STATISTICA, anno LXXIV, n. 3, 2014

OVERVIEW OF NONPARAMETRIC COMBINATION-BASED PERMUTATION TESTS FOR MULTIVARIATE MULTI-SAMPLE PROBLEMS

Rosa Arboretti Giancristofaro [1]
Dept. Land, Environment, Agriculture and Forestry, University of Padova, Padova, Italy
Chiara Brombin
University Centre of Statistics in the Biomedical Sciences (CUSSB), Vita-Salute San Raffaele University, Milano, Italy

1. An overview on nonparametric methodology

In recent years, due to the increasing availability of (big) data and the consequent need to solve more and more complex multivariate problems, there is a growing interest in permutation testing methods.
Actually, permutation tests are essentially exact in a nonparametric conditional framework, where conditioning is on the pooled observed data set, which is generally a set of sufficient statistics in the null hypothesis.

Many complex multivariate problems are difficult to handle outside the conditional framework and in particular outside the nonparametric combination (NPC) of dependent permutation tests method. We present a overview of NPC methodology and develop a SAS code for the method described allowing to solve multivariate multi-sample problems. As pointed out in Pesarin and Salmaso (2010b), despite in the literature permutation tests are mostly derived by means of heuristic arguments (Edgington and Onghena, 2007; Good, 2005), their natural theoretical background must be referred to the principles of conditional inference (Birnbaum, 1962; Edwards, 1972). Since within this framework it can be proved that permutation tests are provided with suitable theoretical properties (Pesarin and Salmaso, 2010b, 2012), whenever permutation tests are correctly applicable, their results may be extended, at least in a weak sense, to population inferences (Pesarin, 2002).
It is worth noting that within a parametric framework the extension from samples to populations is possible only when the data set is randomly selected by well-designed sampling procedures on well-defined population distributions, provided that their nuisance parameters are completely removable (Pesarin, 2002). When

[1] Corresponding Author. E-mail: rosa.arboretti@unipd.it

these conditions fail, especially if selection-bias procedures are used for data collection processes, in general most of the parametric inferential extensions are wrong or misleading. On the contrary, the permutation-based inferential conclusions may be always extended to the reference population even in case of selection-bias sampling (Pesarin, 2002).
Parametric testing methods usually underlying a modelling approach require a set of stringent and often difficult to justify assumptions (Pesarin, 2002), especially when managing real data. Without justification, researchers often assume multivariate normality, random sampling from a population (even in cases of selection-bias), data homoscedasticity also in the alternative (by additive fixed effects), missing and/or censored data independent of treatments, random effects independent of units and/or of errors (the natural deviates), etc., so that consequent inferences can hardly have a real credibility.
On the contrary, nonparametric testing approaches are based on weaker and more realistic foundations, are intrinsically robust; and so related inferences are more credible and easier to interpret.
Roughly speaking, with the term nonparametric test we usually refer to an hypothesis testing procedure that has certainly desirable properties that hold under relative mild assumptions regarding the underlying populations from which the data are obtained (Hollander *et al.*, 2013). Often nonparametric tests are described also as distribution-free methods, meaning that they do not rely on assumptions that the data are drawn from a given probability distribution. Permutation tests are a kind of nonparametric conditional procedures that have the nice property of being exact for whatever, even very small, finite sample size.

The paper is organized as follows. In Section 2 we introduce notation and assumptions underlying NPC methodology for the general case of One-way MANOVA design. Section 3 provides a detailed description of the NPC algorithm and Section 4 is devoted to the choice of proper combining functions. In Section 5 SAS macros are illustrated and in Section 6 applications are shown. A brief discussion is provided in 7.

2. One-way MANOVA design: multivariate C-sample problem

In the following we introduce notation and main assumptions of NPC methodology referring to the general case of a one-way MANOVA design.

(i) Let $\mathbf{X} = \{\mathbf{X}_j,\ j = 1,\dots,C\} = \{\mathbf{X}_{ji},\ i = 1,\dots,n_j,\ j = 1,\dots,C\} = \{X_{hji},\ i = 1,\dots,n_j,\ j = 1,\dots,C,\ h = 1,\dots,V\}$ be a V-dimensional set of data. The symbol $\mathbf{X}$ indicates both the data set and V-dimensional response. The response $\mathbf{X}$ takes its values on the V-dimensional sample space $\mathcal{X}$, for which a σ-algebra $\mathcal{A}$ and a (possibly not specified) nonparametric family $\mathcal{P}$ of non-degenerate distributions are assumed to exist. The data set $\mathbf{X}$ consists of $C \geq 2$ samples or groups of size $n_j \geq 2$, with $n = \sum_j n_j$; the groups are presumed to be related to C levels of a treatment and the data $\mathbf{X}_j$ are supposed i.i.d. with distributions $P_j \in \mathcal{P}$, $j = 1,\dots,C$. To simplify notation, the unit-by-unit representation of data $\mathbf{X} = \{\mathbf{X}(i),\ i = 1,\dots,n;$

$n_1, \ldots, n_C\}$ is used, according to which the first n_1 data vectors belong to the first group, the next n_2 to the second, and so on.

(ii) The null hypothesis refers to equality of multivariate distributions of responses on C groups:

$$H_0 : \{P_1 = \ldots = P_C\} = \left\{\mathbf{X}_1 \stackrel{d}{=} \ldots \stackrel{d}{=} \mathbf{X}_C\right\}.$$

Assume that we are interested in a set of side-assumptions such that H_0 may be properly broken down into a finite set of sub-hypotheses H_{0i} , $i = 1, \ldots, k$, each appropriate for a partial aspect of interest. Therefore, H_0 is true if all the H_{0i} are jointly true; and so it may be written as $\left\{\bigcap_{i=1}^{k} H_{0i}\right\}$. In this sense, H_0 is also called the *global* or *overall null hypothesis*. Actually we may have many sub-hypotheses as the number V of responses, but this is not necessarily true. It must be noticed that H_0 implies that the V-dimensional data vectors in $\mathbf{X}$ are exchangeable with respect to C groups.

The alternative hypothesis, instead, states that at least one of the null sub-hypotheses H_{0i} is not true. Hence, the alternative may be represented by the union of k sub-alternatives,

$$H_1 : \left\{\bigcup_{i=1}^{k} H_{1i}\right\},$$

stating that H_1 is true when at least one sub-alternative is true. In this context, H_1 is called the *global* or *overall alternative*.

(iii) Let $\mathbf{T} = \mathbf{T}(\mathbf{X})$ be a k -dimensional vector of test statistics, in which the ith component $T_i = T_i(\mathbf{X})$, $i = 1, \ldots, k$, represents the non-degenerate i the *partial test* which is assumed to be appropriate for testing sub-hypothesis H_{0i} against H_{1i}. Without loss of generality, in the NPC context all partial tests are assumed to be marginally unbiased, consistent and significant for large values.

2.1. Condition for proper inferential solutions

Partial tests must satisfy the following conditions.

1. All permutation partial tests T_i are marginally unbiased and significant for large values, so that they are both conditionally and unconditionally stochastically larger in H_1 than in H_0.

2. All permutation partial tests T_i are marginally consistent, i.e. as sample sizes tend to the infinity $\Pr\{T_i \geq T_{i\alpha} | H_{1i}\} \to 1$, $\forall \alpha > 0$, where $T_{i\alpha}$, which is assumed to be finite, is the critical value of T_i at level α.

Actually, the NPC methodology relies on a Conditional Monte Carlo (CMC) procedure. Details on the procedure will be presented in Section 3. However, it must be emphasized that, when conditions (i)–(iii) defined in Section 2 are jointly satisfied, the NPC of dependent tests leads to exact solutions.
When these conditions are not completely satisfied, solutions, approximations and related conclusions may be biased and not reliable.

3. Nonparametric Combination algorithm

The NPC methodology is based on:

a) on a decomposition of the global null hypothesis into k, $k > 1$, sub-hypotheses, where for each sub-hypothesis, there exists a suitable partial permutation test statistic;

b) on a simulation procedure, conditional on the set of observed data, which provides an estimate of the null multivariate permutation distribution of the whole set of statistics;

c) on a combination of the partial simulation results into a second-order statistic whose null permutation distribution is estimated by using the same simulation results of the first step.

NPC methodology is essentially a two-stage algorithm allowing to obtain a conditional Monte Carlo (CMC) estimate of the permutation distribution of combined tests. Actually, with reference to point a), the k-dimensional hypothesis test problem is processed in two phases: firstly, we define a suitable set of k, with $k \geq 1$, unidimensional permutation tests called *partial tests*. Each partial test examines the marginal contribution of any single response variable in the comparison made between several treatment groups. The second phase is the nonparametric combination of dependent tests in one *second order combined test*, which is suitable for testing possible global differences between the multivariate distributions of two or more groups. When there is a stratification variable, we expect two combination levels: the partial tests combination in s second order combined tests, $s \geq 1$, within the i-th stratum, $i = 1, \ldots, s$, and a further combination of the tests in a single *third order combined test*.

The first phase concerns with the estimate of the k-variate distribution of $\mathbf{T}$, while the second finds the estimate of permutation distribution of combined test T''_ψ by using the same CMC results of the first phase. Once the combining function ψ has been chosen, the notation T''_ψ is replaced by T''.

In this multivariate setting, simulations from the permutation sample space $\mathcal{X}_{/\mathbf{X}}$ are carried out by means of a CMC method.

■ The first phase of a procedure estimates the distribution of $\mathbf{T}$ including the following steps:

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

b) Consider a random permutation $\mathbf{X}^* \in \mathcal{X}_{/\mathbf{X}}$ of $\mathbf{X}$ and the values of vector statistics $\mathbf{T}^* = \mathbf{T}(\mathbf{X}^*)$. Actually, in multivariate situations the permutation $\mathbf{X}^*$ is obtained by first considering a random permutation $(u_1^*, \ldots, u_n^*)$ of $(1, \ldots, n)$ and then by assignment of related individual data vectors to the proper group; thus, by using the unit-by-unit representation, $\mathbf{X}^* = \{\mathbf{X}(u_i^*) = [X_1(u_i^*), \ldots, X_V(u_i^*)]\ ,\ i = 1, \ldots, n;\ n_1, \ldots, n_C\}$.

c) Carry out B independent repetitions of step (S.b$_k$). The set of CMC results $\{\mathbf{T}_b^*,\ b = 1, \ldots, B\}$ is thus a random sampling from the permutation k-variate distribution of vector test statistics $\mathbf{T}$.

d) The k-variate EDF $\hat{F}(\mathbf{t}|\mathcal{X}_{/\mathbf{X}}) = \left[\frac{1}{2} + \sum_b \mathbb{I}(\mathbf{T}_b^* \leq \mathbf{t})\right]/(B+1),\ \forall \mathbf{t} \in \mathcal{R}^k$, gives a consistent estimate of the corresponding k-dimensional permutation CDF $F(\mathbf{t}|\mathcal{X}_{/\mathbf{X}})$ of $\mathbf{T}$. Moreover, the ESFs

$$\hat{L}_i(t|\mathcal{X}_{/\mathbf{X}}) = \left[\frac{1}{2} + \sum\nolimits_b \mathbb{I}(T_{ib}^* \geq t)\right]/(B+1), \quad i = 1, \ldots, k,$$

give consistent estimates $\forall z \in \mathcal{R}^1$ of the k marginal permutation SLF $L_i(z|\mathcal{X}_{/\mathbf{X}}) = \Pr\{T_i^* \geq t|\mathcal{X}_{/\mathbf{X}}\}$. Thus $\hat{L}_i(T_i^o|\mathcal{X}_{/\mathbf{X}}) = \hat{\lambda}_i$ gives a consistent estimate of the marginal p-value $\lambda_i = \Pr\{T_i^* \geq T_i^o|\mathcal{X}_{/\mathbf{X}}\}$, relative to test T_i.

■ The second phase of the algorithm for simulating a procedure for NPC should include the following steps:

a) The k observed p-values are estimated on the data $\mathbf{X}$ by $\hat{\lambda}_i = \hat{L}_i(T_i^o|\mathcal{X}_{/\mathbf{X}})$, where $T_i^o = T_i(\mathbf{X})$, $i = 1, \ldots, k$, represent the observed values of partial tests and $\hat{L}_i$ is the ith marginal ESF, the latter being jointly estimated by the CMC method on the data set $\mathbf{X}$, in accordance with step S.d$_k$ above.

b) The combined observed value of the second-order test is evaluated through the same CMC results as the first phase and is given by:

$$T''^o = \psi(\hat{\lambda}_1, \ldots, \hat{\lambda}_k).$$

c) The bth combined value of vector statistics (step (S.d$_k$)) is then calculated by

$$T_b''^* = \psi(\hat{L}_{1b}^*, \ldots, \hat{L}_{kb}^*),$$

where $\hat{L}_{ib}^* = \hat{L}_i(T_{ib}^*|\mathcal{X}_{/\mathbf{X}})$, $i = 1, \ldots, k$, $b = 1, \ldots, B$.

d) Hence, the p-value of the combined test T'' is estimated as

$$\hat{\lambda}_\psi'' = \sum\nolimits_b \mathbb{I}(T_b''^* \geq T''^o)/B.$$

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

Of course, if proper routines for exact calculations were available, then the multivariate distribution $F(\mathbf{t}|\mathcal{X}_{/\mathbf{X}})$, the partial p-values $(\lambda_1, \ldots, \lambda_k)$, the distribution of the combined test $F_\psi(t|\mathcal{X}_{/\mathbf{X}})$, and the combined p-value λ''_ψ can be evaluated exactly.

4. Choosing sensible combining functions

For the sake of simplicity, here we focus only on combining functions applied to p -values associated with partial tests. It may be shown that partial tests are permutationally equivalent to their p-values, as a direct consequence of the monotonic non-increasing behaviour with respect to t of significance level functions.

Thus, the NPC in one second-order test

$$T'' = \psi(\lambda_1, \ldots, \lambda_k)$$

is achieved by a continuous, non-increasing, univariate, measurable and non-degenerate real function $\psi : (0, 1)^k \to \mathcal{R}^1$.

Note that the continuity of ψ is required because it has to be defined irrespective of the cardinality of $(\Lambda_1, \ldots, \Lambda_k)$. Moreover, the measurability property of ψ is required because it is used as a test statistic which then must induce a probability distribution on which inferential conclusions are necessarily based.

In order to be suitable for test combination (see Pesarin, 1992, 2001; see also Goutis *et al.*, 1996), all combining functions ψ must satisfy at least the following reasonable properties:

(**P.1**) A combining function ψ must be non-increasing in each argument: $\psi(.., \lambda_i, ..) \geq \psi(.., \lambda'_i, ..)$ if $\lambda_i < \lambda'_i$, $i \in \{1, \ldots, k\}$. Also, it is generally desirable that ψ is symmetric, i.e. invariant with respect to rearrangements of the entry arguments: $\psi(\lambda_{u_1}, \ldots, \lambda_{u_k}) = \psi(\lambda_1, \ldots, \lambda_k)$ where $(u_1, \ldots, u_k)$ is any permutation of $(1, \ldots, k)$.

(**P.2**) Every combining function ψ must attain its supremum value $\bar{\psi}$, possibly not finite, even when only one argument attains zero: $\psi(.., \lambda_i, ..) \to \bar{\psi}$ if $\lambda_i \to 0$, $i \in \{1, \ldots, k\}$.

(**P.3**) $\forall \alpha > 0$, the critical value T''_α of every ψ is assumed to be finite and strictly smaller than $\bar{\psi}$: $T''_\alpha < \bar{\psi}$.

Actually property (**P.1**) is associated with the notion that large values are significant and related to the unbiasedness of combined tests. Instead, properties (**P.2**) and (**P.3**) are related to consistency.

4.1. Some Useful Combining Functions

Here we present some of the most used combining functions. Further details on the combination of one-sided independent tests may be found in Birnbaum, 1954; Oosterhoff, 1969; and Folks, 1984.

(a) The Fisher *omnibus* combining function is based on the statistic

$$T''_F = -2 \cdot \sum_i \log(\lambda_i).$$

If the k partial test statistics are independent and continuous, then in the null hypothesis T''_F follows a central χ^2 distribution with $2k$ degrees of freedom. In practice, T''_F is the most popular combining function.

(b) The Liptak combining function is based on the statistic

$$T''_L = \sum_i \Phi^{-1}(1 - \lambda_i),$$

where Φ is the standard normal CDF. If the k partial tests were independent and continuous, then in the null hypothesis T''_L would be normally distributed with mean 0 and variance k (see Liptak, 1958).

An alternative version of the Liptak function considers logistic transformations of p-values: $T''_P = \sum_i \log[(1 - \lambda_i)/\lambda_i]$.

(c) The Tippett combining function is given by

$$T''_T = \max_{1 \le i \le k}(1 - \lambda_i),$$

significant for large values (the equivalent form $T''_T = \min(\lambda_i)$ is significant for small values). Its null distribution, if the k tests are independent and continuous, behaves according to the largest (smallest) of k random values from the uniform distribution in the open interval $(0, 1)$. Tippett's T''_T was the first combining function reported in the literature.

5. The algorithm and its SAS implementation

To perform the ANOVA permutation test, according to the NPC test theory (Pesarin and Salmaso, 2010b), the `SAS` macro NPC_Csamples `npc(dati, var_byn, var_cat, var_con, dom_byn, dom_con, weights, clas, nsample, strato, paired, unit, missing)` can be used.

The input parameters of the procedure are:

- `dati`: name of the data set;
- `var_byn`: list of binary variables;
- `var_cat`: list of categorical, non-binary variables;
- `var_con`: list of continuous variables;
- `dom_byn`: specify NOTEQ;
- `dom_con`: specify NOTEQ;

- `weights`: list of weights for the variables, firstly specify weights for binary variables and then weights for categorical variables and continuous variables;
- `clas`: variable defining the two groups (alphanumeric variable);
- `nsample`: number of conditional resamplings;
- `strato`: variable defining strata (alphanumeric variable);
- `paired`: paired data (yes/no);
- `unit`: variable identifying paired observations;
- `missing`: presence of missing values (yes/no).

6. Applications

6.1. Wines data: evaluating chemical composition of wines

We present a simple C-sample problem, where free and bound monoterpene and C[13]-norisoprenoid concentrations of Weisser Riesling wines of different vintages and from different regions in South Africa, Germany and Northern Italy were compared (Flury, 1997; Marais *et al.*, 1992). There are a total of nine South African wines, seven German wines (all from Pfalz) and ten from Northern Italy (from both Trentino Alto Adige and Friuli).
Formalizing the testing problem, as shown in Section 2, here we deal with a 15-dimensional data set by $\mathbf{X} = \{\mathbf{X}_j,\, j = 1,2,3\} = \{\mathbf{X}_{ji},\, i = 1,\ldots,n_j,\, j = 1,2,3\} = \{X_{hji},\, i = 1,\ldots,n_j,\, j = 1,2,3,\, h = 1,\ldots,15\}$. The hypotheses are $H_0 : \{\mathbf{X}_1 \stackrel{d}{=} \mathbf{X}_2 \stackrel{d}{=} \mathbf{X}_3\}$ against $H_1 : \{\text{at least one equality is not true}\}$.

According to the CMC procedure, iterations are now done from the pooled data set $\mathbf{X} = \mathbf{X}_1 \biguplus \mathbf{X}_2 \biguplus X_3$, which is still a set of sufficient statistics for the problem in H_0.

A suitable test statistic based on deviance among sampling means is

$$T^*_{hC} = \sum_{j=1}^{C} (\bar{X}^*_{hj} - \bar{X}_{h\bullet})^2 \cdot n_j, h = 1,\ldots,V$$

where $\bar{X}^*_{hj} = \sum_i (X^*_{hji})/n_j$ and $\bar{X}_{h\bullet} = \sum_j \bar{X}_{hj} \cdot n_j/n$. Note that $\bar{X}_{h\bullet}$ is a permutationally invariant quantity, being based on the sum of all observed data. Hence, statistic T^*_{hC} is permutationally equivalent to $T^*_h = \sum_{j=1}^{C} n_j \cdot (\bar{X}^*_{hj})^2$

Results from NPC analysis highlights how globally the three cultures gives wines that are different in terms of their chemical composition, and in particular in terms of 6 out 15 collected variables (see Table 1).

6.2. A Two-Sample Epidemiological Survey: Problem Description

A prospective multicentre epidemiological study known as SETIG involving the surveillance of treatments in severe infections (Arboretti *et al.*, 2000) took place

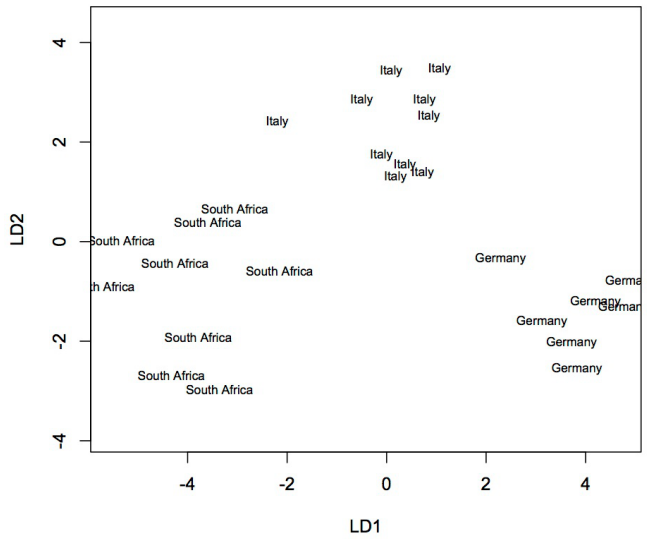

Figure 1 – Standard representation of Wines data through Fisher's Linear Discriminant Analysis. Scatterplot of the first two linear discriminant (LD) functions, showing that Weisser Riesling wines from three countries are well separated.

TABLE 1
Raw and adjusted p-values from NPC analysis.

	p-value	p-FWE
X1	0.6809	0.9892
X2	**0.0078**	0.0772
X3	0.9978	1.0000
X4	0.7705	0.9912
X5	**0.0472**	0.3327
X6	0.2596	0.8140
X7	0.1102	0.5585
X8	0.2767	0.8140
X9	0.9920	1.0000
X10	0.8918	0.9988
X11	**0.0382**	0.2963
X12	**0.0020**	**0.0234**
X13	0.1092	0.5585
X14	**0.0006**	**0.0078**
X15	**0.0008**	**0.0100**
Global		**0.0078**

between 1995 and 1997. Its objective was to compare different diagnostic and therapeutic approaches. The observational nature of the study requires particular care in the analysis because the incorrect use of observational data for evaluating therapy causal effects can produce biased treatment comparisons. In the SETIG survey, the hypothesis of interest concerns the comparative effect of the *specific therapeutic approach* versus an *empirical approach* with respect to several outcomes which constitute the pattern of interest, where 'specific therapeutic approach' means that an antibiotic specific to the particular infection is used, and 'empirical approach' means that a generic wide-spectrum antibiotic is used. The causal pattern of coherent alternatives is the following: compared to patients treated with the empirical therapy, patients receiving the specific therapy should show a reduced death rate, a shorter duration of treatment, a shorter length of stay in hospital, and a higher rate of infection resolutions. It is clear that in statistical terms such a pattern can be reduced to a two-sample multivariate comparison with restricted alternatives and mixed variables with a possible presence of missing values.

The analysis was carried out on 334 patients with sepsis, 154 of whom were treated with empirical therapy and 180 with a specific therapy. The 334 patients were at first stratified into four homogenous strata with respect to possible confounding factors (age, number of concomitant factors, presence of diabetes, presence of surgical intervention, presence of a tumour, type of unit where patient is admitted, geographic area, etc.). For the construction of the four strata $s = 1, 2, 3, 4$, a stratification by *propensity score* was used in accordance with Rosenbaum and Rubin (1983). The propensity score is defined as the probability of being assigned to a particular treatment, given a vector of concomitant variables (i.e. the confounding factors). This score, which may also have a prognostic interpretation, summarizes all the information required to balance the distribution of confounding variables between treatment groups, in order to have strata with homogeneous units. In this study, the propensity score was evaluated by a logistic model (see Arboretti *et al.*, 1999; Arboretti *et al.*, 2000). Four variables were taken into consideration: death (D) and clinical resolution (R), both binary variables with 1 denoting 'yes' and 0 'no'; duration of treatment (U), a binary variable with 1 denoting 'more than 15 days' and 0 '15 days or less'; length of stay (L), a positive integer variable. In the analysis, group 1 contains all subjects treated with a specific therapy and group 2 those treated with an empirical therapy. The multidimensional system of overall hypotheses defining the causal pattern can be written, and correspondingly analysed, either within-strata with respect to variables, as

$$H_0 : \left\{ \bigcap_s \left[(D_{1s} \stackrel{d}{=} D_{2s}) \bigcap (R_{1s} \stackrel{d}{=} R_{2s}) \bigcap (U_{1s} \stackrel{d}{=} U_{2s}) \bigcap (L_{1s} \stackrel{d}{=} L_{2s}) \right] \right\},$$

against

$$H_1 : \left\{ \bigcup_s \left[(D_{1s} \stackrel{d}{>} D_{2s}) \bigcup (R_{1s} \stackrel{d}{<} R_{2s}) \bigcup (U_{1s} \stackrel{d}{>} U_{2s}) \bigcup (L_{1s} \stackrel{d}{>} L_{2s}) \right] \right\},$$

or within-variables with respect to strata, as

$$H_0 : \left\{ \left[\bigcap_s (D_{1s} \stackrel{d}{=} D_{2s}) \right] \bigcap \left[\bigcap_s (R_{1s} \stackrel{d}{=} R_{2s}) \right] \right.$$
$$\left. \bigcap \left[\bigcap_s (U_{1s} \stackrel{d}{=} U_{2s}) \right] \bigcap \left[\bigcap_s (L_{1s} \stackrel{d}{=} L_{2s}) \right] \right\}$$

against

$$H_1 : \left\{ \left[\bigcup_s (D_{1s} \stackrel{d}{>} D_{2s}) \right] \bigcup \left[\bigcup_s (R_{1s} \stackrel{d}{<} R_{2s}) \right] \right.$$
$$\left. \bigcup \left[\bigcup_s (U_{1s} \stackrel{d}{>} U_{2s}) \right] \bigcup \left[\bigcup_s (L_{1s} \stackrel{d}{>} L_{2s}) \right] \right\}.$$

We note that in each stratum this problem presents three binary variables and one quantitative; moreover, all sub-alternatives are one-sided, three in a positive direction and one negative. Variables U and L present some missing values which may be missing not completely at random. However, in this respect, our analysis is performed conditionally since we are mainly interested in the direct effects of two treatments.

As regards the analysis, this is carried out firstly within each stratum and variable and then between strata or between variables. As variables are either binary or quantitative and there are missing values, all partial tests have the form

$$T^*_{hs} = \varphi_h \left(\sum_i X^*_{h1si} \cdot \gamma^*_{h1s} - \sum_i X^*_{h2si} \cdot \gamma^*_{h2s} \right),$$

$s = 1, \ldots, 4,\ h = D, R, U, L$, where $\gamma^*_{hjs} = (\nu^*_{hks}/\nu^*_{hjs})$, $k \neq j = 1, 2$, and the function $\varphi_h(\cdot)$ is $-(\cdot)$ or $+(\cdot)$ according to whether the hth sub-alternative is '$\stackrel{d}{<}$' or '$\stackrel{d}{>}$'.

6.3. Analysis of SETIG Data Using SAS

In this section we briefly revise the analysis of the SETIG data using the software package `SAS` 9.0. In order to carry out the analysis in SAS, the NPC_2samples macro can be used along with the PROC MULTTEST to compute the closed testing procedure within strata, using the exact permutation min$P$ tests (Westfall and Young, 1993). The main instructions to run the macro are:

```
filename mac_npc '...\npc.sas';

dom_byn=great great less, dom_con=great,
clas=group, nsample=10000, strato=strata, paired=no, missing=yes);
```

which requires the following input parameters: data = name of the data set; var_byn = list of binary variables; var_cat = list of categorical variables, non-binary; var_con = list of continuous variables; dom_byn = list of directional

marginal sub-hypotheses for binary variables: if $X_{Au} < X_{Bu}$, specify LESS; if $X_{Au} > X_{Bu}$, specify GREAT; if $X_{Au} \neq X_{Bu}$, specify NOTEQ; dom_con = list of directional marginal sub-hypothesis for continuous variables, see above; weights = list of weights for the variables, firstly specify weights for binary variables, then weights for categorical variables and for continuous variables; clas = variable defining the two groups (character variable); nsample = number of conditional resamplings; strato = variable defining strata (character variable); paired = paired data (character variable) (specify yes/no); unit = variable identifying paired observations; missing = presence of missing values (specify yes/no).

Notice that we have specified D, U, and R as binary variables and L as a continuous variable. Furthermore, we have considered *Group* and *Strata* as variables defining the two groups and the four strata respectively. In order to specify the multidimensional system of hypotheses for each variable within each stratum, a left-tailed test for variable R and three right-tailed tests for the other three variables were considered. This analysis was carried out using $nsample = 10000$ CMC iterations. Since we are dealing with missing data, the option for missing values should be activated.

7. Conclusions

With this work we present an overview of NPC techniques for the analysis of multivariate complex problems.
Applications to real data show how the NPC methodology can be considered as an effective framework for hypothesis testing problems in presence of complex multivariate causal patterns, in both experimental and observational studies, provided that in the null hypothesis the exchangeability of data with respect to groups is assumed.
The proposed approach allows to explore the causal pattern at different levels of the analysis: at the univariate level by means of partial tests; at the multivariate level, either within strata or within variables, by means of second-order combined tests; and at the global level by means of a third order of combination.
NPC tests are relatively efficient and do not require strong underlying assumptions.
Their good properties and power behavior are maintained when applied to high-dimensional and small sample size data set, even when the number of covariates exceeds the number of cases (Pesarin and Salmaso, 2010a). Within the NPC framework, continuous, categorical or mixed variables, with or without missing values, may be easily handled, while the underlying dependence relation structure among variables is nonparametrically and implicitly captured by the combining procedure. Hence, the researcher is not explicitly required to specify the dependence structure of variables. This is a great advantage especially when dealing with correlated data (Pesarin and Salmaso, 2010b; Brombin and Salmaso, 2013).
By making available these `SAS` macros, we hope that researchers dealing with complicated multi-dimensional inferential problems consider NPC approach, that represents a robust nonparametric alternative solution and works under a set of less-stringent assumptions.

The `SAS` macro is available upon request by authors.

REFERENCES

R. Arboretti, F. Pesarin, M. Romero, L. Salmaso (1999). *Sas macro for multivariate and multistrata permutation tests.* In *Proceedings of SUGItalia 99.* pp. 439–51.

R. Arboretti, F. Pesarin, M. Romero, L. Salmaso (2000). *Il progetto setig e la valutazione comparativa delle strategie terapeutiche adottate: metodologia statistica e applicazione.* The Italian Journal of Clinical Pharmacy, 14, pp. 26–36.

A. Birnbaum (1954). *Combining independent tests of significance.* Journal of the American Statistical Association, 49, pp. 559–574.

A. Birnbaum (1962). *On the foundations of statistical inference.* Journal of the American Statistical Association, 57, pp. 269–326.

C. Brombin, L. Salmaso (2013). *Permutation Tests for Shape Analysis.* SpringerBriefs in Statistics.

E. Edgington, P. Onghena (2007). *Randomization Tests, 4th edn.* Chapman and Hall/CRC, London.

A. Edwards (1972). *Likelihood.* Cambridge University Press, Cambridge.

B. Flury (1997). *A First Course in Multivariate Statistics.* Springer NY.

J. L. Folks (1984). *Combination of independent tests.* In P. R. Krishnaiah, P. K. Sen (eds.), *Handbook of Statistics*, North-Holland, Amsterdam, vol. 4, pp. 113–121.

P. Good (2005). *Permutation, Parametric, and Bootstrap Tests of Hypotheses, 3rd edn.* Springer, New York.

C. Goutis, G. Casella, M. Wells (1996). *Assessing evidence in multiple hypotheses.* Journal of the American Statistical Association, 91, pp. 1268–1277.

M. Hollander, D. Wolfe, E. Chicken (2013). *Nonparametric Statistical Methods, 3rd edn.* Wiley Series in Probability and Statistics. Wiley, Hoboken, New Jersey (USA).

I. Liptak (1958). *On the combination of independent tests.* Magyar Tudomanyos Akademia Matematikai Kutato Intezenek Kozlomenyei, 3, pp. 127–141.

J. Marais, G. Versini, C. J. van Wyk, A. Rapp (1992). *Effect of region on free and bound monoterpene and $c_1$3-norisoprenoid concentration in weisser riesling wines.* South African Journal of Enology and Viniculture, 13, pp. 71–77.

J. Oosterhoff (1969). *Combination of one-sided statistical tests.* Mathematical Centre tracts. Mathematisch Centrum.

F. Pesarin (1992). *A resampling procedure for nonparametric combination of several dependent tests.* Journal of the Italian Statistical Society, 1, pp. 87–101.

F. Pesarin (2001). *Multivariate Permutation tests: with application in Biostatistics.* John Wiley & Sons: Chichester-New York.

F. Pesarin (2002). *Extending permutation conditional inference to unconditional one.* Statistical Methods and Applications, 11, pp. 161–173.

F. Pesarin, L. Salmaso (2010a). *Finite-sample consistency of combination-based permutation tests with application to repeated measures designs.* Journal of Nonparametric Statistics, 22, pp. 669–684.

F. Pesarin, L. Salmaso (2010b). *Permutation Tests for Complex Data: Theory, Applications and Software.* Wiley.

F. Pesarin, L. Salmaso (2012). *A review and some new results on permutation testing for multivariate problems.* Statistics and Computing, 22, pp. 639–646.

P. Rosenbaum, D. Rubin (1983). *The central role of the propensity score in observational studies for causal effects.* Biometrika, 70, pp. 41–55.

P. H. Westfall, S. S. Young (1993). *Resampling-Based Multiple Testing.* Wiley, New York.

SUMMARY

Overview of NonParametric Combination-based permutation tests for Multivariate multi-sample problems

In this work we present a review on nonparametric combination-based permutation tests along with `SAS` macros allowing to deal with two-sample and one-way MANOVA design problems, within NonParametric Combination methodology framework. Applications to real case studies are also presented.

Keywords: Multivariate multi-sample problems, One-way MANOVA design, NPC methodology, SAS