. Carry out B independent repetitions of step 2 (Conditional Monte Carlo - CMC). The set of CMC results $\{\mathbf{T}_b^*, b = 1, \ldots, B\}$ is thus a random sampling from the permutation V-variate distribution of $\mathbf{T}$. Hence, $\hat{\lambda}_h = \hat{L}_h(T_h^o|\mathcal{X}_{/\mathbf{X}}) = \frac{[\frac{1}{2} + \sum_b \mathbb{I}(T_{hb}^* \geq T_h^o)]}{(B+1)}$, $h = 1, \ldots, V$ gives a consistent estimate of the marginal p-value $\lambda_h = \Pr\{T_h^* \geq T_h^o|\mathcal{X}_{/\mathbf{X}}\}$, for test T_h.

To test multivariate hypotheses, the p-values associated with the partial tests T_h must be combined using a non-degenerate combining function $\psi : [0,1]^V \rightarrow R^1$ such that ψ is non-increasing in each argument, ψ attains its supremum value, possibly not finite, even when only one argument attains zero, and for every $\alpha > 0$, the critical value of every ψ is assumed to be finite and strictly smaller than the supremum value.
According to Birnbaum (1954, 1955) in order for combined tests to be admissible, their rejection regions must be convex. For this reason, ψ should be chosen in order to satisfy

Table 2: Representation of the CMC method in multivariate tests

$\mathbf{X}$	$\mathbf{X}_1^*$	$\cdots$	$\mathbf{X}_b^*$	$\cdots$	$\mathbf{X}_B^*$
T_1^o	T_{11}^*	$\cdots$	T_{1b}^*	$\cdots$	T_{1B}^*
$\vdots$	$\vdots$		$\vdots$		$\vdots$
T_V^o	T_{V1}^*	$\cdots$	T_{Vb}^*	$\cdots$	T_{VB}^*

Table 3: Representation of nonparametric combination

T_1^o	T_{11}^*	$\cdots$	T_{1b}^*	$\cdots$	T_{1B}^*
$\vdots$	$\vdots$		$\vdots$		$\vdots$
T_V^o	T_{V1}^*	$\cdots$	T_{Vb}^*	$\cdots$	T_{VB}^*
			$\downarrow$		
$\hat{\lambda}_1$	$\hat{L}_{11}^*$	$\cdots$	$\hat{L}_{1b}^*$	$\cdots$	$\hat{L}_{1B}^*$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\hat{\lambda}_V$	$\hat{L}_{V1}^*$	$\cdots$	$\hat{L}_{Vb}^*$	$\cdots$	$\hat{L}_{VB}^*$
			$\downarrow$		
T''^o	$T_1''^*$	$\cdots$	$T_b''^*$	$\cdots$	$T_B''^*$

such a property.
Among many others, some combining functions satisfying all these conditions are Fisher's $T_F'' = -2 \cdot \sum_h \log(\lambda_h)$, Tippett's $T_T'' = \max_{1 \leq h \leq V}(1-\lambda_h)$, and Liptak's $T_L'' = \sum_h \Phi^{-1}(1-\lambda_h)$ where λ_h are the p-values of partial tests.

Once an appropriate combining function has been chosen, the global tests can be formulated to address the global system of hypotheses. The second phase of the algorithm requires the following steps:

a. Consider V observed p-values estimated on the data $\mathbf{X}$ by $\hat{\lambda}_h = \hat{L}_h(T_h^o|\mathcal{X}_{/\mathbf{X}})$.

b. The combined observed value of the second-order test is given by $T''^o = \psi(\hat{\lambda}_1, \ldots, \hat{\lambda}_V)$.

c. The bth combined value of vector statistics (step 4) is then calculated by $T_b''^* = \psi(\hat{L}_{1b}^*, \ldots, \hat{L}_{Vb}^*)$, where $\hat{L}_{hb}^* = \hat{L}_h(T_{hb}^*|\mathcal{X}_{/\mathbf{X}})$, $h = 1, \ldots, V$, $b = 1, \ldots, B$.

d. Hence, the p-value of the combined test T'' is estimated as $\hat{\lambda}_\psi'' = \sum_b \mathbb{I}(T_b''^* \geq T''^o)/B$.

e. If $\hat{\lambda}_\psi'' \leq \alpha$, the global null hypothesis H_0 is rejected at significance level α.

To sum up, the NPC along with the choice of a suitable test statistic allow us to analyze completely randomized single-case designs with multivariate ordinal outcomes.

2.1 Simulation study

The power of the proposed procedure was evaluated under different sample sizes, number of components V, correlations or marginal distributions of the multivariate ordinal outcome. The Anderson-Darling test statistic and the Fisher combining function were adopted.
All simulations were performed in R (version 3.5.0, R Core Team, 2018) and both the

number of simulations and the number of permutations were set to 2000.
The gaussian copula based procedure proposed by Ferrari and Barbiero (2012) was used to generate multivariate ordinal data with a prespecified correlation matrix and marginal distributions (R package GenOrd, Barbiero and Ferrari, 2015).

Let Σ indicate the correlation matrix

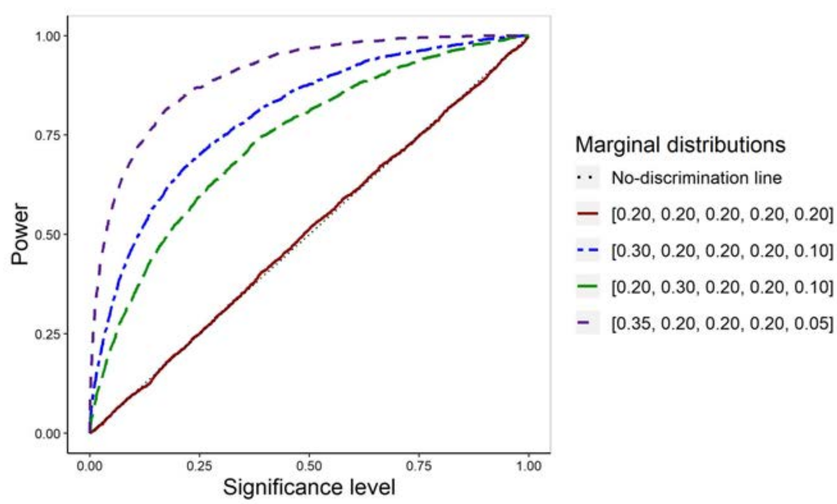
$$\Sigma = \begin{bmatrix} 1 & \sigma & \sigma & \sigma \\ \sigma & 1 & \sigma & \sigma \\ \sigma & \sigma & 1 & \sigma \\ \sigma & \sigma & \sigma & 1 \end{bmatrix},$$

we considered N measurement occasions for 2 treatment levels (A and B) and a V-variate ordinal dependent variable with 5 levels, with the following the marginal distributions:

$$\begin{aligned} P(X_h = 1) = p1 \quad P(X_h = 2) = p2 \quad P(X_h = 3) = p3 \\ P(X_h = 4) = p4 \quad P(X_h = 5) = p5 \quad \forall h = 1, ..., V. \end{aligned} \tag{1}$$

First of all, the power behavior of the proposed procedure under the null hypothesis and the alternative hypothesis was investigated.
With marginal distributions for treatment A equal to $[0.20, 0.20, 0.20, 0.20, 0.20]$ and varying the marginal distributions for treatment B (see Figure 1), it emerged that the proposed testing procedure is reliable under the null hypothesis (marginal distributions for treatment B equal to $[0.20, 0.20, 0.20, 0.20, 0.20]$). In addition, it looks quite powerful with different configurations under the alternative hypothesis.

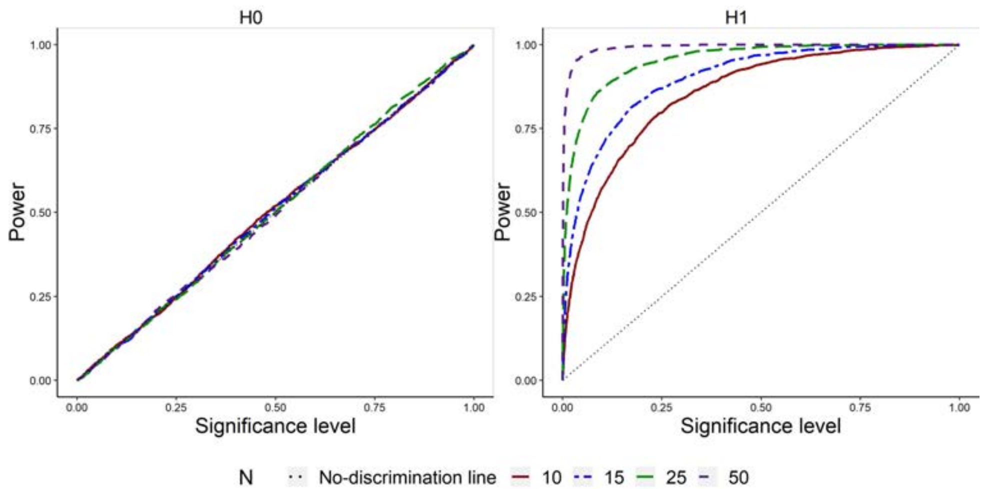
Figure 1: Power behavior with different marginal distributions for treatment B. $N = 15$, $V = 4$, $\sigma = 0.25$



The simulation study showed also some interesting properties of the testing procedure. Let us consider under H_0, marginal distributions equal to $[0.20, 0.20, 0.20, 0.20, 0.20]$ for both the treatments and marginal distributions for treatment B under H_1 equal to $[0.35, 0.20, 0.20, 0.20, 0.05]$.

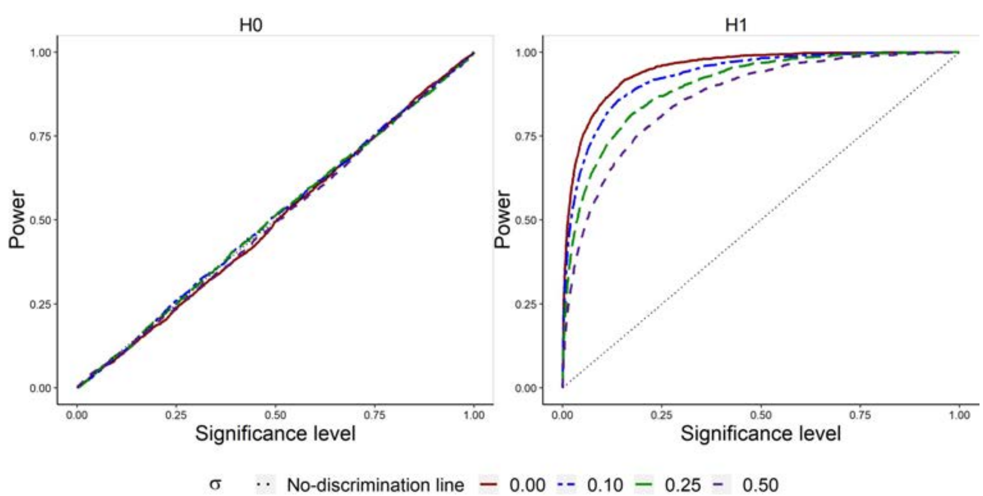
By varying the number of measurement occasions N, the testing procedure respects the significance level under the null hypothesis. On the other hand, under the alternative hypothesis, it can be noticed an increase of the power of the test with the increase of N. However, the power of the test is quite high also for small values of N.

Figure 2: Power behavior with different numbers of measurement instances. $V = 4$, $\sigma = 0.25$



By setting the number of measurement instances equal to 15 and varying the parameter σ, an increase in the correlations among components seems to negatively affect the power of this testing procedure (see Figure 3).
This is an open problem (Arboretti Giancristofaro et al., 2016) related to the NPC procedure, although it takes implicitly into account dependency among components.

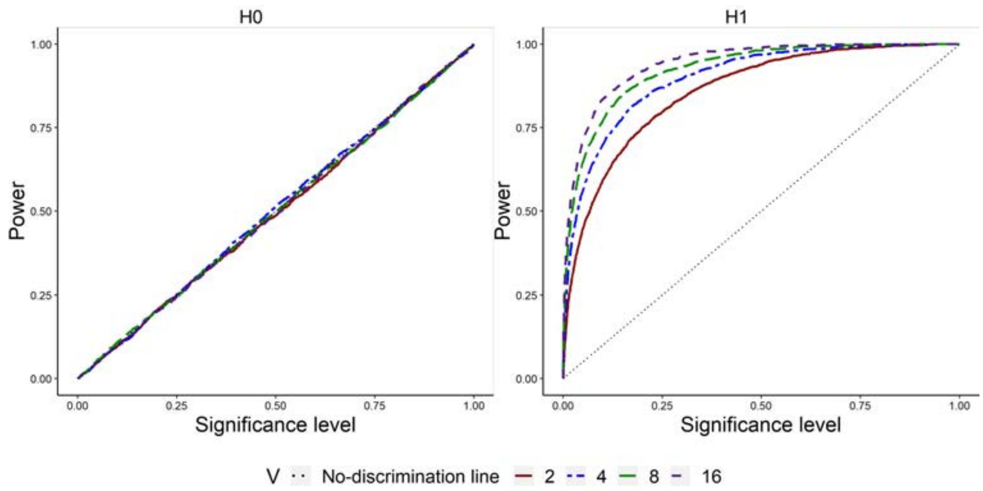
Figure 3: Power behavior with different values of σ. $N = 15$, $V = 4$



Finally, the simulation study has shown how an increase in the number of components looks to positively affect the power under H_1 (see Figure 4).
This is the so called finite-sample consistency property (Pesarin and Salmaso, 2010).

Having achieved promising results regarding the power of the proposed procedure, a real case study was carried out.

Figure 4: Power behavior with different numbers of components. $N = 15$, $\sigma = 0.25$



2.2 A real case study

The investigated problem consisted in assessing eventual differences in performances between two different composite materials for t-shirts (A and B).
A professional athlete was followed for a period of 30 days during his training activities and he was asked to evaluate the perceived heat (y) and humidity (z) wearing these t-shirts (see Table 4).
Each day the t-shirt to be worn was randomly chosen.
The proposed procedure was able to identify a significant difference among the two materials and in particular, testing directional hypotheses, it emerged that A outperformed B (see Table 5).
It is worth noting that to test the directional hypothesis $H_1 : F_A(\mathbf{x}) \leq F_B(\mathbf{x})$, we referred to the following modified version of the test statistic proposed by Pesarin and Salmaso (2010) :

$$T_{AD} = \sum_i [\hat{F}_2^*(X_i) - \hat{F}_1^*(X_i)] \Big/ \{\bar{F}(X_i)[1 - \bar{F}(X_i)]\}^{1/2}$$

Table 4: Scheme

Day 1	Day 2	Day 3	...	Day 30
A	B	B	...	A
y_{A1}	y_{B1}	y_{B2}	...	y_{A15}
z_{A1}	z_{B1}	z_{B2}	...	z_{A15}

Table 5: P-values table

Two-sided hypotheses	Directional hypotheses
0.0005	0.0005

3. Conclusions

In this paper, we presented an extension of permutation tests for the analysis of completely randomized single-case designs with multivariate ordinal outcomes, within the context of the NPC procedure.
The simulation study showed that the proposed procedure is reliable under the null hypothesis and powerful under the alternative hypothesis, also with a small number of measurement instances.
The correlation among components seems to negatively affect the power, but with increasing number of components it is possible to achieve a quite high power thanks to the finite-sample consistency property (Pesarin and Salmaso, 2010).
Further extensions of the presented solution could be of interest in order to explore different types of randomized single-case designs.

References

Arboretti Giancristofaro, Rosa et al. (2016). "Dependency and truncated forms of combinations in multivariate combination-based permutation tests and ordered categorical variables". In: *Journal of Statistical Computation and Simulation*. ISSN: 15635163. DOI: 10.1080/00949655.2016.1177826.

Barbiero, Alessandro and Pier Alda Ferrari (2015). *GenOrd: Simulation of Discrete Random Variables with Given Correlation Matrix and Marginal Distributions*. R package version 1.4.0. URL: https://CRAN.R-project.org/package=GenOrd.

Birnbaum, Allan (1954). "Combining Independent Tests of Significance*". In: *Journal of the American Statistical Association*. ISSN: 1537274X. DOI: 10.1080/01621459.1954.10483521.

— (1955). "Characterizations of Complete Classes of Tests of Some Multiparametric Hypotheses, with Applications to Likelihood Ratio Tests". In: *The Annals of Mathematical Statistics*. ISSN: 0003-4851. DOI: 10.1214/aoms/1177728590.

Ferrari, Pier Alda and Alessandro Barbiero (2012). "Simulating Ordinal Data". In: *Multivariate Behavioral Research*. ISSN: 00273171. DOI: 10.1080/00273171.2012.692630.

Good, P (2001). "Permutation tests: a practical guide to resampling methods for testing hypothesis. 1994". In: *Springer Verlag Inc., New York*.

HaardÖrfer, Regine and Phill Gagné (2010). *The use of randomization tests in single-subject research*. DOI: 10.1177/1088357609350745.

Heyvaert, Mieke and Patrick Onghena (2014). "Randomization tests for single-case experiments: State of the art, state of the science, and state of the application". In: *Journal of Contextual Behavioral Science*. ISSN: 22121447. DOI: 10.1016/j.jcbs.2013.10.002.

Kratochwill, Thomas R., John H. Hitchcock, et al. (2013). *Single-Case Intervention Research Design Standards*. DOI: 10.1177/0741932512452794.

Kratochwill, Thomas R. and Joel R. Levin (2010). "Enhancing the Scientific Credibility of Single-Case Intervention Research: Randomization to the Rescue". In: *Psychological Methods*. ISSN: 1082989X. DOI: 10.1037/a0017736.

Onghena, Patrick (2005). "Single-Case Designs". In: *Encyclopedia of Statistics in Behavioral Science*. American Cancer Society.

Onghena, Patrick and Eugene S. Edgington (2005). "Customization of pain treatments: Single-case design and analysis". In: *Clinical Journal of Pain*. ISSN: 07498047. DOI: 10.1097/00002508-200501000-00007.

Park, Hyun Sook, Leonard Marascuilo, and Robert Gaylord-Ross (1990). “Visual inspection and statistical analysis in single-case designs”. In: *Journal of Experimental Education*. ISSN: 19400683. DOI: 10.1080/00220973.1990.10806545.

Pesarin, Fortunato and Luigi Salmaso (2010). *Permutation Tests for Complex Data: Theory, Applications and Software*. ISBN: 9780470689516. DOI: 10.1002/9780470689516.

Plavnick, Joshua B. and Summer J. Ferreri (2013). “Single-Case Experimental Designs in Educational Research: A Methodology for Causal Analyses in Teaching and Learning”. In: *Educational Psychology Review*. ISSN: 1040726X. DOI: 10.1007/s10648-013-9230-6.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Sierra, Vicenta, Antonio Solanas, and Quera Vicenç (2005). “Randomization tests for systematic single-case designs are not always appropriate”. In: *Journal of Experimental Education*. ISSN: 19400683. DOI: 10.3200/JEXE.73.2.140-160.

Smith, Justin D. (2012). “Single-case experimental designs: A systematic review of published research and current standards”. In: *Psychological Methods*. ISSN: 1082989X. DOI: 10.1037/a0029312.

Solmi, Francesca and Patrick Onghena (2014). “Combining p-values in replicated single-case experiments with multivariate outcome”. In: *Neuropsychological Rehabilitation*. ISSN: 14640694. DOI: 10.1080/09602011.2014.881747.